Università degli Studi di Padova

Head Office: Università degli Studi di Padova

Department of Mathematics

Ph.D. Course in: Brain, Mind and Computer Science

Curriculum: Computer Science and Innovation for Societal Challenges

Series: XXXVI

Value-Aware Recommendation: Algorithms and Applications

Thesis written with the financial contribution of *estilos srl*

Coordinator: Prof. Anna Spagnolli

Supervisor: Prof. Nicolò Navarin

Co-Supervisor(s): Prof. Lamberto Ballan, Prof. Anna Spagnolli

Ph.D. student: Alvise De Biasio

DEDICATED TO THE PEOPLE WHO HAVE ALWAYS SUPPORTED ME
— DURING THIS JOURNEY AND IN LIFE

Abstract

In a high variety of application domains, the amount of data generated daily has grown more and more over time to the point that its use now exceeds the computational capacity of humans. For example, in the case of e-commerce or online streaming platforms with many new joining users and new items marketed every day, it is complex and time-consuming to manually process and exploit hidden information in order to promptly intercept user interests. In these contexts, machine learning algorithms capable of learning from data have been successfully adopted in the industry by all major market players for their ability to identify patterns in user interactions and generate recommendations that can trigger possible purchases or views on the platforms.

These algorithms, known in the literature as recommender systems, are essentially information filtering technologies designed to process a very large number of alternatives in situations of information overload with the aim of funneling the user's attention to a subset of potentially more interesting items. Over time, alongside the development of gradually more complex machine learning models, e.g., based on deep neural networks, these systems have become increasingly effective at predicting users' interests. Intuitively, the underlying assumption is that a higher-performing service that can provide recommendations of greater interest to users will in turn positively impact business goals as well, e.g., in the form of higher customer retention or loyalty. However, although in some cases this assumption holds, in many others the recommendation of products or services despite being of great interest to users may bring only partial benefits to the business, e.g., certain products may be unprofitable for the company while others may encourage the purchase of complementary competing products.

In reality, recommender systems can be designed to target organizational economic goals more directly by incorporating monetary considerations such as profitability and business value aspects into the underlying machine learning models. Such systems, that are denoted in the literature as value-aware recommender systems, are highly relevant because typically organizations aim to generate recommendations of interest to users only as long as they can increase business value performance indicators. However, although these value-aware systems are of great interest for business purposes, research is still highly scattered and composed of many works proposed in isolated contexts, i.e., where such systems are designed to target only certain application domains and their reuse in other contexts requires major readaptations of the underlying models. Hence, more in-depth research is required.

With this thesis we aim to focus on the study of value-aware recommendation systems, investigating benefits and potential harms of using these algorithms in practical business applications. There are three main contributions arising from our work.

The first contribution is the systematic study of existing literature on value-aware recommender systems. In particular, our work yielded the first systematic reviews specialized on value-aware recommendation systems and a particular business-relevant subset of them, which we denoted in our thesis as economic recommendation systems, i.e., value-aware systems that exploit price and profit information and related concepts from economic and marketing theories to optimize any organization's economics. In our studies, we characterized different facets of such value-aware and economic recommender systems, discussing respectively algorithmic approaches, evaluation aspects, application domains, open challenges and future research directions in the field. Both reviews are based on the well-established PRISMA guidelines, i.e., a systematic review methodology known throughout the scientific community for the high rigor and reliability of results, which aims to identify all research that is relevant to a given topic.

The second contribution consists in the adaptation according to a value-aware perspective of four different families of state-of-the-art recommendation algorithms widely used in industry i.e., nearest neighbors, matrix factorization, learning-to-rank and neural models. In particular, we proposed various in-processing approaches to integrate the objective function of such families of algorithms with the goal of optimizing the profitability of recommendations at learning time. The proposed methods have consistently proved effective in generating more

profitable, yet relevant recommendations than (non value-aware) baseline methods and than value-aware post-processing methods that are widely used in the literature. In addition, our in-processing methods have also proven to be computationally much more efficient at prediction time and could therefore be preferred in practical business applications where post-processing methods might be inapplicable because they require a certain overhead to generate recommendations, e.g., considering the case of large-scale production systems with millions of active users and catalog items.

Finally, with our third contribution we aimed to focus on broader issues concerning value optimization for society as a whole. Specifically we argue that, just as it is important for the business to have a recommender system that can optimize business performance indicators, it is of great value for the society to have a recommender system that seeks to maintain high user well-being, e.g., by not encouraging risky or aggressive behavior in the real world to maintain high performance in the platforms. Although not directly measured, this well-being is impacted by certain levels of diversity and fairness in recommendations that if not appropriately calibrated may risk influencing the behavior of certain user groups that are more sensitive than others to certain topics due to repeated exposure mechanisms. With the aim of addressing this problem, we proposed several diversity-based recommendation algorithms to calibrate the exposure of such sensitive users to what we denoted in our work as influential items, focusing our analysis on two case studies on potentially depressed and aggressive users.

Acknowledgments

First of all, I would like to thank Dr. Andrea Bergamo and the *estilos srl* team for giving me the opportunity to develop this work. It was a pleasure and an honor for me to be able to embark on this journey.

Afterwards, a big thank you goes to my advisor, Prof. Nicolò Navarin, for his guidance and support during this doctoral path. Thank you very much for the constructive comparisons, the ideas provided to this work, as well as for the availability demonstrated throughout this period.

Next, I am pleased to thank Prof. Dietmar Jannach, for welcoming me as a visiting researcher at the University of Klagenfurt. You helped me a lot during my journey in Austria, and I'm very proud of the work we carried out.

Also, I need to thank all the people I interacted with during this experience. A great thank goes to Dr. Andrea Montagna, Prof. Fabio Aiolli, Prof. Merylin Monaro, Prof. Luca Oneto, Prof. Lamberto Ballan, Prof. Mirko Polato, and all my *estilos* and *UniPd* colleagues for the encouragement and for contributing positively to this work.

Last but not least, my biggest thanks to my family, my girlfriend and all the people who have always been there for me in the most difficult times. Thank you for supporting me every day: part of this work is also yours.

Then, a special wish to all readers who will venture into this thesis.
— I hope this reading will thrill you and provide you with new ideas for your own journey.

Contents

Listing of figures

Listing of tables

Listing of acronyms

Artificial Intelligence Acronyms:

AI .... Artificial Intelligence
ML .... Machine Learning
DL .... Deep Learning
RL .... Reinforcement Learning
MLP .... Multi-Layer Perceptron
GNN .... Graph Neural Network
SGD .... Stochastic Gradient Descent
KL .... Kullback-Leiber Divergence
LDA .... Latent Dirichlet Allocation
EA .... Evolutionary Algorithms
NSGA-II .... Non-dominated Sorting Genetic Algorithm II
MOABC .... Multi-Objective Artificial Bee Colony

Recommender Systems Acronyms:

RS .... Recommender System
CF .... Collaborative Filtering
CB .... Content-Based Filtering
HS .... Hybrid System
IR .... Information Retrieval
LTR .... Learning To Rank
CARS .... Context-Aware Recommender System
MORS .... Multi-Objective Recommender System
MSRS .... Multi-Stakeholder Recommender System
VARS .... Value-Aware Recommender System
RLRS .... Reinforcement Learning-based Recommender System
MCRS .... Multi-Criteria Recommendation Systems
ECRS .... Economic Recommender System
PARS .... Profit-Aware Recommender System

UCF .... User-Based Collaborative Filtering

ICF .... Item-Based Collaborative Filtering

MF .... Matrix Factorization

CAMF .... Context-Aware Matrix Factorization

BPR .... Bayesian Personalized Ranking

NCF .... Neural Collaborative Filtering

SLIM .... Sparse LInear Method

Mult-VAE .... Variational AutoEncoders with Multinomial Likelihood

VNS .... Value Neighbor Selection

VMF .... Value Matrix Factorization

VNCF .... Value Neural Collaborative Filtering

VBPR .... Value Bayesian Personalized Ranking

HPRS .... Hybrid Perspective Recommender System

CPR .... Constrained Profit Ranking

MOPR .... Multi-Objective Profit Ranking

Metrics Acronyms:

MAE .... Mean Average Error

RMSE .... Root Mean Square Error

HR .... Hit-Rate

Prec .... Precision

Rec .... Recall

F1 .... F1-Score

AUC .... Area Under the ROC Curve

MRR .... Mean Reciprocal Rank

NDCG .... Normalized Discounted Cumulative Gain

IDCG .... Ideal Discounted Cumulative Gain

EP .... Expected Profit

PAH .... Profit At Hit

P-NDCG .... Price-Based NDCG

P-IDCG .... Price-Based IDCG

NDCV .... Normalized Discounted Cumulative Value

IDCV .... Ideal Discounted Cumulative Value

AIR .... Average Influential Ratio

SIB .... Sensitive ImBalance

NP .... Non-Parity Unfairness

Business Concepts Acronyms:

KPI .... Key Performance Indicator

IPV .... Individual Page View

CTR .... Click-Through Rate

CVR .... Click Conversion Rate

GMV .... Gross Merchandise Value

CLV .... Customer Lifetime Value

WTP .... Willingess To Pay

RFM .... Recency Frequency Monetary

RCT .... Rational Choice Theory

MAUT .... Multi-Attribute Utility Theory

Psychology Concepts Acronyms:

OCEAN .... Openness to experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism

NEO-PI .... NEO Personality Inventory

IPIP .... International Personality Item Pool (IPIP)

Systematic Review Acronyms:

PRISMA .... Preferred Reporting Items for Systematic Reviews and Meta-Analyses

DA .... Dimension of Analysis

RQ .... Research Question

SQ .... Search Query

EC .... Eligibility Criteria

SL .... Study Limitation

Part I

Introduction and Background

1
Introduction

This chapter opens the curtain of this thesis by introducing the underlying motivations, main contributions and overall structure.

1.1 Motivations

The exponential growth in the amount of data available in recent years has raised the need for increasingly advanced systems to process and analyze such information. Indeed, in many application domains the amount of data has become so large that it exceeds the capacity of humans to process it. For these reasons, systems based on machine learning algorithms have now become the standard in the industry. These systems are able to learn from data, make predictions or extract knowledge that can be useful for a variety of purposes. In particular, especially in e-commerce, advertising platforms or streaming media sites, machine learning algorithms have been employed successfully for their ability to promptly intercept user interests and provide recommendations that can increase platform adoption or trigger further possible purchases. In the literature, these algorithms are denoted as recommender systems because of their ability to provide recommendations of interest to the user. These recommender systems are essentially information filtering technologies designed to process a very large number of alternatives in situations of information overload with the aim of funneling the user's attention to a subset of potentially more interesting items.

Over time, alongside the development of gradually more complex machine learning models, such recommender systems have become increasingly accurate at analyzing platform interactions, to the point that nowadays they are employed by all major market players. By effectively understanding important aspects of human behavior and cognition, recommender systems enable companies to exploit such information for business purposes. The success of these systems has been mainly attributed to their ability to identify users' latent interests (or needs). Intuitively, the underlying assumption is that a more accurate service able to generate recommendations of greater interest to the user often has a direct positive impact on the company's business. This assumption holds in many circumstances, as a user who is more satisfied with the service often unconsciously rewards the company by taking greater advantage of its services and promoting the brand with other users. Correspondingly, the vast majority of the literature of recommender systems (following machine learning trends in the field) focuses on improving the quality of predictions more and more, e.g., by adopting more advanced machine learning algorithms, specializing systems by application domain or embedding additional knowledge into the underlying models.

However, machine learning in general is not just about building complex models that can provide accurate estimates. Indeed, especially in certain application domains, such as in the case of e-commerce or the others mentioned earlier, there are various commercial aspects that need to be considered when designing algorithms to be successful for users and business. At present (and probably in the future), algorithms are designed to answer to a given input with a certain output. Hence, it is still the burden of humans to design such algorithms so that in input they receive the data that matters and in output they answer with what is really required. Thus, recommender systems, having to address complex business needs first and foremost, should be designed both to provide recommendations that are attractive to the user and to bring some kind of value to the company. Indeed, it might bring little (or no) benefit to the company to recommend unprofitable products or to give more visibility in its platform to competitors' products.

In reality, it is possible to design recommendation systems to target the economic goals of organizations more directly. One way to do this, for example, is to incorporate monetary considerations such as profitability or business value aspects directly within the underlying machine learning models. These systems are known in the literature as value-aware recommender systems and are for obvious reasons of great interest to the industry. However, as the literature has focused primarily on improving the quality of inferences of the underlying machine learning models it has probably unconsciously neglected the business motivations that should be equally important for the success of these systems. In fact, research on value-aware recommender systems is still highly scattered nowadays and composed of many papers with specific peculiarities that have emerged in isolated contexts. For example, many papers have proposed systems targeting only certain application domains, such as taxi drivers or insurance, and reusing the algorithms in other contexts would not be possible without major readaptations.

As we will delve into the next chapters, this gap in the literature is also partly due to the information that is currently public. In fact, many studies in the field are based on proprietary datasets because the most known datasets do not contain business value information that is needed to build such value-aware systems. Publicly disclosing even anonymized subsets of such information could in fact be somewhat detrimental to companies, e.g., competitors could take advantage of economic data on purchases and profitability to study weaknesses in the business model and take away market share. However, especially recently, this area is becoming more open, and several studies based on public datasets have also emerged. It therefore represents a twofold opportunity for research to study in greater depth the special characteristics of these value-aware systems and for industry to be able to adopt technologies better suited to its objectives. In particular, more in-depth research is needed to explore the different facets of these value-aware recommendation systems, adopting a multidisciplinary perspective that combines different aspects from computer science, statistics, economics and psychology. This thesis has been created with this goal in mind.

1.2 Contributions

The aim of this thesis is study value-aware recommendation systems, investigating benefits and potential harms of using these algorithms in practical business applications. The contributions of our work can be grouped into three main branches.

Systematic analysis of existing literature on value-aware recommender systems: our work yielded the first systematic reviews based on PRISMA guidelines specialized on value-aware recommender systems and a particular business-relevant subset that we denoted as economic recommendation systems. In our studies, we analyzed in-depth such systems, by discussing a number of related articles collected from different research streams. In particular, we discussed the technical approaches that can be used to build economic and value-aware recommendation algorithms, offline and online evaluation methodologies, most commonly used datasets and major application domains. We also pointed out current challenges, limitations of today's research and a number of possible future research directions in the field.

Investigation of in-processing approaches to build profit-aware recommendation systems: we extended, according to a value-aware perspective, four different families of state-of-the-art recommendation algorithms widely used in industry i.e., nearest neighbors, matrix factorization, learning-to-rank and neural models. The key idea is to integrate the objective function of such families of algorithms to embed profit awareness at learning time, thus generating more profitable yet relevant recommendations at prediction time. By comparing our in-processing models with some of the most commonly used post-processing methods in three different real-world datasets, we demonstrate that the proposed models may represent viable alternatives to build profit-aware recommendation systems for practical business applications. The proposed methods consistently exhibit comparable or better performance than post-processing methods in generating more profitable recommendations. Moreover, our methods also proved to be computationally more efficient because, by incorporating profit awareness at learning time, they do not require any post-processing overhead at prediction time.

Study of the problem of recommending influential items to sensitive users: focusing on broader issues regarding value optimization for society as a whole, in this thesis we argue that it is of great value for society to have a recommender system that seeks to maintain high user well-being without compromising the performance of the platform. Although not directly measured, the users' well-being may be impacted by recommendations due to repeated exposure mechanisms: if certain levels of diversity and fairness are not appropriately calibrated the recommender may risk influencing over time in a dangerous way the behavior of certain user groups that are more sensitive than others, e.g., by encouraging risky or aggressive behaviors. In our work, with the aim of addressing this problem, we proposed several diversity-based recommendation algorithms to calibrate the exposure of such sensitive users to items that may influence in a negative way their behavior. Specifically, by focusing our analysis on two case studies of potentially depressed and aggressive users, we demonstrate the validity of our methods in mitigating the overexposure of potentially aggressive users to controversial items that may have a negative impact on their behavior and encouraging exposure of favorable items that may positively influence the behavior of potentially depressed users.

1.3 Publications

Part of the content of this thesis has been published in international peer-reviewed journals [105, 106, 313]. Other contributions are currently submitted to other venues. Below we report the complete list.

Journals

J01 **A Systematic Review of Value-Aware Recommender Systems**. Alvise De Biasio, Andrea Montagna, Fabio Aiolli, Nicolò Navarin. Expert Systems With Applications, vol. 226, pp. 120131, 2023.

J02 **On the Problem of Recommendation for Sensitive Users and Influential Items: Simultaneously Maintaining Interest and Diversity**. Alvise De Biasio, Merylin Monaro, Luca Oneto, Lamberto Ballan, Nicolò Navarin. Knowledge-Based Systems, vol. 275, pp. 110699, 2023.

Workshops

W01 **Graph-based Explainable Recommendation Systems: Are We Rigorously Evaluating Explanations?**. Andrea Montagna, Alvise De Biasio, Nicolò Navarin, Fabio Aiolli. Workshop on User Perspectives in Human-Centered Artificial Intelligence.

Submissions

S01 **Economic Recommender Systems — A Systematic Review**. Alvise De Biasio, Nicolò Navarin, Dietmar Jannach. Electronic Commerce Research and Applications.

S02 **Model-Based Approaches to Profit-Aware Recommendation**. Alvise De Biasio, Dietmar Jannach, Nicolò Navarin. Expert Systems With Applications.

1.4 Outline

The thesis is organized in two main parts:

PART I provides a comprehensive background to fully understand the remainder of this thesis. In particular, with Chapter 1, which has already started, we introduce the motivations, outline and main contributions of this work. Then, in Chapter 2 we discuss the state-of-the-art in the field. We first offer a broad introduction on recommender systems in Section 2.2. In particular, we cover the most-known recommendation problems, the various classes of algorithms and offline evaluation aspects. Then we delve into a more specialized area, focusing on the design of a recommender system in accordance with specific purposes in Section 2.3. Specifically, we first discuss how business value can be generated from recommendations and then introduce the two recent classes of value-aware and economic recommender systems which are the focus of this thesis.

PART II groups the original contributions of this thesis. In particular, Chapter 3 [106] and Chapter 4 are based on two systematic reviews specialized on the two novel families of value-aware and economic recommender systems. In these chapters, after categorizing different facets of such systems we discuss algorithmic approaches, evaluation methodologies, application domains, open challenges and future research directions in the field. Next, leveraging the previous findings, Chapter 5 offers a set of experimental studies on profit-aware recommendation systems. In particular, we discuss how the objective function of matrix factorization and three other widely-used methods in the industry can be extended to optimize the profitability of recommendations at learning time. Then, Chapter 6 [105] focuses on the problem of recommending items that can influence the behavior of sensitive users. Specifically, we discuss various methodologies that can be used to diversify recommendations while maintaining high performance, focusing our analysis on two different case studies based on potentially depressed and aggressive users. The thesis ends with a summary of the results achieved by our research and a discussion of possible future directions in Chapter 7.

2
Background

In this chapter, we provide all the background knowledge necessary to understand the remainder of this thesis. We start with some preliminaries on machine learning in Section 2.1. Then we introduce recommender systems in Section 2.2. In particular, we present the learning problem, offline evaluation metrics, most commonly used recommendation algorithms and typical open challenges in the field. Afterward, we cover more advanced recommendation design topics in Section 2.3. Specifically, we discuss how a recommendation system can be adapted to generate recommendations that may improve certain business value performance indicators.

2.1 Machine Learning Preliminaries

Learning from experience is one of the fundamental components of intelligent behavior and what allows humans to adapt to a variety of situations. Everyday activities, such as understanding the mechanisms of our world or making conscious decisions to achieve certain goals, involve learning from experience. Many of these activities that characterize human beings such as reasoning or creativity are complex, and it is very difficult to make a machine succeed in imitating them. However, a number of different algorithms have been proposed in the literature, demonstrating valuable performance in certain specific tasks. This science that aims to enable machines to learn from experience is called *Machine Learning* (*ML*).

The use of machine learning proved to be highly useful for a variety of applications, e.g., concerning image recognition, question answering and recommendations. One of the key characteristics of these applications is that the amount of data collected day by day exceeds the human capability of extracting and exploiting the hidden information. For example, especially in e-commerce, with many new users joining every day and new items being marketed by various vendors, it would be very time-consuming for a human being to provide recommendations of interest to every customer. Instead, an algorithm capable of learning from data, since it can process information much faster, may generate various recommendations of interest that could in turn trigger possible purchases.

2.1.1 Learning From Data

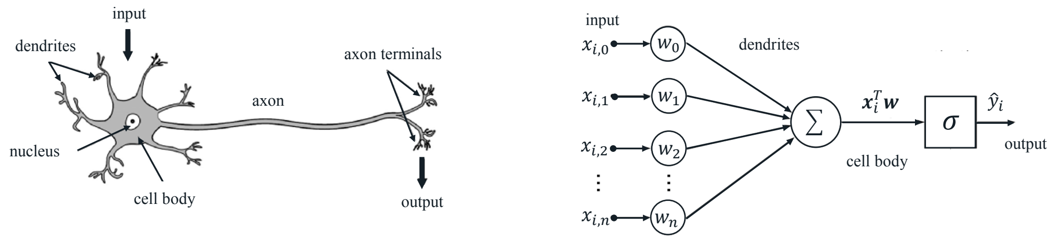
As human beings by our nature, we typically learn general concepts from specific examples. Thereafter we can leverage these concepts in different ways to perform our daily activities. Although it is often difficult to state such activities unambiguously since many concepts that seem simple to us are actually very complex to express in a language that can be understood by a computer program, it is possible in many cases to define a certain specific *task*. This way, we can design computer programs that can learn from experience to perform these tasks and verify their performance through specific metrics. For example, providing recommendations to customers in e-commerce is an ambiguous task because purchasing dynamics often involve logical rational processes (e.g., the usefulness of certain products and services in terms of functionalities) and unconscious irrational processes (e.g., the subjective perception of such usefulness). However, as we will explore in detail in Chapter 2.2, it is possible to design systems that can provide recommendations of interest to customers by learning from experience.

Referring to the previous task, the learning paradigm predominantly used in the literature is denoted as *supervised learning*. In accordance to this paradigm, the system is designed to predict some known expected output (i.e., users' interests) from a set of examples (i.e., mainly user-item interactions). However, other well-known paradigms in the field of machine learning such as *unsupervised learning* or *reinforcement learning* are also sometimes used to build recommender systems. For example, unsupervised learning is often used to extract hidden patterns from the data that can make supervised recommendations more accurate, e.g., by classifying users into clusters with similar purchasing behaviors and feeding the recommendation algorithm with such additional information. Instead, reinforcement learning can be used to optimize certain (long-term) goals, e.g., to generate recommendations able to optimize certain business performance indicators. In this thesis, we focus mainly on supervised learning.

More formally, in the machine learning field, the supervised learning paradigm involves training a model on a set of instances in the form $\mathcal{D} = \{(\mathbf{x}_i, y_i) : i = 1, \dots, m\}$ to predict for each i-th instance $\mathbf{x}_i \in \mathbf{X} \in \mathbb{R}^{m \times n}$ (composed of $n \geq 1$ features) the known output $y_i \in \mathbf{Y} \in \mathbb{R}^m$ (also referred to as label), where $\mathbf{X}$ and $\mathbf{Y}$ are denoted as the matrices of the input features and output labels of the instances. Depending on the type of output, the problem is typically referred to in the literature as *classification* or *regression*. In the first case, the output is categorical, e.g., considering $y_i \in \{-1, 1\}$ in the case of *binary classification* and $y_i \in \{0, \dots, l\}$ in the case of *multi-class classification*. In the second case, instead, the output is numerical, i.e., $y_i \in \mathbb{R}$. The goal of a supervised learning model is to best approximate the function $c(\mathbf{\Theta}) : \mathbf{X} \to \mathbf{Y}$ representing the relationships between $\mathbf{x}_i$(s) and y_i(s), where $\mathbf{\Theta}$ are the model parameters. Through such an approximation function, a prediction $\hat{y}_i$ representing the output of the model for each instance i can be generated. As we will later detail in Section 2.2.1, recommendation systems were proposed in the literature in early studies generalizing from the classification and regression problems.

2.1.2 Perceptron Algorithm

Many of the algorithms that have been proposed in the field of machine learning take inspiration from natural mechanisms of our world. By imitating such mechanisms, algorithms can learn from experience to tackle a given task. One of the most well-known and widely used algorithms from which many other algorithms proposed in the literature have taken inspiration is the *perceptron*. This algorithm was proposed by Frank Rosenblatt in 1958 [379] as a computational model of the McCulloch-Pitts biological *neuron* [300], i.e., the cells in the human



(a) Biological neuron

(b) Mathematical Neuron

Figure 2.1: Biological and mathematical neuron representations [321, 336].

brain that are involved in processing and transmitting electrical and chemical signals. Such computational (or artificial) model of the neuron, similarly to its biological counterpart, receives one or more inputs representing post-synaptic potentials at neural dendrites and use them to generate an output, representing the neuron's action potential which is transmitted along its axon.

As shown in Figure 2.1, the perceptron exploits this artificial neuron structure, which can be viewed primarily as a multi-dimensional vector $\mathbf{w}$ containing a weight for each input feature of $\mathbf{X}$ and by an activation function $\sigma(\cdot)$, to learn from experience to perform a binary classification task. As a first step, the algorithm initializes the weight vector randomly. Next, for each i-th instance of the training set the algorithm uses the weights to predict whether the instance is positive or negative, i.e., if $\hat{y}_i = \sigma(\mathbf{x}_i^\mathsf{T}\mathbf{w}) > 0$. Then, the predictions of the algorithm are compared with the real values y_i of each example. In case of error, i.e., if $\hat{y}_i \neq y_i$, the original weight vector $\mathbf{w}$ is updated with a new version $\mathbf{w}'$ calculated as $\mathbf{w}' = \mathbf{w} + \iota \cdot (y_i - \hat{y}_i) \cdot \mathbf{x}_i$ where ι is a constant denoted in the literature as the learning rate that governs the amount of update in each iteration. Iterating on all the training set, the algorithm allows for finding a linear separator to distinguish instances into positive or negative, if it exists.

As we will see later in this thesis, many linear recommendation algorithms are based on principles similar to that of perceptron (see Section 2.2.5.2 and Section 2.2.5.4). Indeed, such recommendation algorithms typically learn from experience by iteratively updating weights that they use in turn to generate recommendations to users.

2.1.3 The Curse of Dimensionality

One of the most well-known problems in the field of machine learning is denoted in the literature as the *curse of dimensionality*. Typically this problem occurs when the number of dimensions (or features) used is very large. In fact, as the dimensionality of the data increases, typically the *representation* of individual instances becomes increasingly sparse, i.e., only a few features are populated per instance. In these settings, it is difficult for machine learning algorithms to find groups with similar properties on which to base their predictions: as a result, the predictions provided are often not much accurate. This problem is known especially in the field of recommender systems. In large scale settings with millions of active users and millions of items in the catalog, it is indeed very likely that each user interacts with only a few items. Correspondingly, the sparsity of interactions often makes the recommendations ineffective.

To address this problem, usually in the field of machine learning, *dimensionality reduction* techniques are used to find a lower-dimensional representation of the data on which to base predictions. The main advantages of these techniques are mainly related to reducing noise and obtaining datasets that are easier to use by algorithms and more intuitive to understand by humans. One of the most simple yet intuitive dimensionality reduction techniques is known as *feature selection*, which aims to select a subset of the most informative features, discarding those that are unhelpful or unrelated. However, although very popular, this approach tends to suffer in the presence of very sparse data, e.g., as in large-scale recommendation settings. In such settings, more advanced approaches tend to be preferred in the literature, such as *Singular Value Decomposition* (*SVD*) or *Principal Component Analysis* (*PCA*), which exploit matrix decomposition techniques from linear algebra to obtain a dense lower-dimensional representation of features that can be used to generate more accurate predictions.

In the field of recommender systems, dimensionality reduction techniques (and in particular SVD), as we shall see in Section 2.2.5.2 were first used to build the *matrix factorization* approaches that became state-of-the-art in the well-known Netflix Prize competition. The basic idea behind these methods is to determine a dense lower-dimensional representation of users and items that can be used to make more accurate recommendations.

2.1.4 Deep Learning

In many practical circumstances, it is very difficult to model the data in a dense representation that contains all the useful information for a machine learning algorithm designed to address a given task. Feature selection and dimensionality reduction algorithms in general, while useful and often used as the first option may achieve overly approximate solutions by discarding much useful information. To address this problem, recently the academic community has been moving toward the development of *Deep Learning* (*DL*) algorithms that have the ability to autonomously learn the best representation from data. As we will see in the next section, these algorithms based on *deep neural networks* are composed of many layers that encode information at different levels of abstraction, which are then used to address a given task.

Such neural networks have achieved state-of-the-art performance in many application domains, e.g., especially in those concerning multimedia data such as images and text or in those where the number of dimensions used to represent the data is still very large. As we will discuss later in this work, promising results have also been obtained in the field of recommender systems. Deep learning algorithms were initially proposed to be used as an alternative to the well-known matrix factorization approach, which is instead based on linear dimensionality reduction algorithms (Section 2.2.5.5). Recent trends have then moved toward generative models based on what is denoted in the literature as autoencoders (Section 2.2.5.6), i.e., networks capable of encoding multidimensional information into a dense lower-dimensional representation which is in turn used to reconstruct the original information.

2.1.5 Multi-Layer Perceptron

The *Multi-Layer Perceptron* (*MLP*) is one of the most known neural architectures in the deep learning literature. Such architecture is composed of several interconnected artificial neurons (see Section 2.1.2) and can be seen as a simplified computational model of the human brain. As can be seen in Figure 2.2, structurally the MLP is composed of a network of multiple layers of artificial neurons stacked on top of each other - the number of neurons and layers is a hyperparameter of the network. The network layers are divided into input, hidden and

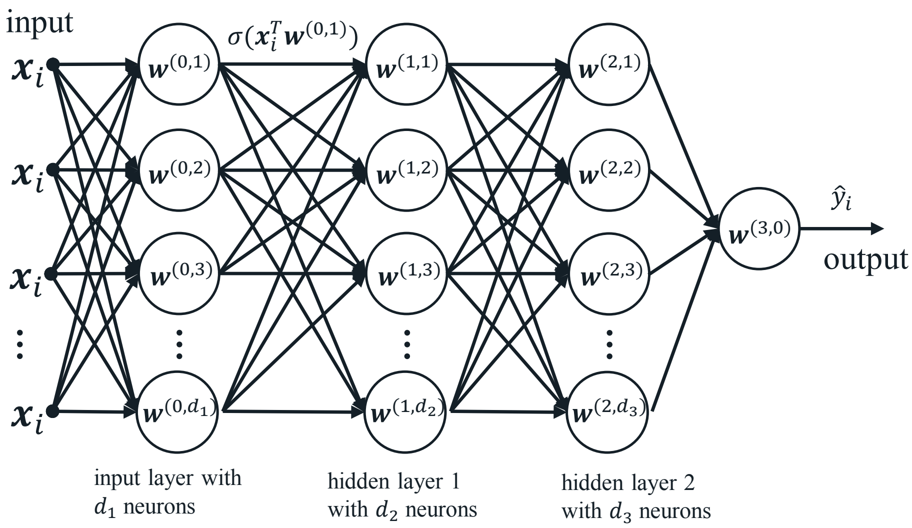


Figure 2.2: Multi-layer perceptron neural architecture.

output layers: data initially flow into the input layers, information then propagates into the hidden layers and finally the output layer provides the predictions. MLP makes it possible to overcome the limitation related to the linearity of the predictions of the original perceptron algorithm, i.e., the predictions being based on a combination of the outputs of multiple individual neurons that are activated based on how the information flows in the network can approximate also complex non-linear mechanisms.

Although the artificial neuron was developed in the 1950's [379], the MLP has only recently become popular in the academic and industrial community. Indeed, the computational costs of training each neuron individually (that is prohibitively expensive in most of practical settings) was overcome only in 1986 [380] when it emerged the now famous *backpropagation* method by which all neural networks are trained today. This method consists mainly of two steps known as forward and backward pass. With the first step, information propagates among the various neurons in the network until it reaches the output layer to compute predictions. Then, by adopting a more generalized version of the previously discussed perceptron learning rule (see Section 2.1.2), based on the difference between the predictions and the expected output, the error is calculated and the weights of the neurons within the network layers are updated sequentially backward. These steps are performed iteratively for various epochs until convergence.

The algorithm by which the training is performed is also known as *Stochastic Gradient Descent* (*SGD*) in that in order to update the weights of neurons in a given layer at a certain interaction, it is necessary to propagate the gradient of the *loss function* in a descending manner between layers. This loss function (also called objective function) is a mathematical function commonly found in (almost) all machine learning algorithms that quantifies the differences between a model's predictions and expected values. This function is typically used to update internal model parameters (e.g., internal MLP weights) in order to obtain more accurate predictions, epoch by epoch.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	5		2					3		4
u_2				5	5		2		3	
u_3	2					4		5		
u_4		3			4			1		
u_5			2			5		4		
u_6		4				2			3	
u_7				1				5		

(a) Explicit feedback

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
u_1	1		1					1		1
u_2				1	1		1		1	
u_3	1					1		1		
u_4		1			1			1		
u_5			1			1		1		
u_6		1				1			1	
u_7				1				1		

(b) Implicit feedback

Figure 2.3: User-item interaction matrices with explicit and implicit feedback.

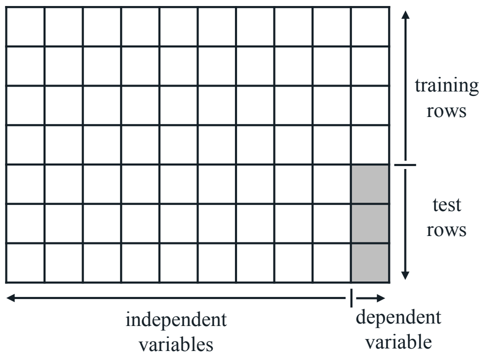
One of the most widely used loss functions for regression problems in the literature is the *squared loss*:

$$\mathcal{L} = \frac{1}{n} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}} (y_i - \hat{y}_i)^2 \tag{2.1}$$

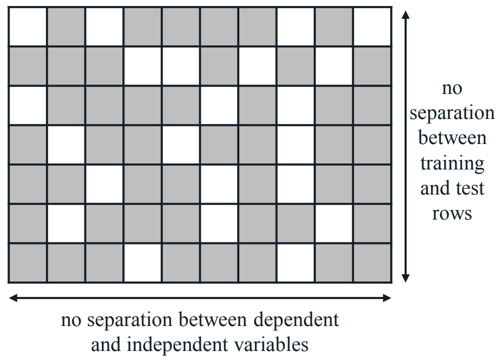
that is, the average of the squared error calculated over all the instances of the training set. This loss function is the basis of many recommendation algorithms such as the well-known matrix factorization that we will describe in Section 2.2.5.2 and its corresponding neural network-based version called neural collaborative filtering that we will describe in Section 2.2.5.5. As we will see throughout this document, the role of the loss function is very important for any algorithm's learning because different losses can configure the internal parameters of the models so that the predictions may give more weight to certain aspects than others. In particular, as part of the original contributions of this thesis, in Chapter 5 we will explore how to modify the loss function of various recommendation algorithms to generate recommendations that optimize certain business performance indicators.

2.2 Introduction to Recommender Systems

A *Recommender system* (*RS*) is an information filtering technology that aims to make suggestions to users [55, 374]. Commonly, the main purpose of an RS is to help users evaluating a large number of alternatives in situations of information overload [58] by proposing subsets of items that could be of greatest interest. In the RSs field, an item is a general term used to denote what the system can suggest. These suggestions, also known as recommendations, can concern: products to buy, news to read, social pages to follow, and other entities depending on the particular application domain (e.g., as in e-commerce [269], online streaming [360] or social networks [89]). In the following, we indicate as $\mathcal{U} = \{u_1, \ldots, u_m\}$ the set of m users that may receive the recommendations and we denote as $\mathcal{I} = \{i_1, \ldots, i_n\}$ the set of n items that can be recommended by the system. Moreover, we denote for convenience as $\mathcal{I}_u^+$ ($\mathcal{I}_u^-$) the set of items with which user u has (respectively: has not) interacted with.



(a) Classification



(b) Rating prediction

Figure 2.4: Comparison of classification and rating prediction tasks.

Recommendations may or may not be personalized, depending on whether the RS recommends different items to different users or whether it recommends the same items to all users. There are cases where a non-personalized recommendation is able to achieve satisfactory results, such as in the case of travel or restaurant recommendation, where the most popular or trending items are usually recommended. However, in most business settings such as in e-commerce a more personalized recommendation that fully captures users' interests tends to obtain better results. Correspondingly in this thesis we focus on this case of greatest interest to the industry.

Typically, recommender systems exploit historical data about user-item interactions to learn how to provide suggestions to users. As shown in Figure 2.3, these interactions can be divided into implicit or explicit, depending on the nature of the feedback provided by the user. Specifically, implicit feedback is that type of interaction which comes from a natural use of the system: e.g., clicks, views, purchases. Instead, explicit feedback represents some preference that a user explicitly provides regarding certain items and unambiguously indicates his or her tastes: e.g., 1-to-5 stars ratings or likes/dislikes. In this document, we mainly focus on implicit feedback-based RSs that are widely used in practice, considering a binary user-item interaction matrix $\mathbf{X} \in \{0, 1\}^{m \times n}$, where each entry $x_{u,i} \in \{0, 1\}$ indicates whether user u interacted with item i or not (e.g., purchased it).

2.2.1 Generalizing Recommendations From Rating Predictions

Depending on the application domain, the system may or may not recommend items that the user has already interacted with in the past. There are cases such as in the food delivery business where restaurants are often recommended from which the user frequently orders. However, in many other contexts such as in e-commerce or online streaming platforms, products or content that the user has not interacted with tend to be preferred for recommendations. In this thesis we focus on the latter case, where most of the literature also focuses. Indeed, the first learning problem formalized in the RSs field is known as the *rating prediction problem* and was initially proposed by considering an explicit type of feedback, e.g., $x_{u,i} \in [0, 5]$. Specifically, given some known (or observed) user preferences in the user-item interaction matrix $\mathbf{X}$, the problem requires predicting the missing (or unobserved) entries through a learning algorithm. Such rating predictions of user interests can be in turn used to generate recommendations to users, e.g., by selecting the top-k items with the highest predicted ratings.

This setting has various similarities and differences with the well-known *classification* and *regression* problems previously discussed. In particular, classification and regression problems require estimating a certain dependent variable from a specified number of independent variables. Instead, in the rating prediction problem, there is no single dependent variable to predict because each column can have missing values. The main difference is that while in classification and regression problems, the missing entries are restricted to a single dependent variable, in the rating prediction problem there is no clear demarcation between dependent and independent variables as each variable plays a dual role (see Figure 2.4). Moreover, there is also no clear distinction between training and test set rows as each row can also have missing values. Hence, the rating prediction problem can be viewed as a generalization of classification and regression problems where predictions are performed entry-wise instead of row-wise.

2.2.2 On the Top-k Recommendation Problem

In practical circumstances (e.g., consider e-commerce), it is not necessary to have a very precise algorithm that can predict all the missing user preferences of the user-item interaction matrix because recommendation systems are typically employed to determine the best k items to recommend to each user. This particular setting is known as the *top-k recommendation problem* and has become highly important especially in recent years [111]. Accordingly, in the most general form a recommendation algorithm is designed to determine an ordered list $\mathbf{y}_{u,k}$ of k items to be recommended to each user by optimizing a generic *utility function* $\tau(\mathbf{y}_{u,k}) : \mathbf{y}_{u,k} \to \mathbb{R}$. More formally:

$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \quad \tau(\mathbf{y}_{u,k}) \tag{2.2}$$

The utility function can be implemented in any arbitrary way, considering the relevance of recommendations for the end user (i.e., how much the recommendations reflect the user's interests), the business value for the service provider (e.g., how much profitability the company may achieve from the recommendations) or other aspects.

Focusing on relevance aspects, the vast majority of algorithms in the RSs literature typically first seek to learn a *scoring function** $c(\mathbf{\Theta}) : \mathbf{X} \to \hat{\mathbf{X}}$ to predict missing entries of $\mathbf{X}$, where $\hat{\mathbf{X}} = \{\hat{x}_{u,i} \in \mathbb{R} : 0 \leq \hat{x}_{u,i} \leq 1\}^{m \times n}$ is the prediction matrix and $\hat{x}_{u,i}$ represents the expected interest of the user toward an item he or she has never interacted with [368]. Correspondingly, the vast majority of the literature in the RSs field (and most of this thesis with the exception of Section 4.2.2), given $\rho_{u,i}$ as the generic utility of the user-item interaction and considering $\rho_{u,i} = \hat{x}_{u,i}$, operationalize the utility function as $\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \rho_{u,i}$ to build almost any recommendation algorithm to directly optimize:

$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \tag{2.3}$$

thus with the aim of recommending the top-k items most potentially interesting to each user. To find these items, in accordance with the optimization problem above, after generating the predicted scores through the scoring function, the algorithms typically sort in descending score order the items for each user and select the top-k.

*The scoring function is parameterized by a set $\mathbf{\Theta}$ of model parameters. For some algorithms, such as the well-known User-Based Collaborative Filtering based on Nearest-Neighbor techniques [326] that we will describe in-depth in Section 2.2.5.1, we assume $\mathbf{\Theta} = \emptyset$ since there are no model parameters.

Notation	Definition
u	user
i	item
m	number of overall users
n	number of overall items
k	number of items to recommend
$\mathcal{U} = \{u_1, \dots, u_m\}$	set of users
$\mathcal{I} = \{i_1, \dots, i_n\}$	set of items
$\mathbf{X}$	user-item interaction matrix
$x_{u,i}$	user-item feedback
$\mathbf{\Theta}$	set of model parameters
$c(\mathbf{\Theta})$	scoring function
$\hat{\mathbf{X}}$	prediction matrix
$\hat{x}_{u,i}$	user-item predicted interest
$\mathbf{y}_{u,k}$	recommendations list
$\tau(\mathbf{y}_{u,k})$	utility function
$\rho_{u,i}$	user-item interaction utility

Table 2.1: Main notation.

However, although this user-focused utilitarian conception is currently the most widely used in the literature, a recommendation provider (or more generally another stakeholder) may have different goals. In particular, in the context of this thesis, as an original contribution we will delve into what we called value-aware recommender systems in Chapter 3 and economic recommender systems in Chapter 4. As we will later discover, these novel classes of RSs differ from traditional systems in that they aim to optimize certain business values for the organization. Hence, in these cases the utility function and optimization problem previously introduced in Eq. (2.2) and Eq. (2.3) will be extended to incorporate those logics.

We summarize the main notation[*] used throughout this thesis in Table 2.1.

2.2.3 Evaluation of Recommender Systems

The most commonly used evaluation method in the RSs literature is to hide some data from a particular dataset (e.g., interactions), train a model on the remaining data, and then predict the hidden data[†] (also referred to as ground truth) [216, 483]. Depending on the particular setting, the known and hidden data can be divided according to different ways [307]. For example, to evaluate the rating prediction problem, typically a certain subset of ratings are randomly assumed as known for training (e.g., considering 60% of the dataset ratings), another subset is used for validation (e.g., considering 20% of the dataset ratings) and then models are tested on the remaining ratings (*random split*). Similarly, to evaluate the top-k recommendation problem, many authors used to directly

[*]Note that we slightly overloaded the notation previously introduced in Section 2.1.1 for convenience.

[†]Note that an underlying problem concerning the evaluation of all recommender systems is that there may be items of interest to the user that he or she has not interacted with and hence are not present in ground truth.

partition the dataset users into training, validation and test set users, considering for each set all related user-item interactions* (*user split*).

A variety of metrics are used in the recommender systems literature for evaluation purposes. On the one hand, *accuracy metrics* are typically used to evaluate the rating prediction problem, e.g., to assess the ability of the algorithm to predict the hidden ratings. On the other hand, *relevance metrics* and *ranking metrics* are used to evaluate the top-k recommendation problem. The former are used to assess the algorithm's ability in generating recommendations that are relevant to users, i.e., that match the ground truth in terms of users' interests. The latter are used to assess the ability of the algorithm in generating recommendations that contain the most relevant items in the highest ranking positions. We briefly introduce some of these metrics below.

2.2.3.1 Accuracy Metrics

When an algorithm is evaluated in terms of how accurate it is at predicting hidden ratings, usually either the *Mean Average Error* (*MAE*) or the *Root Mean Square Error* (*RMSE*) is measured. Considering a certain user-item interaction pair (u, i), the metrics calculate a particular type of error between the predicted ratings $\hat{x}_{u,i}$ and the hidden ground truth $x_{u,i}$. For example, the *Mean Average Error* can be defined as:

$$MAE = \frac{1}{|\hat{\mathbf{X}}|} \sum_{\hat{x}_{u,i} \in \hat{\mathbf{X}}} \left| x_{u,i} - \hat{x}_{u,i} \right| \tag{2.4}$$

Similarly, the *Root Mean Square Error* can be defined as:

$$RMSE = \sqrt{\frac{1}{|\hat{\mathbf{X}}|} \sum_{\hat{x}_{u,i} \in \hat{\mathbf{X}}} \left(x_{u,i} - \hat{x}_{u,i} \right)^2} \tag{2.5}$$

Depending on the needs, one or the other metric can be adopted, as for example, *RMSE* unlike *MAE* highly penalizes large prediction errors.

2.2.3.2 Relevance Metrics

When the emphasis is on evaluating the algorithm's ability to generate recommendations that contain more relevant items, the RSs community tends to measure performance using *Hit-Rate* (*HR@k*), *Precision* (*Prec@k*), *Recall* (*Rec@k*) and *F1-Score* (*F1@k*). The metrics calculate how well an algorithm is able to generate recommendations that match with hidden ground truth user interests. Let $rel^{\mathbf{y}}_{u,j}$ be a relevance variable† [220] that indicates if the item recommended at position j in the ordered ranking $\mathbf{y}_{u,k}$ of k items is relevant or not for user u. Accordingly,

*Considering the user split, since as we will discuss in Section 2.2.4 the system often cannot generate recommendations for users without knowing anything about them, it is common to introduce into the training set a minimum number of known validation and test user interactions (e.g., three or four user-item interactions per user) to avoid what is called as the *user cold-start*.

†In an implicit feedback setting [368], each item's relevance corresponds to its ground truth information, e.g., assuming $x_{u,i} = 1$ if the user actually interacted with the item in the hidden data, and $x_{u,i} = 0$ if not.

for example, *Hit-Rate* at position k is defined as:

$$HR@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \begin{cases} 1 & \text{if} \quad \sum_{j=1}^{k} rel^{\mathbf{y}}_{u,j} \geq 1 \\ 0 & \text{otherwhise} \end{cases} \tag{2.6}$$

and represents the fraction of users for which the recommendations list contains at least one relevant item. Instead, *Precision* at position k is defined as:

$$Prec@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{j=1}^{k} rel^{\mathbf{y}}_{u,j}}{k} \tag{2.7}$$

thus representing number of relevant items in the top-k recommendations over the number of recommended ones. Similarly, *Recall* at position k can be defined as:

$$Rec@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{j=1}^{k} rel^{\mathbf{y}}_{u,j}}{\sum_{i=1}^{n} x_{u,i}} \tag{2.8}$$

hence representing the number of relevant items in the top-k recommendations over the total number of relevant ones in the ground truth. Precision and recall can also be combined into the *F1-Score* at position k as:

$$F1@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{2 \cdot Prec_u@k \cdot Rec_u@k}{Prec_u@k + Rec_u@k} \tag{2.9}$$

that represents their harmonic mean.

2.2.3.3 Ranking Metrics

When the emphasis is on evaluating the algorithm's ability to generate recommendations that contain more relevant items in the top-positions of the ranking, the RSs community tends to measure performance using *Mean Reciprocal Rank* (*MRR@k*) or *Normalized Discounted Cumulative Gain* (*NDCG@k*). In particular, *Mean Reciprocal Rank* at position k is the mean rank of the first relevant item in the recommendations list:

$$MRR@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{j^1_u} \tag{2.10}$$

where, in the equation, j^1_u is the rank (position) of the first item relevant to user u. Instead, *Normalized Discounted Cumulative Gain* at position k:

$$NDCG@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{j=1}^{k} \frac{rel^{\mathbf{y}}_{u,j}}{\log_2(j+1)}}{IDCG_u@k} \tag{2.11}$$

can be defined as an inverse log reward over all the ranking positions with relevant items among the top-k recommended ones. In the equation, $IDCG_u@k = \sum_{j=1}^{|\mathbf{r}_u|} \frac{rel^{\mathbf{r}}_{u,j}}{\log_2(j+1)}$ is usually referred to as the *Ideal Discounted*

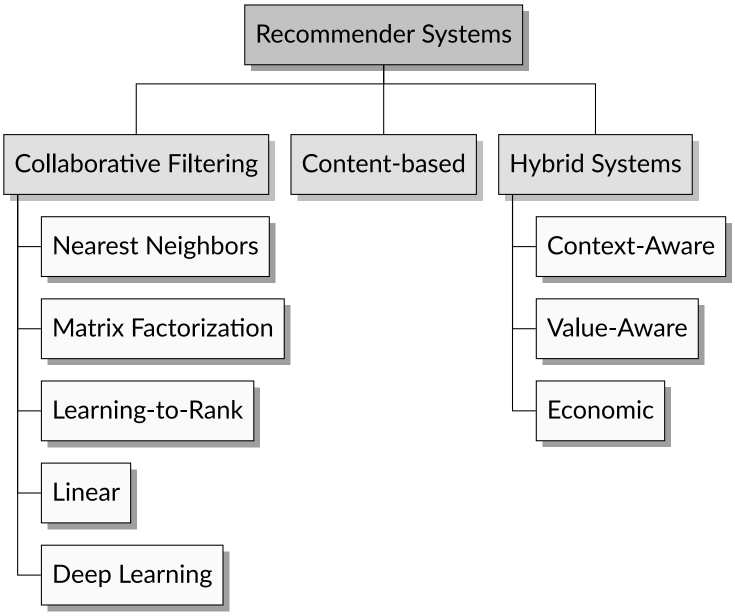


Figure 2.5: Taxonomy of recommender systems. *The taxonomy contains the recommendation algorithm families on which we will focus most throughout this thesis, see Section 2.2.5, and Part II.

Cumulative Gain obtained by sorting all the ground truth items relevant to the user belonging to the ordered list $\mathbf{r}_u$ in descending relevance order. Hence, *NDCG@k* measures how precise an RS algorithm is in recommending the most relevant items to each user in the highest ranking positions.

In the following sections we will focus mostly on ranking metrics because in practical circumstances user attention is often limited and interaction probability likely decreases according to the position of recommended items due to the well-known *position bias* phenomenon, i.e., the user is more likely to interact with items in the highest ranking positions. Specifically, the experiments that we present as part of our original contributions in Chapter 5 and Chapter 6, will primarily use *NDCG@k* to measure the ability of a recommender system to generate recommendations that match users' interests. Moreover, we will also consider another variation of this metric to measure business value related aspects that are more in line with the underlying goals of the value-aware recommender systems class which is the focus of this thesis.

2.2.4 Recommender Systems Taxonomy

Countless recommendation algorithms have been proposed since the early 1990's [372, 373]. Traditionally, they can be mainly divided into collaborative filtering algorithms, content-based approaches, and hybrid systems [9, 219]. *Collaborative filtering* (*CF*) [401] systems base their recommendations solely on knowledge about the past behavior of a community of users, e.g., their previous purchases on an online shop or the feedback they provide on items on a media streaming site (see Figure 2.5). CF systems are nowadays widely used in industry [22, 163] and they are the basis for our work on value-aware recommendation presented in this thesis.

Content-based (*CB*) systems [284] have their roots in *Information Retrieval* (*IR*) [42, 238] and rely on meta-information about the available items and the past content preferences of individual users. Such systems have their place in practice as well, in particular in cases when no large user community exists, which would be required to build a collaborative system. Since both CF and CB systems can have limitations (e.g., considering the well-known cold-start problem) [9], various forms of combining these approaches in a *hybrid system* (*HS*) were proposed in the past [64]. Today, the most common form of such hybrid systems are collaborative systems that use additional knowledge about the items as *side information* [404]. Value-aware recommender systems [106] typically fall into this category as well since they aim to leverage the power of collaborative filtering [401], but they additionally take information about the items' business value [214] into account when creating the final recommendation lists to be presented to users.

Each recommendation algorithms family can have different advantages and disadvantages. For example, collaborative filtering algorithms are the ones currently most widely used in industry settings but relying on customer history to generate recommendations can suffer from the well-known *cold-start* problem. That is, when a *new user* registers in the platform, recommendations cannot be generated because there is no user-item interaction information available. Similarly, when a *new item* is marketed by the company, it will be recommended very infrequently because no user will have interacted with it yet. In comparison, content-based algorithms would not suffer from the new item problem but the recommendations they generate tend to be often obvious and of little interest to users. Hybrid systems such as context-aware algorithms also tend to suffer less from the cold-start problem but they are more difficult to design and maintain over time. In addition, they often rely on user demographics that are not always available for privacy reasons. In the following we will focus mainly on collaborative filtering algorithms and the use of a particular context information concerning the economic value of recommending an item for the organization.

2.2.5 Collaborative Filtering Algorithms

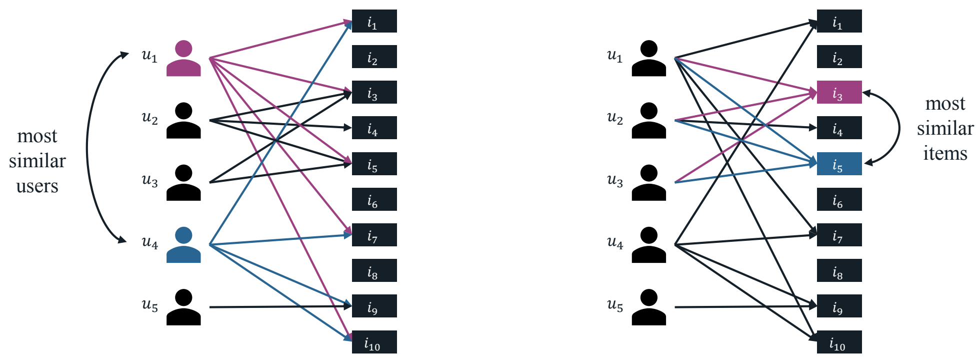
Within the family of collaborative filtering algorithms, the following main approaches can be identified:

- *Nearest neighbors* techniques were used in the earliest recommendation systems both in academia and industry in the 1990's [276, 326, 373]. While such approaches are conceptually simple, it turns out that they can often lead to competitive results in terms of prediction accuracy, at least for small datasets [26].

- *Matrix factorization* approaches were initially explored in the late 1990's [51], and they became the state-of-the-art [241, 242, 243] in the context of the well-known Netflix Prize competition [45]. Despite the recent wave of modern deep learning algorithms [190], matrix factorization methods are still relevant today [370] as they often lead to very good performance in pure collaborative filtering approaches [371].

- *Learning-to-rank* techniques like *Bayesian Personalized Ranking* [368, 369] became popular around the 2010's, when the community increasingly started to more directly target the top-*k* recommendation problem [111, 171, 483]. The goal of such approaches is not to predict the relevance to the user of each item on an absolute scale but rather to find an optimal ranking of the recommended items.

- *Linear* approaches became of interest to the community around 2012 when practical challenges concerning current large scale systems with millions of active users and catalog items gained importance. In particular, these approaches have been used in well-known algorithms such as *Sparse LInear Method* [328] to generate very fast high-quality recommendations.



(a) User-based collaborative filtering

(b) Item-based collaborative filtering

Figure 2.6: Examples of user-based and item-based collaborative filtering algorithms mechanisms.

- *Deep Learning* techniques for collaborative filtering based on *Multi-Layer Perceptron* [190] and *Variational Autoencoders* [270] dates back to 2007 [381] and 2018 [270], respectively. Today, these neural networks-based methods are considered state-of-the-art. One main advantage of such systems in practice is that various types of side information can be easily integrated into the networks [400].

In our thesis, we will exploit each of the above discussed types of collaborative filtering approaches for our original contributions (see Chapter 5 and Chapter 6), proposing adaptations to the original algorithms or integrating additional knowledge to the predictions to more directly target certain goals. In this section, we describe the underlying rationale of the baseline methods.

2.2.5.1 Nearest Neighbors

Collaborative filtering algorithms based on nearest neighbors techniques have been successfully applied in various application domains especially considering explicit feedback-based cases, e.g., where $x_{u,i}$ is a rating in the range $[1, 5]$. The underlying idea behind such algorithms is to provide recommendations to users based on the preferences and behaviors of other users who are similar to them. There are two main nearest neighbors variants that are well known in the literature, i.e., user-based and item-based, both of which are based on certain assumptions and exploit certain similarity criteria to make predictions (see Figure 2.6).

In particular, the *User-Based Collaborative Filtering* (*UCF*) [326, 373] assumes that similar users tend to give similar ratings to the same item. The algorithm consists of the following steps. At first, the similarity $sim(u, v)$ between each pair of users (u, v) is calculated exploiting the user-item interaction matrix. Typically, Pearson's correlation coefficient is used to calculate such similarity:

$$sim(u, v) = \frac{\sum_{i \in \mathcal{I}_u^+ \cap \mathcal{I}_v^+} (x_{u,i} - \bar{x}_u) \cdot (x_{v,i} - \bar{x}_v)}{\sqrt{\sum_{i \in \mathcal{I}_u^+ \cap \mathcal{I}_v^+} (x_{u,i} - \bar{x}_u)^2} \cdot \sqrt{\sum_{i \in \mathcal{I}_u^+ \cap \mathcal{I}_v^+} (x_{v,i} - \bar{x}_v)^2}} \tag{2.12}$$

where $\mathcal{I}_u^+ \cap \mathcal{I}_v^+$ represents the set of items that were mutually rated by u and v, and $\bar{x}_u = \frac{\sum_{i \in \mathcal{I}_u} x_{u,i}}{|\mathcal{I}_u|}$ and $\bar{x}_v$ denote the average rating of those users. Next, to compute the predicted scores $\hat{x}_{u,j}$ of an item j that user u has never interacted with, the algorithm identifies a certain neighborhood $\mathcal{P}(u,j)$ of users most similar to the target user who rated such an item (where l is an algorithm parameter) and calculates a weighted sum of similarities:

$$\hat{x}_{u,j} = \bar{x}_u + \frac{\sum_{v \in \mathcal{P}(u,j)} sim(u,v) \cdot (x_{v,j} - \bar{x}_u)}{\sum_{v \in \mathcal{P}(u,j)} |sim(u,v)|} \tag{2.13}$$

Since each user can rate items in a subjective way, thus having a different scale, the individual user's average rating $\bar{x}_u$ is added to the weighted similarity computation as a heuristic criterion to more properly center predictions*.

A well-known variant of the user-based algorithm is the *Item-Based Collaborative Filtering* (*ICF*) which is based on the assumption that users tend to give similar ratings to similar items. The approach is similar to the previous one with the difference that similarity is calculated among items and not among users, thus considering the columns and not the rows of the user-item interaction matrix. As earlier, first the similarity $sim(i,j)$ between each items pair (i,j) is calculated. In this case, similarity is often calculated using Cosine similarity instead of Pearson's correlation coefficient:

$$sim(u,v) = \frac{\sum_{u \in \mathcal{U}_i \cap \mathcal{U}_j} (x_{u,i} - \bar{x}_i) \cdot (x_{u,j} - \bar{x}_j)}{\sqrt{\sum_{u \in \mathcal{U}_i \cap \mathcal{U}_j} (x_{u,i} - \bar{x}_i)^2} \cdot \sqrt{\sum_{u \in \mathcal{U}_i \cap \mathcal{U}_j} (x_{u,j} - \bar{x}_j)^2}} \tag{2.14}$$

where $\mathcal{U}_i \cap \mathcal{U}_j$ represents the set of users who mutually rated item i and item j, and $\bar{x}_i = \frac{\sum_{u \in \mathcal{U}_i} x_{u,i}}{|\mathcal{U}_i|}$ and $\bar{x}_j$ denote the average rating of those items. Accordingly, the algorithm identifies a certain neighborhood $\mathcal{Q}(u,j)$ of l items rated by u most similar to item j and calculates the predicted scores based on a weighted sum of similarities:

$$\hat{x}_{u,j} = \frac{\sum_{i \in \mathcal{Q}(u,j)} sim(i,j) \cdot x_{u,j}}{\sum_{i \in \mathcal{Q}(u,j)} |sim(i,j)|} \tag{2.15}$$

However, unlike the user-based variant, in this case the algorithm exploits the *own* user ratings to generate predictions. Therefore, although it depends highly on the underlying data, the item-based algorithm tends to provide more effective predictions in practice compared to the user-based variant since it does not require any heuristic criteria to center predictions around the user's mean ratings.

Overall, the approach underlying nearest neighbors collaborative filtering systems is simple and intuitive, and the algorithms are easy to implement and debug. In addition, the recommendations provided are easily explainable in that they can be interpreted, for example, as generated based on the interests of the users most similar to the target user when considering the user-based variant. It is also possible to create approximate versions of the approaches to enable incremental training, thus avoiding to restart training from scratch every time a new feedback needs to be incorporated. However, the main drawback is that the algorithms are often impractical in large-scale settings in terms of space and computational time required during the training phase. Moreover, often the recommendations provided may have limited coverage due to sparsity reasons, e.g., considering the user-based algorithm

*In implicit feedback settings typically authors used to ignore the rating values and avoid prediction centering, i.e., considering the user-based algorithm, $\hat{x}_{u,j} = \sum_{v \in \mathcal{P}(u,j)} sim(u,v)$. On the contrary, considering each user-item interaction $x_{u,i} \in \{0,1\}$, the nominator in Eq. (2.13) would always be equal to zero since both the user-item interaction $x_{v,j}$ of neighbor v and the average rating $\bar{x}_u$ of user u would be equal to one.

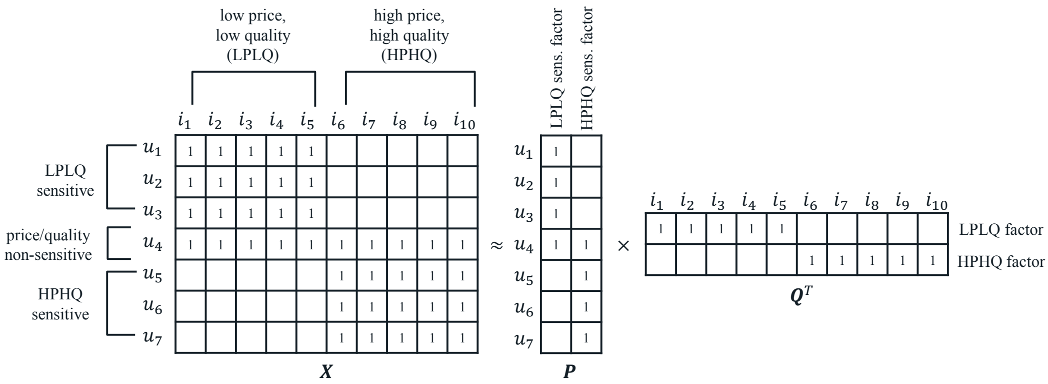


Figure 2.7: Example of user-item interaction matrix factorization into $d = 2$ latent factors representing users' sensitivity to products with high or low price and quality.

it is not possible to recommend a certain item to a given user if at least one of the users similar to the target user has not interacted with that item. Furthermore, sparsity can also create issues when calculating similarity, e.g., when only a few items have been mutually rated by users the similarity computation may become unreliable.

2.2.5.2 Matrix Factorization

Various *Matrix Factorization* (*MF*) [241, 242, 243] methods have been successfully used in the RSs literature to obtain better recommendations than neighborhood-based methods in terms of quality and efficiency. In particular, the pairwise similarity estimation adopted by nearest neighbors methods tends not to be robust in highly sparse datasets because in such cases users have mutually rated only a few items. Correspondingly, algorithm predictions of user interests tend to be inaccurate. Matrix factorization methods, by contrast, would provide a more efficient and reliable calculation. The basic idea of these methods is to project the user-item interaction matrix into a lower dimensional space by exploiting *dimensionality reduction* algorithms. In this space, unlike the user-item interaction matrix, the representation of users and items is dense and can be used to directly calculate user interests more precisely.

In particular, if all entries of the user-item interaction matrix are observed, it is possible to approximately factorize $\mathbf{X} \in \mathbb{R}^{m \times n}$ into two matrices $\mathbf{P} \in \mathbb{R}^{m \times d}$ and $\mathbf{Q} \in \mathbb{R}^{d \times n}$ of lower dimensions (where d is the embeddings' size) such that:

$$\mathbf{X} \approx \mathbf{P}\mathbf{Q}^{\mathsf{T}} \tag{2.16}$$

where $\mathbf{p}_u \in \mathbb{R}^d$ and $\mathbf{q}_i \in \mathbb{R}^d$, respectively u-th and i-th row of $\mathbf{P}$ and $\mathbf{Q}$, are the d-dimensional dense representations of user u and item i, also known as *embeddings* or *latent factors*. It thus follows that the preference of a user u for an item i can be estimated:

$$\hat{x}_{u,i} = \mathbf{p}_u \cdot \mathbf{q}_i \tag{2.17}$$

through the dot product of such d-dimensional embeddings. Conceptually, as can be intuitively inferred by observing Figure 2.7, these embeddings represent the affinity of a certain user or item toward some higher-level abstractions, e.g., users' sensitivity to products of high or low price and quality.

The possibility of factorizing $\mathbf{X}$ is a property of linear algebra, and there are various methods that can be used in the literature to obtain an exact solution. One of such methods for example is Singular Value Decomposition (*SVD*) which requires the vectors composing $\mathbf{P}$ and $\mathbf{Q}$ to be orthogonal to each other. However, in order to be performed, these algebraic methods require that the matrix $\mathbf{X}$ be fully observed. In certain cases, if the matrix is not very sparse, it is possible to fill in the missing entries with the average of the ratings so that the algebraic factorization can be directly performed. However, in most practical cases where the sparsity is very high this would strongly compromise the reliability of the representations and consequently also the accuracy of the final predictions. Hence, the vast majority of methods proposed in the literature exploit the well-known *Stochastic Gradient Descent* (*SGD*) method to learn users and items representations. In particular, considering that the error of approximately factoring $\mathbf{X}$ into $\mathbf{P}\mathbf{Q}^\intercal$ corresponds to $\|\mathbf{X} - \mathbf{P}\mathbf{Q}^\intercal\|^2$, i.e., the sum of the squares of the entries of the matrix $(\mathbf{X} - \mathbf{P}\mathbf{Q}^\intercal)$, it is possible to define a *squared loss* function* to be optimized to obtain an approximate representation of users and items:

$$\mathcal{L} = -\sum_{(u,i)\in\mathcal{D}} (x_{u,i} - \mathbf{p}_u \cdot \mathbf{q}_i)^2 = -\sum_{(u,i)\in\mathcal{D}} (x_{u,i} - \hat{x}_{u,i})^2 \tag{2.18}$$

where $\mathcal{D} = \{(u,i) : i \in \mathcal{I}_u^+\}$ is the set of *known* interactions for each user. Typically to obtain this approximate representation, SGD proceeds by running several iterations until it reaches convergence, calculating the loss function and updating accordingly the embeddings $\mathbf{p}_u$ and $\mathbf{q}_i$ for every user and item at each iteration.

Besides the basic algorithm described above, currently matrix factorization is a family of state-of-the-art algorithms that includes several variants. For example, implicit and explicit feedback can be handled simultaneously to generate more accurate predictions, and approximate variants for incremental learning are also possible. Overall, the main advantages of matrix factorization and more generally of any model-based recommendation algorithm over nearest neighbors methods can be attributed to: (i) better space efficiency since there is no need to save the similarity matrix for predictions; (ii) faster training and inference since, for example, latent factors allow predictions to be calculated by simple multiplication; (iii) improved handling of the overfitting issue because additional regularizers can be used in the objective function to increase generalization. Instead, the main drawback is related to the explainability of the recommendations. Indeed, while the predictions of nearest neighbors methods can be easily interpreted based on user/item similarity criteria, matrix factorization predictions, depending on a model, are not so intuitive to explain. However, literature variants are also available that at the expense of a slight loss of accuracy allow for better explainability.

2.2.5.3 Bayesian Personalized Ranking

Sometimes the sparsity of the user-item interaction matrix is so high that the effectiveness of MF is compromised. The main problem is that matrix factorization was originally designed to directly address the rating prediction problem. Accordingly, MF's primary goal is to predict hidden user interests. These predictions are then used to

*In implicit feedback settings, a *binary cross-entropy loss* $\mathcal{L} = -\sum_{(u,i)\in\mathcal{D}} x_{u,i} \log \hat{x}_{u,i} + (1 - x_{u,i}) \log(1 - \hat{x}_{u,i})$ is typically used in the RSs literature as a more effective alternative of the squared loss function.

generate recommendations by selecting the top-k items with highest predicted score. However, in practical cases where sparsity is very high, it is very difficult to obtain a sufficiently accurate prediction of all user interests to be used to generate good quality recommendations. In these contexts, it is much more important to rely on a relative order among user interests that would allow the most relevant items to be recommended to the user.

To address the high sparsity problem and more directly target the top-k recommendation problem, an algorithm referred to as *Bayesian Personalized Ranking* (*BPR*) [368, 369] was proposed in the literature. This algorithm is typically applied on-top of matrix factorization and is used exclusively in implicit feedback settings (recall Section 2.2). Specifically, the MF objective function is replaced with a *pairwise loss* function that approximates the well-known *Area Under the ROC Curve* (*AUC*) ranking statistic:

$$\mathcal{L} = - \sum_{(u,i,j) \in \mathcal{D}} \ln \sigma(\hat{x}_{u,i} - \hat{x}_{u,j}) \tag{2.19}$$

where $\sigma(\cdot)$ is a sigmoid function and $\mathcal{D} = \{(u, i, j) : i \in \mathcal{I}_u^+ \wedge j \in \mathcal{I}_u^-\}$ is a set of pairwise training examples consisting of pairs of positive and negative items for each user. The term $\sigma(\hat{x}_{u,i} - \hat{x}_{u,j})$ represents the probability that a user u prefers an item i over an item j. Since matrix factorization is the main underlying model of BPR, $\hat{x}_{u,i}$ and $\hat{x}_{u,j}$ generally represent the predicted scores for a positive item $i \in \mathcal{I}_u^+$ and a negative item $j \in \mathcal{I}_u^-$. By minimizing the loss function $\mathcal{L}$, the score of positive items becomes higher than the score of negative ones. This way the algorithm can be trained to identify a relative ordering of user interests, where the most relevant items will thus have higher predicted scores. Then such relative order among user interests can be used to generate the top-k recommendations.

Overall when compared with MF, BPR allows for higher quality recommendations because the algorithm is designed to directly address the top-k recommendation problem which unlike the rating prediction problem is more important in practical settings. However, in terms of computational efficiency of training and prediction, both methods tend to suffer when the number of registered users in the platform and the number of items available to recommend are very high.

2.2.5.4 Sparse LInear Method

Although MF and BPR may enable very accurate predictions, often in large-scale systems with millions of active users and catalog items and high sparsity training and inference may be slow and not feasible in practice. To handle the typical sparsity of the user-item interaction matrix more efficiently, a linear method referred to as *Sparse LInear Method* (*SLIM*) [328] based on the well-known *elastic net* regularization algorithm has been studied in the literature. From a practical standpoint, this method is most appropriate for use in implicit feedback settings (assuming missing entries equal to zero) although it can also be adapted to handle explicit feedback. Indeed, in these settings, regularization would make it easier to shrink the regularization parameters to zero and to obtain predictions much faster than previous latent factors based methods.

Specifically, SLIM calculates the predicted score of an item i with which user u has not yet interacted as:

$$\hat{x}_{u,i} = \mathbf{x}_u^\intercal \mathbf{w}_i \tag{2.20}$$

thus multiplying the size-n row vector of user interactions $\mathbf{x}_u$ by a size-n column vector of coefficients $\mathbf{w}_i$, which contains a coefficient for each item. Overall, the model used by SLIM to generate the predictions can be presented as $\hat{\mathbf{X}} = \mathbf{XW}$ where $\mathbf{W}$ is a matrix of sparse coefficients $n \times n$. Considering the error $\hat{\mathbf{X}} - \mathbf{X}$ that is incurred by approximating $\mathbf{X}$ as $\mathbf{XW}$, the coefficient matrix can then be learned by minimizing the following regularized optimization problem:

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{X} - \mathbf{XW}\|_F^2 + \frac{\chi}{2}\|\mathbf{W}\|_F^2 + \omega\|\mathbf{W}\|_1 \\ \text{subject to} \quad & \mathbf{W} \geq 0, \quad diag(\mathbf{W}) = 0 \end{aligned} \tag{2.21}$$

where χ and ω are regularization parameters and $\|\mathbf{W}\|_1$ and $\|\mathbf{W}\|_F$ are the L_1-norm and the Frobenius L_F-norm of the matrix of coefficients $\mathbf{W}$. The two constraints $\mathbf{W} \geq 0$ and $diag(\mathbf{W}) = 0$ are set respectively to learn only positive relationships between items and to avoid trivial solutions, i.e., where $\mathbf{W}$ is an identity matrix and $\|\mathbf{X} - \mathbf{XW}\|_F^2$ is always equal to zero.

Overall, SLIM when compared to MF and BPR allows for predictions of comparable quality but much faster. Specifically from the point of predictions, it is not necessary to calculate the predicted scores of all items to obtain a user's recommendation because the regularization shrinks many coefficients to zero. Moreover, since the columns of $\mathbf{W}$ are independent, the training of the algorithm can be easily parallelized and is also more efficient than that of MF and BPR.

2.2.5.5 Neural Collaborative Filtering

With the increased adoption of online platforms (e.g., e-commerces, multimedia streaming) and more generally with the growth of data volume, various deep learning algorithms have been proposed in the literature with the aim of more effectively exploiting such information. In particular, although matrix factorization is widely used in industry, the algorithm is linear in that it is based on a fixed inner-product of the latent factors of items and users. Hence, especially in large scale systems, the underlying linearity of MF may not allow the algorithm to capture some complex user-item relationships, thus affecting the quality of the predictions. One way to increase the predictive power of MF might be to increase the size of embeddings. However, especially in highly sparse settings this could lead to the well-known problem of overfitting. To increase the generalization of matrix factorization in the literature, an algorithm referred to as *Neural Collaborative Filtering* (*NCF*) [190] has been proposed. This algorithm conceptually exploits the same basic MF idea of predicting user interests from a dense representation of users and items. However, NCF, unlike MF, by mounting on top of the dense representation a neural architecture based on *Multi-Layer Perceptron* (*MLP*), might also be able to generalize non-linear data relationships, thereby obtaining potentially more accurate predictions.

The neural architecture of NCF is shown in Figure 2.8. The input layer consists of a sparse representation of a particular user $\mathbf{x}_u$ and a certain item $\mathbf{x}_i$, obtained from the user-item interaction matrix $\mathbf{X}$. Above the input layer, a fully connected layer projects this sparse representation into a dense representation of users and items. This dense representation is analogous to the latent factor representation based on embeddings $\mathbf{p}_u$ and $\mathbf{q}_i$ of matrix factorization, albeit in this case the embeddings dimensions are $m \times d$ and $n \times d$, respectively - as in MF, d remains a configurable embeddings' size parameter of the algorithm. Both embeddings are fed into a multi-layer perceptron neural architecture. Typically, a three-layer structure is used whose size depends on the size of the embeddings, i.e., where the layers have the following sizes $\{2 \cdot d, d, \frac{d}{2}\}$. The model is then trained by optimizing the same loss function of matrix factorization through stochastic gradient descend to predict the user's interest in

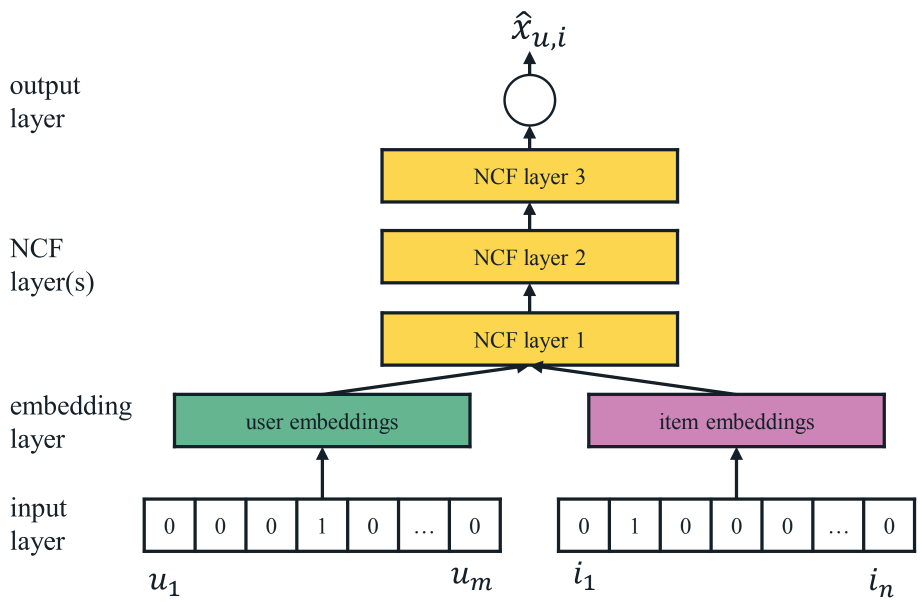


Figure 2.8: Neural collaborative filtering architecture.

a certain item:

$$\hat{x}_{u,i} = \sigma(\mathbf{p}_u^\mathsf{T}\mathbf{x}_u, \mathbf{q}_i^\mathsf{T}\mathbf{x}_i|\mathbf{p}_u, \mathbf{q}_i, \boldsymbol{\Theta}) \tag{2.22}$$

where $\boldsymbol{\Theta}$ denotes the overall set of parameters of the neural model and $\sigma(\cdot)$ the inference function.

Overall, although MF continues to demonstrate state-of-the-art performance in medium-sized datasets, NCF has shown encouraging results in large scale settings. In addition, the neural model, unlike MF would allow to exploit additional context features (if also managed as embeddings and connected to the MLP architecture) to further increase the relevance of predictions.

2.2.5.6 Multinomial Likelihood Variational Autoencoder

An alternative solution to MF for large scale recommendation settings based on a deep learning architecture has recently emerged. In particular, *Multinomial Likelihood Variational Autoencoder* (*Mult-VAE*) [270]is a generative algorithm based on variational autoencoders that has been proposed in the literature for collaborative filtering systems with implicit feedback data. Similar to NCF, this neural network-based model would allow greater generalization when compared to latent factor models such as MF, because of its ability to model complex nonlinear relationships, thus obtaining potentially more accurate predictions.

Mult-VAE is based on variational autoencoders, which unlike traditional autoencoders, is designed to reconstruct input data using a probabilistic inferential in-the-middle model that approximates the underlying variational distribution. The algorithm is based on the following generative process. Initially the model samples for each user u a latent d-dimensional representation $\mathbf{p}_u$ from a standard Gaussian prior. Then, $\mathbf{p}_u$ is transformed into a probability distribution $f(\mathbf{p}_u)$ over n items through a non-linear multilayer perceptron with parameters $\boldsymbol{\Theta}$

and softmax activation. The process assumes that the vector $\mathbf{x}_u$ of user interests is drawn from a multinomial distribution $Mult(|\mathbf{x}_u \neq 0|, f(\mathbf{p}_u))$ with $|\mathbf{x}_u \neq 0|$ sum of the number of interactions for the user. To determine $\mathbf{\Theta}$, the model has to estimate the posterior distribution $g(\mathbf{p}_u|\mathbf{x}_u)$. However, $g(\mathbf{p}_u|\mathbf{x}_u)$ is intractable. Therefore, such distribution is approximated by a diagonal Gaussian distribution $h(\mathbf{p}_u) = \mathcal{N}(\mu_u, diag(\sigma_u^2))$ with $\{\mu_u, \sigma_u^2\}$ free variational parameters such that the Kullback-Leiber divergence $kl(h(\mathbf{p}_u) \mid\mid g(\mathbf{p}_u|\mathbf{x}_u))$ between the two distributions $h(\mathbf{p}_u)$ and $g(\mathbf{p}_u|\mathbf{x}_u)$ is minimized. However, the number of parameters $\{\mu_u, \sigma_u^2\}$ to optimize grows with the number of users and items in the dataset and can become a bottleneck in real-world applications. Thus, the variational autoencoder replaces the variational parameters of the function $h(\mathbf{p}_u)$ by turning it into a function $h_{\mathbf{\Phi}}(\mathbf{p}_u|\mathbf{x}_u) = \mathcal{N}(\mu_{\mathbf{\Phi}}(\mathbf{x}_u), diag(\sigma^2_{\mathbf{\Phi}}(\mathbf{x}_u)))$ parameterized by $\mathbf{\Phi}$ that, if optimized, approximates the intractable posterior $g(\mathbf{p}_u|\mathbf{x}_u)$. The model attempts to minimize through SGD the distance between the distributions by optimizing an evidence lower bound:

$$\mathcal{L} = \mathbb{E}_{h_{\mathbf{\Phi}}(\mathbf{p}_u|\mathbf{x}_u)}[\log g_{\mathbf{\Theta}}(\mathbf{x}_u|\mathbf{p}_u)] - \psi \cdot kl(h_{\mathbf{\Phi}}(\mathbf{p}_u|\mathbf{x}_u) \mid\mid g(\mathbf{p}_u)) \tag{2.23}$$

that is interpreted as composed of a first reconstruction error and a second regularization term, where $\psi \in [0, 1]$ is a regularization variable.

Overall, Mult-VAE proved to be an alternative solution for large scale recommendation settings showing superior results to SLIM, MF and NCF in various well-known datasets.

2.2.5.7 Adapting Recommendation Algorithms

Various adaptations of the recommendation algorithms described in the previous sections have been proposed in the literature. Often these adaptations depend on the particular goal to be achieved through the recommendation [164, 213]. For example, under certain circumstances a recommendation provider may desire to convey certain business strategies through the platform in order to optimize certain KPIs [155, 214, 419]. In other cases, it may be necessary to avoid potential discrimination against certain user or item groups [303, 399, 466].

Most of the adaptations proposed in the literature tend to fall into three main families of methodologies, i.e., *pre-processing*, *in-processing* and *post-processing* [106, 467]. The use of a given methodology involves the adaptation of a specific stage of the learning process of a recommendation algorithm. Pre-processing methodologies focus on the first stage of model training, as the algorithms operate directly on the input data [255, 397]. Indeed, by changing the input data with sampling, perturbation or feature engineering operations, it is possible to modify accordingly the internal representations on which the algorithms rely to make the predictions. In-processing methodologies, conversely, mainly operate during the training phase [66, 69, 436, 462]. In particular, it has been proposed in many works in the literature to introduce specific regularizers into the algorithms' objective function to target certain domain-specific goals. Finally, post-processing methodologies are typically applied after model training, as they operate directly on the output predictions [240, 291, 395, 458]. For example, many papers in the literature rely on re-ranking techniques to adjust the order of recommended items according to certain objectives.

In-processing and post-processing methodologies are the basis of the original contributions of this thesis. In particular, we proposed three different post-processing algorithms in Chapter 6 to target specific objectives. In addition, we proposed to integrate the objective function of four different families of recommendation algorithms in Chapter 5 in order to optimize certain metrics of interest.

2.3 Purpose-Driven Recommender Systems Design

In the previous section, we focused on collaborative filtering-based recommender systems designed to recommend the top-k items most relevant to users. However, recommender systems, as mentioned in Section 2.2.2, can be designed to serve both user [58, 84, 195, 209, 389] and organizational purposes [97, 164, 213, 295, 341, 342, 349, 386, 421, 454]. In the following we mainly focus on a subset of such systems that more directly target organizational interests with the aim of optimizing the *economic value* of recommendations.

2.3.1 An Economic Perspective on the Meaning of Value

From early academic definitions in the mid-1950s, the term *value* had multiple meanings. These meanings depend mainly on the particular interpretation in the literature. There are mainly two interpretations, a more quantitative one from the economics literature (e.g., related to profits and benefits), and a more recent one involving also emotional dynamics from the marketing literature (e.g., related to the subjective customer's perception).

In particular, Miles [309] in earlier economic studies denoted the value of a product or service into four main components, i.e., *cost value*, *use value*, *exchange value*, and *estimated value*. As reported in the author's research: (i) cost value is referred to as the cost for the company to produce a certain product or to provide a specific service, e.g., the cost to produce every component and assemble a mobile phone; (ii) use value relies more on functionalities, e.g., considering the previous mobile phone, its use value may be the ability to make a phone call; (iii) exchange value considers a possible value increase over time, e.g., when the mobile phone after ten years may be worth more as a vintage item; (iv) estimated value also concerns the attractiveness of an item to customers, e.g., a mobile phone with a colour display may be more attractive than one with a black-and-white display. Subsequently, the term value was redefined in later economic studies according to a utilitarian interpretation [24, 25]. Accordingly, value is defined as the expected benefit that a buyer receives as a function of the price paid, e.g., if the purchase of a product or service generates certain savings, its value lies in the difference between the savings and the price paid. Finally, especially in recent marketing studies [245, 277, 319, 469], this rational and utilitarian conception tends to be complemented by a more emotional and subjective dimension related to the customer's perception, e.g., for two distinct customers, the same product or service might have a different value depending on the perceived benefits, emotions and feelings it generates.

2.3.2 On the Business Value of Recommendations

Even if the meaning of the term value may sometimes appear blurry or subjective depending on the particular definition, the sale of a more valuable product or service can often be linked to the economic results of the company [48, 67, 123]. In particular, such value improvement can be measured according to various business *Key Performance Indicators* (*KPIs*). These KPIs in the RSs literature typically include [216]:

- The number of *user clicks* on the recommendations, often measured by the *click-through rate* (*CTR*);
- The degree of *user adoption* of the system, often measured by the *conversion rate* (*CVR*);
- The overall *revenue* generated from the *sales* of the firm's products and services;

- The possible effects on the *sales distribution* of the items sold, e.g., shifting toward more profitable items;
- The overall degree of *user engagement* with the platform, as indicator of *customer satisfaction*.

The effectiveness of recommendations for organizations [9, 22] depends mainly on the specific business value categories that the systems are able to optimize [21, 163]. The type of business value optimized may depend on various variables [216], often related to particular *business strategies* that the company may desire to implement [85, 194, 198, 372]. In some cases, it may be advantageous for the organization to maximize the conversion rate of recommendations [165, 181], e.g., to increase the number of consumers in the platform. In other circumstances, it may be helpful to optimize for user engagement [163, 295], e.g., to retain the acquired consumers and guarantee stable cash flow levels. Moreover, most of the times the particular business strategy implemented depend on the *revenue model* of the company (e.g., transaction-based, advertising, subscription) [85, 194, 198, 263, 304, 362, 372]. For example, in the case where the revenue model is primarily transaction-based (e.g., Walmart), since there is a direct link between purchases and revenue, the company might be interested in shifting the customer behavior towards the purchase of the more profitable items [343]. In contrast, in case the organization's revenue model is based on ads (e.g., YouTube), the company may be interested in increasing the number of clicks [104] as this is directly related to the consumption of ads that providers pay to see their brand advertised. Finally, a company might also be interested in optimizing user engagement [163] in the case of subscription-based models (e.g., Netflix) as this positively correlates with retention.

2.3.3 An Introduction to Value-Aware Recommender Systems

Research on RSs traditionally focused on users [375]. Indeed, if recommendations were not able to meet user needs, they would not be as successful. However, in real-world circumstances, in addition to suggesting items of interest, the reasons a service provider may want to implement a recommendation system may vary and typically are related to the increase in certain business KPIs [216, 218]. In particular, there are various ways in which an RS can generate value for a business [9, 21, 22, 43, 163, 385], considering economics and marketing aspects [184, 250, 334, 366, 412]. One of such ways is to design an RS to optimize one or more of the business value categories mentioned in the previous section (e.g., CTR, CVR, and others). In this thesis we focus on such recommendation approaches that are referred in the literature as *Value-Aware Recommender Systems* (*VARSs*) [22, 65] and can be informally characterized as follows:

> *A Value-Aware Recommender System (VARS) is an RS that is designed to optimize the business value of recommendations.*

The first studies in the VARSs field [316] dates back to 2007. However, the first explicit reference to the term *value-aware* is found in the work of Amatriain and Basilico [22] where the authors denoted VARSs as future research direction for industrial applications. The importance of VARSs research was subsequently brought to the attention of the academic community in the *Workshop on Value-Aware and Multi-Stakeholder Recommendation* (*VAMS 2017*) [65]. After VAMS 2017, there has been an increase in the number of specialised articles on VARSs, proposing novel algorithms to address industry needs. However, most research remains scattered and composed of many application-specific approaches proposed in isolated contexts (e.g., in insurance or taxi drivers domains).

A first attempt to formalize this broad and highly important branch of research is proposed as an original contribution of this thesis (Chapter 3). In particular, we will characterize various approaches in the literature that can be used to build VARSs, which we will divide according to the particular algorithmic method used. Some approaches which we will call *post-processing*, typically exploit certain heuristic criteria to re-rank the recommendations of an existing RS with the aim to optimize a set of business KPIs. Instead, other algorithms which we will call *in-processing*, aim to design an RS so that it learns to optimize the business value directly at training time thereby gaining computational advantages at prediction time. Moreover, we will also explore the main commercial applications, as well as the open challenges and the future research directions of such value-aware recommendation algorithms.

2.3.4 From Value-Aware to Economic Recommendation

While there are various ways in which an RS can create value for users and providers, and while there are several KPIs that firms might seek to optimize, ultimately, the provision of a recommendation service almost always serves some economic goal of the organization such as profit and growth (i.e., long-term profit). However, we note that some forms of business value improvements are more directly targeting profitability aspects than others. An increase in revenue through recommendations or a shift in the sales distribution toward the most lucrative items is almost *directly* reflected in a profit improvement [82, 200, 353]. Instead, a growth in user engagement, as in the case of Netflix [163], with more customers joining and fewer leaving, is sometimes only *indirectly* reflected in higher long-term profits for the organization.

As another original contribution of this thesis (Chapter 4), we will focus on the first type of the described recommendation approaches that can be seen as a subset of VARSs, i.e., those that target profit effects in a more direct way. Typical examples in this context are: RSs that consider company profit and customer relevance in a balanced way [69, 82, 96, 240, 320]; systems that leverage discounts and pricing algorithms to trigger purchases [8, 217, 223, 225, 481]; or methods that consider customers' price sensitivity to recommend items more in line with their price preferences [78, 153, 476, 489, 490]. We call such systems *Economic Recommender Systems* (*ECRSs*), and we informally characterize them as follows:

> *An Economic Recommender System (ECRS) is an RS that exploits price and profit information and related concepts from marketing and economics to directly optimize an organization's profitability.*

Since most recommender systems may at least indirectly target some profit-related or growth-related goal, the boundaries between an ECRS, a VARS and a "traditional" RS may sometimes appear blurry. However, a clear distinction can often be made depending on the underlying revenue model of the company [85, 198, 372]. For example, click-through rate maximization may be seen as a VARS method in case it is only about increasing site interactions [176, 449]. However, it may also be considered as an ECRS method in case there is some revenue associated with each click event (e.g., commissions suppliers pay to marketplaces for each generated impression), as in the case the company is based on an advertising revenue model [291, 411, 480].

Later in this work, we will identify five key approaches from the literature to build ECRSs, which we divide into customer and organization-oriented ones, depending on the focus of the underlying algorithms. Customer-oriented approaches in the literature, for instance, integrate purchasing behavior mechanisms (e.g., price sensitivity) into the models to generate more relevant recommendations that will automatically lead to more profit. Organization-oriented ones, on the other hand, apply particular organizational strategies (e.g., profit awareness, promotional pricing) to optimize profit. Moreover, we will also delve into the evaluation methodologies that can be used to assess the quality of such economic recommendation approaches.

2.3.5 Behavioral Harms of Value-Aware Recommendations

It is important to note that considering certain types of economic information to an inappropriate extent may also lead to *unintended negative effects* and *behavioral harms* of value-aware recommendations [13, 121, 188]. Specifically, it is vital to ensure that an VARS does not negatively impact the user's trust [272] in the organization [39, 155, 200, 343]. Indeed, trust is one of the most important factors driving adoption [44, 239] and purchase intention [327, 350]. Recommendations that are irrelevant [76, 324, 433, 473], manipulative [11, 13, 101, 169, 254, 417, 455], or poorly explainable [99, 443, 472] because they are too biased towards the profitable items [442] can harm trust, leading customers to reactance [137, 461] or churning.

Besides trust, there are also other possible harms that may emerge in case the recommendation strategy is oriented too strongly toward business value. While algorithms are often designed to improve sales diversity [10, 252] or to stimulate the sales or consumption of niche items [297, 464], they in practice might sometimes nudge users to buy the most popular ones [138, 139, 140, 201, 257, 258]. Such effects may in turn have business value implications considering that popular items sometimes have lower margins [147]. Finally, competition effects [156, 264, 493] may also be important to consider, since rewarding higher-margin items could push sellers to increase prices [491], thus impacting customers' willingness-to-pay [12], and market demand [29, 475].

Various approaches have been proposed in the recommender systems literature to increase the effectiveness of the systems by considering how users typically behave when they receive recommendations. Although most of these approaches are not value-aware because the underlying algorithms are not designed to maximize business value, inspiration can be taken from such methods to mitigate the potential behavioral harms of value-aware recommendations. Below we offer a concise overview of such approaches that considers especially *diversity* and *fairness* aspects.

2.3.5.1 Diversity Aspects of Recommendations

Considering the diversity of recommendations is very important when designing a value-aware recommender system since the long-tail items usually have the highest profit margins. A very large number of approaches have been proposed in the literature to optimize the diversity* of recommendations [59, 95, 252, 422, 424]. Indeed, besides the previous business value implications, it has been shown that more diverse recommendations can be used to address the well-known over-fitting problem [208, 414] and are associated with higher levels of user satisfaction [126, 149]. With a more diverse recommendation, the user is less likely to interact with obvious items that might bore him or her and is more likely to interact instead with items he or she did not know before that might surprise him or her (and potentially triggering a purchase).

Similarly to the value-aware methods that we will describe in Chapter 3, most of the algorithms proposed in the literature to increase the diversity levels of recommendations relied on post-processing approaches [1, 10, 57,

*Diversity in the RSs literature has been defined in various ways [252]. The main definition of diversity [59] is based on the concept of dissimilarity between pairs of items in the result set. According to this definition, a rank that minimizes the similarity of items recommended to a user achieves a higher level of diversity.

61, 92, 197, 252, 359, 440, 498]. Such approaches typically perform re-ranking operations to increase the diversity of recommendations, while seeking to maintain high relevance by exploiting hyperparameters. Other approaches perform diversity optimization at training time and therefore can be called in-processing. These include some adaptations of the well-known collaborative filtering algorithms based on nearest neighbors [281, 365] or NCF [271], evolutionary algorithms [441], graph-based methods [260] and other probabilistic methods [80].

2.3.5.2 Fairness Aspects of Recommendations

If customers would perceive the recommendations as unfair, e.g., because they are too biased toward the items of greatest interest to the business, they could leave the platform and the company could lose important revenue streams. In general, the problem of making a fair* recommendation, besides the previous commercial motivations, is of great interest nowadays especially in many areas of our lives when people need to make critical decisions (e.g., bank loans, legal processes, job selection) [303]. Indeed, it is essential that these decisions do not reflect discriminatory behavior that could be harmful to people. One of the most notorious biases that has traditionally been studied in the literature of recommender systems is demographic bias [127]. This particular type of bias occurs when the recommender system discriminates against users from a particular group (e.g., women/men, young/elderly).

To mitigate the effect of bias and thus to eliminate discrimination in model predictions, recommender systems typically exploit algorithmic fairness methods [237, 356, 467, 468]. Three main scopes have been identified depending on whether the RS should not discriminate: users/consumers (*C-fairness*), items/providers (*P-fairness*) or both (*CP-fairness*) [66, 453]. Techniques are distinguished into pre-, in-, or post-processing according to when fairness is introduced into the learning process [71, 351]. For example, post-processing methods have been applied to various recommenders by re-ranking the predicted scores [399, 458], introducing fairness constraints [466] or considering temporal aspects to amortize fairness on series of multiple rankings [50, 395]. Instead, in-processing methods, have been used to introduce algorithmic fairness in matrix factorization [462] or in SLIM [66] by adjusting the objective function of the algorithms.

2.3.6 Related Areas in Value-Aware Recommendation Research

Value-aware recommender systems are related to other important research areas, including the following:

- *Multi-Stakeholder Recommender Systems* [3, 4]: where the system is designed to meet the interests of multiple stakeholders (e.g., consumers, providers, suppliers);
- *Multi-Objective Recommender Systems* [18, 488]: where the system is designed to optimize several objectives simultaneously (e.g., accuracy, diversity);
- *Fair Recommender Systems* [107, 351, 356, 467, 468]: where the system is designed to avoid possible discrimination against certain user or item groups.

*There are multiple definitions of fairness in the RSs literature [303]. One of the most known definitions is *Demographic Parity* [226]. Consider a generic predictor that is required to assign a class to an individual. A predictor is said to achieve demographic parity if a specific outcome is equally likely to be assigned to individuals from different groups.

The relationships between VARSs and these other areas can be characterized as follows. Regarding multi-stakeholder RSs, we note that probably any VARS in practice does not *exclusively* focus on business value but considers the interests of other stakeholders—in e-commerce, in particular, those of consumers or suppliers as well [69, 82]. Such multi-stakeholder considerations mean that VARSs in practice are multi-objective RSs that consider different competing objectives, e.g., business vs. consumer value [96, 155] or short-term vs. long-term profits. However, not every multi-stakeholder RS necessarily is a value-aware one, e.g., considering that an RS may also be designed to recommend users to other users (e.g., on dating platforms). Likewise, a multi-objective RS could also optimize non-economic goals, e.g., popularity, which may in turn have a direct inverse relationship with profitability under certain circumstances [147]. Finally, in terms of fairness, when building a VARS there is always the possibility that by designing a system too biased [79] toward highest-value items [82, 316], the organization might risk being perceived as unfair by consumers. However, there are various other application areas of fair recommender systems, which are not related to economic aspects, e.g., when the recommender system is designed to avoid discrimination of underrepresented groups in the recommendations.

Various surveys and theses have been published in the mentioned areas of multi-stakeholder [3, 4] and multi-objective [18, 488] RSs, and on related topics such as fairness [351, 356, 467, 468], diversity [252], trust [116], and explainability [413, 477]. We refer the readers to these works for in-depth coverage of the respective topics. The present thesis has certain affinities with a previous review on and price- and profit-aware RSs [214] and with a number of papers that addressed profit-aware algorithmic aspects, e.g., [316, 353]. It however differs from these previous works in various ways. First, our study contains the first systematic reviews of value-aware and economic recommender systems based on PRISMA guidelines [340] (see Chapter 3 and Chapter 4). In particular, we identified a number of important algorithmic approaches, while also discussing methodological questions (e.g., performance evaluation methods) that were not covered in previous works. Moreover, as part of the original contributions of this thesis, we also proposed three novel model-based profit-aware recommendation approaches (Chapter 5). The proposed models proved effective in generating more profitable yet relevant recommendations and offer a more efficient alternative to existing re-ranking approaches. Finally, still part of our original contributions is the study of the problem of recommending items that can influence sensitive users' behavior (Chapter 6). With such study we aim to focus on broader issues, moving from the optimization of the value for the business to the optimization of the value for the society as a whole.

Part II

Original Contributions

3
A Systematic Review of Value-Aware Recommender Systems

As previously discussed in Section 2.3.2, although suggesting products and services of interest to customers is a fundamental requirement for the sustainability of any business, an organization often decides to adopt an RS to improve its business performance. Correspondingly, for these reasons, in the past few years there has been increased interest in value-aware RSs, see Section 2.3.3. Differently from traditional RSs, VARSs are designed to directly optimize the business value of recommendations, e.g., increasing user engagement, optimizing sales revenue and improving profitability aspects.

However, as previously mentioned in Section 2.3.6, although value-aware RSs are highly important for industry, most of research is scattered and composed of many papers proposed in isolated contexts, e.g., where algorithms are designed to target specific application domains. Moreover, although there were some existing surveys that had explored related topics on multi-stakeholder RSs, multi-objective RSs, and fairness aspects, not much research had yet explored in depth the business value optimization aspects of recommendations. Hence, as our first contribution in the field, in this chapter we propose to investigate the existing literature through a systematic review approach. Following this process, we identified over a hundred relevant articles. From the analysis of these articles we have determined the main families of algorithms, as well as the main commercial applications and business value categories typically optimized. In addition, again as part of the contribution we identified several open challenges in the field and some promising research directions for the future.

The main contributions of this work can be summarized as follows:

- We provided a systematic literature review focused on value-aware recommender systems by discussing a number of related articles collected from different research streams.
- We described the technical approaches that can be used to build VARSs algorithms and the business value categories that are traditionally optimized.

- We discussed the main application domains, the most commonly used datasets and pointed out current challenges and possible future research directions in the field.

The remainder of the chapter is organized as follows. In Section 3.1 we present the methodology used for the systematic review. In Section 3.2 we introduce the various families of algorithms we identified. In Section 3.3 we discuss the main commercial applications and business value types typically optimized. In Section 3.4 we introduce the datasets available in the literature. In Section 3.5 we discuss open challenges and future research directions in the field. Finally, in Section 3.6 we end the chapter with a summary of findings.

The article entitled "*A Systematic Review of Value-Aware Recommender Systems*" [106] was published in the journal *Expert Systems With Applications* (2022 impact factor of 8.5).

3.1 Systematic Review Methodology

The present study follows a systematic review process based on *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (*PRISMA*) [340] guidelines. The PRISMA article selection process is recognized throughout the scientific community as a rigorous and reliable methodology. The process aims to identify, evaluate, and interpret all available research relevant to a particular research question, topic area, or phenomenon of interest while ensuring high reproducibility of results. Below, we report the research questions behind the study, the information sources queried, the search strategy used to identify the articles, the eligibility criteria used for selection, the overall selection process, and the limitations of the study.

3.1.1 Research Questions

The goal of our work is to review the state-of-the-art of current VARSs research. More specifically, the present survey aims to answer the following *research questions* (*RQs*):

- **RQ1**: What are the main value categories typically optimized in value-aware recommender systems?
- **RQ2**: What are the main techniques used to design value-aware recommender systems?
- **RQ3**: What are the main applications of value-aware recommender systems?
- **RQ4**: What are the main datasets used in the literature of value-aware recommender systems?
- **RQ5**: What are the main state-of-the-art challenges and future research directions?

3.1.2 Search Query

As mandated by the PRISMA guidelines, our survey aims to answer previous RQs by systematically querying online libraries. In particular, we queried Elsevier Scopus, IEEE Xplore, Springer Link, and ACM Digital Library to identify relevant articles. We identified all articles from *Jan 1, 2006* to *Dec 31, 2022* resulting from the following *search query* (*SQ*):

- **SQ**: *((”recommender system” OR ”recommendation system”) AND (”value” OR ”revenue” OR ”sales” OR ”click” OR ”profit” OR ”price” OR ”customer” OR ”product” OR ”optimization” OR ”maximization” OR ”aware”)).*

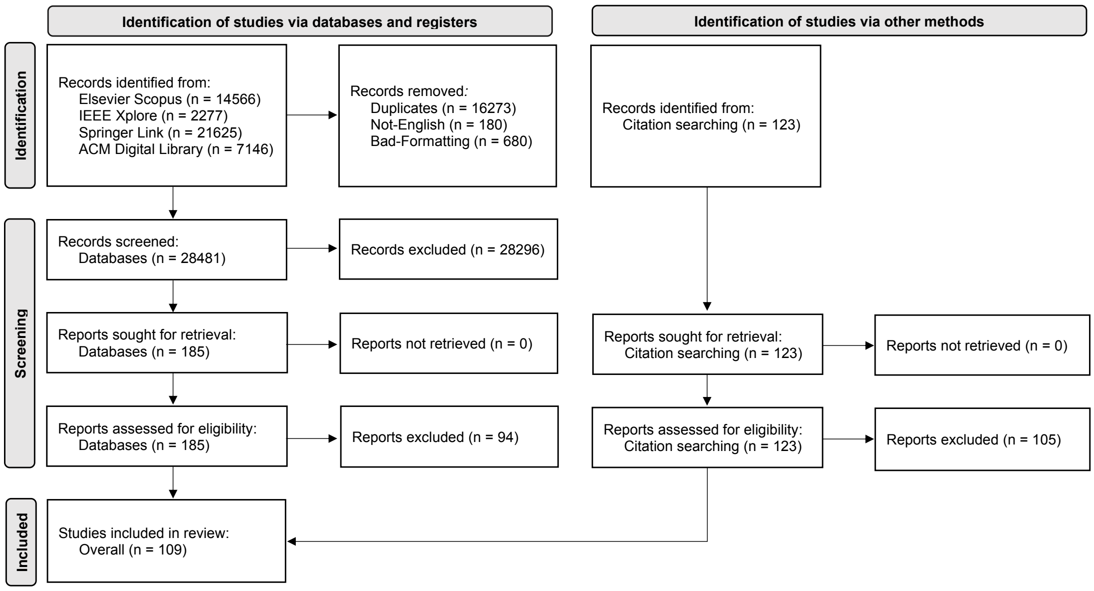


Figure 3.1: Systematic review PRISMA flow diagram of value-aware recommender systems surveyed literature.

To stay below the maximum number of search results that could be extracted by querying the various databases, it was necessary to implement operational arrangements, i.e., breaking the search query into different subqueries, each executed in a distinct time range.

3.1.3 Eligibility Criteria

To be included in the review, articles must pass a rigorous analysis process. Specifically, articles must meet the following *eligibility criteria* (*EC*):

- **EC1**: Articles should focus on value-aware recommender systems.
- **EC2**: Articles must be in English and the full content of the article must be accessible by the authors.
- **EC3**: Articles must be unique, and any duplicate copies of the same article are not included.
- **EC4**: Articles must be peer-reviewed by journals or conferences.
- **EC5**: Graduate theses and doctoral dissertations are not included.

3.1.4 Article Selection Process

In accordance with the PRISMA guidelines, we followed a multi-stage process to identify all the relevant resources included in this review. At each stage we applied the eligibility criteria defined in Section 3.1.3 to discard non-relevant documents. As shown in the PRISMA flow diagram in Figure 3.1, a total of 14,566 articles from Elsevier

Scopus, 2,277 articles from IEEE Xplore, 21,625 articles from Springer Link, and 7,146 articles from ACM Digital Library were identified in the first research phase. We identified 16,273 duplicate records, 180 non-English articles, and 680 records that exhibited formatting problems in the title and metadata that were removed as a result of applying EC2, EC3, EC4 and EC5 criteria. In the screening stage, the titles and abstracts of 28,481 articles were analysed, and 28,296 records were excluded after applying EC1 criterion because the topics covered were not relevant to our study. A total of 185 articles were first sought for retrieval and then assessed for eligibility. At this stage, 94 articles were excluded after reading the full text as they did not met EC1 criterion. From this subset of eligible articles, an additional 123 articles were identified by searching for references in their bibliography, then sought for retrieval and finally assessed for eligibility. In this last stage, 105 records were excluded after reading the full text as a result of applying EC1 criterion. At the end of this overall process, a total of 109 studies were included in the review.

3.1.5 Study Limitations

The possible *study limitations* (*SLs*) are the following:

- **SL1**: Articles were selected primarily from IEEE Xplore, Elsevier Scopus, ACM Digital Library, and Springer Link and from reference searches in the bibliographies of articles that passed the screening stage.
- **SL2**: Unpublished articles, non-English articles, articles whose content was not accessible, graduate theses, doctoral dissertations, commercial products, and demos were not included.
- **SL3**: Since research on VARSs is still in its infancy, any future applications or techniques not yet addressed by the literature are outside the scope of this review.

3.2 Value-Aware Recommendation Algorithms

In this section, we introduce the main algorithms in the literature on VARSs. These algorithms leverage different technical approaches and, in some cases, depend on the nature of the recommended content. Although other taxonomies based, for example, on business KPIs or value dimensions, are available in the literature (see Section 2.3.2), we provide a classification of VARSs according to the technical approaches used, highlighting the different mechanisms underlying the various algorithms. As indicated in Figure 3.2, VARSs algorithms can first be divided into in-processing and post-processing based on the time at which value-driven optimization of recommendations occurs - although pre-processing methods may also exist, none have been found in the literature. Then, the approaches can be further divided into value objective, value reinforcement, value ranking, and value policy according to the specific technique used. In the following, we introduce each of these approaches.

3.2.1 Value-Aware Post-Processing Algorithms

Post-processing algorithms can be applied to any recommendation algorithm (treated as a black box) to optimize the business value of recommendations. As described in Section 2.2.2, in traditional scenarios, a recommender

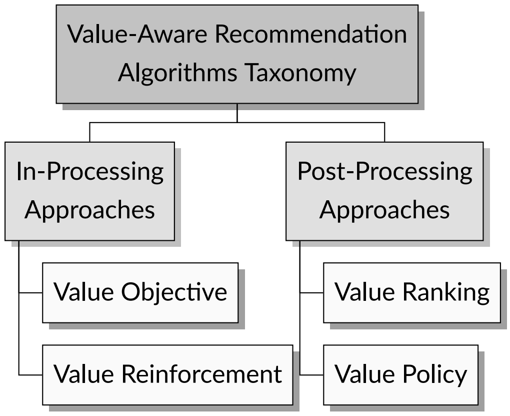


Figure 3.2: Value-aware recommendation algorithms taxonomy.

system suggests to a user u a rank $\mathbf{y}_{u,k}$ of k items that maximizes the expected interest by sorting the predicted scores of the items the user has never interacted with in descending order and selecting the top k. Post-processing methods rely on predicted scores and other economic information to re-rank the output of the original algorithm.

Value Ranking

This class of post-processing methods extends the approach in Eq. (2.3) by incorporating economic value information into the objective function to re-rank the output of the original algorithm. Given a generic business value $val_i \in \mathbb{R}$ associated with each item (e.g., product profit), a strategy commonly used by these systems [33, 82, 103, 108, 214, 240, 286, 291, 316, 434, 480] is to recommend the list $\mathbf{y}_{u,k}$ of items that maximize the weighted expected interest:

$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \quad \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot val_i \tag{3.1}$$

selecting the first k items with the highest $\hat{x}_{u,i} \cdot val_i$. As noted in some studies [82, 316], in this way, it is possible to provide more profitable recommendations overall than those generated by a traditional RS at the cost of some reduction in accuracy. However, as also noted [214, 240, 291], the interests of customers and organizations must be balanced appropriately. Clients may feel dissatisfied with a system that recommends only high-profit, irrelevant items, and the organization may risk losing loyal customers. To mitigate this drawback, several studies [33, 240, 275, 286, 322, 448, 480] have proposed simple extensions of Eq. (3.1) to account for the perspectives of different stakeholders and determine the best trade-off between economic value for the organization and customer interests using regularization parameters to control the equation. Some variants [103, 108, 291, 434] of the approach have also used constraints to match certain conditions such as the user budget.

Value Policy

Advanced post-processing approaches that are more complex than simple value ranking ones have also been proposed. We refer to these methods as value policies to indicate that they are based on specific policies consisting of multiple steps to optimize the economic value derived from the entire recommendation process. In particular, various studies [189, 200, 266] have proposed such multiple-step process-based approaches to optimize economic value. For example, one study [200] proposed an algorithm that recommends relevant items to gain customer trust and then recommends profitable items once trust is gained to increase business value. More sophisticated models have also been studied [266] by incorporating various factors such as price, profitability, product competition, and saturation effects to improve profitability over a finite time horizon. In recent work [189] it has been proposed also a probabilistic approach to optimize multiple strategic parameters (e.g., click-through rate, user engagement) one at a time considering that optimizing one parameter could have positive effects on other value indicators as well. Furthermore, other works [41, 230, 318, 481] have proposed methodologies optimizing the value of recommendations by integrating dynamic pricing algorithms. For example, some works [230] have proposed to optimize the discount of recommended items by exploiting multi-armed bandits. By contrast, a more recent work [318] has proposed personalizing the price of recommended products based on customer willingness to pay to simultaneously optimize service provider profit and customer surplus.

3.2.2 Value-Aware In-Processing Algorithms

While the methods presented above optimize the business value of recommendations after the learning process, in-processing algorithms aim to modify existing or to introduce new algorithms to generate recommendations that optimize business value without the need to perform subsequent operations.

Value Objective

This class of methods contains approaches that integrate the objective function of known algorithms to generate more valuable recommendations. For example, in some work [69, 345, 346, 423] it has been proposed to modify the well-known *nearest neighbors* approaches that we previously described in Section 2.2.5.1. In particular, a recent research [69] has proposed to modify the neighbour selection procedure, by selecting the most profitable (instead of the most similar) neighbors to increase the overall profitability of recommended items while maintaining accuracy under shilling attacks, i.e., attacks from malicious users who generate biased ratings to influence recommendations for their own interests. Similarly, another study [423] proposed increasing sales diversity by recommending users to items by reversing the original algorithm. Moreover, further research [151, 197, 436, 478, 482] extended the well-known *matrix factorization* algorithm [242, 243] that we previously introduced in Section 2.2.5.2 by incorporating economic value information into the objective function.

Value Reinforcement

Recent studies have proposed value-aware recommendation algorithms that exploit *reinforcement learning* (*RL*) [408]: a learning approach that aims to automatically learn an optimal policy based on the sequential interactions between an agent and the environment through trial and error to maximize a reward. In particular, given the sequential nature of user interaction with an RS, *reinforcement learning-based recommender systems* (*RLRSs*) [14] have emerged as alternative approaches based on RL techniques to generate recommendations. Much of the recent literature on VARSs [176, 182, 221, 229, 268, 353, 411, 449, 484, 485, 500] exploits this methodology to maximize

Application Domain	Most Frequently Used Technique	Typically Optimized Value	Details
Product Recommendation	Value Objective	Sales and Revenue	Table 3.2
Advertising Recommendation	Value Reinforcement	Sales and Revenue	Table 3.3
News Recommendation	Value Reinforcement	User Engagement	Table 3.4
Media Recommendation	Value Objective	Sales Distribution	Table 3.5

Table 3.1: Main application domains of value-aware recommender systems identified in the surveyed literature.

the long-term value of recommendations. For example, in one of the earliest studies in the field [411] focused in the advertising domain, it was proposed to exploit RLRSs to maximize the *customer lifetime value* (CLV) of recommendations, i.e., the total value generated by the customer throughout his or her history, considering that a certain economic value can be associated to each click. Similarly, in a more recent study [353] focused on the e-commerce domain, it was proposed exploit such algorithms to generate recommendations that maximize the economic value of each user action, e.g., not only clicks but also purchases.

3.3 Value-Aware Recommender Systems Application Domains

Recent years have witnessed a growing interest in VARSs. Since algorithms are often designed based on domain-dependent characteristics, in this section, we review the literature on VARSs in various application domains. As indicated in Table 3.1, these include the recommendation of products, advertising, news and media. This analysis is proposed because each type of application has distinctive characteristics that lead to a preference for certain methodologies and for optimizing certain types of economic value. The following sections refer to the detailed tables linked to the main table for a more in-depth discussion of individual research works.

3.3.1 Product Recommendation

Many VARSs have been developed to optimize product sales, see Table 3.2. Below, we provide an overview of the main topics addressed in the literature, including the accuracy-profitability trade-off, the optimization of multiple objectives simultaneously from a multi-stakeholder perspective, the usefulness of recommendations for the customer, the long-term implications of value-aware recommendations, and real-world studies.

Profitability-Relevance Trade-Off. Business interest in leveraging RSs to increase revenue or other key performance indicators of global e-tailers existed since the 2000's. In early work, Chen et al. [82, 316] proposed a methodology to weight the recommendations of a collaborative filtering algorithm with product profitability factors. This approach allows the system to meet the customer's needs and achieve higher profit margins for the organization. However, as observed by the authors, focusing excessively on profitability could rapidly degrade the accuracy of recommendations. Hence, to balance the potentially conflicting interests of multiple stakeholders

Ref	Year	Author	Technique Used	Optimized Value	Dataset
[316]	2007	Mu-Chen Chen et al.	Value Ranking	Sales and Revenue	Foodmart
[82]	2008	Chen et al.	Value Ranking	Sales and Revenue	Foodmart
[200]	2008	Hosanagar et al.	Value Policy	Sales and Revenue	N/A
[434]	2009	Wang and Wu	Value Ranking	Sales and Revenue	Self-collected
[103]	2009	Das et al.	Value Ranking	Sales and Revenue	N/A
[16]	2010	Akoglu and Faloutsos	Value Objective	Sales and Revenue	N/A
[230]	2011	Kamishima and Akaho	Value Policy	Sales and Revenue	MovieLens
[436]	2011	Wang and Zhang	Value Objective	User Engagement	Self-collected
[223]	2012	Jiang and Liu	Value Objective	Sales and Revenue	Self-collected
[206]	2013	Huang et al.	Value Objective	Sales and Revenue	Foodmart
[181]	2013	Hammar et al.	Value Objective	Sales and Revenue	Self-collected
[266]	2014	Li and Lakshmanan	Value Policy	Sales and Revenue	Self-collected
[481]	2015	Zhao et al.	Value Policy	Sales and Revenue	Self-collected
[225]	2015	Jiang et al.	Value Policy	Sales and Revenue	Self-collected
[41]	2016	Beladev et al.	Value Policy	Sales and Revenue	Self-collected
[343]	2016	Panniello et al.	Value Policy	Sales and Revenue	Self-collected
[478]	2016	Zhang et al.	Value Objective	User Engagement	Self-collected
[482]	2017	Zhao et al.	Value Objective	User Engagement	Self-collected
[214]	2017	Jannach and Adomavicius	Value Ranking	Sales and Revenue	MovieLens
[459]	2017	Yang et al.	Value Objective	User Engagement	Foodmart, ChainStore, Amazon
[229]	2017	Ju et al.	Value Reinforcement	Sales and Revenue	Dunnhumby
[322]	2017	Nguyen et al.	Value Ranking	Sales and Revenue	Self-collected
[69]	2019	Cai and Zhu	Value Objective	Sales and Revenue	Book-Crossing
[202]	2019	Hosein et al.	Value Objective	Sales and Revenue	MovieLens
[285]	2019	Louca et al.	Value Objective	Sales and Revenue	Self-collected
[289]	2019	Ma et al.	Value Objective	Sales and Revenue	SPMF
[151]	2019	Ge et al.	Value Objective	User Engagement	Amazon
[353]	2019	Pei et al.	Value Reinforcement	Sales and Revenue	REC-RL
[112]	2019	Desirena et al.	Value Ranking	Sales and Revenue	Self-collected
[275]	2019	Lin et al.	Value Objective	User Engagement	EC-REC
[170]	2020	Gu et al.	Value Objective	User Engagement	JD
[62]	2020	Brodén et al.	Value Reinforcement	Sales and Revenue	Self-collected
[240]	2022	Kompan et al.	Value Ranking	Sales and Revenue	Self-collected
[39]	2021	Basu	Value Policy	Sales and Revenue	Self-collected
[268]	2021	Li et al.	Value Reinforcement	Sales and Revenue	Self-collected
[221]	2021	Ji et al.	Value Reinforcement	User Engagement	Self-collected
[29]	2022	Aridor and Gonçalves	Value Policy	Sales and Revenue	N/A
[72]	2022	Cavenaghi et al.	Value Ranking	Click-Through Rate	Self-collected
[155]	2022	Ghanem et al.	Value Ranking	Sales and Revenue	MovieLens
[259]	2022	Lee et al.	Value Objective	Sales and Revenue	Self-collected
[491]	2021	Zhou and Zou	Value Policy	Sales and Revenue	N/A

Table 3.2: Articles of the surveyed literature that focus on value-aware product recommendations.

some techniques based on constrained optimization [16, 103, 181, 214, 266, 434] or multi-objective algorithms [29, 69, 112, 155, 155, 170, 206, 240, 275, 285, 289, 322] have been proposed in the literature. As reported in many studies, e.g., [214], designing algorithms to recommend more valuable items could actually increase business KPIs while maintaining the relevance of recommended products for the end user. However, above a certain threshold, the purchase probability drops dramatically, and the business value generated as a result is reduced [155]. Hence, being aware of this trade-off, algorithms should be fine tuned accordingly to maximize the business value.

On the Utility of Customer Recommendations. A different research perspective [151, 436, 459, 478, 482] finds that the utility of customer recommendations can be directly related to the sales performance of the recommendation system. In fact, according to leading economic theories, a rational customer would choose products that maximize his or her utility. For example, based on this perspective, Wang and Zhang [436] developed a recommendation algorithm that maximizes the marginal utility of recommended products for the customer. Similarly, Yang et al. [459] proposed an adaptive association rule mining algorithm to recommend the highest utility products. Moreover, Zhang et al. [478] designed a recommendation system that jointly optimizes the utility of customers and sellers in an online marketplace. Furthermore, Zhao et al. [482] proposed maximizing the utility of recommendations considering complementarity and substitutability aspects of the products being recommended to each customer compared to those he or she already purchased. Finally, Ge et al. [151] proposed optimizing the utility of recommended products by considering also customer budget constraints.

Long-Term Implications of Value-Aware Recommendations. Thus far, the discussed works have focused mostly on optimizing short-term sales performance. However, as argued by Jannach and Adomavicius [214] and Ghanem et al. [155], the performance of an RS also depends to a large extent on the long-term effects of recommendations. Purely profit-oriented strategies are overly biased towards the organization's short-term interests and can lead to long-term customer churn. Instead, strategies that balance profit with customer utility, and thus are more oriented to the customer's perspective, would likely lead to sustained profitability in the long run due to higher customer satisfaction levels. Considering this aspect, for example, Hosanagar et al. [200] argued that a recommendation system should first try to maintain a certain level of trust by proposing products that are relevant to the customer before optimizing profitability. Seeking to embed these considerations into recommendation models, more recent works [62, 202, 229, 268, 353] proposed directly optimizing the long-term performance of recommender systems by exploiting probabilistic approaches [202] or reinforcement learning [62, 229, 268, 353] algorithms. The latter have been used, for example, to maximize the cumulative value of recommendations [353] or to optimize customer lifetime value in cold-start scenarios [221].

Static vs. Dynamic Pricing. The majority of research on VARSs is based on algorithms that keep prices static. However, an alternative approach is represented by systems that integrate recommendations with dynamic pricing [41, 206, 223, 230, 481]. According to this philosophy, Kamishima and Akaho propose a system that strategically adjusts the price of items recommended to customers based on the type of customer visiting the system. If the customer would purchase the product at a discounted price, the system would propose a favourable price to obtain additional revenue. A different approach was proposed by Jiang and Liu [223], who designed a system that optimizes the discount of promotional products to increase the overall profitability of non-promotional ones. The authors proposed exploiting intra/cross-category effects to stimulate customers to

Ref	Year	Author	Technique Used	Optimized Value	Dataset
[411]	2015	Theocharous et al.	Value Reinforcement	User Engagement	Self-collected
[480]	2017	Zhang et al.	Value Ranking	Sales and Revenue	Self-collected
[283]	2018	Long et al.	Value Policy	Sales and Revenue	Package, NBA
[291]	2019	Malthouse et al.	Value Ranking	Sales and Revenue	Self-collected
[182]	2019	Han et al.	Value Reinforcement	Sales and Revenue	MovieLens
[484]	2020	Zhao et al.	Value Reinforcement	User Engagement	Self-collected
[176]	2021	Guo et al.	Value Reinforcement	User Adoption	Self-collected
[189]	2022	He et al.	Value Policy	All Business Values	Amazon

Table 3.3: Articles of the surveyed literature that focus on value-aware ads recommendations.

purchase non-discounted products. Additionally, regarding the topic of personalized promotions, Zhao et al., [481] proposed customizing the discount of recommended products based on customer willingness to pay predictions, while Beladev et al. [41] propose recommending product bundles by pricing them to maximize revenue.

3.3.2 Advertising Recommendation

Several value-aware systems have been proposed to optimize the business value of advertising systems, see Table 3.3. Below, we provide an overview of traditional systems in this field and more recent perspectives that aim to optimize customer lifetime value.

Traditional Advertising Strategies. In advertising systems, the sponsored space is traditionally sold through auctions, where different advertisers compete for customers' attention. The systems often work as follows [30, 134, 176]: the advertiser first defines a subset of potential target customers based on certain demographic and/or purchasing characteristics; subsequently, he or she selects an objective to optimize through sponsored recommendations (e.g., number of clicks, add-to-carts or gross merchandise volume); finally, the advertiser defines a bid price that he or she will pay when the objective is reached. Therefore, a common strategy used by service providers to maximize system revenues is to sort advertisers' products into sponsored space by weighting the bid price by click-through rate or click-conversion rate. As a result, much of the literature in the field of computational advertising [30, 75, 134, 166, 191, 302, 339, 494] investigates algorithms to predict the previous metrics as accurately as possible from the characteristics of the recommended items. Early work [166] by Microsoft proposed a Bayesian algorithm based on a regression model to predict CTR in a Microsoft Bing sponsored search. Subsequent work focused on the ads systems of Google [302], Facebook [191] and Yahoo [75], Etsy [30], and Alibaba [134, 339, 494] as well as the algorithms used for CTR prediction.

Considering User Interest to Generate Higher Returns. Although conventional advertising strategies are widely adopted, alternative approaches have been proposed to optimize other aspects of advertising, particularly considering users' interests [283, 291, 480]. Indeed, as previously discussed, indiscriminately promoting high-profit items that do not match users' interests could push users away from the system. To consider both the interests of the organization and the end users, Zhang et al. [480] proposed a methodology to balance the revenue generated from the ads of an app store and the overall quality of the recommendations. Adopting a similar perspective: Long et al. [283] developed an algorithm that optimizes the overall profitability of a promotional campaign while maintaining a certain number of satisfied customers; Malthouse et al. [291] proposed a multi-stakeholder system that jointly optimizes ad revenue and user utility. Considering the user's interests in recommendations would increase customer lifetime value and improve many business value categories at the same time.

Maximizing Customer Lifetime Value and Advertiser Revenue. As previously observed, the interests of multiple stakeholders should be balanced appropriately to maximize customer lifetime value. Trying to increase short-term profitability with overly biased recommendations could negatively impact an organization's reputation. If the trust relationship is broken, some customers may decide to purchase from competitors, and the company may lose valuable sources of revenue. To address this problem, some works [176, 182, 411, 484] have studied how to optimize the long-term performance of an advertising system. Instead of recommending to customers ads that have the highest probability of being clicked, Theocharous et al. [411] and Han et al. [182] proposed leveraging reinforcement learning techniques to optimize customer lifetime value and, more generally, cumulative reward for the platform. Zhao et al. [484] further adapted the approach in the case of sequential recommendations by proposing an approach that maximizes cumulative user engagement by balancing longer browsing sessions and the click-through rate. Moreover, in addition to the interests of service providers and customers, the system should consider the interests of advertisers. According to the latter perspective, Guo et al. [176] proposed a system based on multi-armed bandits to recommend the best advertising strategy to advertisers. The system aims to encourage the adoption of the platform by helping advertisers define customer targets and bid prices to improve the performance of marketing campaigns by reducing the cost of trial and error.

3.3.3 News Recommendation

Some value-aware recommenders have been proposed to optimize the business value of news systems, see Table 3.4. Below, we provide an overview of conventional news recommendation strategies, the existing relationship between click-through rate and user engagement, and the optimization of long-term metrics to generate greater returns for the service provider.

Conventional News Recommendation Strategies. The reputation of a news company is directly related to the impact of the information it provides on society [495]. The business model may be subscription based, advertising based, or both. Conventionally, the number of clicks or views a certain news obtains during its overall lifespan is directly related to the organization's returns. Since the click-through rate is directly related to a news service provider revenue, a common goal is to maximize the number of clicks. Therefore, traditional news RSs [133, 231] use CTR as a primary indicator to determine which articles most closely match the reader's interests. The systems generate news candidates with the highest probability of being clicked by the users.

On the CTR/User Engagement Relationship. As for advertising, although the CTR measures the probability of clicks in the current step, it does not capture the engagement that may occur due to the action itself. In fact, even if a user clicks on an article simply for curiosity, he or she might not necessarily be interested in

Ref	Year	Author	Technique Used	Optimized Value	Dataset
[262]	2010	Li et al.	Value Reinforcement	Click-Through Rate, User Engagement	Self-collected
[47]	2016	Besbes et al.	Value Ranking	Click-Through Rate, User Engagement	Self-collected
[449]	2017	Wu et al.	Value Reinforcement	User Engagement	Self-collected
[485]	2018	Zheng et al.	Value Reinforcement	Click-Through Rate	Self-collected
[499]	2019	Zihayat et al.	Value Ranking	User Engagement	Self-collected
[500]	2019	Zou et al.	Value Reinforcement	Click-Through Rate, User Engagement	Self-collected
[286]	2020	Lu et al.	Value Ranking	User Engagement	Self-collected
[398]	2022	Spyridou et al.	Value Ranking	Click-Through Rate, User Engagement	Self-collected

Table 3.4: Articles of the surveyed literature that focus on value-aware news recommendations.

reading it. Consequently, a growing body of work [47, 286, 499] has considered the relationship between CTR and user engagement by proposing to optimize the latter. Besbes et al. [47] formulated a heuristic methodology that examines the probability of clicking on a news item and the engagement effect that it triggers. Through this formulation, a certain news is proposed to a user also considering the future navigation paths of the contents. Moreover, as observed by Lu et al. [286] and Spyridou et al. [398], news recommendation differs from many traditional recommendation domains, such as e-commerce or entertainment, in that news organizations have a clear responsibility to society to provide high-quality information. Algorithms should first and foremost consider the civic role of journalism for an informed citizenry and optimize the editorial value of news (i.e., a mix of serendipity, dynamism, diversity, and coverage) rather than looking solely at CTR.

Optimizing Long-Term Metrics. As with other value-aware systems, the relationship between value and time should not be underestimated. In some cases optimizing exclusively for short-term CTR may prove counterproductive if the news provided is not of interest for the user. Taking this aspect into consideration, several works [262, 449, 485, 500] have proposed methodologies to optimize long-term metrics. For example, Wu et al. [449] propose optimizing long-term user engagement by maximizing the total number of clicks per period using a multi-armed bandit system. The model also considers that, in some cases, the user may abandon the system due to incorrect recommendations. A similar approach based on contextual bandits was originally proposed by Li et al. [262] to maximize the total number of user clicks. Furthermore, more advanced approaches based on reinforcement learning have been proposed by Zheng et al. and Zou et al. [485, 500] to optimize both CTR and long-term user engagement while considering the user's return pattern on the platform in addition to click information.

3.3.4 Media Recommendation

Some value-aware recommender systems have been designed to optimize the value of multimedia services, see Table 3.5. Below, we provide an overview of the main topics in the literature mainly concerning the optimization

Ref	Year	Author	Technique Used	Optimized Value	Dataset
[211]	2006	Iwata et al.	Value Objective	Sales and Revenue	Self-collected
[212]	2008	Iwata et al.	Value Objective	Sales and Revenue	Self-collected
[346]	2008	Park and Tuzhilin	Value Objective	Sales Distribution	MovieLens
[33]	2013	Azaria et al.	Value Ranking	Sales and Revenue	Self-collected
[345]	2013	Park	Value Objective	Sales Distribution	MovieLens, Book-Crossing
[197]	2014	Ho et al.	Value Objective	Sales Distribution	MovieLens
[423]	2014	Vargas and Castells	Value Objective	Sales Distribution	Netflix Prize, Million Song
[441]	2016	Wang et al.	Value Ranking	Sales Distribution	MovieLens, Netflix Prize, Jester
[27]	2017	Antikacioglu and Ravi	Value Ranking	Sales Distribution	MovieLens, Netflix Prize
[180]	2019	Hamedani and Kaedi	Value Ranking	Sales Distribution	MovieLens, Netflix Prize
[318]	2021	Najafabadi et al.	Value Policy	Sales and Revenue	Self-collected
[475]	2021	Zhang et al.	Value Policy	Sales and Revenue	Self-collected

Table 3.5: Articles of the surveyed literature that focus on value-aware media recommendations.

of user engagement and the effects on the sales distribution of items with which the user interacts.

On The Effects of Optimizing User Engagement on Sales Distribution. In contrast to ordinary goods (e.g., physical products), movies, music and other digital goods are referred to as information goods because their production and distribution costs are negligible and they can be copied, shared, rented or resold easily [318]. As with news systems, the main revenue models of companies providing multimedia services are based on either subscriptions or advertising. Thus, especially for companies in the entertainment industry, user engagement is directly related to profits. As a result, RSs are traditionally designed with the goal of providing the user with the content of greatest interest [163]. However, given the considerably large amount of content available, RSs tend to recommend the most popular items, risking boring the users with poorly tailored recommendations [139, 257, 258]. To keep users engaged, one of the main techniques is to optimize the distribution of recommended items with the goal of helping the user discover surprisingly new and relevant items. This can be done, for example, by increasing the diversity [252] of recommendations [27, 180, 423] or promoting long-tail items [197, 345, 346] that tend to be proposed less frequently by RSs because of popularity bias.

Optimizing Sales Revenue According to the Revenue Model. In addition to user engagement, research on media value-aware recommenders have proposed approaches to optimize other business value indicators. Some works [33, 211, 212] have proposed domain-specific approaches to recommend films that have the highest probability of maximizing system sales revenue. Azaria et al. and Iwata et al. proposed two different variants of their approach depending on whether the customer pays a subscription (*subscription-based revenue model*) to have the opportunity to watch several movies in a given time frame [33, 211] or a fixed price (*transaction-based revenue model*) for individual movies [33, 212]. The importance of the value-aware approach on the overall revenues of a movie provider based on an on-demand revenue model has also been studied in detail in two recent papers [318, 475]. In particular, as already observed in similar literature domains, according to Zhang et al. [475], recommendation systems that aim solely at profit optimization could risk driving customers away from the company. Instead, according to Najafabadi et al. [318], by personalizing the prices would allow the offer to be more tailored to the customer's willigness to pay and may simultaneously create more profit for the sellers.

3.4 Available Datasets

In many studies, VARSs have been trained and evaluated on public datasets. These datasets frequently contain certain kind of economic value information. As shown in Table 3.6, most datasets are related to the product or media application domain. For example, some studies that focused on product recommendations [82, 316, 459] have exploited the Foodmart dataset [329]. This is a Microsoft SQL Server sample database of a supermarket. The dataset contains sales data (e.g., prices, costs, profitability) and master data about customers (e.g., country) and products (e.g., brand). Similarly, other studies that focused on product recommendations [69, 189, 266, 459] have exploited datasets crawled from Amazon [298, 325] and Epinions [376]. These datasets are primarily based on product review data from various product categories and contain customer ratings, text reviews, and product metadata (e.g., brand, category, price). Further works have also exploited supermarket transaction [229, 289, 459] or e-commerce [170, 275, 322, 353, 500] datasets for their analyses such as Dunnhumby [330], SPMF [141], Chainstore [354], EC-REC [275], REC-RL [353] and JD [170], which contain customer, product, and purchase transaction information. Instead, many studies that are more focused on media recommendations [27, 180, 182, 189, 197, 202, 214, 345, 346, 441] relied on the well-known MovieLens dataset [186]. This is a very popular dataset that is used extensively in RSs research. Unlike the previous datasets, MovieLens does not explicitly contain economic value data. Therefore, in several studies [182, 189, 202, 214], some methodologies based on probability calculations have been used to randomly integrate this information. Instead, other studies [27, 180, 197, 345, 346, 441] have used MovieLens to design algorithms capable of optimizing product distributions without the need to add economic information. Also further research works that focused on media recommendations [27, 180, 345, 346, 423, 441] have adopted a similar philosophy and are based on well-known datasets that do not contain economic information, such as Netflix Prize [45], Book-Crossing [498], Million Song [301], and Jester [160].

3.5 Open Challenges and Future Research

Value-aware recommendation systems offer many business benefits over traditional systems. However, optimizing value brings new challenges. In this section, we discuss some of these challenges to guide future research directions.

Balancing the Interests of Multiple Stakeholders. As previously discussed, although value-aware recommendations can in many cases improve some business KPIs, a VARS that always recommends irrelevant highest-value items could hurt the company's reputation by driving customers away [214, 240, 491]. To address

Ref	Dataset	Domain	Content	Availability
[329]	Foodmart	Product	Contains transaction data, product metadata and customer demographics of a supermarket chain	https://github.com/julianhyde/foodmart-data-hsqldb
[325]	Amazon	Product	Contains product review data and metadata crawled from Amazon e-commerce site	https://nijianmo.github.io/amazon/index.html
[170]	JD	Product	Contains data collected from the recommender systems logs of the JD Chinese e-commerce site	https://github.com/guyulongcs/CIKM2020_DMT
[330]	Dunnhumby	Product	Contains transaction data from a subset of households that make frequent purchases from a retailer	https://www.dunnhumby.com/sourcefiles
[141]	SPMF	Product	Contains customer transaction data from a Belgian retail store	https://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php
[354]	ChainStore	Product	Contains transaction data and product metadata from a supermarket chain in California	http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html
[275]	EC-REC	Product	Contains records of impressions, clicks and purchases from a large-scale e-commerce platform	https://drive.google.com/open?id=1rbidQksa_mLQz-V1d2X43WuUQQVa7P8H
[353]	REC-RL	Product	Contains user interaction data collected from a real-world e-commerce platform	https://github.com/rec-agent/rec-rl
[376]	Epinions	Product	Contains who-trust-whom online social network data from the Epinions consumer review site	https://snap.stanford.edu/data/soc-Epinions1.html
[186]	MovieLens	Media	Contains movie ratings collected over various time periods from the MovieLens web site	https://grouplens.org/datasets/movielens/
[45]	Netflix Prize	Media	Contains anonymous movie ratings from subscribers to the Netflix online movie rental service	https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data
[498]	Book-Crossing	Media	Contains anonymized data of implicit/explicit book ratings from the Book-Crossing community	http://www2.informatik.uni-freiburg.de/~cziegler/BX/
[301]	Million Song	Media	Contains audio features and metadata for over a million contemporary popular music tracks	http://millionsongdataset.com/
[160]	Jester	Media	Contains anonymous ratings of jokes by users of the Jester Joke Recommender System	https://eigentaste.berkeley.edu/dataset/

Table 3.6: Main datasets used to build value-aware recommender systems in the surveyed literature.

this issue, many studies [47, 69, 103, 155, 200, 266, 283, 285, 289, 291, 434, 480] propose recommendation algorithms that seek to determine the best trade-off between relevance and business value. More in general, the interests of multiple stakeholders (i.e., consumers, suppliers and organizations) should be balanced properly in the recommendation process. However, although several studies have addressed this issue, the proposed algorithms are often proposed in isolated contexts targeting a particular type of industry (e.g., retail, entertainment, insurance) [268, 318] or revenue model (e.g., transaction-based, subscriptions) [211, 212]. Hence, the algorithms are not always applicable in different contexts without major adaptations. Therefore, generalizing algorithms to be suitable to many application contexts could be an interesting research direction for the future.

Long-Term Business Value Optimization Algorithms. The majority of algorithms proposed in the VARSs literature rely on post-processing approaches to optimize the short-term business value of recommendations without considering the long-term effects [155, 200, 214]. However, although widely employed in the literature, these approaches are risky because if a potential client notices that the recommendations are biased, he or she may lose trust in the organization and decide to purchase from competitors. To address this issue, a few recent studies proposed to exploit reinforcement learning techniques [176, 182, 221, 229, 268, 353, 411, 449, 484, 485, 500]. Indeed, the recommendation process can be modelled as a sequential decision problem in which an agent interacts with customers to maximize a cumulative business value reward for the organization. However, to date, not many studies that exploit reinforcement learning for value-aware recommendations are present in the literature. Hence, this may be a promising research direction for the future.

Dynamic Pricing for Business Value Improvement. Another important point to consider when designing a VARS is that the price of a product is one of the variables that may influence most customers' purchasing decisions [72, 492]. To date, VARSs literature has primarily studied how to generate more valuable recommendations while keeping prices static. Some specialized works [41, 72, 223, 230, 318, 481] have instead proposed to integrate dynamic pricing algorithms into the recommendation model to increase the revenue and overall profitability of the organization. Currently, the study of the integration of dynamic pricing algorithms in value-aware systems is still in its infancy but could be a valuable future research direction.

Value-Aware Performance Evaluation Methodologies. To evaluate the VARSs performance [171, 216, 218], many studies [39, 343] relied on online A/B tests. However, performing these tests is costly in terms of both time and money for organizations. Often, an A/B test can last several months if long-term aspects are evaluated and unexpected effects can sometimes occur, e.g., due to particular world events that may affect the results. In addition, a poorly performing recommendation system could cause significant financial damage to the organization by making performance evaluation very risky. Thus, given the complexity and cost of conducting field tests, most studies on VARSs [69, 82, 189, 202, 214, 266, 275, 289, 316, 322, 353, 459, 494] exploit offline approaches based on public datasets for performance evaluation. However, as previously discussed, the majority of popular public datasets [45, 186] do not contain economic information (e.g., prices, profits), making it difficult to measure the potential business value generated by a VARS. Moreover, another important limitation is that it is often unclear under what circumstances the data were obtained. The experimental results of various studies could therefore be affected by bias, e.g., due to certain population characteristics that may lead to erroneous conclusions. Hence, future research may address these issues to provide more reliable performance evaluation methods.

Trustworthy Value-Aware Recommender Systems. Finally, like other AI-based systems, value-aware recommenders should be designed in compliance with trustworthy AI principles [233], including alignment with human values, robustness and safety, privacy preservation, fairness [445], explainability [428] and transparency, reproducibility, and accountability. Studying each of these aspects in detail could be a valuable research direction. For example, investigating how to explain value-aware recommendations without degrading business value or studying the reproducibility of major algorithms in the literature could be interesting for the future.

3.6 Summary of Findings

In this chapter, we provided a systematic review of value-aware recommender systems. These systems are highly important for industry since they aim to optimize the business value of recommendations, e.g., increasing customer lifetime value, maximizing user engagement and improving profitability aspects (see Section 2.3.2).

Analyzing the surveyed literature, we identified various algorithmic approaches that we mainly divided into post-processing and in-processing, depending on the time the business value optimization occurs. Post-processing approaches can be applied on top of any RS to re-rank the recommendations of the baseline according to certain economic criteria. In-processing approaches, instead, typically incorporate business value aspects directly at learning time, either by extending the objective function of existing supervised learning algorithms or by exploiting reinforcement learning techniques.

Moreover, we found many application domains where value-aware recommender systems can provide benefits for organizations, i.e., concerning product, advertising, news and multimedia recommendations. Depending on the application domain, certain value-aware algorithms tend to be more frequently used and some business value categories are more likely to be optimized.

However, although VARSs can bring great value to the business, the optimization of such value often brings new challenges, such as balancing the interests of different stakeholders (consumers, suppliers, and organizations) and maintaining high recommendation performance in the short and long term. More in-depth research is required especially to improve the reliability of performance evaluation methodologies and to design higher-performing systems following recent trustworthy AI principles.

4
Economic Recommender Systems – A Systematic Review

As mentioned earlier in this document in Section 2.3.4, the optimization of any business value category is beneficial to the company only if it helps to increase, to some extent, the economics of the business. Indeed, a value-aware system may only be partially useful to the company if the firm revenue model does not allow for capitalizing on certain types of user interaction. For example, if the company relies on a transaction-based revenue model, i.e., if it earns revenue when a user makes a purchase, it could result in a potential missed opportunity to employ a value-aware system designed to optimize uniquely the click-through rate. By contrast, if the company generates revenue predominantly through advertising mechanisms, the optimization of the number of clicks is directly related to the optimization of business economics since advertisers typically incur a cost for each user click to promote their products on any e-commerce platform. For this reason, since the optimization of corporate economics is an extremely crucial part of the overall business value optimization, in this chapter we decided to focus on an important subtype of value-aware value-aware systems that we called economic recommender systems. In contrast to our previous contribution on VARSs (see Chapter 3), in this chapter we offer a more in-depth examination of the technical approaches underlying economic recommendation systems, which we divided according to different dimensions of analysis by exploiting a systematic review approach. Moreover, we also focused on the methodological aspects of offline and online performance evaluation that were not addressed previously since the spectrum was too broad to delve into such topics.

The main contributions of this work can be summarized as follows:

- We offered a systematic literature review focused on economic recommender systems which should serve researchers and practitioners alike as a starting point to understand the state-of-the-art.
- We categorized existing works into different dimensions of analysis to describe technical approaches, evaluation methodologies, limitations of today's research and possible future directions.

The remainder of the chapter is organized as follows. In Section 4.1 we describe the systematic review methodology. In Section 4.2 we discuss existing ECRSs technical approaches. In Section 4.3 we analyze evaluation methodologies that can be used to evaluate such systems. In Section 4.4 we point out some open challenges and future research directions. Finally, Section 4.5 ends the chapter with a summary of findings.

The manuscript entitled "*Economic Recommender Systems – A Systematic Review*" has been submitted and is currently under review in the journal *Electronic Commerce Research and Applications* (2022 impact factor of 6).

4.1 Systematic Review Methodology

Similarly to the previous chapter (see Section 3.1), for this study we followed a systematic review process based on PRISMA guidelines [340] to identify all available research relevant to our purposes. However, we further specialized the research by decomposing the study of economic recommender systems into different dimensions of analysis. Below we discuss how we decomposed the study of ECRSs into various dimensions, the underlying research questions, the search queries used, the eligibility criteria for article inclusion, the overall article analysis and selection process, and the possible limitations of the survey.

4.1.1 Decomposing Economic Recommender Systems in Various Dimensions of Analysis

Economic recommender systems can be characterized by several interrelated topics. To identify relevant articles, we therefore followed an inductive process starting from two related surveys [106, 214], decomposing ECRSs into different *dimensions of analysis* (*DAs*). As Figure 4.1 shows, we identified five types of approaches that can be divided into customer and organization-oriented ones, depending on their main focus. Customer-oriented approaches aim to integrate RSs models with purchasing behavior mechanisms to generate more relevant recommendations that could in turn lead to more value for the firm. Instead, organization-oriented ones make use of specific organizational strategies to directly or indirectly optimize business KPIs. Below, we explain the rationale behind each of them.

- **DA1**: **Price Sensitivity** approaches aim to explicitly consider customers' price preferences in the recommendation process. In fact, price is one of the variables that most strongly influence customers' buying behavior [19, 273]. For example, customers are often willing to pay more for certain types of items based on presumed greater utility, better aesthetics, brand prestige, supplier reliability, or a combination of various factors [217]. By considering customers' price sensitivity in the algorithms [214], more accurate and relevant recommendations could directly increase the probability of purchase and thus lead to higher sales revenue for the organization.
- **DA2**: **Economic Utility Modeling** approaches aim to explicitly consider the utility of recommendations for the customer in accordance with an economic perspective. There are many utilitarian dynamics [388] related to the particular type of purchased products [151, 436]. For example, if a customer has just purchased a computer or a smartphone, it is very likely that he or she will not purchase the same or a similar product again within a short time. Conversely, there are other products, such as dog food or diapers,

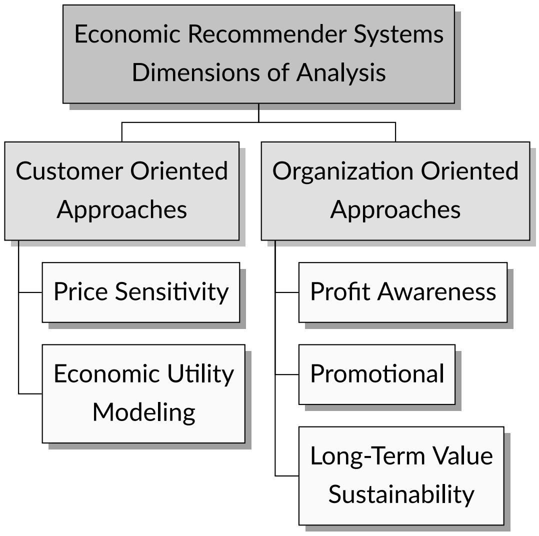


Figure 4.1: Economic recommender systems dimensions of analysis.

for which he or she is very likely to continue buying repeatedly for an extended period of time. Generating more relevant recommendations by considering the customer's utilitarian behavior could increase conversion rates and generate more profits for the firm.

- **DA3**: **Profit Awareness** approaches aim to directly incorporate profit information into the recommendation models. In fact, profit (i.e., sales revenue minus costs) is one of the most important business KPIs for a successful enterprise [410]. Depending on the particular level of this indicator, a company may or may not invest in research and development to grow the business, attract investors to finance its operations, obtain possible financing from banks, and many other issues of strategic interest to entrepreneurs and managers [154]. Overall, generating more profitable recommendations by explicitly considering profit information could directly optimize the organization's economic goals.

- **DA4**: **Promotional** approaches generate recommendations while strategically setting the prices of certain products or focusing the customer's attention on certain brands or promotions. For example, the company can offer certain products at a discounted price (individually or in bundles) to incentivize impulsive buying behaviors [159, 314]. Similarly, the firm can make customers aware of certain products that they would be unlikely to discover on their own and indirectly trigger a possible purchase in the future [246]. Both approaches can be integrated into the recommendation process to optimize profit.

- **DA5**: **Long-Term Value Sustainability** approaches aim to generate recommendations considering a long-term economic perspective. In fact, long-term sustainable business growth is one of the most important aspects for a company [288, 337, 363]. For example, a company may be interested in making customers progressively purchase more and more products and services over time to increase their customer lifetime value. Generating recommendations by considering such long-term economic goals of the company thus has the potential to stimulate business growth in a sustainable way over time.

ID	Dimension	Search Query	Scopus	IEEE	Springer	ACM	Total
DA1	Price Sensitivity	((“recommender system”) AND (“price preference” OR “price sensitivity” OR “price elasticity” OR “willingness to pay” OR “price-aware”))	670	2	469	57	**1198**
DA2	Economic Utility Modeling	((“recommender system”) AND (“economic”) AND (“utility theory”))	104	0	188	17	**309**
DA3	Profit Awareness	((“recommender system”) AND (“multi-stakeholder” OR “profit-aware” OR “value-aware”))	351	5	290	63	**709**
DA4	Promotional	((“recommender system”) AND (“dynamic pricing” OR “price personalization” OR “product bundling”))	483	0	423	29	**935**
DA5	Long-Term Value Sustainability	((“recommender system”) AND (“customer lifetime value” OR “RFM” OR “cumulative profit” OR “long-term value”))	619	4	450	35	**1108**

Table 4.1: Search queries and results divided by online database of the different dimensions of analysis on which this article focuses. Queries were run on *May 12, 2023* looking for all documents published since *January 1, 2000*.

4.1.2 Research Questions

Having identified these dimensions of analysis, we aim to review the state-of-the-art of ECRSs research by answering the following *research questions* (*RQs*):

- **RQ1**: What technical approaches are used to build ECRSs?
- **RQ2**: What evaluation methods are used to assess the performance of an ECRS?
- **RQ3**: What are the main challenges and future research directions in the area of ECRSs?

4.1.3 Search Query

As mandated by the PRISMA systematic review process, in our survey we aim to answer previous RQs by querying online libraries such as Elsevier Scopus, IEEE Xplore, Springer Link, and ACM Digital Library to identify all available yet relevant articles. We created a *search query* for each of the previous DAs by analyzing the most recurring key terms identified in a series of specialized articles extracted from the literature of two related surveys [106, 214]. In Table 4.1, we report the used search queries and the number of identified documents.

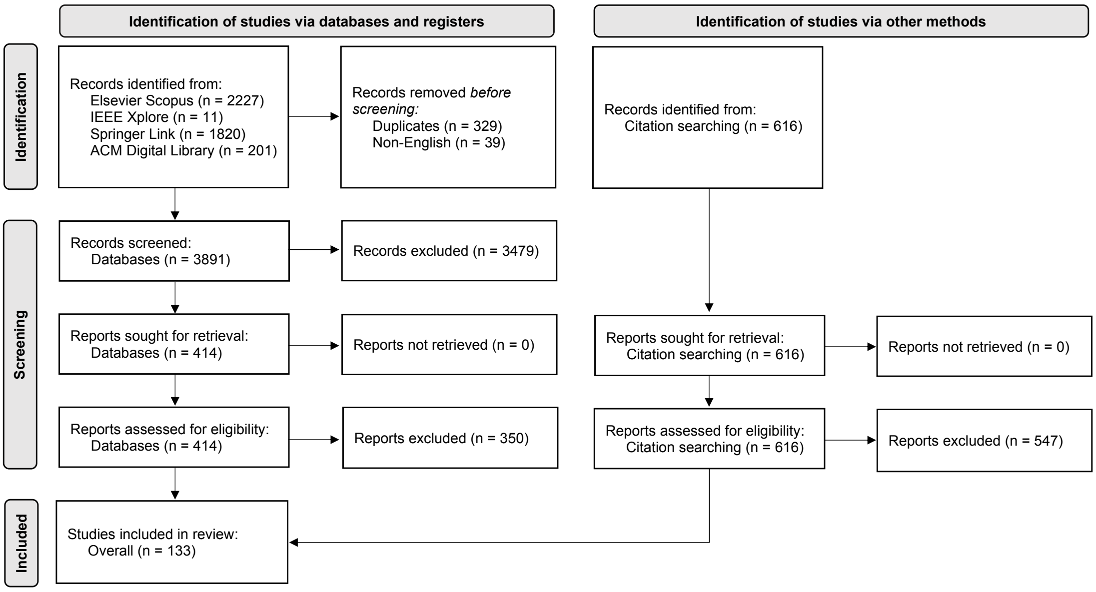


Figure 4.2: Systematic review PRISMA flow diagram of economic recommender systems surveyed literature.

4.1.4 Eligibility Criteria

To be included in the review, articles must meet the following *eligibility criteria* (*EC*):

- **EC1**: Articles must focus on research questions related to one of the dimension of analysis of ECRSs.
- **EC2**: Articles must explicitly mention the business KPIs included in the search queries.
- **EC3**: Articles must be unique, written in English, and the full content must be accessible to the authors.
- **EC4**: Articles must be peer-reviewed by either scientific journals or conferences.
- **EC5**: Graduate theses and doctoral dissertations are not included.

4.1.5 Article Selection Process

As shown in the PRISMA flow diagram in Figure 4.2, we followed a multi-stage process to identify all the relevant resources included in this review. At each stage we applied the eligibility criteria defined in Section 4.1.4 to discard non-relevant documents. In the first identification phase 2227 articles from Elsevier Scopus, 11 articles from IEEE Xplore, 1820 articles from Springer Link, and 201 articles from ACM Digital Library were identified for subsequent analyses. In this phase, 329 duplicated records and 39 non-English articles were identified and removed as a result of applying EC3, EC4 and EC5 criteria. In the second screening phase, the titles and abstracts of the remaining 3891 articles were analyzed, and 3479 records were removed after applying EC1 and EC2 criteria because the covered topics were not relevant to the present review. In this phase, 414 articles were then sought for

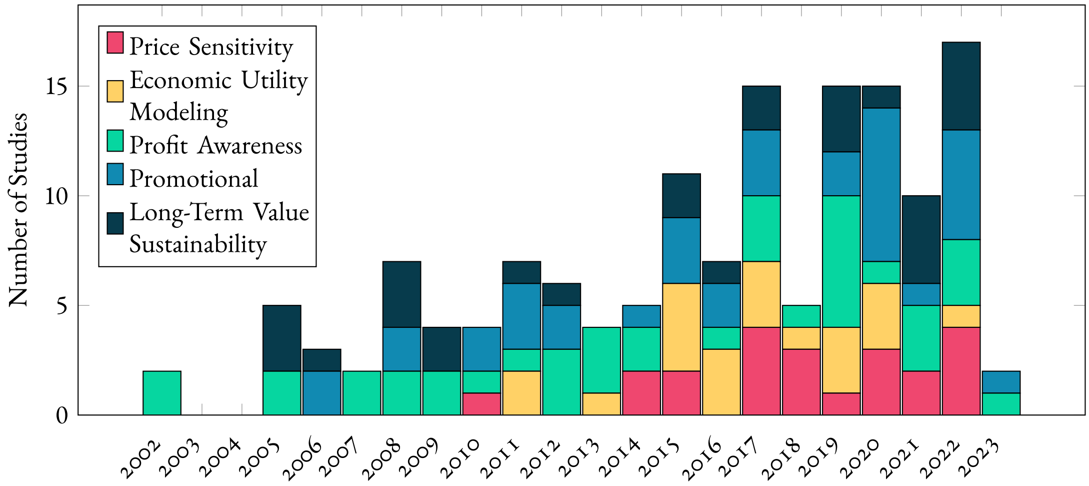


Figure 4.3: Distribution of surveyed papers per year divided by dimension of analysis.

retrieval and assessed for eligibility, excluding 350 articles after full text reading as they did not met EC1 and EC2 criteria. From this subset of 64 eligible articles, an additional 616 articles were identified by searching for references in their bibliographies. These articles were then assessed for eligibility, removing 547 records after reading the full text as a result of applying EC1 and EC2 criteria. At the end of this overall process, 133 studies were included in the review. In Figure 4.3, we show some statistics of the references obtained at the end of the analysis process by reporting the distribution by year of surveyed papers divided by subdimension of analysis. As can be seen from the figure, there is a growing interest in the literature for all ECRSs dimensions.

4.1.6 Study Limitations

The possible study limitations (SL) are the following:

- **SL1**: Articles were primarily selected from Elsevier Scopus, IEEE Xplore, Springer Link, ACM Digital Library, and from reference searches in the bibliographies of articles that passed the screening stage. Additional online libraries may be considered in future research.

- **SL2**: The study does not cover preprints, non-English articles, non-accessible articles, graduate theses, doctoral dissertations, industry products, and demos.

- **SL3**: Other dimensions of analysis of ECRSs beyond those identified in Section 4.1.1 are left for future extensions of this work.

Approach	Price Sensitivity	Economic Utility Modeling	Profit Awareness	Promotional	Long-Term Value Sustainability
In-Processing	[77, 78, 87, 153, 167, 168, 293, 383, 430, 431, 447, 476, 489, 490]	[109, 151, 293, 431, 436, 456, 478, 479, 482]	[16, 60, 69, 88, 96, 119, 165, 206, 268, 275, 289, 320, 355, 361, 437, 438, 439, 448]	[15, 32, 36, 73, 74, 110, 146, 148, 150, 157, 230, 248, 249, 280, 296, 383, 406, 431, 452]	[176, 189, 211, 212, 221, 229, 353, 411, 449, 470, 471, 484, 500]
Post-Processing	[34, 72, 175, 217, 240, 394, 429, 459]	[7, 102, 367, 459, 481, 486, 487]	[33, 82, 103, 155, 205, 240, 285, 287, 290, 291, 316, 366, 390, 434, 435, 459, 480]	[8, 34, 35, 41, 108, 131, 177, 217, 223, 224, 225, 235, 390, 481, 497]	[39, 200, 202, 343]

Table 4.2: ECRSs studies divided by dimension of analysis and algorithmic approach.

4.2 Economic Recommendation Algorithms

In this section, we discuss the underlying algorithmic approaches to each of the ECRSs dimensions of analysis introduced in Section 4.1.1, i.e., price sensitivity, profit awareness, promotional, long-term value sustainability, and economic utility modeling, referring to the notation proposed in Section 2.2.2.

In Table 4.2 we report the studies identified by the present survey that propose technical approaches, categorized by dimension of analysis and algorithmic method. The approaches can be divided into in- and post-processing* methods, depending on the time the economic value optimization occurs. In-processing approaches aim to incorporate economic aspects directly into the models, either by extending the objective function of known algorithms (e.g., by introducing new variables or regularizers) or by developing entirely new algorithms. The underlying algorithms may be based, for example, on supervised or reinforcement learning paradigms. Post-processing approaches, on the other hand, can be mounted on top of any recommender and aim to transform the recommendations generated from the baselines by applying specific heuristic economic criteria. These may incorporate the economic value by simply re-ranking the output of the original algorithm or by exploiting also additional learning models.

Analyzing the distribution of the studies in Table 4.2, we can make some observations. In particular, it can be noted that there are several relevant works for all the dimensions of analysis. In addition, in-processing and post-processing methods are equally used across all dimensions. This implies that the research field is broad and that there are various lines of active research. Overall, given that there is a substantial number of works in each dimension of analysis, we are confident that our categorization scheme properly reflects the various types of activities in this research area.

*Pre-processing methods may also exist in industry, e.g., when a recommendation provider wants to rule out certain unprofitable items. Our literature search, however, did not surface such approaches.

4.2.1 Price-Sensitivity Algorithms

Price is one of the variables that most influence customers' buying behavior [19, 273]. Accordingly, many studies in the literature [78, 153, 167, 240, 490] propose algorithms to explicitly consider customers' price sensitivity, as more accurate and relevant recommendations (i.e., in terms of being in the right price range) could directly increase the probability of purchase and thus lead to higher sales revenue for the organization. Below, we give some insights on how these methods work by discussing a set of selected articles.

In-Processing Price-Sensitivity Methods

Most of the approaches used to generate price-sensitive recommendations are based on in-processing algorithms. The main characteristic of these algorithms is that price sensitivity is incorporated directly into the model.

In particular, this methodology proved particularly flexible to be applied to the well-known *Matrix Factorization* (*MF*) [242, 243] model. The original model estimates the expected interest of the user toward a given item via the dot product of latent factor vectors. These are traditionally learned through a dimensionality reduction algorithm applied to the user-item interaction matrix. Considering price-sensitive methodologies based on MF, for example, one paper [153] proposes to incorporate cost factors* into the model's objective function to generate more accurate travel tour recommendations. The experiments reported by the authors indicate that explicitly incorporating cost factors improves the overall accuracy of the recommendations when compared with a plain MF model. Also, extending MF, other papers in the literature [77, 78] propose incorporating customers' price preferences explicitly into the objective function through the use of particular regularizers. However, whereas previously the purpose was to enhance the overall performance of the system, here the study is about the use of price preferences to make recommendations in product categories that the user has never explored (*transfer learning*). In particular, according to the authors, generating recommendations for customers' unexplored product categories can cause significant performance drops (-40%) if traditional algorithms are used, since the learned product user preferences are difficult to transfer from one category to another. Instead, explicitly incorporating customers' price preferences into the objective function can help to significantly improve (+43%) performance on unexplored categories compared to state-of-the-art baselines.

Other studies in the literature [167, 168] propose incorporating customers' price preferences within existing *context-aware* recommendation algorithms [251]. According to an experimental study with real customers in the food & beverage field [168], explicitly incorporating discount sensitivity into the algorithms can help to significantly improve performance in a coupon recommendation task when compared to the CAMF method [37], i.e., a context-aware variant of matrix factorization. Specifically, in the domain of location-based deals, the analysis shows that the most important feature for predicting purchase probability is the discount-to-distance ratio: the higher the discount offered by the store, the more likely the customer is to travel longer distances to obtain it. However, as is well known in the literature, context variables often depend on the considered business domain. In particular, eBay.com has some unique characteristics [167]. In this multi-seller platform, the same products are offered at various prices simultaneously by various sellers with different reputation scores. According to a study [167], in this business domain, incorporating customers' *willingness-to-pay* (*WTP*), discounting, and seller repu-

*Note that here we respect the original paper's terminology by referring to the cost, but actually the cost for the user is simply the item's price.

Ref	Re-Ranking Method	Description
[394]	$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot prs_{u,i} \qquad (4.1)$$	Recommendation of the top-k items with the highest price-sensitivity weighted predicted scores.
[429]*	$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} w_1 \cdot \hat{x}_{u,i} + w_2 \cdot prs_u \qquad (4.2)$$	Recommendation of top-k items that best balance customer's expected interest and price sensitivity.
[240]	$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot \left(\left(1 + \log_{10}\left(0.1 + \frac{0.9 \cdot prc_i}{cst_i}\right)\right)^{\zeta} + \right. \\ \left. + \left(1 + \log_{10}\left(0.1 + \frac{0.9 \cdot prc_i}{\bar{\mathbf{prc}}_u}\right)\right)^{\pi} \right) \qquad (4.3)$$	Recommendation of the top-k items that best balance the customer's expected interest and price sensitivity with the organization's profitability.

Table 4.3: Price-sensitive re-ranking methods. *The formula captures the main essence of the described approaches.

tation features into a context-aware recommender can help to significantly improve the accuracy of predictions, with an 84% improvement over MF models.

In addition, recent studies [476, 489, 490] propose incorporating customers' price preferences into algorithms by exploiting Graph Neural Networks (GNNs) [143]. Specifically, in two related studies [489, 490], it is proposed to construct a GNN-based recommender by building a heterogeneous graph consisting of different types of nodes: customers, items, prices, and product categories. The key idea is to propagate price influence from prices to users by leveraging items as a bridge so that price preferences are implicitly encoded into the embeddings. The use of price-sensitive GNNs is also exploited in the field of session-based recommendations [476]. For all studies based on GNNs [476, 489, 490], the models are able to generate slightly more relevant recommendations than the baselines. However, as various authors pointed out, it is difficult to handle heterogeneous information and model complex relationships underlying customer buying behavior, and research still offers many opportunities to develop better-performing models that can fully exploit the potential of GNN-based algorithms.

Post-Processing Price-Sensitivity Methods

A number of price-sensitive recommendation algorithms also make use of post-processing methods. The latter are primarily re-ranking algorithms, which can be applied on top of any price-agnostic recommender baseline.

In this domain, it is proposed, for example, to generate recommendations by weighting the expected interest $\hat{x}_{u,i}$ by the price-sensitivity $prs_{u,i}$. The latter is a particular variable, learned through a different model, indicating how price-sensitive a given user is to a given item (see Eq. (4.1)) [394]. A similar approach is also proposed in another study [429]. However, in this case, the price-sensitivity variable prs_u depends only on the customer and not on the item (see Eq. (4.2)). In addition, it is necessary to use another regression model to learn how to properly weigh (through w_1, w_2 coefficients) the price-sensitivity with the user's expected interest. Both studies show that through the use of price-sensitivity algorithms, more relevant recommendations can be obtained.

Recently, another research [240] proposes a hybrid approach (see Eq. (4.3)) combining the price-sensitive and the profit-aware* subdomains. This approach weighs the expected interest $\hat{x}_{u,i}$ by balancing a user price preference factor $\frac{prc_i}{\bar{\mathbf{prc}}_u}$ with a profitability factor $\frac{prc_i}{cst_i}$, where $\zeta, \pi \in [-1, 1]$ in Eq. (4.3) are regularization parameters. In particular, considering $\bar{\mathbf{prc}}_u$ as the average user price, the first factor captures the difference between the customer's typical price level and the actual item's price. The second factor, $\frac{prc_i}{cst_i} = 1 + \frac{prf_i}{cst_i}$, captures how much an item's sale is able to repay the underlying cost and bring profit to the organization. In this way, it becomes possible to effectively balance the interests of customers with those of the organization because the increase in profitability that traditionally adversely affects the relevance of recommendations is more than offset by the increase in the latter due to the influence of price preferences.

In Table 4.3, we formally characterize the three discussed price-sensitive re-ranking methods.

4.2.2 Economic Utility Modeling Algorithms

In the economic literature [299], user behavior is often modeled using utilitarian theories to construct systems that can describe and/or optimize certain dynamics. According to the *Rational Choice Theory* (*RCT*), at each time instant, a rational user, when faced with a set of alternatives, will choose those with the highest utility for him or her [388]. Accordingly, many studies in the literature [151, 436, 482] propose algorithms that explicitly consider the customer's utilitarian behavior to generate more useful recommendations that can in turn increase conversion rates and profitability. Below we give some insights on how these methods work by discussing a few selected articles focused, respectively, on multi-attribute, repurchase, and complementary recommendations.

In the field of RSs, many studies in the literature assume that the utility $\rho_{u,i}$ of a product to a customer depends on his or her purchase history [436]. Most existing RSs recommend for each user u a list $\mathbf{y}_{u,k}$ consisting of the top-k items (see Eq. (2.3)) with the highest predicted scores $\hat{x}_{u,i}$. The list $\mathbf{y}_{u,k}$ is traditionally selected from a set of items with which the user has never interacted before. Interpreting this assumption from the perspective of economic utility theory (see Eq. (4.4)) [436], then, the utility $\tau(\mathbf{y}_{u,k})$ of a recommendation $\mathbf{y}_{u,k}$ is nothing but the sum of the predicted scores, i.e., $\rho_{u,i} = \hat{x}_{u,i}$. In this case, a recommendation $\mathbf{y}_{u,k}$ generated by optimizing the total utility of a set of k recommended items optimizes the expected user interest estimated by any recommendation algorithm.

However, in addition to the previous utility definition, alternative definitions are recently emerging in the literature. For example, in the field of *Multi-Criteria Recommendation Systems* (*MCRSs*) [17], in the presence of a set $\mathcal{A}$ of attributes associated with items, various studies in the literature [109, 120, 207, 387, 486] propose to generate recommendations by exploiting the *Multi-Attribute Utility Theory* (*MAUT*) [234]. MAUT is one of the most widely used utilitarian theories in decision making, which aims to weigh a set of relevant variables to determine the overall utility of each alternative. In the context of recommendations, in particular, the overall optimized utility (see Eq. (4.5)) in this case depends on the utility $\rho_{i,a}$ of the single attribute a of item i, and a weight $w_{u,a}$ that each user can provide to indicate the importance of that attribute.

Other studies focus on the problem of repeated purchase recommendations [151, 436, 478, 482]. Unlike traditional RSs, algorithms developed for this task generate recommendations by also considering items that the user

*We discuss profit-aware methods in Section 4.2.3.

Ref	Name	Utility Function	Description
[436]	Standard	$$\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \rho_{u,i} \quad (4.4)$$	Function that determines the utility of top-*k* recommended items based on the user expected interest.
[207]	Multi-Attribute	$$\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \sum_{a \in \mathcal{A}} w_{u,a} \cdot \rho_{i,a} \quad (4.5)$$	Function that determines the utility of top-*k* recommended items based on the utility of each item's feature to the user.
[436]*	Constant Elasticity of Substitution	$$\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \rho_{u,i} \cdot qnt_{u,i}^{\xi_i} \quad (4.6)$$	Function that determines the utility of top-*k* recommended items based on the quantity of items purchased by the user up to a specific time.
[478]	King-Plosser-Rebelo	$$\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \rho_{u,i} \cdot \ln(1 + qnt_{u,i}) \quad (4.7)$$	A variant of the previous function that determines the utility of top-*k* recommended items based on the quantity of purchased items.
[482]	Multi-Product	$$\tau(\mathbf{y}_{u,k}) = \frac{1}{\lvert \mathbf{y}_{u,k} \rvert} \sum_{i,j \in \mathbf{y}_{u,k}: i \neq j} \Big(c_{i,j} \cdot qnt_{u,i}^{1-b_{i,j}} + (1 - c_{i,j}) \cdot qnt_{u,j}^{1-b_{i,j}} \Big)^{\frac{1}{1-b_{i,j}}} \quad (4.8)$$	Function that determines the utility of top-*k* recommended items by considering the complementary and substitutability relationships of a potential purchase for the user.
[151]	Marginal Utility per Dollar	$$\tau(\mathbf{y}_{u,k}) = \sum_{i \in \mathbf{y}_{u,k}} \frac{\tanh\left(\rho_{u,i}\right) \cdot rsk_{u,i}}{(1 + qnt_{u,i}) \cdot prc_i} \quad (4.9)$$	Function that determines the utility of top-*k* recommended items by considering the customer's risk aversion.

Table 4.4: Economic utility functions from rational choice theory. *The formulas capture the main essence of the described approaches.

already purchased in the past. In particular, it is observed that the repurchase cycle of some products may follow the *Law of Diminishing Marginal Utility* [299, 436]. According to this theory, many products have decreasing utility for the user as the quantity of purchased products increases (e.g., computers, cell phones), while others, instead, are likely to be purchased frequently over time (e.g., baby diapers, pet food). Using the standard utilitarian criterion in Eq. (4.4) it is not possible to model this behavior. Indeed, in this case, the usefulness of recommendations for the user highly depends on the quantity $qnt_{u,i}$ of item i purchased by him or her until a specific time. In this context, promising results can be obtained by modeling the repurchase cycle through the *Constant Elasticity of Substitution Utility Function* [420]. This allows the decreasing marginal utility of product i to be properly modeled through a parameter $\xi_i \in [0, 1]$ associated with item i (see Eq. (4.6)). This parameter can be learned by extending the MF objective function. In this way, the algorithm can explicitly consider the decreasing utility of certain products for the user and generate more relevant recommendations.

With similar methodologies, other utilitarian functions are also used in the literature to model customer purchasing behavior [151, 478, 482]. However, these studies focus on different objectives. For example, one study [478] proposes three different business cases (i.e., e-commerce, P2P lending, freelancing) that exploit the *King-Plosser-Rebelo Utility Function* (see Eq. (4.7)) to optimize the *Total Surplus*, i.e., an indicator that considers both the usefulness of the recommendations for the customer and the profit for the producer. Another study [482] in contrast propose to use the *Multi-Product Utility Function* (see Eq. (4.8)) in order to also consider any complementary and substitutability relationships among the recommended products. In the equation, the variables $b_{i,j}$ and $c_{i,j}$ are additional parameters that the recommendation algorithm can jointly learn with the latent factors to model the indifference curves between pairs of products, i.e., how much the increase in one product affects the relative marginal utility of another product. Finally, one study [151] proposes using the *Marginal Utility per Dollar Function* (see Eq. (4.9)). This function considers the price prc_i of item i and a risk attitude coefficient $rsk_{u,i}$ to model customers' risk-aversion, i.e., the tendency of consumers to spend only a small portion of their total wealth on a single purchase.

In Table 4.4, we formally characterize the utility criteria discussed above.

4.2.3 Profit-Awareness Algorithms

Profit is one of the most important business KPIs for a successful enterprise [410]. Accordingly, many studies in the literature [82, 96, 165, 240, 268, 320] propose profit-aware recommendation algorithms to directly optimize the firm's profitability. Below we give some insights on how these methods work by discussing a few selected articles.

In-Processing Profit-Aware Methods

Profit-aware in-processing approaches in the literature are quite heterogeneous, scattered, and there are several parallel lines of research. Below, we offer a brief overview of major research directions in this area.

Some early studies [206, 355, 459] exploit *Association Rules* [196]. According to this particular methodology [344], recommendations are generated through a frequentist approach based on statistical support and confidence constructs [338]. One of Amazon's most prominent recommenders, i.e., "*customers who bought this item also bought*", is seemingly based on association rules. In particular, many studies in the literature [88, 437, 438, 439] propose to generate association rules while also optimizing profitability. The main methods incorporate profit considerations when weighting the rules [68]. However, unlike modern RSs based on collaborative and content-based filtering algorithms, association rules [439] are not personalized, i.e., different users do not get different recommendations. In addition, association rules may generally face challenges when the total number of recommendable items is very large.

Other earlier studies [16, 60, 361] propose graph-based approaches. In particular, one research [16] focuses on social networks. The proposed algorithm is designed to explicitly optimize the value of recommendations in customer-product graphs. However, in the study, profit is operationalized through non-monetary metrics. Another relatively recent approach based on graphs [361] is developed specifically for the taxi industry. In this particular application domain, if we assume an hourly rate, a taxi driver's profit depends solely on the hours billed to customers: simply put, it is critical for a taxi driver to minimize the distance to find a customer and maximize the distance traveled with a customer on board. The proposed algorithm recommends pick-up points for taxi drivers in order to maximize the profit of driving routes while balancing the potential congestion resulting from multiple requests from different customers at the same location.

More recently, a study [69] proposes a profit-aware RS based on collaborative filtering. The algorithm is based on an extension of the well-known neighbor selection criterion of the user-based nearest-neighbor collaborative filtering model [326]. The original algorithm calculates the predicted score based on a weighted sum of similarities between users belonging to a given neighborhood. The authors of [69] instead propose to calculate the predicted scores by selecting the neighbors that would allow the generation of the highest value-weighted expected interest. Although the focus of the paper is on shilling attacks, i.e., attacks by malicious users who generate biased ratings to influence recommendations for their interests, the subprocedure for selecting the most valuable users can be used to generate more profitable recommendations.

Other recent approaches [268, 275, 448] are based on *Learning To Rank* (*LTR*) [93]. This is a well-known technique in *Information Retrieval* (*IR*) [42, 238]. IR algorithms aim to help users to find the most relevant items based on specific search queries. In particular, one study [448] uses this methodology in a product search application context. The proposed algorithm integrates the price into the objective function in order to optimize the overall sales revenue of an e-commerce. This technique is later applied [268, 275] also to generate recommendations without the need to anchor them to an underlying search query. For example, one study [275] proposes a multi-objective algorithm that is able to optimize multiple objective functions simultaneously through LTR. The algorithm is Pareto-efficient, i.e., it optimizes each objective (e.g., CTR and GMV*) one at a time, with the constraint that no single objective can be further improved without affecting others.

Finally, some studies [96, 320] propose using profit-aware multi-objective *Evolutionary Algorithms* (*EAs*) [199, 488]. One of these [320] is based on *Non-dominated Sorting Genetic Algorithm II* (*NSGA-II*). A more recent one [96] is based on *Multi-Objective Artificial Bee Colony* (*MOABC*). In these cases, the optimization target is a combination of the item's profit and the user's expected interest. Both algorithms obtained very promising offline results on the overall profit improvement, although the comparison was performed exclusively with a traditional user-based collaborative filtering algorithm [326].

*We provide the definition of the most frequently used online metrics in Table 4.8.

Post-Processing Profit-Aware Methods

In the context of this survey, many profit-aware approaches rely on post-processing re-ranking methods. As mentioned earlier, these approaches consider the recommender baseline as a black box and generate recommendations by exploiting a combination of certain heuristics.

All examined profit-aware approaches are based on a simple but important assumption [106, 214]: the items most relevant to the user are often not those of the highest business value to the organization. Consequently, prioritizing the highest-profit items in recommendations would allow for increased business profitability as a result of actual user purchases of those items. In one of the earliest approaches [82, 316] it is proposed to weight the probability of purchase (i.e., the estimated expected interest) by profitability in order to maximize an average expected profit (see Eq. (4.10)). This approach should make it possible to provide more profitable recommendations than those generated by traditional RSs. Experiments in a synthetic dataset about a subset of groceries transactions show encouraging results: the proposed algorithm was able to increase profitability without excessively impacting the relevance of recommendations. However, as also reported by the authors [82, 316], the interests of customers and the organization must be balanced appropriately. In fact, the organization could risk losing loyal customers should they feel dissatisfied with overly biased recommendations toward higher-value items and decide to leave the platform.

To mitigate this drawback, and thus to avoid providing completely irrelevant recommendations, various studies propose more or less straightforward extensions of Eq. (4.10) based on constrained optimization techniques. One of the earliest papers [103] proposes a constrained re-ranking method based on the *Dice* coefficient (see Eq. (4.11)). This can help prevent the system from providing recommendations that are too dissimilar from the original ones based on a threshold ν. However, the study is based on various simplifying assumptions and does not provide an empirical evaluation of the approach. In two related studies [434, 435] instead, it is proposed to maximize profitability under customer satisfaction and budget constraints (see Eq. (4.12)), where γ and ϱ_u are two thresholds used to keep the probability of purchase and the price of items within certain ranges, respectively. In particular, an expert system is proposed where different optimization goals can be specified in order to optimize profitability or balance profitability and satisfaction in order to achieve a win-win situation for suppliers and customers. A similar variant of this approach (see Eq. (4.13) and Eq. (4.14)) is also proposed in two related studies [155, 214] where the short- and long-term profit-relevance tradeoff is investigated through the use of simulations. In Eq. (4.14), $\delta \in [0, 1]$ is a regularization parameter.

In addition, other studies [290, 291, 480] propose algorithms to address the problem of sponsored recommendations. In this scenario, a supplier who decides to sponsor its products pays the platform for each user interaction. One study [291] in particular proposes a multi-objective post-processing re-ranking algorithm (see Eq. (4.15)). In the equation, $y_{u,i}$ is a decision variable ($y_{u,i} = 1$ iff $i \in \mathbf{y}_{u,k}$), and $spn < k$ is the maximum number of sponsored items that can be included in the recommendation list. The algorithm is designed to balance the recommendation of high ad revenue sponsored items with the user's interests.

In Table 4.5, we formally characterize the profit-aware re-ranking methods discussed above.

Ref	Re-Ranking Method		Description
[82]	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot prf_i$	(4.10)	Recommendation of the top-k items with the highest profit-weighted predicted scores.
[103]	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot prf_i$ s.t. $dice(\hat{\mathbf{x}}_u, \hat{\mathbf{x}}_u^\intercal \mathbf{prf}) \geq \nu$	(4.11)	Recommendation of the top-k items with the highest profit-weighted predicted scores under the *Dice* similarity constraints.
[435]*	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} prf_i$ s.t. $\hat{x}_{u,i} \geq \gamma, \quad prc_i \leq \varrho_u$	(4.12)	Recommendation of the top-k most profitable items under customer expected interest and budget constraints.
[214]*	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \hat{x}_{u,i} \cdot prf_i$ s.t. $\hat{x}_{u,i} \geq \gamma$	(4.13)	Recommendation of the top-k items with the highest profit-weighted predicted scores, keeping the expected interest above a certain threshold.
[155]	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} \partial \cdot \hat{x}_{u,i} + (1 - \partial) \cdot prf_i$	(4.14)	Recommendation of the top-k items that best balance consumer and business value.
[291]*	$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \partial \cdot y_{u,i} \cdot \hat{x}_{u,i} + (1 - \partial) \cdot y_{u,i} \cdot prf_i$ s.t. $\sum_{i \in \mathcal{I}} 1(y_{u,i} \cdot prf_i > 0) \leq spn \quad (\forall u \in \mathcal{U})$	(4.15)	Recommendation of the top-k items that best balance consumer and business value, keeping the maximum number of sponsored items below a certain threshold.

Table 4.5: Profit-aware re-ranking methods. *The formulas capture the main essence of the described approaches.

4.2.4 Promotional Algorithms

Promotional methods [308, 314] aim to increase sales figures by promoting products and services to the most appropriate customer segments. We identify three main strategies in the RSs literature that can be used to optimize profit and related business KPIs. *Pricing methods*, can be used to offer products at a discounted price or to strategically adjust prices in order to increase the market demand. *Bundling methods*, are special pricing methods that are applied to product bundles. *Brand-awareness methods*, finally, can be used to focus customers' attention on the organization's products in order to generate extra sales. Below we give some insights on how these methods work by discussing a few selected articles in each category.

Pricing Methods

As discussed earlier in Section 4.2.1, price is one of the most influential variables of customer buying behavior, and considering this variable explicitly would allow for recommendations more in line with customers' interests. However, while the previous section focuses on customer-oriented methodologies that integrate price sensitivity as additional information in order to generate more relevant recommendations, in this section we instead discuss promotional techniques that an organization might want to apply to incentivize the purchase of certain products by strategically setting the prices [46, 157]. In the following, we describe two organizational strategies referring to: (a) *occasional discounting*; (b) *personalized dynamic pricing*.

One of the most commonly used promotional strategies to incentivize product purchases is to offer occasional discounts [86], for example at certain times of the year (e.g., winter sales) or special events (e.g., Black Friday). In the context of RSs many studies [177, 217, 223, 225, 383, 431, 452] aim to generate recommendations while considering discounts. Some studies propose, for example, to use re-ranking algorithms [217] to promote products on sale or in-processing methods [431] based on adaptations of MF-based models to explicitly consider customers' discount sensitivity [383]. Another method is proposed in two related studies [223, 225]. In particular, as noted by the authors, there may be inter- and cross-category effects when discount products are bought. Thus, especially in e-commerce, organizations can exploit RSs to incentivize customers to buy discount products but also those products that are related to them but not on sale (e.g., camera on sale and full-price lens). A similar analysis is also made [177] to determine the optimal shipping-fee discount to attract customers to the platform and encourage them to purchase products related to the discounted ones.

While discounts may be occasional and the same for all customers, some methodologies are proposed in the RSs literature to generate dynamic customer-specific prices in order to strategically promote certain products and generate higher profits. In this context, some initial studies [34, 35] propose to use survey-based techniques (*conjoint analysis*) to estimate customer WTP (recall Section 4.2.1) and filter items that are priced higher than WTP in the ranking. The authors also discuss some possible configurations of the algorithm to set the prices based on WTP in order to generate more profit for the organization. However, the proposed pricing model is only theoretical as it is not validated by empirical experiments. Another study [230] proposes a system that classifies customers based on whether they would buy products only if discounted or not. Based on the type of customer, the system can offer a discount in order to incentivize purchases. However, as discussed later [296], the study is based on assumptions that are not feasible in practice: all products have the same price; only two price values are available (i.e., standard and discounted price). Another work [481] proposes a different methodology. The study focuses on a lottery-based mechanism that aims to obtain the exact WTP for one subset of products and then to exploit this information to predict the WTP for another subset of products. In this way, the system can offer a personalized promotion to increase the conversion rate of the latter product subset. The authors report significant results on the potential ability of this system to increase profit over conventional systems. However, the experiments are based on a user study with a low number of users. Finally, another study [8] proposes a dynamic personalized pricing recommender system for information goods (e.g., digital movie rentals). These goods differ from physical goods in that their production and distribution costs are negligible and they can be copied, rented, and resold easily. In this context, traditional markup-based pricing methods (i.e., cost plus margin) are not effective because there is no true underlying unit cost. The proposed system first classifies customers according to their WTP and quality sensitivity (e.g., whether they prefer a premium version of the same product). Then it calculates a personalized price to incentivize purchase.

Bundling Methods

One frequently used promotional strategy [426] to increase sales revenue of certain products is to offer them at a discount if purchased in bundles [185]. In the literature [457] it is proposed, for example, to include in the bundles: (a) products that are complementary to each other in order to incentivize cross-selling; (b) products that are uncorrelated, for example, to clear the stock in the warehouse; (c) the same product in multiple quantities (e.g., 2x1 promotion). Specifically, in RSs research [265], one branch of the literature focuses on recommending bundles to optimize profit by exploiting price modeling techniques. The other branch, in contrast, does not exploit such techniques and focuses solely on optimizing relevance*. In this review, we focus only on bundling approaches that aim to explicitly optimize business KPIs.

Concerning price modeling bundle recommendation techniques, two related earlier studies [146, 148] focus on the development of a shopbot (i.e., comparison shopping agents) capable of offering bundles at a discounted price based on an integer linear programming model. The proposed algorithm is validated using data from Amazon.com and Buy.com reporting significant results from the perspective of potential economic savings of price-sensitive bundle purchasing customers. However, the data sample used is very small, and optimization of business KPIs is not explicitly considered. In contrast, two other studies [224, 497] leverage similar integer programming-based approaches to recommend bundles with the goal of optimizing profitability [224] or any business objective [497]. In particular, considering the case where the bundle can be created directly by the customer by selecting the products of his or her preference, the first study [224] proposes a multistage approach that can dynamically determine the price of the added products in real-time with the goal of maximizing profits for the organization. In contrast, the second study [497] investigates how to incorporate product compatibility and potential cost savings to generate bundles that, if recommended, could optimize certain business objectives (e.g., profitability, revenue, and others). Both studies report results regarding the potential ability of the proposed systems of increasing profitability and conversion rates. In addition, two other approaches [41, 131] are proposed recently. The first approach [41] is based on a collaborative filtering algorithm that integrates demand estimation and price modeling techniques to make recommendations with the goal of jointly maximizing purchase probability and sales revenue considering the customer WTP. The second approach [131] is based on an algorithm that can recommend bun-

*Relevance-based bundling algorithms [406] can be based for example on association rules [132, 227, 460], graph-based approaches [36, 110, 150, 280, 280], GNNs [15, 73, 74] and transformers [32].

dles with customized discounts to customers considering also inventory levels. However, in the former case, the bundle does not offer an additional discount over the full price of the individual products. Instead, the bundle is created exclusively so that the total price of the products inside it is aligned with the customer's WTP to meet his or her price preferences. In the latter case, on the other hand, the evaluation is based on a simulation focused on the aviation industry with a large number of assumptions.

Brand-Awareness Methods

Some methods in the literature can be used to promote the organization's products and services, raise brand awareness, and increase profitability in the long run. These methods can be interpreted by referring to the sales funnel [425]. The sales funnel is a theoretical model that describes the customer journey in different stages according to the type of customer interaction with the organization [347]. Depending on the status of the customer in the sales funnel, it might be advisable to design an RS with different purposes.

If the customer has not yet made the first purchase (which is referred to as the prospect state), it might be promising to maximize the conversion rate by closing the first deal as quickly as possible [181]. At this early stage, recommending the most popular products may not be the best strategy. Since many popular products are commonly purchased together, customers would discover them on their own without the need of a recommendation. Instead, it could be more beneficial to present still popular but unrelated products, optimizing coverage. In this way, it may be possible to attract more customers to the platform and increase the probability they make their first purchase.

Once the customer has made the first purchase, the company can exploit mechanisms to optimize profits in the long run [56, 165]. One option could be to mainly recommend items with high consumer ratings [219]. However, similarly to the previous case, this may not be the best choice either, as many customers might search for and buy such items anyway. Instead, it might be more valuable to stimulate the purchase of products of possible interest that are likely unknown to the customer [56], e.g., products that do not fall in the top-k but have medium-high ranking positions. This way, the company may get both the revenue from the purchases of products that the customer would discover on their own without the recommendations, and an additional revenue through the purchases that were triggered by the recommendations. With similar objectives, it might also be worthwhile for the company to leverage an RS [165] to launch a marketing campaign with the purpose of promoting new products in the market. Such a system could be designed to select a set of seed consumers for the marketing campaign such that if these seed consumers provide relatively high ratings, the number of other consumers to whom the new product is recommended is maximized.

4.2.5 Long-Term Value Sustainability Algorithms

It is very important for organizations to grow sustainably over time [288, 363]. Accordingly, a number of studies in the literature [200, 212, 353] propose recommendation algorithms that consider temporal dynamics to optimize long-term business value. Many of them rely on the *Customer Lifetime Value* (*CLV*) [49, 53] and other related conceptual models (e.g., *Recency Frequency Monetary - RFM*) from the business literature. CLV represents the expected business value of all future cash flows attributed to a specific customer discounted to the present time.

Similarly to what is found for bundling methods (in Section 4.2.4), some RSs studies propose to exploit CLV to optimize long-term profit [212, 353] while others exploit it solely to optimize relevance* [393, 409]. In this review, we focus only on algorithms that aim to optimize long-term business KPIs. Below we give some insights on how these methods work by discussing a few selected articles. In particular, we first discuss in- and post-processing methods based on supervised learning and then we delve into recent algorithms based on reinforcement learning.

Post-Processing and Supervised Learning Methods for Long-Term Business Value Optimization

Some studies [39, 200, 202, 343] propose post-processing algorithms to maximize the long-term business value of recommendations by exploiting heuristic criteria. In particular, Hosanagar et al. [200] propose an algorithm following this simple but effective intuition: when a customer trusts an RS, the system should bias the recommendations to increase profitability; instead, when the customer trust is below a certain threshold, the system should recommend the most relevant products to restore trust at the expense of profitability. The original study [200] proposes only a theoretical assessment of the profit surplus that can be generated using this algorithm. However, the algorithm's performance is also evaluated in an online study [343] and in a recent post-hoc econometric analysis [39]. These recent studies demonstrate both the effectiveness of the proposed methods in generating higher sales revenue than a content-based filtering algorithm [343] and how trust is positively correlated with higher sales revenue [39].

Other approaches based on supervised machine learning algorithms are also studied to explicitly optimize the long-term business value of recommendations. In particular, in two related studies [211, 212], a recommendation system is proposed to explicitly maximize CLV. The algorithm is designed specifically for subscription-based [211] and transaction-based [212] revenue models. In particular, survival analysis techniques are used to identify frequent purchasing patterns among higher CLV users. Then, recommendations are generated to match those patterns as closely as possible. The algorithms are evaluated using real data from a mobile cartoon provider with a subscription-based revenue model [211] and an online music provider with a transaction-based revenue model [212], both from Japan. However, although results regarding the improvement of the subscription period and the number of items purchased over time are reported, the evaluation is only based on a simulation system of user purchasing behavior.

Reinforcement Learning Recommendation Methods for Long-Term Business Value Optimization

Recent studies propose methodologies based on *Reinforcement Learning* (*RL*) for optimizing the long-term business value of recommendations [408]. RL is a learning approach that aims to learn an optimal policy (i.e., recommendation strategy) based on the sequential interaction between an agent and the environment through trial and

*Typically, RSs that rely on CLV-related models to optimize relevance [53, 91, 232, 278, 279, 392, 393, 409, 444] follow a common workflow. Algorithms first group users into similar customer value segments. Then they generate recommendations through association rules or collaborative filtering leveraging this additional information.

error to maximize a reward. This methodology is used many times in the literature [176, 189, 221, 229, 353, 411, 449, 484, 500] to optimize the customer lifetime value.

A few studies propose algorithms to directly optimize profit [229, 353]. These studies focus on the transaction-based revenue model where each customer purchase brings a certain profit to the organization. Specifically, in this context, one study [353] considers that a certain profit share can be allocated to each user action (i.e., click, add-to-cart, pay). Hence, the overall profitability can be maximized by optimizing the sum of the profit allocated to each user action considering the probability that such an action will occur given the recommendations. Other studies [176, 221, 411, 449, 484, 500], in contrast, propose algorithms to optimize user engagement, or more generally some strategic interrelated business indicators [189]. One study [411] is based on the advertising revenue model. In this particular context, advertisers are used to pay the platform a certain monetary amount for each click or conversion generated. Hence, in this case, by optimizing user engagement, profit is directly optimized. Instead, other works [176, 221, 449, 484, 500], although they similarly propose to optimize user engagement, are not based on advertising revenue models. Therefore, in these cases, the relationship with profitability is indirect, as user engagement positively correlates with retention.

4.3 Economic Recommendation Evaluation

In this section, we review the evaluation methodologies used in the surveyed papers. First, we give some insights into the different methodologies that are used to evaluate algorithms. Next, we discuss the metrics used in offline evaluation. Then we discuss the results that have been obtained in the real world from ECRSs algorithms by analyzing in detail those studies that report online performance. Finally, we analyze related topics concerning public datasets and the current level of reproducibility.

4.3.1 Evaluation Approaches

In the field of RSs, several methods are proposed to evaluate the performance of algorithms and systems. Depending on the objective of the study, the evaluation may vary in order to assess specific aspects of the recommendations and the system. We identify five methods that are used in the surveyed literature. Some of these are used for offline evaluation (e.g., static predictions, simulation studies, and econometric analyses) [483], while others are used for online evaluation (e.g., user studies, and A/B tests) [83]. While offline methods aim to give a plausible estimate of the performance the system could achieve under real circumstances if certain assumptions are verified, online ones are instead based on real user interactions. In Figure 4.4 we report the distribution of evaluation methods in the literature according to the subdomain of analysis. As can be seen, offline methods are used more frequently than online ones. Moreover, among offline methods, static predictions is the most frequently used method.

Static Predictions. The most commonly used evaluation method in the RSs literature is to hide some data (e.g., ratings, interactions) from a particular dataset, train a model on the remaining data, and then predict the hidden data [216, 483]. After constructing a dataset that contains all the necessary information, the adopted standard is to measure the performance of the system with respect to some underlying objectives [20, 213] with the help of certain metrics. In terms of metrics, given the underlying purposes of ECRSs, the surveyed literature

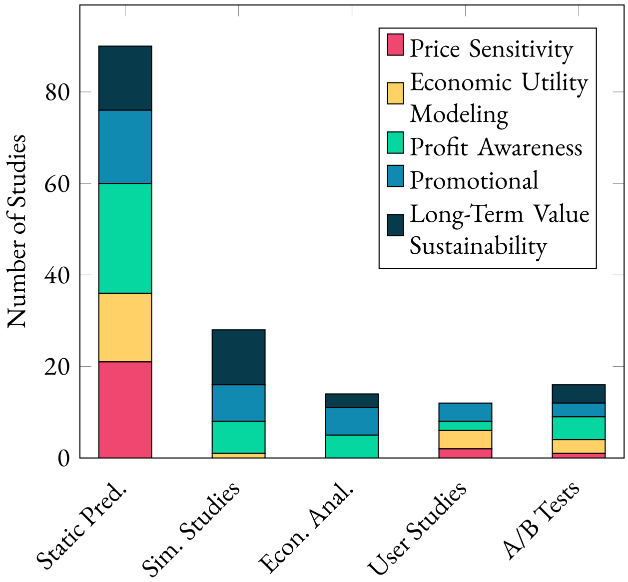


Figure 4.4: Distribution of evaluation methods in the surveyed literature organized by dimension of analysis.

often measures not only relevance prediction metrics (e.g., precision, MRR, NDCG) [171], but also business value metrics* (e.g., profit, revenue) [214, 216, 291].

Simulation Studies. While static predictions methods [483] are mainly used to obtain an estimate of RSs performance in the short term, other studies propose to use dynamic simulations to assess long-term performance [155]. The methodology first involves building a simulator to mimic user behavior [353, 411, 471]. Next, the simulator is used to train and test RSs algorithms on the simulated behavior [177, 449]. Simulators are often adopted to evaluate the performance of reinforcement learning-based recommendation algorithms [14] (e.g., RecoGym [378], RecSim [210]). Moreover, in the surveyed literature there are also simulators [155, 212, 287] created to evaluate supervised learning algorithms. One of these [155], based on agent-based modeling, is designed to realistically mimic customer behavior considering various factors known in the literature to have a high correlation with purchase probability (e.g., trust).

Econometric Analyses. For some algorithms in the surveyed literature [103, 200, 225], performance is assessed with the help of econometric analyses [6, 335]. These are quantitative approaches based on statistical or mathematical methods used to estimate the impact of the system on certain variables of interest (e.g., profit [200]), considering some underlying assumptions. For example, one study [200] investigates the impact of recommendations on corporate profit and consumer welfare by modeling the behavior of a system that considers the simplified case in which the company can sell only two products.

*We discuss the most frequently used offline metrics in Section 4.3.2.

User Studies. In many cases, the impact of the system on certain factors (e.g., user satisfaction) is difficult to model through offline methods. This occurs because in some cases it is not possible to find a good proxy for the target variable, while in other cases it would be necessary to use a large number of assumptions. Especially when the factors are qualitative and the response is subjective (e.g., perceived fairness), the literature adopts user studies as a research methodology [83]. These methods typically involve recruiting a group of users (e.g., through emails or through crowdsourcing platforms like Amazon Mechanical Turk), randomly splitting them into distinct groups, requiring them to perform a particular task (e.g., interacting with an RS designed for the study), observing their (objective) behavior, and asking them about their subjective perceptions. In the surveyed literature these methods are used [33, 343] for example to determine the impact of algorithms on profitability and user trust.

A/B Tests. When it is necessary to measure the performance of recommender systems in real-world circumstances, A/B tests are often performed [83]. In such tests, two (or more) versions of a system are deployed for a certain period of time and users either interact with one or the other version [216]. Although these tests are often complex to execute and require significant effort, the main advantage is that they are able to directly measure business KPIs (e.g., revenue, profit) [83] and to compare different algorithms in production. These tests are used many times in the surveyed literature* since algorithms are often designed to optimize such KPIs. For example, A/B tests are used to measure the effects of a profit-aware algorithm deployed on Alibaba's AliOS appstore [480] and the CTR of a reinforcement learning-based algorithm deployed on a large e-commerce platform [353].

4.3.2 Metrics Used in Offline Evaluations

A variety of metrics are used in the literature in offline evaluations, including both relevance metrics introduced in Section 2.2.3 as well as other metrics aimed to investigate the business value of recommendations. We report in Table 4.6 the business value metrics we identified in our review. In the table we indicate for each metric the reference, the formula and its definition.

The general principle is the same for all value metrics. Similarly to relevance prediction metrics, first a list of top-k recommendations is generated for each user. Then the recommendations are compared to the ground truth and certain value-related aspects are collected. Those value-related aspects are connected to the price and profit (and in more general terms also to the utility) of each recommended item. In particular, differently from prediction relevance metrics, value ones do not only count the hits but multiply that hit by the items' price and profit. We briefly introduce the most frequently used value metrics as follows:

- *Revenue@k* (see Eq. (4.16)) [33, 290, 291] indicates the total revenue from the sale of recommended products actually purchased by users;
- *Profit@k* (see Eq. (4.17)) [96, 155, 214, 320] indicates the total profit from the sale of recommended products actually purchased by users;
- *EP@k* (see Eq. (4.18)) [69] indicates the *statistical* expected profit from the recommendation. *EP@k* compared to *Profit@k* is referred to as statistical because the probability of the user accepting the recommendations is considered rather than the ground truth information;

*We discuss results of A/B tests in Section 4.3.3.

Refs	Metric		Definition
[291]*	$Revenue@k = \sum_{u \in \mathcal{U}} \sum_{j=1}^{k} rel^{y}_{u,j} \cdot prc_j$	(4.16)	*Revenue* at position k is the revenue from relevant items in the recommendations list.
[214]*	$Profit@k = \sum_{u \in \mathcal{U}} \sum_{j=1}^{k} rel^{y}_{u,j} \cdot prf_j$	(4.17)	*Profit* at position k is the profit from relevant items in the recommendations list.
[69]*	$EP@k = \sum_{u \in \mathcal{U}} \sum_{j=1}^{k} \hat{x}_{u,j} \cdot prf_j$	(4.18)	*Expected Profit* at position k is the statistical profit it is expected to achieve by the recommendations considering the expected user interest $\hat{x}_{u,j}$. $EP@k$ is referred to as statistical profit (compared with $Profit@k$ in Eq. (4.17)), because the probability that the user accepts the recommendations instead of the actual ground truth relevance information is considered.
[240]**	$PAH@k = \frac{1}{\lvert\mathcal{U}\rvert} \cdot \frac{Profit@k}{Volume@k}$	(4.19)	*Profit-at-Hit* at position k is the average profit per user from relevant items in the recommendations list.
[285]*	$P\text{-}NDCG@k = \frac{1}{\lvert\mathcal{U}\rvert} \sum_{u \in \mathcal{U}} \frac{\sum_{j=1}^{k} \frac{rel^{y}_{u,j} \cdot prc_j}{\log_2(j+1)}}{P\text{-}IDCG_u@k}$	(4.20)	*Price-Based Normalized Discounted Cumulative Gain* at position k is defined as $NDCG@k$, where the gain is given by the items price. In the equation, $P\text{-}IDCG_u@k$ is the *Price-Based Ideal Discounted Cumulative Gain* obtained by sorting the prices of all relevant items to the user in descending order.

Table 4.6: Most frequently used offline business value metrics in the surveyed literature. *Note that for the sake of notation we used prc_j and prf_j as variables to indicate the price and profit of the recommended item at position j, but these variables depend only on the item and not on the position. **The formulas capture the main essence of the metrics.

- *PAH@k* (see Eq. (4.19)) [240] indicates the overall profit generated by the recommendation per user divided by the number of items sold;
- *P-NDCG@k* (see Eq. (4.20)) [275, 285] indicates the total revenue generated on average per user from the recommendation compared to the theoretically achievable maximum revenue. *P-NDCG@k*, like *NDCG@k* (see Eq. (2.11)) [171], gives more importance to the higher-priced items positioned on the top of the ranking*.

However, analyzing the surveyed articles, some open issues can be identified. In particular, we observe that the literature is mostly scattered, application-specific, and there are no well-defined standards in offline assessment of business value [106, 214]. Often the same metric is referred to by different names (e.g., *Price-Based NDCG* [285], vs. *G-DCG* [275]). Other times, researchers report results that are not comparable to each other because application-specific metrics are proposed in the article to investigate certain types of value (e.g., perishability [390], marginal utility per dollar [151]). In fact, under certain circumstances, it would not even be possible to use certain metrics. For example, in the case where the underlying dataset carries only price information and not profit information (e.g., Amazon [325], Tmall [496]), the metrics related to the latter would not be computable without using synthetic profit distributions of the dataset†. Finally, in cases where simulations are used, the calculation of value metrics may be based on assumptions. The main assumption that can be found [155, 214] is that in some studies the user is supposed to always buy at least one item of the top-k recommended ones. In these cases, since the user may not have actually purchased any of the recommended items if his or her purchase history is analyzed, the underlying ground truth information may be unrealistic.

4.3.3 Real-World A/B Tests and User Studies

Many authors evaluate the performance of ECRSs algorithms using A/B tests or user studies. As is known in the literature, offline evaluation results are not necessarily a valid indicator of online performance [215, 216]. This is often due to the fact that different metrics are used for the two types of experimental evaluation [115, 389]. While offline metrics are often used to measure relevance prediction accuracy (e.g., Precision, NDCG), online metrics are used instead to measure business value (e.g., CTR, GMV, Revenue) [145, 236]. Companies are usually much more interested in assessing how algorithms impact real-world business KPIs exploiting online metrics.

In Table 4.7 we list the studies in the surveyed literature that measure the performance of the proposed systems through A/B tests or user studies. In Table 4.8 we then briefly summarize the meaning of each online metric that is considered for the analysis‡ (i.e., IPV, CTR, CVR, GMV, Revenue, Profit). We refer readers to a recent survey [216] on this topic for further insights into online metrics.

Analyzing Table 4.7 we can make some interesting observations. Some considerations depend on the nature of the particular evaluation methodology (i.e., A/B test vs. user study). For example, considering the recommendations channel and the number of subjects, we note that user studies typically involve few users recruited through

*Note that, as in IR [42, 238], value metrics can be rank-agnostic (e.g., *Revenue@k*, *Profit@k*) or rank-aware (e.g. *P-NDCG@k*), depending on whether the position of the recommended items in the ranking is considered for evaluation or not.

†We discuss the synthetic profit issue in Section 4.3.4.

‡Some niche metrics used to measure certain application-specific factors reported in the studies are not considered

Ref	Year	Eval.	Channel	Subjects	Durat.	Baseline	Δ%IPV	Δ%CTR	Δ%CVR	Δ%GMV	Δ%Rev.	Δ%Prof.
[293]	2022	A/B Test	E-Commerce Platform (Walmart)	36M sessions	-	Walmart Ranker				+0.71%		
[72]	2022	A/B Test	Booking Platform	1M searches	20 days	Platform Ranker		[-0.50%, +2.00%]				
[15]	2022	A/B Test	E-Commerce Platform	-	-	Co-Purchase					+35.0%	
[268]	2021	A/B Test	Online Insurance Platform	-	1 week	LogReg			+[1.05%, 3.98%]		+[2.7%, 16.2%]	
[221]	2021	A/B Test	E-Commerce Platform (Taobao)	-	1 week	Vanilla-CTR	+[6.25%, 8.67%]			+[12.31% 18.03%]		
[110]	2020	A/B Test	Video Game Platform (NetEase)	-	1 year	Platform Ranker			+60.0%	+15.0%		
[353]	2019	A/B Test	E-Commerce Platform	1M users	1 week	Item KNN	+8.80%	+8.20%		+27.90%		
[275]	2019	A/B Test	E-Commerce Platform	-	3 days	LETORIF	+23.76%	+13.80%		+3.62%		
[480]	2017	A/B Test	Appstore (Alibaba AliOS)	1M users	2 weeks	LinDP			-6.00%			+32.0%
[343]*	2016	User Study	Mail Campaign	260 users	9 weeks	CBF					+94.39%	+137%
[481]	2015	User Study	Amazon Mechanical Turk	79 users	-	Amazon Price						+[241%, 248%]
[497]	2014	User Study	Mail Campaign	few users	1 week	Markov Model		+7.43%	+48.92%			
[33]	2013	User Study	Amazon Mechanical Turk	245 users	-	Pers. NonCF					+28.57%	

Table 4.7: Results of real-world A/B tests and user studies in the surveyed literature. In the table we report: the channel used to convey the recommendations; the number of subjects (i.e., users, searches, or sessions); the overall duration of the study (e.g. 20 days, 1 year); the baseline against which the proposed algorithm in the study is compared; and the relative improvements in online metrics of the proposed algorithm compared to the baseline. *The relative improvements are determined by analyzing the sentence "*Overall revenue generated during the experiment was €428 for the content-based group, €832 for the profit-based group*" and Figure 11b in the original paper [343].

Metric	Meaning
IPV	*Individual Page View* is the overall number of clicked items on the platform.
CTR	*Click-Through Rate* is the number of user clicks divided by the number of items shown.
CVR	*Click-Conversion Rate* is the number of purchases (or other events) divided by the number of clicks.
GMV*	*Gross Merchandise Value* is the number of items sold multiplied by their price.
Revenue*	*Revenue* is equal to GMV minus any commissions from item sellers.
Profit**	*Profit* is equal to Revenue minus any item costs.

Table 4.8: Most frequently used online metrics in the surveyed literature [216]. *GMV and Revenue almost always indicate the same measure except in B2C marketplaces like eBay. **Depending on the type of subtracted cost (e.g., raw materials, marketing), profit can be gross, net, or have additional nuances (e.g., EBITDA, EBIT).

e-mail campaigns [343, 497] or Amazon Mechanical Turk [33, 481]. Instead, A/B tests are typically performed on a large scale, exploiting existing systems with large customer bases [353, 480], some of well-known brands (e.g., Walmart, Taobao, Alibaba, NetEase) [110, 221, 293, 480]. Moreover, from a performance point of view, all the studies, whether they are based on user studies or A/B tests, show that ECRSs are able to potentially bring huge business value to the firm. In fact, increases in online metrics are reported in all studies. In some cases, the authors report significant performance improvements* (e.g., +48.92% CVR [497], +35% revenue [15], +32% profit [480]).

However, there may be some limitations regarding the insights we can get from the studies. For example, most of the A/B tests last a very short time, i.e., less than three weeks† [33, 72, 221, 268, 275, 353, 480, 481, 497]. In some cases, the baselines are proprietary algorithms and their internal mechanisms are unknown [33, 72, 110, 293] (e.g., Walmart Ranker). In other cases, results depend on assumptions. For example, a study [481] based on Amazon Mechanical Turk uses synthetic profit information, as the authors did not have product costs available. Another study [343] uses some proxies for offline purchases in addition to explicit purchase data from the firm's online site to measure revenue. In that specific context, offline purchases cannot be connected to the online identities of experiment participants. In particular, the authors treated items that received high ratings by users after they clicked on the "*see more details*" link as purchases to calculate profit.

4.3.4 Available Datasets

Analyzing the ECRSs literature, our survey reveals that many studies report results based on proprietary datasets. This is mainly due to the fact that certain types of information (e.g., prices, profits, purchases, demographics) are of strategic importance to companies, and uncontrolled sharing could create significant economic damage. For example, some information is sensitive to the user, and non-anonymized sharing could have major legal implications due to privacy laws, as well as significant impact on brand reputation. In addition, competitors could make use of economic data related to purchasing and profitability to study weaknesses in the business model and take away market share. However, especially recently, several studies also report results based on public datasets. In this section we extend our previous analysis of the datasets in the literature that have been used to build VARSs (see Section 3.4) by focusing on the information available in these datasets that can be used to build ECRSs, such as prices and profits.

In Table 4.9 we report the most frequently used public datasets in the ECRSs surveyed literature. Specifically, in addition to statistical information such as the number of users, items, interactions, and the density of the dataset, we also report the type of event/interaction (e.g., click, add-to-cart, purchase, rating), and the presence of relevant features for ECRSs algorithms, i.e., date, user demographics, product category, price, and profit.

---

*To ensure evaluation reliability, many authors test the proposed algorithm in different configuration environments reporting different results for each of them [268]. In these cases, Table 4.7 shows a range instead of a single value in metrics improvement.

†Performing long-term A/B tests on a real platform is complex [216] and significant effort is required both in the planning and analysis phases. Often the test could cause financial damage to the brand as users could lose trust in the company due to ineffective recommendations. Other times, it is necessary to re-run the test because of bugs. Or again, certain events (e.g., Easter, Super Bowl) or global macroenomic circumstances (e.g., 2020 COVID-19 crisis, 2022 Ukranian war) may impact performance.

Ref	Dataset	#User	#Item	#Inter	Density	Event	Date	Dem.	Cat.	Price	Prof.
[333]	Cosmetics	1.64×10^6	5.46×10^4	2.07×10^7	0.023 %	View, Add-to-Cart, Remove-From-Cart, Purchase	✓		✓	✓	
[331]	Diginetica	2.05×10^5	1.84×10^5	9.93×10^5	0.002 %	Query, Click, Purchase	✓		✓	✓	
[325]	Amazon(2018)*	-	1.55×10^7	2.33×10^8	-	Review, Ratings	✓		✓	✓	
[474]	Yelp(Full)*	5.56×10^6	5.39×10^5	2.89×10^8	0.009 %	Review, Ratings	✓		✓	✓	
[124]	Yahoo!Music	1.95×10^6	9.82×10^4	1.16×10^7	0.006 %	Ratings					
[203]	Ta-Feng	3.23×10^4	2.38×10^4	8.18×10^5	0.106 %	Purchase	✓	✓	✓	✓	
[186]	MovieLens(20M)*	1.38×10^5	2.73×10^4	2.00×10^7	0.529 %	Ratings	✓	✓	✓		
[45]	Netflix Prize	4.80×10^5	1.78×10^4	1.00×10^8	1.177 %	Ratings	✓				
[141]	SPMF	-	1.65×10^4	8.82×10^4	-	Purchase	✓			✓	
[275]	EC-REC	-	-	-	-	View, Click, Purchase	✓			✓	
[498]	Book-Crossing	1.05×10^5	3.41×10^5	1.15×10^6	0.003 %	Ratings		✓			
[329]	Foodmart	8.84×10^3	1.56×10^3	2.61×10^5	1.894 %	Purchase	✓	✓	✓	✓	✓
[301]	Last.fm	1.89×10^3	1.76×10^5	9.28×10^4	0.027 %	Listen	✓		✓		
[376]	Epinions	2.27×10^5	2.32×10^5	1.13×10^6	0.002 %	Ratings, Graph	✓				
[160]	Jester	7.34×10^4	1.01×10^2	4.14×10^6	55.779 %	Ratings					
[348]	Steam	2.57×10^6	3.21×10^4	7.79×10^6	0.009 %	Purchase	✓		✓	✓	
[496]	Tmall	9.64×10^5	2.35×10^6	4.45×10^7	0.001 %	View, Add-to-Cart, Add-to-Wishlist, Purchase	✓		✓		
[353]	REC-RL	4.90×10^7	2.00×10^8	7.63×10^8	7.79×10^{-6} %	Click, Add-to-Cart, Add-to-Wishlist, Purchase	✓			✓	
[330]	Dunnhumby	2.50×10^3	9.24×10^4	2.60×10^6	1.125 %	Purchase	✓	✓	✓	✓	
[354]	Chainstore	-	4.61×10^4	1.11×10^6	-	Purchase				✓	
[332]	NetEase	1.85×10^4	1.24×10^5	1.13×10^6	0.049 %	Playlist					

Table 4.9: Most used datasets in the surveyed literature. *Since there may be multiple versions of the same dataset, we report the statistics of the most recent one.

Analyzing the reported information, we can make some observations. First, both the datasets' density and size, i.e., the number of interactions, vary greatly. Some of them are quite sparse (e.g., REC-RL [353]), whereas others are dense (e.g., Jester [160]). Some are quite small (e.g., Foodmart [329]), while others are large (e.g., Amazon [325]). In addition, as expected, most of the datasets contain economic information related to actual purchases, as well as prices and possibly profit of products (e.g., in the Cosmetics [333], Diginetica [331], Ta-Feng [203], and Tmall [496] datasets). Indeed, as discussed earlier, economic information is typically used for both algorithmic and evaluation purposes. However, as we also noted in Section 3.4, some datasets do not contain prices (e.g., MovieLens [186], Netflix Prize [45], Book-Crossing [498], Epinions [376], Last.fm [301]) and currently only Foodmart [329] contains profit. In particular, although profit is very important especially to train profit-aware models, there are various studies [16, 60, 69, 96, 155, 214, 287, 320, 355] assuming some synthetic profit distribution, e.g., normal [155], or random [320]. This assumption would allow to overcome the profit availability issue. However, as reported in almost all the studies, this also constitutes an important limitation. In fact, under real circumstances, the profit distribution could be very different from the synthetic one used for the experiments, and the results could vary considerably.

4.3.5 Reproducibility Aspects

The impact of reproducibility on the progress of science is undeniable. However, although there has generally been an increase in reproducible papers in AI over the years [172], many of them are still not sufficiently well documented to reproduce the results of the reported experiments [178]. This problem is observed several times in the field of RSs [40, 100], with well-known cases regarding articles that proposed neural algorithms [135, 370], highlighting for example: non-uniform and lax standards in adopting the correct experimental evaluation methodologies [405]; questionable choices on the use and fine-tuning of baselines for comparative experiments [136].

In particular, by reviewing the ECRSs literature, we note several limitations concerning the reproducibility of the studies. As reported in Table 4.10, only a very small subset of 15 articles, out of 133 (11.27%) identified by the present systematic review share the implementation code*. Notably, as can be seen from the table, we find no article that publicly share the code prior to 2019. In addition, the level of reproducibility is quite uneven when considering the different subdomains of ECRSs. In particular, we note the following critical issues: there are many articles published in the *profit-awareness* subdomain but only two of them share the code; all the articles published in the field of *promotional* strategies refer to relevance-based bundling methods (i.e., there is no code shared about brand-awareness and pricing methods); the code of articles concerning *price-sensitivity* and *long-term value* methods is published only for the most recent and advanced GNN- and RL-based algorithms. Consequently, it would be beneficial and significantly accelerate progress in this field if researchers would pay special attention to increasing the level of reproducibility.

*We did not dive into the code details because even if the code is shared, it was found earlier in the RSs literature [100, 136, 405] that in many cases important information is missing to ensure reproducibility (e.g., preprocessing code).

Ref	Year	Dimension	Link
[74]	2023	Promotional	https://github.com/cjx0525/BGCN
[476]	2022	Price-Sensitivity	https://github.com/Zhang-xiaokun/CoHHN
[447]	2022	Price-Sensitivity	https://github.com/PCNet-Code
[155]	2022	Profit-Awareness	https://github.com/nadaa/rec-strategies-abm
[32]	2022	Promotional	https://github.com/tzoof/BRUCE
[15]	2022	Promotional	https://github.com/muhanzhang/SEAL
[470]	2021	Long-Term Value Sustainability	https://github.com/google-research/google-research/tree/master/recs_ecosystem_creator_rl
[489]	2020	Price-Sensitivity	https://github.com/DavyMorgan/ICDE20-PUP
[456]	2020	Economic Utility Modeling	https://github.com/zhichaoxu-shufe/E-commerce-Rec-with-WEU
[152]	2020	Economic Utility Modeling	https://github.com/TobyGE/Risk-Aware-Recommnedation-Model
[102]	2020	Economic Utility Modeling	https://github.com/xydaisjtu/U-rank
[73]	2020	Promotional	https://github.com/cjx0525/BGCN
[353]	2019	Long-Term Value Sustainability	https://github.com/rec-agent/rec-rl
[275]	2019	Profit-Awareness	https://github.com/weberrr/PE-LTR
[151]	2019	Economic Utility Modeling	https://github.com/TobyGE/Maximizing-Marginal-Utility-per-Dollar-for-Economic-Recommendation

Table 4.10: Studies in the surveyed literature that provide the code.

4.4 Open Challenges and Future Research

In this section we extend our previous analysis of open challenges and future research of VARSs (see Section 3.5) by focusing on the current challenges and possible future research directions of ECRSs.

Comparing Different Algorithmic Approaches. A multitude of algorithmic approaches for optimizing corporate economics are proposed in the literature. In this paper, we categorize them at a high level into in-processing and post-processing methods [106] considering five dimensions of analysis. However, most of the approaches are never compared with each other and may have specificities that may make them preferable in certain circumstances over others. For example, no study has yet compared in-processing with post-processing approaches. In addition, different types of in-processing algorithms are found in the literature. In particular, it is proposed for example to extend the objective function of MF [77, 78, 153, 383], or to use GNNs [476, 489, 490] to generate price-sensitive recommendations. Moreover, value neighbor selection [69], graph-based [16, 60, 361] or evolutionary [96, 320] profit-aware algorithms are proposed as well. However, some types of methods are applied only to certain dimensions of analysis. For example, although feasible in practice, no profit-aware MF objective function extensions or GNNs were surfaced through our study. Similarly, no neighbor selection or evolutionary price-sensitive algorithm was found so far. Therefore, it might be useful for the future both to compare in-processing and post-processing approaches and to implement theoretically feasible algorithms not yet found in the literature, comparing them with existing ones.

Optimizing Economics Trade-Offs. Economics optimization is complex, and the systems must consider multiple trade-offs [106, 214] in the optimization process. For example, considering real-world businesses based on an advertising revenue model (e.g., YouTube, Alibaba's AliOS), it is very important to find the right balance between the ad revenue generated by sponsored items and the actual interests of the user [291, 480]. In particular, special care must be taken not to compromise user trust [200, 343]. In fact, it is shown both through simulations [155], in user studies [327] and subsequent A/B tests [343] that trust is positively correlated with propensity to purchase. A system that is too biased toward higher-profit items that provides irrelevant recommendations to the user [106, 214] could risk impacting the organization's reputation and driving away customers. To address this issue various studies [33, 240, 287, 291, 434, 480] propose algorithms with the goal of balancing the interests of multiple stakeholders [3, 4], particularly considering the profitability/relevance trade-off [214], and optimizing short- or long-term profit [202]. Furthermore, as various studies pointed out, algorithms should take care also of explainability [413, 477], fairness [351, 356, 467, 468], and diversity [252, 343] since they are directly related to trust [110]. However, the current literature has not thoroughly investigated the impact of many of these factors on corporate economics. Hence, providing efficient algorithms to simultaneously optimize multiple economics trade-offs (e.g., profit, fairness, and trust) could be a valuable research direction for the future.

Comprehensive Purpose-Oriented Offline and Online Evaluation. Evaluating ECRSs often requires the use of methods that are different from those used for traditional RSs [83, 483]. As a result, there are still many open challenges in order to to be able to evaluate ECRSs in a comprehensive, purpose-oriented way [20, 213] (i.e., that considers the purposes for which the system is designed). Several of these challenges follow from the analysis presented in this work and from the analysis offered in our previous contribution in Section 3.5. For example, in offline evaluation, it is necessary to use business value metrics besides the widely adopted relevance prediction metrics [171]. Studies often exploit a variety of metrics [69, 177, 214, 240, 285] albeit with similar objectives, and the results reported are not comparable with each other. In addition, offline evaluation methodologies are not standardized and often are designed ad-hoc according to specific needs [200]. Moreover, besides a few exceptions [151, 155, 275, 353, 456, 470, 476], most studies are difficult to reproduce and are often based on proprietary datasets or public datasets with synthetic data [96, 320]. In fact, most datasets [45, 186, 203, 325, 333, 376, 498] do not contain information such as profitability [329], which is however needed for model training. Regarding A/B tests on the other hand, many of them last for a short time [268, 275, 353] and involve a small set of users [33, 343, 481, 497] to avoid potential economic risks [216, 343] for the organization hosting the test. Hence, there could be several future research directions in the field of evaluation. For example, it is necessary to develop better offline value metrics that are indicative of online performance in a given (prototype) scenario. In addition, large-scale A/B tests (i.e., involving many users) and reproducibility studies are also required.

Design of Holistic Algorithmic Methods. In this work, decomposing the literature on ECRSs into five different dimensions of analysis, various algorithmic approaches for optimizing business value are explored. However, most of the existing methods [78, 82, 151, 353, 481] focus exclusively on one of the five perspectives. There are a few exceptions [240, 293] that involve more than one dimension of analysis that study, for example [240], how to combine price-sensitivity with profit-awareness to generate more profit while keeping relevance high. A very small subset of studies [108, 131, 390], on the other hand, provide broader reasoning by also discussing inventory management techniques that might be useful for analogous purposes. Currently, the literature lacks holistic methods capable of leveraging multiple approaches simultaneously [106, 214] complementing each other to optimize different nuances of business value [216] while also considering the interrelationship [189] between them. In addition, it is also necessary to consider the relationship between sales and marketing processes with operational [108, 131, 390] and financial processes so as to propose methods for improving the entire business ecosystem, e.g., reducing raw material costs, minimizing logistics delays, or optimizing cash flows.

4.5 Summary of Findings

With this chapter, we reviewed the existing literature on economic recommender systems. Unlike traditional RSs, economic ones aim to directly optimize profitability by exploiting purchase information (e.g., price and profit) and related concepts from economics and marketing. This topic is highly important because organizations aim to optimize (long-term) profit.

In this survey, we identified a number of relevant works addressing a multitude of related issues on economic RSs. In particular, although the literature is highly scattered, five different approaches that jointly consider the interests of customers and organizations are discussed, i.e., price sensitivity, economic utility modeling, profit awareness, promotional and long-term value sustainability.

In addition, we also delved into the methodological aspects of evaluating the performance of such economic recommender systems. In particular, we identified various metrics that are currently used in the literature to assess the expected business value of recommendations, e.g., *Revenue@k*, *Profit@k* and *PAH@k*. We also discussed the main A/B tests and user studies in the field noting that ECRSs have the potential to bring huge business value to the firm.

Finally, we reported some open challenges and interesting research directions for the future. Besides comparing different algorithmic approaches and designing optimization methods able to consider multiple business value trade-offs (e.g., profitability, fairness and diversity), in the future it may worth to study more comprehensive evaluation methods and holistic algorithms that combines different approaches together to further improve performance.

5

Model-Based Approaches to Profit-Aware Recommendation

As mentioned in the previous chapter in Section 4.4, although various approaches to build ECRSs have been proposed in the literature, most of them have never been compared with each other and may have specifics that might make them preferable in certain circumstances over others. In particular, in this work we focus on the profit-aware approaches that were described earlier in Section 4.2.3 that take the value perspectives of the different stakeholders into account by creating recommendation lists that contain items that are both relevant for consumers and highly profitable for the provider. Many of these profit-aware approaches rely on *re-ranking* techniques [82, 103, 155, 291, 316, 434, 435], where a given baseline recommendation list, which is optimized for consumer relevance, is post-processed to promote items with higher profitability. Commonly, certain guardrails are implemented in the re-ranking process to avoid that items of too little consumer relevance appear in the highest places of the re-ranked lists [103, 434, 435]. A general advantage of such post-processing techniques is that any recommendation model can be used to generate the relevance-optimized baseline list [10]. On the other hand, on high-traffic e-commerce sites, post-processing every single recommendation list may easily lead to a significant computational overhead [459]. Moreover, the effectiveness of the re-ranking process may be limited when the guardrails are set too narrow [155].

In this work, we instead explore the use of new *model-based* (or: in-processing) [106] approaches for building profit-aware recommender systems [214]. In such approaches [69, 268, 355, 361], the task of balancing the competing goals of consumer relevance and provider profitability [155] is embedded directly in the learning process. Specifically, we propose novel profit-aware loss functions for three important families of collaborative filtering techniques: matrix factorization [242, 243], learning-to-rank [368, 369], and neural models [190]. Moreover, we consider a profit-aware variant of the model-free nearest neighbors recommendation approach [326] that was recently proposed in a different context in [69]. Experiments on three real-world e-commerce datasets [325, 329, 474] reveal that our novel model-based approaches are effective in balancing consumer and provider values. Moreover, we compare our models with recent post-processing techniques [82, 155, 214]. This additional comparison reveals our novel model-based approaches can be a favorable alternative to existing re-ranking approaches because they exhibit comparable or better performance in recommending more profitable yet relevant items to the users (with respect to the baselines) but lower prediction times. Because of this greater efficiency, in-processing models might therefore be preferred in practical cases where post-processing methods might be inapplicable, e.g., considering large-scale production systems with millions of active users and catalog items.

The main contributions of this work can be summarized as follows:

- We studied how to optimize the profitability of recommendations by defining four novel objective functions inspired by different families of state-of-the-art recommender systems widely used in industry, i.e., nearest neighbors, matrix factorization, learning-to-rank and neural models.

- We compared our novel in-processing models with some of the most commonly used re-ranking approaches in three different real-world datasets, demonstrating how the proposed models may represent more efficient alternatives with comparable or better performance.

The rest of the chapter is organized as follows. In Section 5.1 we formalize the problem we aim to address. The technical approaches for embedding profit-awareness for different families of collaborative filtering techniques are presented in Section 5.2. Section 5.3 and Section 5.4 describe the experimental setting of our research and the outcomes of the evaluation. The paper ends with a discussion of the findings in Section 5.5 and an outlook on future works in Section 5.6. Finally, Section 5.7 concludes the work with a summary of findings.

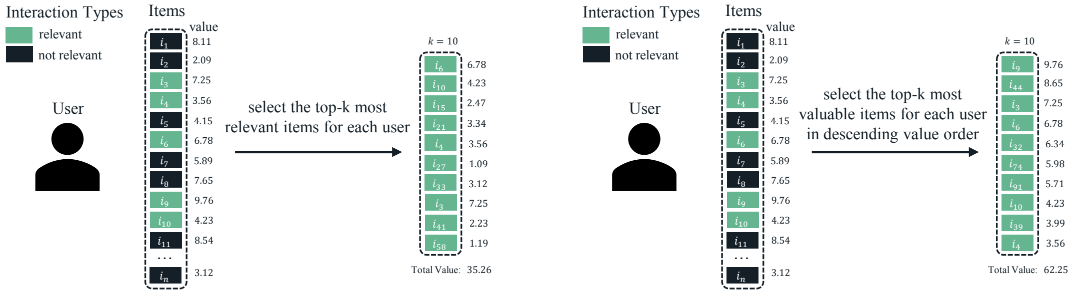
The manuscript entitled "*Model-Based Approaches to Profit-Aware Recommendation*" has been submitted and is currently under review in the journal *Expert Systems With Applications* (2022 impact factor of 8.5).

5.1 Top-k Value Maximization Problem

In this paper, we target an integrated problem formulation which we call *top-k value maximization problem*. As Figure 5.1 shows, the top-k value maximization problem, differently from the well-known top-k recommendation problem described in Section 2.2.2 [368, 483], aims to identify the most valuable, yet relevant items for each user (i.e., that the user may consider for future purchases) in descending value order*. Below we focus on the optimization of *short-term profit* as a particular business value category†.

*Similarly to other settings in the RSs literature [155], the top-k value maximization problem considers that consumer attention is often limited and purchase probability likely decreases according to the position of recommended items (*position bias*).

†By "short-term profit" optimization we refer to the optimization of the overall profit from a unique recommendation of k items to each user, disregarding the potential harm that an overly profit-oriented recommendation may in some cases have on the future cumulative business value (see Section 2.3.5 and Section 4.2.5). The study can be extended to include possible long-term value optimization perspectives in future work [106].



(a) Top-k recommendation problem.

(b) Top-k value maximization problem.

Figure 5.1: Comparison of the top-k recommendation and top-k value maximization problems. (a) Considering an implicit feedback setting, with $x_{u,i} = 1$ if an item is relevant to a user and $x_{u,i} = 0$ if not, the top-k recommendation problem aims to determine the top-k most relevant items for each user. To target this problem, a traditional RS may first rank the items by predicted relevance and then select the top-k as described in Section 2.2.2. (b) Considering for example a generic business value $val_i \in [0, 10]$ associated to each item, the top-k value maximization problem aims to find the top-k most valuable, yet relevant items for each user in descending value order. To target this problem, a profit-aware RS may exploit in-processing or post-processing approaches described in Section 5.2.1 and Section 5.2.2.

5.2 Algorithmic Approaches

Profit-aware recommender systems [214] have recently emerged in the literature with the goal of directly optimizing profit using different techniques. As described in Section 4.2.3, it is possible to divide the underlying approaches mainly into two macro-categories [106] called *post-processing* and *in-processing*, depending on when the profit optimization step is applied. Post-processing methods can be applied to the outputs of any recommendation algorithm, treating it as a black box. In-processing methods, on the other hand, can be used to optimize profit directly at learning time. Using the notation in Section 2.2.2, in the next sections we explore how profit-aware post-processing and in-processing algorithms [106, 214] can be used to properly address the top-k value maximization problem described in the previous section.

5.2.1 Post-Processing Profit-Aware Recommendation Algorithms for the Top-k Value Maximization Problem

As described in Section 2.2.2, in typical circumstances [219, 368], where an RS is designed to recommend the most relevant items (*top-k recommendation problem*), an *ordered list* $\mathbf{y}_{u,k}$ of k items for each user can be obtained by choosing the items with the highest scores (see Eq. (2.3)). However, as argued in Section 3.2, the most relevant items may not necessarily be the most valuable ones. Therefore, an RS algorithm based on Eq. (2.3), although it may optimize the relevance of recommendations for the end users, does not guarantee to optimize the profit for the firm as well.

Analyzing the literature of profit-aware recommendations (Section 4.2.3) [106, 214], we can observe that re-ranking approaches are viable methods to address the top-k value maximization problem. For example, these methods [82, 316] may weigh the consumer's expected interest with the items' profit val_i to rank higher the items of higher value for the company (see Eq. (4.10)). Some algorithms, e.g., as in [214], may also exploit constrained variations of this approach to consider only those items having the predicted scores above a certain threshold $\gamma \in [0, 1]$, as these may be the ones most interesting for the consumers (see Eq. (4.13)). Similarly, other algorithms, e.g., as in [155], may also consider to balance business and consumers' interests by exploiting an additional regularizer $\partial \in [0, 1]$ (see Eq. (4.14)). Note that, considering the referred equations, when $\gamma = 0$ Eq. (4.13) falls back to Eq. (4.10), whereas when $\partial = 1$ Eq. (4.14) falls back to the base case in Eq. (2.3).

5.2.2 In-Processing Profit-Aware Recommendation Algorithms for the Top-k Value Maximization Problem

In-processing algorithms [106] can be theoretically used for the same purposes [213] as post-processing ones. However, these algorithms are typically proposed in isolated contexts [69], often tailoring certain industries (e.g., taxi drivers) [361] and cannot be used in practical applications to properly target the top-k value maximization problem without major adaptations (Section 4.2.3) [106, 214].

To study how in-processing algorithms can be used to more properly address our problem, in the following we focus on four collaborative filtering algorithms. Specifically, we first describe how to embed profit-awareness into the neighbors selection procedure of a *User-Based Nearest Neighbors* algorithm [326] by referring to a recently proposed paper [69] (Section 5.2.2.1). Then, we propose profit-aware loss functions for three well-known model-based algorithms: *Matrix Factorization* [243], *Neural Collaborative Filtering* [190], and *Bayesian Personalized Ranking* [369] (Sections 5.2.2.2, 5.2.2.3).

5.2.2.1 Profit-Aware Nearest Neighbors Adaptations

User-Based Collaborative Filtering (*UCF*) [326, 373] is a well-known nearest neighbors algorithm that has been successfully applied in various application domains [401] (see Section 2.2.5.1). Considering an explicit feedback context (e.g., where $x_{u,i}$ is a rating in the range $[1, 5]$), the algorithm calculates the predicted score $\hat{x}_{u,i}$ of an item i that user u has never interacted with based on a weighted sum of similarities between users belonging to a given neighborhood $\mathcal{P}(u, i)$ (see Eq. (2.13)). Thus, by selecting the neighbors most similar to each user:

$$\underset{\mathbf{y},\mathcal{P}}{\operatorname{argmax}} \quad \frac{\sum_{i \in \mathbf{y}_{u,k}} \sum_{v \in \mathcal{P}(u,i)} sim(u, v) \cdot (x_{v,i} - \bar{x}_u)}{\sum_{v \in \mathcal{P}(u,i)} |sim(u, v)|} \tag{5.1}$$

it is possible to generate a list $\mathbf{y}_{u,k}$ of k recommendations. However, although the algorithm can be used to determine the most potentially interesting items for each user, it may not optimize the profit for the business.

Instead, as noted in a recent study focused on a different context [69], by extending Eq. (5.1) and selecting a set of similar but more profitable neighbors $\mathcal{V}(u, i)$:

$$\underset{\mathbf{y},\mathcal{V}}{\operatorname{argmax}} \quad \frac{\sum_{i \in \mathbf{y}_{u,k}} val_i \sum_{v \in \mathcal{V}(u,i)} sim(u, v) \cdot (x_{v,i} - \bar{x}_u)}{\sum_{v \in \mathcal{V}(u,i)} |sim(u, v)|} \tag{5.2}$$

it may be possible to increase profitability while still keeping the relevance of recommendations high. Intuitively, it is possible to determine the set of profitable neighbors for each user, by selecting those with higher similarity-weighted cumulative profit, where the cumulative profit can be calculated considering the k most profitable items from each neighbor purchase history.

In this paper, we propose to adapt Eq. (5.2) for the class of implicit filtering based CF algorithms employed in various practical applications:

$$\underset{\mathbf{y},\mathcal{V}}{\operatorname{argmax}} \sum_{i \in \mathbf{y}_{u,k}} val_i \sum_{v \in \mathcal{V}(u,i)} sim(u,v) \tag{5.3}$$

by avoiding mean centering and normalization. On the contrary, considering each user-item interaction $x_{u,i} \in \{0,1\}$, the k items included in the list $\mathbf{y}_{u,k}$ may not be the optimal ones. In fact, in an implicit feedback setting, both the user-item interaction $x_{v,i}$ of neighbor v and the average rating $\bar{x}_u$ of user u would be equal to one. Hence, the nominator in Eq. (5.2) would always be equal to zero.

5.2.2.2 Profit-Aware Matrix Factorization and Neural Collaborative Filtering Extensions

Matrix Factorization (*MF*) [241, 242, 243] is a well-known latent factors model for recommendation (see Section 2.2.5.2). The algorithm aims to estimate the expected interest of user u in item i through the dot product between lower-dimensional embeddings (see Eq. (2.17)). The user and item embeddings, $\mathbf{p}_u \in \mathbb{R}^d$ and $\mathbf{q}_i \in \mathbb{R}^d$, are traditionally learned through a dimensionality reduction algorithm, applied to the user-item interaction matrix. The model can handle both explicit and implicit feedback, albeit with some adaptation. If the feedback is implicit (i.e., $x_{u,i} \in \{0,1\}$), as in this work, the learning algorithm typically optimizes a *binary cross-entropy loss* function. By optimizing such loss the algorithm is trained to recommend the most relevant items for each user.

Adopting the underlying principles of profit-aware in-processing approaches presented in Section 4.2.3, in this work we propose to extend the loss function of MF as follows to optimize profitability and relevance:

$$\mathcal{L} = -\sum_{(u,i)\in\mathcal{D}} val_i \cdot x_{u,i} \log \hat{x}_{u,i} + (1 - x_{u,i}) \log(1 - \hat{x}_{u,i}) \tag{5.4}$$

where val_i is the profit of the item. This way the algorithm can give more weight to higher-profit items in the learning process. In addition, we also note that other adaptations can be proposed by considering, for example, explicit feedback scenarios (e.g., $x_{u,i} \in [0,5]$). In these cases, the widely employed *squared loss* function can be weighted as well:

$$\mathcal{L} = -\sum_{(u,i)\in\mathcal{D}} val_i \cdot (x_{u,i} - \hat{x}_{u,i})^2 \tag{5.5}$$

for the item's profit to optimize overall profitability.

The two proposed profit-aware loss functions in Eq. (5.4) and Eq. (5.5) can be used also to plug-in the business value into *Neural Collaborative Filtering* (*NCF*) [190], a deep learning variant of matrix factorization [243] that uses the same loss functions but replaces the user-item dot product with a multi-layer perceptron to learn any arbitrary pattern from data (see Section 2.2.5.5). Similar extensions can be proposed considering also other neural recommendation algorithms (e.g., considering variational autoencoders in Section 2.2.5.6).

5.2.2.3 Profit-Aware Bayesian Personalized Ranking Adaptations

Bayesian Personalized Ranking (*BPR*) [368, 369] is a state-of-the-art optimization framework applicable to various algorithms to generate recommendations in implicit feedback settings (see Section 2.2.5.3). Typically, BPR is applied on top of matrix factorization [243], exploiting a *pairwise loss* function that approximates the AUC ranking statistic (see Eq. (2.19)). By minimizing such loss function $\mathcal{L}$, the score of positive items becomes higher than the score of negative ones. This way, the algorithm can be trained to recommend the most relevant items.

Inspired by the principles behind profit-aware in-processing approaches presented in Section 4.2.3, in this paper we propose to modify BPR's objective function as follows to optimize profitability and relevance:

$$\mathcal{L} = - \sum_{(u,i,j) \in \mathcal{D}} val_i \cdot \ln \sigma(\hat{x}_{u,i} - \hat{x}_{u,j}) \tag{5.6}$$

by weighting the probability $\sigma(\hat{x}_{u,i} - \hat{x}_{u,j})$ of user u preferring positive item i over negative item j by the profit val_i of the positive item. This way, the algorithm can give more weight to higher-profit items that are relevant to the user, thus guiding the overall learning process. In addition, we also note that similarly to the profit-aware re-ranking methods (considering Section 4.2.3), there can be alternative variations of the loss function:

$$\mathcal{L} = -v \sum_{(u,i,j) \in \mathcal{D}} \ln \sigma(\hat{x}_{u,i} - \hat{x}_{u,j}) - (1 - v) \sum_{(u,i) \in \mathcal{D}} val_i \tag{5.7}$$

thus weighting the profit of positive items val_i and the probability $\sigma(\hat{x}_{u,i} - \hat{x}_{u,j})$ according to a regularization parameter $v \in [0, 1]$ to balance consumer utility and business value at learning time.

5.3 Experimental Setting

In this section we describe the experiment preparation phase. Below we discuss the datasets, evaluation metrics, compared algorithms, and hyperparameter tuning.

Experiments are executed on an on-premise server running Ubuntu 20.04 OS equipped with 12 vCPUs, 32 GB RAM, and 2 NVIDIA GeForce RTX 2080 Ti GPUs based on CUDA® 11.6 architecture. The code is developed in Python 3.9.15, is based on TensorFlow 2.11.0, and extends LibRecommender 0.10.2 [*].

5.3.1 Data Preparation

To ensure a comprehensive evaluation, we chose three real-world datasets to run the experiments. Each dataset comes from a different application domain and correspondingly has certain distinctive characteristics:

- **Amazon**[†] [325]: this dataset contains product reviews and corresponding metadata (e.g., price, brand) obtained from Amazon.com. Each review is associated with a rating on a $[1, 5]$ scale. The data are organized into different categories (e.g., Books, Fashion, Electronics). Similarly to another study [490], since

[*] https://github.com/massquantity/LibRecommender

[†] https://nijianmo.github.io/amazon/index.html

Dataset	# Users	# Items	# Interactions	Density
Amazon	$7.039 \cdot 10^3$	$56.365 \cdot 10^3$	$182.379 \cdot 10^3$	0.046 %
Foodmart	$4.115 \cdot 10^3$	$1.559 \cdot 10^3$	$212.547 \cdot 10^3$	3.313 %
Yelp	$1.959 \cdot 10^3$	$9.392 \cdot 10^3$	$58.065 \cdot 10^3$	0.316 %

Table 5.1: The number of users, items, interactions, and the corresponding density of datasets used for experiments [325, 329, 474] after the data preparation phase.

the number of categories is very large, we limit our analysis by selecting only the *Tools and Home Improvement* category. In addition, in accordance with many real-world business cases, e.g., [193, 382], we assume *markup pricing* is used, associating the item price with a proportional profit, i.e., profit equal to 20% of the price.

- **Foodmart*** [329]: this dataset contains a sample of sales transactions from various consumers of a supermarket chain - the dataset is usually exploited in Microsoft SQL Server as a test sample [306, 316]. Each product belongs to a different category (e.g., Food, Drink, Non-Consumable). Since the dataset is not very large, we consider all the categories for the experiments. Moreover, given that each transaction includes the product price and its corresponding cost for the firm, similarly to other studies [82, 316], we use this information to calculate the profit of each item, i.e., subtracting the item's cost from the price.
- **Yelp**† [474]: this dataset contains the user reviews of various real-world businesses organized into different categories (e.g., Shopping, Automotive, Medical). Similarly to Amazon, each review is associated with a rating on a [1, 5] scale. As in two other studies [489, 490], we consider the *Restaurants* category where the price bucket of each item is indicated using a different number of dollar symbols (from $ to $$$$). For this dataset, we associate this price bucket indicator with a proportional economic value, considering a hypothetical case in which the profit of the item is difficult to estimate a priori with certainty, e.g., due to highly variable costs.

Before performing the experiments, some preliminary data preparation is carried out. In accordance with the objectives of the top-k value maximization problem described in Section 5.1, we prepare the various datasets for an implicit feedback recommendation task [368] as follows: every purchase transaction in *Foodmart* is considered as a positive user-item interaction; every review associated with a rating greater or equal than four in *Amazon* and *Yelp* is considered as a positive user-item interaction. All users who have not positively interacted with at least 20 items are excluded, as done in the well-known MovieLens 20M dataset [186], as we are not focusing on cold-start situations. However, no cold-start item is excluded because often unpopular items are those associated with the highest business value [155]. Instead, we exclude all items with null, zero, or negative economic value. In fact, although in real circumstances there may exist occurrences with negligible or even negative profit, this may occur only as a result of specific business strategies [22, 163], e.g., unprofitable popular items may be used as loss leaders to stimulate the purchase of complementary higher-margin niche items [147]. In the following, we do not assume any of those cases for the datasets considered.

The statistics of the datasets after the data preparation phase are shown in Table 5.1. As can be seen from the table, *Amazon* is the least dense dataset with the largest number of items, *Foodmart* is the most dense dataset with

*https://github.com/julianhyde/foodmart-data-json

†https://yelp.com/dataset

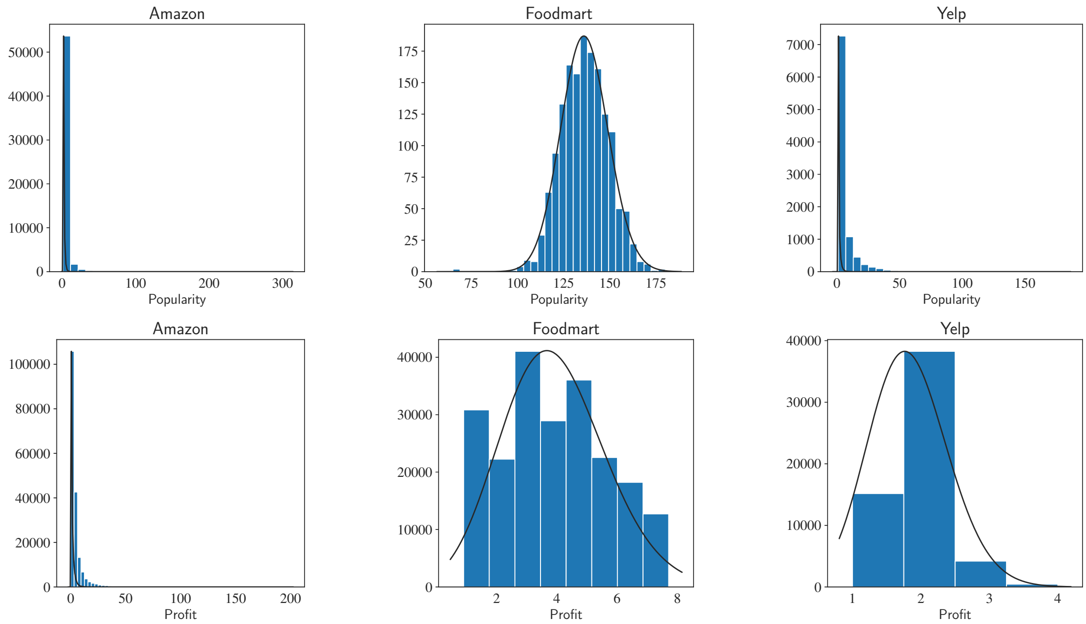


Figure 5.2: Popularity and profit histograms (in blue) with best-fit gamma distributions (in black) of datasets used for experiments [325, 329, 474] after the data preparation phase.

the highest number of interactions, while *Yelp* has an intermediate density and the lowest number of interactions. Popularity and profit histograms of the datasets with fitted gamma distributions are shown in Figure 5.2. As can be noted, considering popularity, both *Amazon* and *Yelp* show a long-tail distribution. Instead, *Foodmart* exhibits a normal popularity distribution. On the other hand, considering profit, *Amazon* shows a long-tail distribution where the profit of most items is very low and very few items are highly profitable. On the contrary, *Foodmart* shows a distribution similar to a normal one with most of the profit generated by the central bins. Finally, *Yelp* shows a left-skewed distribution with the majority of items of medium-low profit. Analyzing Pearson's correlation between popularity and profit, we can note that for Amazon ($corr = -0.03486$) and Foodmart ($corr = 0.00720$) there is no correlation while for Yelp ($corr = 0.20893$) there is a low positive correlation. This fact could have an impact on the experimental results since generally, RSs tend to recommend the most popular items more frequently (*popularity bias*) [2].

5.3.2 Evaluation Metrics

To evaluate the performance of profit-aware algorithms according to the goals of the top-k value maximization problem defined in Section 5.1, we select two metrics that can be used to measure different aspects of recommendations. Using *Normalized Discounted Cumulative Value* (*NDCV@k*), we aim to assess the ability of the algorithms to place the most profitable items actually purchased by each user in the highest positions of the ranking. In addition, using the more widely-known *Normalized Discounted Cumulative Gain* (*NDCG@k*), we want to measure how any increase in profitability might adversely affect the relevance of recommendations for the consumers. Given that *NDCG@k* and *NDCV@k* measure partly competing aspects (i.e., consumer vs. business value) we expect that optimizing one metric will likely result in some reduction in the other.

Differently from *NDCG@k* that evaluates only relevance aspects of recommendations, *NDCV@k* is a metric that aims to evaluate both consumer relevance and organizational value. The idea underlying this metric is taken from a recent paper that measures performance using the *Price-Based NDCG@k* [285], i.e., a variant of the *NDCG@k* where the gain is given by the item's price (see Section 4.3.2). In our context, instead of explicitly considering the price, we consider a *generic business value* (e.g., short-term profit) [214, 216] that the company may aim to optimize in accordance with the purposes of value-aware RSs [106]. Hence, considering val_j as the value an organization obtains if an item recommended at position j is purchased by a user*, we define the *Normalized Discounted Cumulative Value* at position k:

$$NDCV@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{j=1}^{k} \frac{rel^{y}_{u,j} \cdot val_j}{\log_2(j+1)}}{IDCV_u@k} \tag{5.8}$$

as an inverse value-based log reward over all positions with valuable yet relevant items among the top-k recommended ones. In the equation, similarly to the $IDCG_u@k$, we refer to $IDCV_u@k$ as the *Ideal Discounted Cumulative Value* obtained by sorting all the items relevant to the user in descending value order. Therefore, *NDCV@k* can be used to measure how precise an RS algorithm is in recommending the most valuable yet relevant items actually purchased by each user in the highest ranking positions.

Note that, although other metrics from the literature on profit-aware RSs can also be used to measure business value (e.g., overall profit or expected profit as described in Section 4.3.2) [69, 214], these are not *rank-aware*, i.e., they do not consider the items' ranking positions for evaluation purposes. Therefore those metrics are limited in terms of measuring the ability of an RS to recommend the most valuable items in descending profit order for each user as mandated by the top-k value maximization problem. In addition, note also that unlike other papers in the profit-aware RSs field [155, 214], we do not unrealistically assume that the user must always buy one item among the recommended ones (i.e., *guaranteed purchase* simulation discussed in Section 4.3.2). Instead, we rely on the actual consumer purchasing history for performance evaluation.

5.3.3 Compared Algorithms

Various algorithms are compared in the experiments. In particular, we select representative profit-aware in-processing algorithms belonging to the main classes described in Section 5.2.2, namely:

- **Value Neighbor Selection (VNS)**: a UCF variant that selects the most profitable neighbors to generate recommendations as defined in Eq. (5.3).
- **Value Matrix Factorization (VMF)**: an MF variant we propose in this paper that exploits the profit-aware cross-entropy loss defined in Eq. (5.4).
- **Value Neural Collaborative Filtering (VNCF)**: an NCF variant we propose in this paper that exploits the profit-aware cross-entropy loss defined in Eq. (5.4).

*For the sake of notation val_j is used to indicate the value of the recommended item at position j, but typically the value depends only on the item and not on its ranking position.

- **Value Bayesian Personalized Ranking (VBPR)**: a BPR variant we propose in this paper that exploits the profit-aware pairwise loss defined in Eq. (5.6).

Moreover, to denote the profit-aware post-processing algorithms presented in Section 5.2.1 we refer to:

- **Hybrid Perspective Recommender System (HPRS)** [82]: a profit-aware re-ranking algorithm that recommends the top-k items with the highest profit-weighted predicted scores as defined in Eq. (4.10).
- **Constrained Profit Ranking (CPR)** [214]: a constrained variant of HPRS that generates recommendations by considering only items with an expected interest above a certain threshold $\gamma \in [0, 1]$ as in Eq. (4.13).
- **Multi-Objective Profit Ranking (MOPR)** [155]: a multi-objective variant of HPRS that balances consumer and organizational interests with a regularizer $\delta \in [0, 1]$ as in Eq. (4.14).

5.3.4 Hyperparameter Tuning

The hyperparameter tuning procedure proceeds as follows. The users in each dataset are split into training, validation and test sets (60% / 20% / 20%) ensuring that users in one set do not appear in any other set. For each validation and test set user, 4 items are kept as known positive interactions to avoid cold-start situations. The remaining positive interaction items are used as the only relevant ones to evaluate performance. For each model, a grid search is performed by optimizing the *NDCV* on the validation set to find the best hyperparameters. All the models are trained for a maximum of 1000 epochs using early stopping with a *patience* of 10 epochs. Experiments are performed considering a different number of recommended items $k \in \{10, 20\}$. Results are averaged across 3 random splits of users using different seeds. In the following experiments we report the mean and the standard deviation over the different runs.

The following hyperparameter ranges are explored in the grid search. In particular, for UCF and VNS, the number of neighbors is selected from $\{1, 2, 3, 5, 8, 10, 25, 50\}$. Regarding MF and BPR, and their profit-aware variants VMF and VBPR, embeddings sizes are selected in $\{32, 64, 128, 256\}$ and learning rates in $\{10^{-3}, 10^{-4}\}$ are explored while fixing the batch size at 128. As for NCF and VNCF, embeddings sizes in $\{16, 32, 64, 128, 256\}$, learning rate in $\{10^{-3}, 10^{-4}\}$, and batch sizes in $\{64, 128, 256\}$ are explored while setting the multi-layer perceptron hidden units as suggested in the original paper [190] to $\{2 \cdot d, d, \frac{d}{2}\}$, where d is the embedding size. In addition, concerning post-processing approaches, CPR's threshold γ in Eq. (4.13) is varied in the range $\{0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$ and MOPR's regularizer δ in Eq. (4.14) is varied in the range $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

5.4 Experimental Results

In this section we discuss the results of the experiments. We first analyze the performance of our proposed in-processing profit-aware algorithms, comparing them with relevance-based baselines. Then we compare the results of the most widely known post-processing approaches with the baselines and our proposed in-processing algorithms.

5.4.1 Analysis of Profit-Aware In-Processing Methods Results

In Table 5.2 we report the results obtained by applying four in-processing methods (i.e., VNS, VMF, VNCF, and VBPR), one from the literature (i.e., VNS) and three proposed in this paper (i.e., VMF, VNCF, and VBPR). Each method is related to an underlying baseline recommendation model. Three real-world datasets with different characteristics are considered in the experiments. The number of recommended items is varied according to two widely used settings in the literature (i.e., $k \in \{10, 20\}$). Given that the four underlying models (i.e., UCF, NCF, MF, and BPR) are widely used in industry and one model may be preferred over another by a firm for different aspects (e.g., explainability, cost of ownership) with this experiment we do not want to identify one model superior to all others in terms of performance, e.g., in fact a firm might want to optimize the business value of a specific model that is already in production. Instead, what we aim to demonstrate is how any baseline model can be adapted to optimize profitability by exploiting the methodologies discussed in Section 5.2.

As can be seen, all three proposed model-based methods (i.e., VMF, VNCF, and VBPR) proved successful in improving *NDCV@k* over the baselines for all the datasets and the number of recommended items considered. The VNS algorithm that was proposed earlier in [69] proved also effective, except for the Amazon dataset. This indicates that profit-aware algorithms are generally able to balance consumer relevance and profit as well, by recommending higher profit items yet relevant to the users compared to the baselines. Clearly, as expected (see Section 5.3.2), an increase in *NDCV@k* almost always results in a corresponding decrease in *NDCG@k*, because profit-aware algorithms give more weight to the company's interests in the learning process. In fact, similarly to what is observed in other studies in the literature [82, 155, 214, 316], although profit-aware algorithms are able to include in the top-*k* recommendations list items of higher profit, this occurs at the expense of a more or less significant loss of relevance.

In addition, we note that from a computational point of view, the overall prediction time of the model-based algorithms is comparable to the various baselines. Only VNS reports slightly higher prediction times because it performs some more computational operations compared to the UCF baseline. This indicates that in-processing methods are overall efficient and no particular computational overhead (e.g., due to re-ranking operations) is required at prediction time to optimize profit. Moreover, our proposed in-processing algorithms do not lead to additional hyperparameters to tune compared to the baseline models.

5.4.2 Analysis of Profit-Aware Post-Processing Methods Results

In Table 5.2 we also report the results obtained by applying three post-processing methods (i.e., HPRS, CPR, and MOPR) on top to four baseline recommendation models.

As can be observed, by looking at *NDCV@k* the results are mixed. In many cases, considering for example the Foodmart dataset, an improvement over the baselines is found, but not always. We note that this is different from our modeling approaches, which always proved effective. For example, the Amazon and Yelp datasets seem particularly challenging for post-processing algorithms, e.g., considering results for MF and BPR.

Moreover, considering each post-processing algorithm individually, we found also different behaviors. For example, as for HPRS, the algorithm almost never succeeds in improving *NDCV@k* over the baseline (except for Foodmart and a few other cases). In contrast, CPR and MOPR algorithms often succeed in improving performance. However, differently from HPRS and our proposed modeling methods, these post-processing methods re-

			NDCV@10		*NDCG*@10		*NDCV*@20		*NDCG*@20		*Pred. Time* (s)*	
Dataset	Model	Algorithm	mean	std	mean	std	mean	std	mean	std	mean	std
Amazon	UCF	Base	0.0326	0.0062	0.1802	0.0087	0.0325	0.0041	0.1920	0.0077	3.53	0.09
		VNS	0.0139	0.0021	0.0273	0.0032	0.0180	0.0022	0.0357	0.0042	5.45	0.01
		HPRS	0.0207	0.0058	0.0400	0.0071	0.0246	0.0023	0.0507	0.0079	7.18	1.63
		CPR	0.0133	0.0013	0.1059	0.0070	0.0130	0.0020	0.1060	0.0070	7.25	0.38
		MOPR	**0.0345**	0.0078	0.1693	0.0074	**0.0360**	0.0045	0.1762	0.0060	8.12	0.03
	NCF	Base	0.0060	0.0013	0.0336	0.0143	0.0090	0.0050	0.0466	0.0287	6.47	4.04
		VNCF	**0.0150**	0.0045	0.0249	0.0055	**0.0237**	0.0069	0.0339	0.0107	9.53	7.09
		HPRS	0.0021	0.0009	0.0013	0.0007	0.0042	0.0016	0.0020	0.0009	40.15	2.52
		CPR	0.0029	0.0011	0.0058	0.0039	0.0052	0.0052	0.0048	0.0056	30.60	3.31
		MOPR	0.0057	0.0021	0.0263	0.0165	0.0094	0.0045	0.0431	0.0306	36.24	2.57
	MF	Base	0.0141	0.0020	0.0869	0.0104	0.0217	0.0021	0.1050	0.0115	4.79	0.05
		VMF	**0.0205**	0.0005	0.0880	0.0107	**0.0309**	0.0004	0.1061	0.0088	3.71	0.91
		HPRS	0.0029	0.0022	0.0018	0.0011	0.0036	0.0035	0.0017	0.0012	31.14	0.53
		CPR	0.0164	0.0029	0.0322	0.0124	0.0269	0.0051	0.0287	0.0054	20.54	0.67
		MOPR	0.0101	0.0009	0.0723	0.0147	0.0171	0.0014	0.0849	0.0087	31.38	0.95
	BPR	Base	0.0260	0.0037	0.1291	0.0034	0.0375	0.0032	0.1464	0.0031	4.94	0.28
		VBPR	**0.0327**	0.0007	0.1223	0.0023	**0.0428**	0.0068	0.1289	0.0174	5.02	0.30
		HPRS	0.0035	0.0011	0.0021	0.0007	0.0051	0.0020	0.0024	0.0007	31.10	0.51
		CPR	0.0210	0.0032	0.0318	0.0031	0.0329	0.0053	0.0485	0.0033	20.62	0.19
		MOPR	0.0181	0.0025	0.0740	0.0051	0.0277	0.0025	0.0932	0.0064	30.81	0.66
Foodmart	UCF	Base	0.0202	0.0013	0.1223	0.0088	0.0267	0.0020	0.1670	0.0080	3.20	0.11
		VNS	0.0213	0.0016	0.0781	0.0072	0.0276	0.0007	0.1102	0.0025	7.63	4.00
		HPRS	0.0299	0.0028	0.1160	0.0098	0.0392	0.0023	0.1560	0.0057	7.32	0.02
		CPR	0.0272	0.0009	0.1175	0.0089	0.0319	0.0022	0.1601	0.0099	7.30	0.06
		MOPR	**0.0303**	0.0031	0.1101	0.0109	**0.0389**	0.0028	0.1482	0.0079	7.36	0.04
	NCF	Base	0.0099	0.0008	0.0730	0.0030	0.0145	0.0014	0.1099	0.0049	1.89	0.33
		VNCF	0.0191	0.0039	0.0750	0.0136	0.0249	0.0030	0.1033	0.0096	1.81	0.38
		HPRS	0.0212	0.0013	0.0760	0.0044	**0.0280**	0.0026	0.1049	0.0097	4.21	0.12
		CPR	0.0075	0.0116	0.0278	0.0392	0.0278	0.0025	0.1059	0.0076	4.22	0.22
		MOPR	**0.0213**	0.0007	0.0757	0.0031	0.0269	0.0024	0.1029	0.0095	4.17	0.22
	MF	Base	0.0112	0.0021	0.0768	0.0134	0.0153	0.0015	0.1105	0.0130	0.14	0.01
		VMF	0.0195	0.0002	0.0800	0.0026	0.0261	0.0002	0.1166	0.0038	0.15	0.01
		HPRS	**0.0205**	0.0003	0.0735	0.0022	**0.0284**	0.0010	0.1069	0.0055	2.92	0.05
		CPR	0.0196	0.0005	0.0697	0.0022	0.0275	0.0002	0.1029	0.0023	2.88	0.04
		MOPR	0.0196	0.0019	0.0722	0.0083	0.0284	0.0012	0.1076	0.0048	2.92	0.09
	BPR	Base	0.0164	0.0006	0.1044	0.0015	0.0223	0.0009	0.1461	0.0043	0.14	0.00
		VBPR	0.0215	0.0016	0.0828	0.0036	0.0273	0.0018	0.1125	0.0058	0.11	0.04
		HPRS	**0.0260**	0.0004	0.0921	0.0010	0.0352	0.0016	0.1302	0.0041	2.96	0.07
		CPR	0.0249	0.0021	0.0912	0.0029	**0.0354**	0.0011	0.1311	0.0036	2.87	0.10
		MOPR	0.0255	0.0010	0.0894	0.0060	0.0351	0.0016	0.1295	0.0049	2.87	0.05
Yelp	UCF	Base	0.1725	0.0045	0.4552	0.0065	0.2025	0.0042	0.4773	0.0045	0.89	0.10
		VNS	0.1755	0.0051	0.4153	0.0130	0.2064	0.0063	0.4444	0.0090	1.41	0.07
		HPRS	**0.1802**	0.0056	0.4180	0.0044	**0.2133**	0.0044	0.4450	0.0041	1.83	0.07
		CPR	0.1532	0.0056	0.3989	0.0068	0.1616	0.0066	0.4171	0.0068	1.79	0.04
		MOPR	0.1785	0.0026	0.4478	0.0007	0.2101	0.0045	0.4637	0.0082	1.83	0.03
	NCF	Base	0.0578	0.0278	0.1947	0.0951	0.0699	0.0339	0.2212	0.1018	0.91	0.15
		VNCF	**0.0657**	0.0267	0.2043	0.0679	0.0789	0.0310	0.2379	0.0685	0.79	0.17
		HPRS	0.0367	0.0111	0.0963	0.0229	0.0386	0.0157	0.1079	0.0343	1.81	0.11
		CPR	0.0248	0.0089	0.1033	0.0121	0.0091	0.0079	0.0575	0.0518	1.44	0.37
		MOPR	0.0650	0.0343	0.1918	0.0933	**0.0821**	0.0318	0.2186	0.0938	1.76	0.03
	MF	Base	0.1239	0.0043	0.3540	0.0030	0.1451	0.0051	0.3857	0.0047	0.20	0.01
		VMF	**0.1378**	0.0047	0.3792	0.0084	**0.1622**	0.0044	0.4115	0.0051	0.17	0.00
		HPRS	0.0555	0.0040	0.1412	0.0057	0.0691	0.0023	0.1740	0.0045	1.35	0.05
		CPR	0.0849	0.0048	0.2437	0.0158	0.0905	0.0050	0.2655	0.0046	1.01	0.05
		MOPR	0.1027	0.0058	0.2744	0.0096	0.1217	0.0014	0.3143	0.0053	1.35	0.03
	BPR	Base	0.1578	0.0052	0.4234	0.0125	0.1810	0.0031	0.4500	0.0125	0.17	0.02
		VBPR	**0.1600**	0.0048	0.4126	0.0206	**0.1902**	0.0015	0.4428	0.0144	0.18	0.02
		HPRS	0.0740	0.0094	0.1870	0.0140	0.0921	0.0034	0.2202	0.0091	1.34	0.04
		CPR	0.0764	0.0091	0.2270	0.0134	0.0677	0.0085	0.2401	0.0083	0.83	0.02
		MOPR	0.1435	0.0069	0.3969	0.0123	0.1716	0.0039	0.4153	0.0260	1.35	0.10

Table 5.2: Results (i.e., NDCV, NDCG, and overall prediction time in seconds) of different profit-aware in-processing (i.e., VNS, VNCF, VMF, VBPR) and post-processing algorithms (i.e., HPRS, CPR, MOPR) compared to their actual baseline recommendation models (i.e., UCF, NCF, MF, BPR) for different datasets (i.e., Amazon, Foodmart, Yelp) by varying the number of recommended items (i.e., $k \in \{10, 20\}$). *The prediction time is not reported for different cut-off length k because the time needed to compute a recommendation list, which is the focus here, is independent of how many items are used to compute a certain metric.

quire additional hyperparameters to tune how much to weigh users' and organizational interests in the re-ranking process.

Finally, observing the overall prediction time, we note that post-processing methods take longer to generate recommendations compared to the baselines. For example, considering the Amazon dataset and the underlying model BPR, the baseline and the VBPR in-processing algorithm take about 5 seconds to generate predictions, while the HPRS and MOPR post-processing methods take about 30 seconds and the CPR* method takes about 20 seconds. This behavior is expected because, unlike our proposed model-based algorithms, after generating recommendations, post-processing methods need to perform a subsequent re-ranking step that may involve significant computational overhead. This limitation is highly important to consider because in practical cases it could make post-processing methods completely inapplicable for large-scale production systems with millions of active users and very large item catalogs.

5.5 Discussion

As mentioned in the previous sections, in-processing and post-processing methods can both theoretically be used for generating profit-aware recommendations. However, these methods have scarcely been compared in the profit-aware literature and may have some peculiarities that may make them more suitable for being used in certain contexts rather than others. For example, it is known in the literature [106, 214] that post-processing methods are flexible and can be mounted on top of various recommender systems. Moveover, although in-processing methods are typically tailored to specific RSs families, they are potentially more efficient since they avoid re-ranking overhead. Below we discuss the performance and computational aspects of both profit-aware post-processing and in-processing approaches.

5.5.1 Performance Aspects of Profit-Aware Algorithms

What emerges from the experiments presented in this paper is that our three proposed model-based algorithms (i.e., VMF, VNCF, and VBPR) proved successful in improving *NDCV@k* in all the considered cases. The adaptation of the VNS algorithm [69] for the implicit feedback setting proved also effective, except for the Amazon dataset. This may depend on the particular characteristics of this dataset (see Section 5.3.1). In particular, Amazon is very sparse and exhibits long-tail distributions of both popularity and profits. Moreover, given that there is also no correlation between popularity and profit, by selecting the most profitable neighbors to generate recommendations instead of those most similar to the current user (see Section 5.2.2.1), much relevance is lost, thus negatively impacting *NDCV@k*.

As for post-processing methods, they were also effective, but not always. In particular, the Amazon and Yelp datasets proved especially challenging. This behavior may be due to the fact that post-processing methods exploit heuristic criteria to re-rank recommendations from an underlying model (see Section 5.2.1). In the case of

*Note that the prediction times of the various post-processing algorithms (i.e., HPRS, CPR, MOPR) are comparable to each other given the same dataset and underlying model. Only CPR shows slightly lower prediction times because it performs re-ranking not on the entire item spectrum, but on a subset of items with predicted scores above a certain threshold (see Eq. (4.13)).

Amazon and Yelp datasets, for example, the sparsity is high (see Table 5.1), and since post-processing algorithms perform re-ranking operations on the entire spectrum of items, this may negatively affect the subsequent quality of recommendations, including in the final ranking items that are highly profitable but not relevant to users.

5.5.2 Computational Aspects of Profit-Aware Algorithms

From a computational point of view, differently from post-processing methods, in-processing algorithms do not show any computational overhead at prediction time. In practice, post-processing methods could have major limitations in many commercial applications as the high prediction time could be prohibitive in large scale production systems with millions of active users and very large item catalogs. Instead, in-processing algorithms may be preferable to save computational resources or where it is necessary to instantly provide recommendations to users.

Moreover, considering the actual implementation of the methods, the CPR and MOPR post-processing algorithms use an additional hyperparameter to balance consumer utility and provider profits thus requiring also more time to train with respect to the various in-processing algorithms. In fact, in many cases, especially for the Amazon dataset, the HPRS post-processing algorithm, which did not require the use of additional hyperparameters, failed to improve *NDCV@k* performance over the baseline.

5.6 Future Work and Limitations of the Study

In this chapter we addressed what we called the top-*k* value maximization problem, comparing in-processing and post-processing approaches that we used to build profit-aware recommender systems (Section 5.2). A variety of extensions of our work are possible in the future. Below we discuss some possible future algorithms adaptations and comparative analyses.

5.6.1 Possible Algorithms Extensions for Future Studies

We identified several research directions for possible future algorithms extensions.

First, in the chapter, we considered to incorporate profitability aspects through in-processing methods into major RSs algorithmic classes such as nearest neighbors, matrix factorization, learning-to-rank, and neural algorithms. In future works, we might consider embedding profit-awareness in other algorithmic classes (e.g., based on linear models, graph neural networks or association rules mining techniques) [328, 439, 450] or in other algorithms belonging to the same class (e.g., neural algorithms) [90, 174].

In addition, in this chapter we mainly focused on model-based approaches (i.e., algorithms based on MF, BPR and NCF). In the future, it might also be interesting to study in more detail how to improve the performance of the VNS algorithm [69]. For example, by modifying the neighbors selection criterion with an additional hyperparameter, we may adjust the number of profitable neighbors with that of similar neighbors. This way, the algorithm may be able to increase the profitability of recommendations without losing too much relevance, thus making it possible to perform well even for the Amazon dataset.

Moreover, in the chapter we chose to focus on an implicit feedback setting. Therefore, although it is possible to extend the proposed algorithms in various ways, in the experiments we compared only profit-aware in-processing algorithms exploiting the loss functions in Eq. (5.3), Eq. (5.4) and Eq. (5.6). Hence, we leave the comparison of any algorithms designed to handle explicit feedback possibly exploiting the losses in Eq. (5.2) and Eq. (5.5) for the future. Moreover, also the comparison of other variants of in-processing profit-aware algorithms that may use additional hyperparameters, such as in Eq. (5.7), could be incorporated in future experiments.

Furthermore, considering the current experiments, we chose to compare the in-processing methods we designed with three post-processing algorithms we found in the literature. In the future it could be interesting to supplement experiments with the comparison of pre-processing methods that although not found in the literature may be possible alternatives for generating profit-aware recommendations, e.g., completely ruling out unprofitable items with a static threshold before the training phase. Moreover, considering computational aspects, the analysis pointed out some limitations of current post-processing approaches. Hence, a future research direction may be to design more effective or efficient post-processing algorithms that may reduce computational overhead by applying re-ranking only to the most potentially relevant items, thus avoiding to consider the complete spectrum of items. Also studying the possibility of combining pre-, in-, and post-processing approaches to achieve better results could be an interesting future research direction.

5.6.2 Possible Future Comparative Studies

We identified the following research directions for possible future comparative analyses.

First, in this chapter, we focused on the optimization of short-term profit as a particular business value category using collaborative filtering algorithms that are widely used in practice. Given this objective, we left to future work the study of algorithms that may consider temporal dynamics to optimize the long-term business value (e.g., based on reinforcement learning) [212, 353, 411]. In addition we also left for the future the possible study of niche methods [16, 96] or applications (e.g., considering the taxi drivers domain) [268, 361]. An interesting research direction for the future could be also to compare promotional approaches [214] that may increase profitability by incentivizing impulsive purchasing behaviors (e.g., dynamic pricing or bundling methods) [8, 131, 148, 481].

In addition, we recall that in our work the *NDCV@k* metric was mainly used to evaluate the ability of a recommendation algorithm to place the most profitable, yet relevant items to the users in descending profit order in the ranking. However, the metric gives equal weight to relevance and profitability during evaluation. In the current experiments, considering an implicit feedback setting, we normalized both profitability and relevance before calculating the metric. Nevertheless, in possible future studies that may consider an explicit feedback setting (e.g., where the relevance range can be $[0, 5]$), the width of the profitability and relevance ranges before normalization may have an impact on the final results. Correspondingly, a future research direction might be to consider these factors to investigate how to evaluate profit-aware recommendation algorithms offline, which may also involve the design of additional metrics.

Finally, three datasets with different characteristics were used in the paper to evaluate the algorithms. However, the distribution of popularity and profitability and the correlation between these two factors may have an impact on the final results. Since more popular items are generally more relevant to the users, they would be more likely to bring an increase in profitability in the ranking if they also positively correlate with profits. Thus, a future research direction might be to study the relationship between popularity and profitability (and related cold-start aspects) in more depth to understand the contribution of both factors in the final performance of algorithms.

5.7 Summary of Findings

In many practical applications, such as in electronic commerce, a highly important goal for companies is to use a recommender system to generate business value, e.g., increasing profit, conversion rates, or customer retention. Correspondingly, in this chapter, we investigated the use of modeling approaches for major families of collaborative filtering models, i.e., nearest neighbors, matrix factorization, learning-to-rank and neural algorithms, as viable alternatives to the more common re-ranking approaches to build profit-aware recommender systems.

All three proposed model-based approaches proved consistently effective in generating more profitable, yet relevant recommendations without requiring any computational overhead at running time for three datasets with different characteristics. Hence, model-based approaches can be considered viable alternatives to re-ranking approaches, which are currently more popular but have several limitations. Many extensions of this work are possible in the future. We hope that our study will foster further research in this area, for example, considering alternative approaches to those proposed or studying the optimization of long-term business value.

6

On the Problem of Recommendation for Sensitive Users and Influential Items: Simultaneously Maintaining Interest and Diversity

As discussed in Section 4.4, while optimizing for certain business values a value-aware RS should consider multiple trade-offs [106, 214], taking special care to diversity and fairness aspects that may affect the user trust in the platform. In this study we concentrate on this problem considering an alternative definition of value and focusing on broader issues about the optimization of the value for the society as a whole.

In particular, while RSs have traditionally been used as information filtering technologies to help users choose from a large number of alternatives [58], it has been argued in the literature that the tendency of these systems to encourage selective user exposure to a subset of content in order to maximize performance could result in controversial effects. Especially in the social networks field, the repeated exposure to certain contents (e.g., in the form of social pages that deliberately promote certain positions such as the uselessness of vaccines, incitement to violence and drug use) would lead to the occurrence of echo chambers [94], i.e., environments in which users reinforce their position on certain topics due to repeated exposure to similar contents. According to the theory of group polarization [407], this repeated exposure would lead users in a group to develop extreme positions without evaluating alternative positions. Therefore, while in some cases the effects of polarization are limited to extremization of public opinion or to the reinforcement of beliefs and bias, in other cases, they may be associated with an increased risk of violent and aggressive behaviors [130] or with the spread of unhealthy and risky behaviors (e.g., drug use, self-injury). Correspondingly, in this chapter we argue that just as it is important for the business to have RSs that can optimize business KPIs, it is of great value for the society to have RSs that do not encourage risky or aggressive behaviors.

By focusing on the optimization of such kind of value for the society, we address the problem of optimizing the performance of a recommender system that diversifies the distribution of certain items to positively affect the behavior of some users who may be more sensitive than others to specific topics. In particular, we refer to as *sensitive*, the users whose behavior can be influenced by certain items (e.g., depressed and aggressive users). Similarly we refer to as *influential*, the items that can influence the behavior of sensitive users. Based on the effect of influential items on the behavior of sensitive users, we distinguish two subgroups of items: *controversial* and *favorable* items. Moreover, we define as *controversial* the items that could have negative consequences on the behavior of sensitive users (e.g., violence, delinquency, and weapons content), and as *favorable* the items that could have a positive impact (e.g., sports, and hobbies content). Considering the above definitions, in the following we propose various diversity-based recommendation algorithms to mitigate the overexposure to controversial items that can negatively impact the behavior of sensitive users and instead encourage the exposure to favorable items that can positively affect their behavior.

In particular, we propose two approaches that take inspiration from both the research fields of diversity and fairness (see Section 2.3.5.1 and Section 2.3.5.2). The first technique we propose is a redesign of a well-known [399] calibration algorithm. The original algorithm allows the topic proportions of a user's recommendations to be calibrated based on the topic proportions of the ground truth. To suit the method to our context, we modified the objective function in order to calibrate the proportions of influential items for sensitive users based on the distribution of the same items for non-sensitive users. The second technique we propose takes inspiration from an existing algorithm [466] that aims to maximize the value of a ranking under a set of constraints. In this case, we redesigned the algorithm to maximize the expected value of recommendations to sensitive users according to constraints based on target percentages of influential items. Moreover, we also propose a joint approach that can be used to combine the outputs of the two techniques and any additional ranker together to further improve the results taking inspiration from the literature of hybrid recommenders [63] and rank fusion [253]. Compared to the current literature [252, 303, 356] that traditionally proposes very general methods that only consider item or user characteristics separately, the methods we propose consider them simultaneously. This allows for appropriate diversification of recommendations, avoiding potentially negative consequences on sensitive users behavior while simultaneously leading to potentially positive implications for society, without significantly affecting the overall performance of the recommender system.

We evaluated our approaches by exploiting a subset of a real-world dataset containing the social data of 92,255 users who completed a self-reported psychological questionnaire to determine their personality profile according to the Big Five personality theory [418] (see Section 6.1.1). We used correlations between personality traits and respectively depressive disorders [247] and violent behaviors [38] to conduct a comprehensive case analysis on two subsets of *sensitive* users that can be negatively affected by the echo chamber effect, i.e. *potentially depressed* and *potentially aggressive* users. We compared the results obtained with a strong baseline algorithm [498] in the diversity literature, which was the algorithm best suited in the literature to address the problem proposed in this paper. All proposed techniques proved successful in diversifying the distribution of influential items in sensitive user recommendations while maintaining high overall performance.

The main contributions of this work can be summarized as follows:

- We formalized the problem of maximizing the performance of a recommender system that diversifies the recommendations of influential items for sensitive users with the aim of not encouraging risky or aggressive behaviors.

- We introduced two techniques that could be used to address the proposed problem taking inspiration from diversity and fairness studies and a joint approach that can be used to combine the output of any recommendation technique together to achieve better results.
- We conducted a full case analysis about potentially depressed and aggressive users based on a real-world dataset to test the proposed techniques.

The remainder of this chapter is organized as follows. In Section 6.1 we formalize the problem. In Section 6.2 we describe the techniques used to address the problem. In Section 6.3 and Section 6.4 we describe the experimental setting of our research and results obtained on a real-world dataset. In Section 6.5 and Section 6.6 we discuss the results and highlight some future research directions. Finally, Section 6.7 summarizes our findings.

The article entitled "*On the Problem of Recommendation for Sensitive Users and Influential Items: Simultaneously Maintaining Interest and Diversity*" [105] was published in the journal *Knowledge-Based Systems* (2022 impact factor of 8.8).

6.1 Problem Statement

In real-world circumstances, recommender systems tend to recommend items that belong to topics of interest to the user to maximize performance [127]. However, this reduces user exposure to a narrow subset of content leading to the well-known echo chamber phenomenon [94]. Echo chambers are environments in which users reinforce their opinions on certain topics due to repeated exposure to content of similar positions. Consequently, according to the theory of polarization [407], this will lead users belonging to the same group to increasingly reinforce their beliefs toward extreme positions without valuing different opinions. According to various studies, these extreme positions can trigger a number of different issues, including an increased risk of local or international violent conflict [130]. It is important that platforms, especially social ones, address this problem and promote depolarization interventions with the aim of transforming conflicts into more constructive forms [117].

In this paper, we consider the consequences of polarization on the behavior of different users. Among all users, there are some who are more sensitive than others to certain topics. We define *sensitive*, as those users whose behavior can be influenced by certain types of items. Some examples of sensitive users may include depressed or aggressive users. The behavior of these users can be influenced by some particular content. For these properties, we define *influential*, as those items that could influence the behavior of sensitive users. In turn, influential items can be further divided into controversial or favorable items, depending on the type of behavioral influence. We define *controversial*, as those items that could have a dangerous influence on the behavior of sensitive users. Controversial items may include items associated with violence, alcohol and weapons. Over-recommending controversial items to aggressive users could result in a potential spread of verbal and physical aggression, self-injury acts, depressive symptoms, fears, anxiety, and delinquency. On the other hand, we define *favorable*, as those items that might have a positive influence on sensitive users behavior. Favorable items may include items associated with sports, support groups and hobbies. Recommending enough favorable items could positively affect depressed users by supporting people in difficulty, promoting emotional balance, or encouraging healthy lifestyles.

Personality Trait	High Level	Low Level
Openness (O)	wide interests, imaginative, intelligent, original, insightful, curious, sophisticated	commonplace, narrow interests, simple, shallow, unintelligent
Conscientiousness (C)	organized, thorough, planful, efficient, responsible, reliable, dependable	careless, disorderly, frivolous, irresponsible, slipshot, undependable, forgetful
Extraversion (E)	talkative, assertive, active, energetic, outgoing, outspoken, dominant	quiet, reserved, shy, silent, withdrawn, retiring
Agreeableness (A)	sympathetic, kind, appreciative, affectionate, soft-hearted, warm, generous	fault-finding, cold, unfriendly, quarrelsome, hard-hearted, unkind, cruel
Neuroticism (N)	tense, anxious, nervous, moody, worrying, touchy, fearful	stable, calm, contented, unemotional

Table 6.1: Adjectives describing the two polarities (high level vs. low level) of each Big Five personality trait [228].

It is indeed fundamental to correctly balance controversial and favorable items in sensitive user recommendations. Some strategies have been introduced to degenerate the feedback loop of recommendations by promoting the diversity of items [222]. However, these strategies are quite generic, do not consider user sensitivity, and could significantly decrease system performance. Therefore, the aim of this paper is to address the problem of maximizing the performance of a recommender system while diversifying the items suggested for sensitive users. Recommendations that meet the latter criterion could avoid the potential negative consequences that come from item distributions too skewed toward controversial item sets. Moreover, these could simultaneously carry positive consequences by promoting distributions that are more skewed toward favorable item sets.

In the following section, we formalize the problem and we define a probability criterion that a RS should satisfy to preserve influential items' diversity in order to positively affect sensitive users' behavior.

6.1.1 Personality Traits, Sensitive Users and Social Issues

One way to identify certain types of sensitive users may be based on their personality traits. Personality is defined as a set of cognitive and behavioral patterns that account for individual differences: a personality trait leads to a specific behavioral response that is repeated with a certain constancy of time, regardless of the stimulus that causes it. In particular, the *Big Five* theory is a taxonomy of personality traits originally hypothesized by Tupes and Christal [418] in the 1960s and subsequently developed in the 1980s and 1990s by different authors [98, 161, 228]. Such theory defines five basic dimensions through which personality can be described. Researchers have identified a set of adjectives that can be grouped into five clusters that are capable of describing permanent traits of human behavior: Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN model). Each trait can be viewed as a continuous dimension between two polarities, i.e., high level vs. low level (see Table 6.1). Self-reported questionnaires are currently the gold-standard for assessing personality in psychology. Different questionnaires have been proposed in literature for the measurement of the five main personality traits and their sub-dimensions according with the Big Five theory. Two of the most popular are the *NEO Personality*

Inventory (*NEO-PI*) [98] and the *International Personality Item Pool* (*IPIP*) [162]. In addition to questionnaires, more recently other methods have been proposed to determine personality, such as using machine learning models [5, 23, 52, 113, 125, 144, 183, 244, 274, 294, 305, 402, 403, 446, 463, 465].

In the field of recommender systems, personality traits have been used mainly as additional features [261, 427] to address the well-known cold-start problem [114, 415]. Indeed, since users tend to exhibit similar behaviors based on these features, personality traits can be exploited in the learning phase as additional data to guide inference of recommender systems in the space of the most likely solutions [267, 317]. However, besides addressing the cold-start, personality traits have been exploited also to enhance the effectiveness of recommendations [128, 204, 416], to make cross-domain recommendations exploiting transfer learning mechanisms [432] or to adjust the degree of diversity in recommendations according to the users' openness to experience [81, 414, 451].

However, in this work we argue that personality traits can be also used to identify certain subsets of users whose behavior may be problematic for the society. In particular, one of the most important social problems today involves mental disorders. A significant part of the population suffers from clinical conditions such as anxiety, depression, and substance abuse. In particular, depression has a high prevalence, affecting around 7% of the population [31] with major socio-economic impacts (e.g., increased mortality, direct costs of medications and hospitalization, generation of indirect costs for absence from work, turnover and disability compensation) [312]. Another relevant social problem concerns aggressive behaviors, such as bullying, racial violence, physical and verbal abuse, minor and major crimes, and anti-social behaviors. Although in recent years there has been a reduction in major crimes in many countries*, aggressive behaviors persist and are also evolving in the world of social networks (e.g., cyberbullying, hate speech) [129]. Different studies revealed that being the victim, as well as the perpetrator, of cyber-aggressions is related to lower well-being and mental health [28, 310]. In the following, a brief discussion on the relationships between personality traits and these social issues is reported.

6.1.1.1 Personality Traits and Depressive Disorders.

The correlation between personality traits and mental disorders is known in the literature. A meta-analysis [384] examined the relationship between the Big Five traits and personality disorders defined in *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) [31]. The study reports that each personality disorder is associated with a particular pattern of personality traits. For example, personality disorders characterized by emotional distress (e.g., paranoid, schizotypal, borderline, avoidant, and dependent disorders) show a positive correlation with Neuroticism; histrionic and narcissistic disorders show a positive correlation with Extraversion while schizoid, schizotypal, and avoidant disorders show a negative correlation with this trait. Disorders characterized by difficulties in relationships (e.g., paranoid, schizotypal, antisocial, borderline, and narcissistic) show negative correlations with Agreeableness. Furthermore, obsessive-compulsive disorder seems to be positively correlated with Conscientiousness, while antisocial and borderline disorders show a negative correlation with this trait. Other studies have investigated specific mental disorders in more depth. A meta-analysis of 175 studies published from 1980 to 2007 [247] found that depressive, anxiety, and substance abuse disorders in adults are predominantly correlated with traits of high Neuroticism and low Conscientiousness. People affected by dysthymic disorder and social phobia show low levels of Extraversion. Similar results on depression were also reported by another study [179]. A

*https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Crime_statistics meta-analysis of 10 cohort studies suggests that depressive symptoms and personality traits are prospectively related. Personality traits are associated with the development of depressive symptoms, while depressive symptoms are associated with temporary or persistent personality changes.

6.1.1.2 Personality Traits and Aggressive Behavior.

There is strong scientific evidence supporting that violent acts are highly correlated with specific personality profiles [122]. The personalities that are more strongly associated with a higher risk of committing criminal behavior are antisocial personality disorder and psychopathy. The literature also demonstrated the link between specific personality profiles and the risk of committing mild violence, such as bullying. For example, a meta-analysis revealed that studies that assessed personality through the Big Five personality model consistently reported an association between lower levels of Agreeableness and Conscientiousness and higher levels of Neuroticism and Extraversion with both bullying and victimization. On the contrary, cognitive and affective empathy was negatively associated only with bullying behavior [311]. Consistently, other studies [158, 192] found that Agreeableness is negatively correlated with deliquency and aggressive behavior. Sharpe and Desai [391] found also that Neuroticism tends to be positively correlated with aggressive behavior while Conscientiousness tends to be negatively correlated. Another research [38] distinguished between physical aggression and violent behavior, showing that the former is directly and indirectly related to Openness, Agreeableness, and Neuroticism, while the latter is indirectly related to Openness, and Agreeableness. Recent studies have also confirmed the relationship between the Big Five traits and other forms of aggression, such as relational aggression [364] and sexual aggression [70].

6.1.2 Recommending Influential Items to Sensitive Users

Consider the task of recommending top-k items to a user and the notation in Section 2.2.2. Let $sns_u \in \{0,1\}$ be a variable indicating whether user u belongs to a set $\mathcal{S} \subseteq \mathcal{U}$ of sensitive users. Let $nfl_i \in \mathcal{Z} = \{0,1\}$ be a binary variable defined over $\mathcal{Z}$ categories indicating whether the item i belongs to a set $\mathcal{L} \subseteq \mathcal{I}$ of influential items. We define the set of influential items as a set of items that can potentially affect the behavior of sensitive users. These include a combination of controversial and favorable items depending on the type of influence on sensitive user behavior. Let $fav_i \in \{0,1\}$ and $con_i \in \{0,1\}$ be two variables indicating whether the item i belongs to a set $\mathcal{F} \subseteq \mathcal{L}$ of favorable items or $\mathcal{C} \subseteq \mathcal{L}$ of controversial items, respectively. For the case studies analyzed in our paper and to simplify the problem, the influential item set will contain only favorable ($\mathcal{F} = \mathcal{L} \wedge \mathcal{C} = \emptyset$) or controversial ($\mathcal{C} = \mathcal{L} \wedge \mathcal{F} = \emptyset$) items. Finally, considering $\mathbf{y}_{u,k}$ as the list of k recommended items to the user, let $y_{u,i} \in \{0,1\}$ be a variable indicating whether the item i belongs to $\mathbf{y}_{u,k}$. According to our objectives, a recommender satisfies a first diversity criterion with respect to the sensitive user group $\mathcal{S}$ and the influential item set $\mathcal{L}$ if:

$$\mathbb{P}(y_{u,i} = 1 | sns_u = 1, nfl_i = 1) = \mathbb{P}(y_{u,i} = 1 | sns_u = 0, nfl_i = 1) \tag{6.1}$$

The criterion requires an equality between the probability that an influential item is selected in the top-k recommendations for a sensitive user and the probability that the same item is selected in the recommendations for a non-sensitive user.

To give an example of how this affects a real-world scenario, let us examine two different cases. First, consider a set of potentially aggressive sensitive users and a set of influential items composed of social pages about controversial topics (e.g. weapons, alcohol). Then, assume to have a recommender system that tends to over-recommend these types of item to sensitive users compared to the non-sensitive group. According to echo chamber theory, in these circumstances, the system could influence the users, potentially making them even more aggressive. On the contrary, a recommendation system that meets the diversity criterion in Eq. (6.1) would balance the distributions of controversial items of potentially aggressive users while reducing the echo chamber effect. Second, consider a set of potentially depressed sensitive users and assume to have a recommender system that tends to under-recommend sets of favorable social pages (e.g., sports, support groups) to this group of users. In this case, the system under normal circumstances could keep users in their depressive state without providing any kind of help. In contrast, if it meets the diversity criterion, it could positively impact these people's mental state. Indeed, the exposition to positive stimuli (e.g., sports, social activities), could affect their emotional state and encourage them to engage in activities (e.g., practicing sports, seeking social or clinical support) that promote their well-being, leading them a step forward to get out of their depression condition. Note that while the diversity criterion in Eq. (6.1) may be useful in some cases, it may not always be sufficient. Promoting an imbalanced item distribution to highly sensitive users could bring greater benefits on their behavior. For highly depressed users, it might be desirable to promote more favorable items compared to non-sensitive users. Conversely, for highly aggressive users, it might be desirable to recommend even less controversial items. Therefore, in these cases, a more general formulation of the previous criterion might be useful:

$$\begin{cases} \mathbb{P}(y_{u,i} = 1 | sns_u = 1, fav_i = 1) \geq \alpha \\ \mathbb{P}(y_{u,i} = 1 | sns_u = 1, con_i = 1) \leq \beta \end{cases} \tag{6.2}$$

with $\alpha > \mathbb{P}(y_{u,i} = 1 | sns_u = 0, fav_i = 1)$ as a lower bound on the probability of recommending favorable items to sensitive users and $\beta < \mathbb{P}(y_{u,i} = 1 | sns_u = 0, con_i = 1)$ as an upper bound on the probability of recommending controversial items. The recommender system that meets the diversity criterion in Eq. (6.2) is able to recommend even fewer controversial items and even more favorable items to sensitive users compared to the non-sensitive group. Note that α and β are defined in the range $[0, 1]$. Thus, by setting $\alpha = 0$ or $\beta = 1$ it is possible to enforce just one of the two constraints defined in the above equation. The algorithmic implementations we propose in the next section will rely on the criteria in equations (6.1) and (6.2).

6.2 Algorithmic Approaches

In the following, we introduce the techniques designed to solve the problem introduced in Section 6.1. We took inspiration from some existing methods used in the field of fairness and redesigned them to suit our context (see Section 2.3.5.1 and Section 2.3.5.2).

6.2.1 Calibrating Influential Item Distribution of Sensitive Users

We developed a first possible solution by modifying a well-known calibration approach [399]. The original algorithm was designed to solve a class imbalance problem to reflect the full spectrum of ground truth interests of

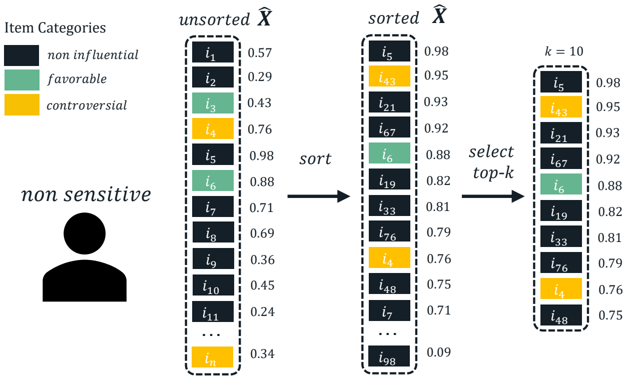


Figure 6.1: Behavior of Calibration and Re-ranking for non-sensitive users: the algorithms sort the predicted scores in descending order and return the top-k items.

the users in the recommendations. We redesigned the algorithm to calibrate the recommendations for each sensitive user based on a target distribution of items from all non-sensitive users. A brief description of the proposed approach follows based on the notation introduced in Section 6.1.2.

Given the categorical probability distribution $g(\mathcal{Z}|i)$ of categories $\mathcal{Z}$ for each item i, we define the distribution $g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ over categories $\mathcal{Z}$ of the set of items recommended over all non-sensitive users as:

$$g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) = \frac{\sum_{u \in \mathcal{U} \setminus \mathcal{S}} \sum_{i \in \mathbf{y}_{u,k}} w_i \cdot g(\mathcal{Z}|i)}{\sum_{u \in \mathcal{U} \setminus \mathcal{S}} \sum_{i \in \mathbf{y}_{u,k}} w_i} \tag{6.3}$$

where a_i is a weight associated with item i that can be used to weight the distribution. Some possible choices to define w_i can be the predicted score $\hat{x}_{u,i}$, the position of item i in the ranking $\mathbf{y}_{u,k}$ or others. For our experiments, we weighed all items uniformly i.e., $w_i = 1$. Then, we define the probability distribution $h(\mathcal{Z}|u \in \mathcal{S})$ over categories $\mathcal{Z}$ of the set of items recommended to a single sensitive user as:

$$h(\mathcal{Z}|u \in \mathcal{S}) = \frac{\sum_{i \in \mathbf{y}_{u,k}} w_i \cdot g(\mathcal{Z}|i)}{\sum_{i \in \mathbf{y}_{u,k}} w_i} \tag{6.4}$$

For the sake of notation, in the following we omit the category and user dependence in the distributions g and h if clear from the context.

We can now define an utility function (inspired by the one proposed by Steck [399]), to find the optimal set $\mathbf{y}^*_{u,k}$ of k items to recommend to the sensitive user ($sns_u = 1$) as:

$$\underset{\mathbf{y}_{u,k}}{\text{argmax}} \quad (1-\lambda) \sum_{i \in \mathbf{y}_{u,k}} sns_u \cdot \hat{x}_{u,i} - \lambda \cdot kl(g \,||\, h) \tag{6.5}$$

where $\lambda \in [0, 1]$ is a regularization parameter. The algorithm optimizes the predicted interest while calibrating the distribution (6.4) of each individual sensitive user to make it as close as possible to the target distribution (6.3)

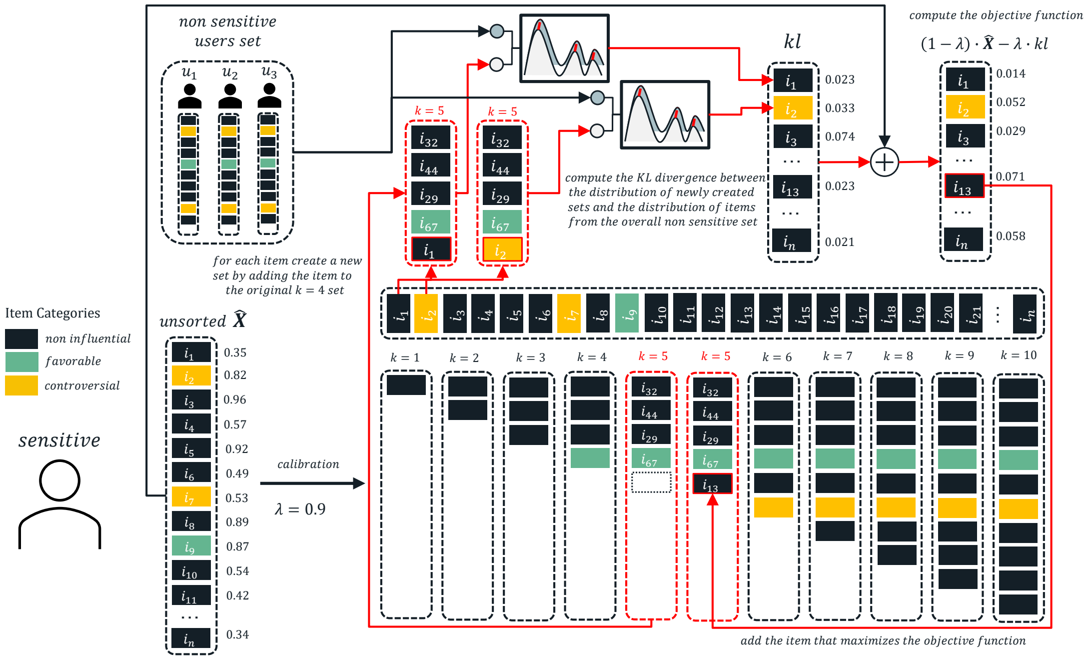


Figure 6.2: Execution of Calibration (Algorithm 6.1) on a sensitive user: at each iteration (e.g., 5-th iteration), it is evaluated which item to be added based on the utility function in Eq. (6.5) that considers the predicted score $\hat{x}_{u,i}$ and the KL divergence in Eq. (6.6).

defined over all non-sensitive users. The term $kl(g \mid\mid h)$ is the Kullback-Leiber divergence, defined to quantify the distance between the two distributions as:

$$kl(g \mid\mid \tilde{h}) = \sum_{\mathcal{Z}} g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) \log \frac{g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})}{\tilde{h}(\mathcal{Z}|u \in \mathcal{S})} \tag{6.6}$$

with $\tilde{h}(\mathcal{Z}|u \in \mathcal{S}) = (1-\varepsilon)\cdot h(\mathcal{Z}|u \in \mathcal{S}) + \varepsilon \cdot g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ and $\varepsilon = 0.01$ as an approximation of the distribution $h(\mathcal{Z}|u \in \mathcal{S})$ to avoid that the KL function diverges when $h(\mathcal{Z}|u \in \mathcal{S}) = 0$ and $g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) > 0$.

The problem is solved with a greedy approach that operates as follows (Algorithm 6.1). For each user: if the user is non-sensitive, the algorithm sorts the predicted scores in descending order and returns the top-k items (Figure 6.1); if the user is sensitive, the algorithm returns a set of k items by exploiting an iterative procedure, depicted in Figure 6.2 and described in the following.

For any sensitive user, the algorithm starts from an empty set $\mathbf{y}^*_{u,k} \leftarrow \emptyset$ (line 14) and iteratively adds items until a set of the required size is obtained (line 15). At each iteration, for each item $j : j \in \mathcal{I}, j \notin \mathbf{y}^*_{u,k}$ the algorithm computes the KL divergence in Eq. (6.6) between the distribution $h_{\mathbf{y}'_u}(\mathcal{Z}|u \in \mathcal{S})$ of the set $\mathbf{y}'_u = \mathbf{y}^*_{u,k} \cup j$ and the distribution $g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ of the top-k items of all non-sensitive users (line 21). The algorithm then adds the item j^{best} to $\mathbf{y}^*_{u,k}$ (line 29) that maximizes the objective function in Eq. (6.5) (line 22).

In Figure 6.2 we provide an illustrative example of the execution of the Calibration algorithm on a sensitive user to obtain a list of 10 items. We focus on the 5-th iteration. For each item $j : j \in \mathcal{I}, j \notin \mathbf{y}^*_{u,k}$ the algorithm

Algorithm 6.1 Calibration

Input:
u: user identifier;
$\mathcal{S}$: set of sensitive users;
$\hat{\mathbf{x}}_u$: scores predicted by the backbone model for user u;
$g(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$: categorical distribution of recommended items over all non-sensitive users;
γ: objective function regularization parameter;
k: number of items to be recommended;
Output:
$\mathbf{y}^*_{u,k}$: top-k items to recommend to user u;
Procedure:
if $u \notin \mathcal{S}$
  return argsort$(\hat{\mathbf{x}}_u, \text{order} = \textit{descending})[0:k]$
else
  $\mathbf{y}^*_{u,k} \leftarrow \emptyset$
  while $|\mathbf{y}^*_{u,k}| < k$
    $obj^{best} \leftarrow -\inf$
    $j^{best} \leftarrow -1$
    for $j \in \hat{\mathbf{x}}_u.items$
      if $j \notin \mathbf{y}^*_{u,k}$
        $\mathbf{y}^{'}_u \leftarrow \mathbf{y}^*_{u,k} \cup j$
        $kl_{\mathbf{y}^{'}_u} \leftarrow kl(g \,||\, h_{\mathbf{y}^{'}_u})$
        $obj \leftarrow (1-\lambda)\sum_{i \in \mathbf{y}^{'}_u} sns_u \cdot \hat{x}_{u,i} - \lambda \cdot kl_{\mathbf{y}^{'}_u}$
        if $obj > obj^{best}$
          $obj^{best} \leftarrow obj$
          $j^{best} \leftarrow j$
        end if
      end if
    end for
    $\mathbf{y}^*_{u,k} \leftarrow \mathbf{y}^*_{u,k} \cup j^{best}$
  end while
  return $\mathbf{y}^*_{u,k}$
end if

defines the set $\mathbf{y}_u^{'} = \{i_{32}, i_{44}, i_{29}, i_{67}, j\}$. Thus, the algorithm first calculates the KL divergence between the set $\mathbf{y}_u^{'}$ and the top-k items of the set of non-sensitive users and then the objective function. Item i_{13} is added to $\mathbf{y}^*_{u,k}$ because it's the one that maximizes the value of the objective function.

6.2.2 Recommending Influential Items to Sensitive Users Under Constraints

Another simple yet effective solution was designed taking inspiration from a well-known methodology in the literature on fairness [142, 466]. The original algorithm aimed to eliminate discrimination in rankings by achieving demographic parity through fairness constraints. The algorithm we propose allows to determine the list of top-k items that maximize the expected value of sensitive users recommendations under constraints on the percentage of controversial and favorable items. Below is a brief description of the proposed methodology based on the notation defined in Section 6.1.2.

The list of top-k items $\mathbf{y}_{u,k}$ recommended to the sensitive user ($s_u = 1$) satisfies the diversity constraint in Eq. (6.2) if:

$$\begin{cases} \frac{\sum_{i \in \mathbf{y}_{u,k}} sns_u \cdot fav_i}{k} \geq \alpha \\ \frac{\sum_{i \in \mathbf{y}_{u,k}} sns_u \cdot con_i}{k} \leq \beta \end{cases} \tag{6.7}$$

where α and β in the range $[0, 1]$ are, respectively, a lower and an upper bound on the percentage of favorable and controversial items allowed in the top-k list.

The optimization problem we propose aims to find the optimal set $\mathbf{y}^*_{u,k}$ of k items that maximizes the predicted value for the sensitive user:

$$\underset{\mathbf{y}_{u,k}}{\operatorname{argmax}} \quad \sum_{i \in \mathbf{y}_{u,k}} sns_u \cdot \hat{x}_{u,i} \tag{6.8}$$

subject to the constraints in Eq. (6.7).

The problem is solved with an efficient greedy algorithm that works as follows (Algorithm 6.2). As for the previous approach: if the user is non-sensitive, the algorithm returns the k items with the highest predicted scores (Figure 6.1); if the user is sensitive, the algorithm exploits an iterative procedure to determine the set of top-k items (Figure 6.3).

For each sensitive user, the algorithm starts from an empty set $\mathbf{y}^*_{u,k} \leftarrow \emptyset$ (line 14) and sorts the items according to the predicted scores $\hat{\mathbf{x}}_u$ in descending order (line 15). The algorithm then cycles through the ordered list $\hat{\mathbf{x}}_u^{sorted}$ (line 16). At each iteration the item j that matches the constraints in Eq. (6.7) (lines 18-19) is addeed to the set $\mathbf{y}^*_{u,k}$ (line 21) until a list of the required size is obtained.

In Figure 6.3 we provide an illustrative example of the execution of the Re-ranking algorithm on a sensitive user to obtain a list of 10 items. In the first 6 iterations, the algorithm adds the items in order of predicted scores because they all match the constraints in Eq. (6.7). In the 7-th iteration, the algorithm discards the controversial item i_2 because it does not match the constraint $\beta = 0$ and replaces it with the next item i_{56}. Finally, at the 10-th iteration, the algorithm discards the non-influential item i_{65} because it does not match the constraint $\alpha \geq 0.2$ and replaces it with the next available favorable item i_{59}.

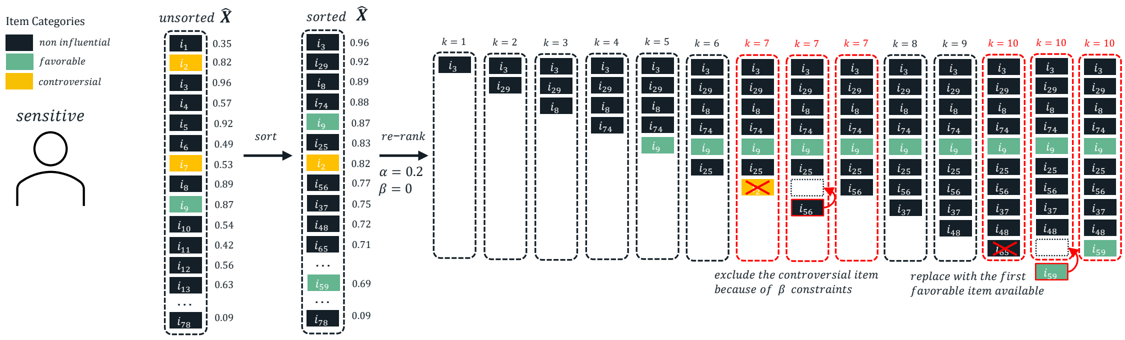


Figure 6.3: Execution of Re-ranking (Algorithm 6.2) on a sensitive user: after sorting the items according to the predicted scores in descending order, at each iteration the algorithm adds the item that matches the constraints in Eq. (6.7) until a list of k items is obtained.

Algorithm 6.2 Re-ranking

Input:
u: user identifier;
$\mathcal{S}$: set of sensitive users;
$\hat{\mathbf{x}}_u$: scores predicted by the backbone model for user u;
α: minimum percentage of favorable items allowed;
β: maximum percentage of controversial items allowed;
k: number of items to be recommended;
Output:
$\mathbf{y}^*_{u,k}$: top-k items to recommend to user u;
Procedure:
if $u \notin \mathcal{S}$
    return argsort($\hat{\mathbf{x}}_u$, order $= descending$)$[0:k]$
else
    $\mathbf{y}^*_{u,k} \leftarrow \emptyset$
    $\hat{\mathbf{x}}_u^{sorted} \leftarrow$ sort($\hat{\mathbf{x}}_u$, order $= descending$)
    for $j \in \hat{\mathbf{x}}_u^{sorted}.items$
        if $|\mathbf{y}^*_{u,k}| < k$
            $constraint_\alpha \leftarrow ((fav_j + \sum_{i \in \mathbf{y}^*_{u,k}} sns_u \cdot fav_i)/k \geq \alpha)$
            $constraint_\beta \leftarrow ((con_j + \sum_{i \in \mathbf{y}^*_{u,k}} sns_u \cdot con_i)/k \leq \beta)$
            if $constraint_\alpha \wedge constraint_\beta$
                $\mathbf{y}^*_{u,k} \leftarrow \mathbf{y}^*_{u,k} \cup j$
            end if
        else
            return $\mathbf{y}^*_{u,k}$
        end if
    end for
end if

6.2.3 Combining Different Methods for Influential Items Recommendation

In the previous section, we presented two algorithms for recommending influential items for sensitive users. As it will be detailed in Section 6.4, it is difficult to decide a-priori which algorithm will perform best on a given problem setting. A possible approach to relieve a user from the need to select a single algorithm is to develop a technique that can automatically combine multiple strategies.

A solution for combining different approaches to recommend influential items to sensitive users together was designed by taking inspiration from the literatures of hybrid recommenders [63] and rank fusion [253]. The former is a branch of recommender systems research [55, 374] that deals with combining the outputs of different recommenders to achieve greater performance. The latter, on the other hand, is a branch of information retrieval research [292] related to that of recommenders concerned with mixing the ranks generated by different IR systems. The algorithm we propose is inspired by the classes of algorithms implemented in these literatures to determine the best rank generated from a set of different rankers. Below is a brief description of the proposed methodology based on the notation defined in Section 6.1.2.

Let $\mathcal{R}$ be a set of rankers that can be applied to recommend $\mathbf{y}'_{u,k}$ lists of items to sensitive users (e.g., exploiting Algorithm 6.1 or Algorithm 6.2). Let AIR* be a metric indicating the percentage of influential items in any $\mathbf{y}_k$ list of k recommended items. The optimization problem we propose aims to find the optimal set $\mathbf{y}^*_{u,k}$ of k items that minimize the deviation in AIR between the recommendations for the single sensitive user $u \in \mathcal{S}$ and those for all non-sensitive users belonging to the set $\mathcal{U} \setminus \mathcal{S}$:

$$\underset{\mathbf{y}'_{u,k}}{\operatorname{argmin}} \quad |AIR(\mathbf{y}'_{u,k}) - AIR(\mathbf{y}_{\mathcal{U} \setminus \mathcal{S},k})| \tag{6.9}$$

The problem can be solved by the following iterative greedy algorithm (Algorithm 6.3). First, the recommendations for non-sensitive users $\mathbf{y}^*_{\mathcal{U}\setminus\mathcal{S},k}$ are determined by sorting the predicted scores $\hat{\mathbf{X}}_{\mathcal{U}\setminus\mathcal{S}}$ in descending order and selecting the top-k items for each user (line 10).

Then, starting with an empty matrix $\mathbf{y}^*_{\mathcal{S},k} \leftarrow \emptyset$ the algorithm proceeds by iterating the following steps for each sensitive user (line 12). Different candidate lists $\mathbf{y}'_{u,k}$ are generated (line 16) from a set $\mathcal{R}$ of rankers (line 15). For each of the candidate lists, the corresponding $|\Delta_{AIR@k}|'$ (line 17) is calculated based on: the AIR that would be obtained if the list $\mathbf{y}'_{u,k}$ is added to the recommendations already selected for sensitive users at the current iteration; the AIR of recommendations for non-sensitive users. Then, the candidate list $\mathbf{y}^{best}_{u,k}$ that minimizes the $|\Delta_{AIR@k}|'$ (line 20) is selected and added to $\mathbf{y}^*_{\mathcal{S},k}$ (line 23). Finally, the algorithm returns $\mathbf{y}^*_{\mathcal{U}\setminus\mathcal{S},k} \cup \mathbf{y}^*_{\mathcal{S},k}$ joining the recommendations for sensitive and non-sensitive users (line 25).

Note that by design, since the algorithm is incremental, for the first few iterations the term $AIR(\mathbf{y}^*_{\mathcal{S},k} \cup \mathbf{y}'_{u,k})$ used to compute $|\Delta_{AIR@k}|'$ (line 17), tends to vary because it is influenced more by the rank of the individual user $\mathbf{y}'_{u,k}$ than the rank of all sensitive users $\mathbf{y}^*_{\mathcal{S},k}$. However, as the number of sensitive users processed increases, $\mathbf{y}^*_{\mathcal{S},k}$ has more weight in the calculation and consequently $|\Delta_{AIR@k}|^{best}$ tends to stabilize and decrease.

*We will define this metric later in the paper in Section 6.3.2 devoted to experimental evaluation.

Algorithm 6.3 Combination

Input:
$\mathcal{U}$: set of all users;
$\mathcal{S}$: set of sensitive users;
$\hat{\mathbf{X}}$: scores predicted by the backbone model for all users;
$\mathcal{R}$: set of rankers to be combined;
k: number of items to be recommended;
Output:
$\mathbf{y}_k^*$: top-k items to recommend to all users;
Procedure:
$\mathbf{y}^*_{\mathcal{U}\setminus\mathcal{S},k} \leftarrow \text{argsort}(\hat{\mathbf{X}}_{\mathcal{U}\setminus\mathcal{S}}, \text{order} = \textit{descending})[0:k]$
$\mathbf{y}^*_{\mathcal{S},k} \leftarrow \emptyset$
for $u \in \mathcal{S}$
    $|\Delta_{AIR@k}|^{best} \leftarrow +\inf$
    $\mathbf{y}^{best}_{u,k} \leftarrow \emptyset$
    for ranker $\in \mathcal{R}$
        $\mathbf{y}'_{u,k} \leftarrow \text{ranker}(\hat{\mathbf{x}}_u)$
        $|\Delta_{AIR@k}|' \leftarrow |AIR(\mathbf{y}^*_{\mathcal{S},k} \cup \mathbf{y}'_{u,k}) - AIR(\mathbf{y}^*_{\mathcal{U}\setminus\mathcal{S},k})|$
        if $|\Delta_{AIR@k}|' < |\Delta_{AIR@k}|^{best}$
            $|\Delta_{AIR@k}|^{best} \leftarrow |\Delta_{AIR@k}|'$
            $\mathbf{y}^{best}_{u,k} \leftarrow \mathbf{y}'_{u,k}$
        end if
    end for
    $\mathbf{y}^*_{\mathcal{S},k} \leftarrow \mathbf{y}^*_{\mathcal{S},k} \cup \mathbf{y}^{best}_{u,k}$
end for
return $\mathbf{y}^*_{\mathcal{U}\setminus\mathcal{S},k} \cup \mathbf{y}^*_{\mathcal{S},k}$

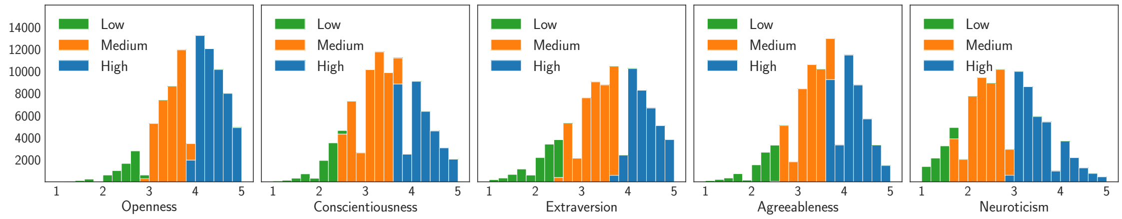


Figure 6.4: The *myPersonality* dataset [244, 315, 358] distribution of user Big Five personality traits.

6.3 Experimental Setting

In this section we introduce our experimental setting, discussing the dataset used and the evaluation metrics.

The experiments were deployed on Google Cloud Platform instances running Debian 10 OS equipped with 8 vCPUs and 64 GB RAM optimized with the TensorFlow Enterprise 2.3 environment and accelerated with the Intel® MKL-DNN/MKL library. The experiments have been coded in Python 3.9.7 and are based on the RecTorch 0.9.0 [357] and the Scikit-learn 1.0.2 [352] libraries. An NVIDIA Tesla T4 GPU was used for experiments with variational autoencoders [270].

6.3.1 Data Preparation

One of the most famous datasets that has been collected with the aim of studying the relationship between personality and activity in social networks is *myPersonality*[*] [244, 315, 358]. Many studies on user personality information in recommender systems are based on this dataset (see Section 6.1.1). The dataset contains data from about 4.3 millions of Facebook users who contributed to psychological research between 2007 and 2012 by filling out a personality questionnaire through a social application. In particular, the dataset contains personal information of anonymized users (e.g., gender, age, etc.), information about their activities on Facebook, personality traits, and other psychological information. Personality data were collected through a 20-item mini-IPIP questionnaire [118] that allows the determination of the Big Five personality traits. According to the main standards, the personality traits of the users in the dataset are represented as a value defined in a range $[1, 5]$ for each trait. Information is also present in discrete form. Each variable can be associated with a value in the set $\{Low, Medium, High\}$ according to threshold criteria indicating the influence of each trait. In addition, the dataset contains information about which Facebook pages the user likes. These data were collected in the form of a topical decomposition of the users resulting from a 600-component *Latent Dirichlet Allocation* (*LDA*) where each user was treated as a document containing the words from its own dictionary of likes. Such LDA [54] is a well-known probabilistic topic modeling technique in natural language processing that allows to represent a document from a set of underlying topics. The model is based on a two-step Bayesian generative process where each document is considered as a set of words that combined together compose one or more subsets of latent topics, each of which is characterized by a particular word distribution. In the dataset, each LDA topic is represented by a set of 5 distinct pages and the user's preference for the topic is expressed through a value in the range $[0, 1]$. The same page can be found

[*] http://mypersonality.org/

on multiple topics. Each user can thus be represented by a weighted combination of topics, the interpretation of which may indicate a particular taste in films and music groups, sexual and religious orientation, or a political view. The dataset contains a total of 4,282,857 users and 6,171,599 pages.

For the experiments we present in the following sections, we considered a subset containing all the users of the dataset that have associated the Big Five personality traits and the information of the pages the user likes in the LDA format. Moreover, since the same pages could be found multiple times in different LDA topics, we decomposed the topics into individual pages through an averaging operation. As a result, we obtained a subset of 92,255 users and 1,836 pages. The distribution of the Big Five personality traits of the users is reported in Figure 6.4. As we can see, for each personality trait, there are three subsets indicating the influence of the trait (low vs. medium vs. high). The criteria defining the membership of the subset are variable according to the trait. For example, when comparing Openness and Neuroticism we observe that the range indicating a low trait influence is wider for Openness while the range indicating a high trait influence is wider for Neuroticism. Generally, it would also appear that for each trait, the subset indicating a low trait influence receives fewer users than the subset indicating a medium or high influence.

6.3.2 Evaluation Metrics

We evaluated the relevance of recommendations using the well-known *Normalized Discounted Cumulative Gain* (*NDCG@k*) [220] defined in Section 2.2.3. Moreover, we employed two other metrics, i.e., *Average Influential Ratio* (*AIR@k*) and *Sensitive ImBalance* (*SIB@k*), to measure other aspects of recommendations according to the goals of the problem we defined in Section 6.1. We define *AIR@k* and *SIB@k* below.

In particular, we exploited *AIR@k* as a measure of the number of influential items in user recommendations. The $AIR_u@k$ for the user u can be defined as:

$$AIR_u@k = \frac{1}{k} \sum_{i \in \mathbf{y}_{u,k}} nfl_i \tag{6.10}$$

and the overall *AIR@k* is given by the average of $AIR_u@k$ on all users of the test set.

Moreover, we used $SIB_i@k$ as an item-level measure that indicates how frequently certain items are recommended on average to sensitive users compared to non-sensitive ones. We use this metric exclusively to provide an interpretation of the pages recommended the most and least frequently. $SIB_i@k$ was developed by repurposing a widely adopted metric in the field of fairness to suit our context, that is, *Non-Parity Unfairness* (*NP*) [462]. Let $\frac{1}{|S|} \sum_{u \in S} bin(\hat{x}_{u,i})$ be the average predicted score of item i from the sensitive user group and $\frac{1}{|U \setminus S|} \sum_{u \in U \setminus S} bin(\hat{x}_{u,i})$ the average predicted score from the non-sensitive one with $bin(\hat{x}_{u,i}) = 1$ if $i \in \mathbf{y}_{u,k}$ as a binarization function. Binarization has been used to normalize the predicted scores of different recommender systems. Thus, $SIB_i@k$ for item i can be calculated as:

$$SIB_i@k = \frac{1}{|S|} \sum_{u \in S} bin(\hat{x}_{u,i}) - \frac{1}{|U \setminus S|} \sum_{u \in U \setminus S} bin(\hat{x}_{u,i}) \tag{6.11}$$

The results of the experiments are reported by aggregating all the metrics by user group and item set.

6.4 Experimental Results

In this section, we present the results of applying the algorithmic approaches presented in Section 6.2 for the problem described in Section 6.1 on two real-world case studies.

- In the first experiment (Section 6.4.1) we first checked if the Big Five personality information was present in the LDA-format pages associated with the users and then we gave a preliminary interpretation by studying the most correlated pages with the various personality traits. Although the literature seemed to confirm the initial hypothesis [282], we preferred to verify it. This was done because our dataset, unlike those used in previous studies, was based on LDA-format pages and had a lower number of available data points. The most correlated pages were used to select the influential item sets used in the next experiment.
- In the second experiment (Section 6.4.2) we studied the performance of recommender systems based on two different case studies: recommend favorable items to potentially depressed users and recommend controversial items to potentially aggressive users. The subsets of users were selected based on the Big Five personality traits most correlated with the depressive disorders and aggressive behaviors described in Section 6.1.1. Item sets have been selected manually from the most correlated pages identified in the previous experiment according to the topic of the page (e.g. sports, hobbies, weapons, alcohol, and others).
- In the third experiment (Section 6.4.3), given the tendency of SLIM and Mult-VAE to over-recommend controversial items and under-recommend favorable items, we exploited the algorithms proposed in Section 6.2.1 and 6.2.2 to improve the recommendations for sensitive users and we measured the variations in performance.
- In the fourth experiment (Section 6.4.4), we compare the results obtained from the calibration and re-ranking approaches proposed in Section 6.2 with the algorithms from the diversity literature (Section 2.3.5.1).
- In the fifth experiment (Section 6.4.5), we explore the results obtained from the proposed *combination* approach presented in Section 6.2.3, that combines the outputs of different rankers.

We used two representative state-of-the-art recommendation algorithms as backbones to generate recommendations, i.e., SLIM (see Section 2.2.5.4) and Mult-VAE (see Section 2.2.5.6).

6.4.1 Preliminary Analysis of Recommendation to Sensitive Users

In this section, we present the results from an experiment designed to evaluate if LDA users' preferences are predictors of Big Five personality traits. In addition, we give a preliminary interpretation of the pages that are most and least correlated with each personality trait.

The experiment was carried out using the following methodology. We randomly split the dataset into training and test sets (80% / 20%). For each personality trait, we trained *Lasso* [187] to predict the personality score from users' LDA preferences. We performed a 5-fold grid search cross-validation on the training set to find the best hyperparameters of the model by optimizing the *Root Mean Square Error* (*RMSE*) [187]. The search space for Lasso was defined by exploring its regularization coefficient in the $[-6, -2]$ logarithmic range. Subsequently, Lasso was re-trained for each personality trait in the full training set with the hyperparameters found in the previous step and *RMSE* was evaluated in the test set. We then exploited the $\mathbf{w}$ coefficients of the fine-tuned models to give a qualitative evaluation of the most and least predictive pages for each trait. In the following, we discuss the results obtained.

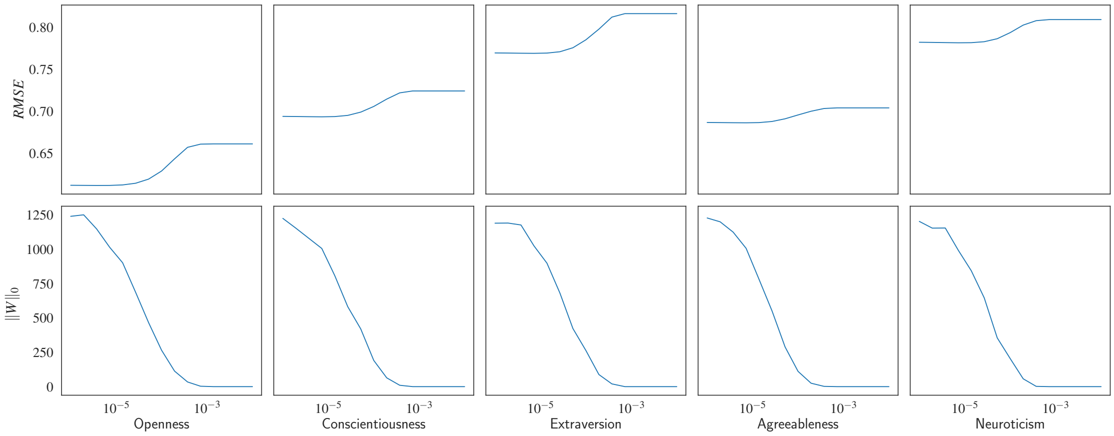


Figure 6.5: The average $RMSE$ and the corresponding number of non-zero $\mathbf{w}$ coefficients of Lasso along the cross-validation search space divided by Big Five personality trait.

Metric	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
$RMSE$	0.6096	0.6979	0.7703	0.6836	0.7828
$\lVert\mathbf{w}\rVert_0$	1,100	954	964	966	952

Table 6.2: The RMSE and the corresponding number of non-zero $\mathbf{w}$ coefficients of Lasso evaluated on the test set divided by Big Five personality trait.

In Table 6.2 we report the results for the fine-tuned models evaluated in the test set. Results indicate that it is particularly difficult to predict Neuroticism and Extraversion. Openness is associated with higher performance, while Conscientiousness and Agreeableness show intermediate results. In Figure 6.5 we show the average $RMSE$ along the cross-validation search space and the corresponding number of non-zero $\mathbf{w}$ coefficients ($\lVert\mathbf{w}\rVert_0$) for the different Big Five personality traits. As expected for all the personality traits the average $RMSE$ is increasing as the number of non-zero coefficients decreases until a saturation point is reached. The results are in line with those presented by Liu et al. (2016) [282] where, differently from our work, LDA topics are defined on Facebook user status updates.

In Table 6.3 we show the pages most positively and negatively associated with each personality trait according to the $\mathbf{w}$ coefficients of the models fine-tuned with the experimental procedure described above. Analyzing the results, we can see that interests in acting, drawing, philosophy and poetry are positively associated with Openness and can be interpreted as indicators of creativity and curiosity. TV shows such as *Survivor* and *Cake Boss*, on the other hand, are negatively associated and can be interpreted as indicators of commonplaces and narrow interests. Regarding conscientiousness, some positively associated interests are *Cappex.com* and *QuikTrip* or sports such as running that may be indicators of organization, reliability, and self-control. Interests in marijuana, manga, and video games such as *The Sims 3* are negatively related and can be interpreted as a symptom of irresponsibility. Some of the positive associations with Extraversion are dancing and acting, brands like *Victoria's Secret* or singers like *Lil Wayne*, *Rihanna* and *Michael Jackson*, these may be interpreted as indicators of socialization, energy and activity. Interests in anime, manga, and video games like *The Sims 3* and *Zynga*, on the other hand, are negatively associated

Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
Score	Page	Score	Page	Score	Page	Score	Page	Score	Page
7.32	Acting	8.94	Running	13.25	Dancing	8.21	The Bible	6.02	Alice in Wonderland
6.04	Philosophy	3.63	Cappex.com	7.79	Acting	5.36	Toy Story	5.07	Juno
5.27	Drawing	3.57	Grey's Anatomy	7.61	Victoria's Secret Pink	3.99	Camping	4.72	Pedigree Adoption
4.99	Writing Poetry	3.04	Criminal Minds	7.36	Lil Wayne	3.00	Friendship	4.36	Glass
4.00	The Princess Bride	2.89	Jesus Daily	5.26	Superbad	2.74	Everything	3.99	The Sims 3
3.88	Singing	2.77	Cooking	4.49	Everything	2.57	Cuddling	3.70	Hot Topic
3.77	The Alchemist	2.70	HGTV	3.90	Wiz Khalifa	2.49	Chris Tomlin	3.49	Evanescence
3.64	The Boondocks	2.65	QuikTrip	3.86	DJ Pauly D	2.31	Hiking	3.25	My Chemical Romance
3.45	Learning	2.65	Victoria's Secret	3.50	Rihanna	2.31	God	3.14	The Vampire Diaries
3.44	Astrology	2.61	Camping	3.49	Michael Jackson	2.27	Chase Community	3.04	Twilight
...	...	...	...	...	...	...	...	...	...
-2.55	Everybody Loves Raymond	-2.57	Glass	-3.78	Everybody Loves Raymond	-2.09	Paintball	-2.87	The Patriot
-2.56	Cake Boss	-2.58	Hot Topic	-3.82	Anime	-2.25	Animal Farm	-2.99	Michael Jordan
-2.62	The Notebook	-3.20	Billy Mays	-4.01	Linkin Park	-2.32	Natalie Portman	-2.99	Everything
-2.63	Sports	-3.22	The Sims 3	-4.05	StumbleUpon	-2.54	Halloween	-3.04	Hip hop music
-2.69	Hockey	-3.35	Manga	-4.10	The Sims 3	-2.55	Kim Kardashian	-3.84	Sports
-2.74	Buffalo Wild Wings	-3.49	My Phrases	-4.39	Manga	-2.69	Best Quotes	-3.85	Soccer
-2.91	Paintball	-3.64	Marijuana is Safer	-4.81	Alice in Wonderland	-2.74	Alice in Wonderland	-4.29	Running
-3.05	Texas Hold'em Poker	-3.64	Food	-5.13	Evanescence	-2.79	Scarface	-4.39	Superbad
-3.16	Dr Pepper	-3.91	Social Interview	-5.47	Zynga RewardVille	-2.90	Urban Dictionary	-4.58	Hiking
-3.79	Survivor	-5.07	Ray William Johnson	-7.66	NCIS	-3.39	Marilyn Manson	-5.03	Snowboarding

Table 6.3: The LDA pages associated with the top-10 highest and lowest values of L_1 coefficients divided by Big Five personality trait.

and can be indicators of shyness and introversion. As for Agreeableness, positively associated interests are *Bible*, *God*, friendship, and cuddling, which can be interpreted as indicators of kindness, generosity, and affection. Sports like paintball, movies like *Scarface* and singers like *Marilyn Manson* are negatively related and can be interpreted as indicators of cruelty, harshness, and coldness. Finally, as for Neuroticism, positive associations can be found with singers such as *Evanescence* and *My Chemical Romance*, TV shows such as *The Vampire Diaries* and films such as *Alice in Wonderland*. These associations can be interpreted as symptoms of tension, anxiety, and moodiness. On the other hand, sports such as running, soccer, snowboarding, and hiking are negatively correlated and can be interpreted as indicators of emotional stability and control.

6.4.2 Analysis of Influential Items Recommendations to Sensitive Users

In this section, we study the performance of recommender systems for two different case studies. We also give an interpretation of the pages that the algorithms recommend most and least frequently.

As a preliminary step, we created a binary user-item interaction matrix from the dataset presented in Section 6.3.1. For the experiments, we considered the top-k LDA pages for each user as a binary rating measure. This procedure was performed because the RecTorch library recommendation algorithms [357] did not accept real numbers as input, but only binary ratings*. After some empirical tests aimed at selecting only the most relevant pages for 92,255 users, we chose $k = 100$ obtaining a total of 9,225,500 ratings.

Then, we selected two subsets of users from the dataset by filtering their Big Five personality profile. Potentially depressed and potentially aggressive users were selected, respectively, based on the correlation of personality traits with depressive disorders and aggressive behaviors discussed in Section 6.1.1. The subset of potentially de-

*This will be discussed in more detail in Section 6.5.

Case Study	Sensitive	Non-Sensitive
Depression	1,403 (1.52 %)	90,852 (98.48 %)
Aggression	5,904 (6.40 %)	86,351 (93.60 %)

Table 6.4: The dataset distribution of sensitive and non-sensitive users divided by case study.

Case Study	Item Set	Sensitive	Non-Sensitive
Depression	Influential	3,138 (2.24 %)	316,451 (3.48 %)
	Non-Infl.	137,162 (97.76 %)	8,768,749 (96.52 %)
Aggression	Influential	55,554 (9,41 %)	683,124 (7,91 %)
	Non-Infl.	534,846 (90,59 %)	7,951,976 (92,09 %)

Table 6.5: The dataset distribution of influential and non-influential items in the ground truth preferences of sensitive and non-sensitive users divided by case study.

pressed users was defined considering users with high Neuroticism, low Extraversion, and low Conscientiousness [247]. The subset of potentially aggressive users was defined by considering users with high Neuroticism and low Agreeableness [38]. For the results we present below, we will refer to both subsets of users as sensitive users, while the rest of the users will be defined as non-sensitive. Next, two subsets of items, favorable and controversial, respectively, were manually selected from the available pages based on an analysis of the topics of the pages. Items were chosen arbitrarily by selecting some of the most correlated pages identified in the previous experiment. This choice is meant to be illustrative of our experiments, but in real-world applications it must be regulated according to well-defined criteria. The subset of favorable items was selected from pages related to sports, sports teams, famous sportsmen, and sports channels. The subset of controversial items was selected instead from pages related to violent sports, war games, alcoholic drinks, and death metal bands. For the results that follow, we will refer to both subsets of items as influential items while the rest of the items will be defined as non-influential items. Next, we studied the performance of two recommenders where items are the pages, for two different case studies. In the first case, we will study the recommendation of favorable items to potentially depressed users, while in the second case, we will study the recommendation of controversial items to potentially aggressive users. The distribution of sensitive users is shown in Table 6.4. The set of potentially aggressive users is higher than the set of potentially depressed users compared to the total: 6.40% of 92,255 instead of 1.52%. The distribution of influential items in ground truth user preferences is shown in Table 6.5. Potentially depressed users tend to put fewer likes on favorable item pages compared to the non-sensitive group: 2.24% instead of 3.48%. Instead, potentially aggressive users tend to put more likes on controversial item pages than the non-sensitive group: 9.41% compared to 7.91%.

The experimental methodology proceeds as follows for each of the two case studies presented above. We randomly split the dataset vertically on the users into training and test sets (60% / 40%). In the vertical split procedure, users who appear in the training set are not included in the test set. The proportion of sensitive users in both sets was balanced using a stratification procedure. We used 20% of the items per user in the test set as known ratings to avoid cold-start, and the remaining 80% to compute the metrics. We trained two state-of-the-art

Depression				Aggression			
SLIM		Mult-VAE		SLIM		Mult-VAE	
Score	Page	Score	Page	Score	Page	Score	Page
0.0728	Naruto Shippuuden	0.1111	Naruto Shippuuden	0.0358	Seether	0.0367	Nine Inch Nails
0.0721	Bleach	0.0936	Vocaloid	0.0306	Dimebag Darrell	0.0350	Seether
0.0709	Patrick Star	0.0936	Naruto	0.0290	Metallica	0.0339	Pantera
0.0698	Daft Punk	0.0912	Gaia Online	0.0287	Superbad	0.0338	Breaking Benjamin
0.0635	deadmau5	0.0882	zOMG!	0.0286	Evanescence	0.0326	Godsmack
0.0582	Courage Wolf	0.0880	deviantART.com	0.0259	Linkin Park	0.0319	Tool
0.0560	Naruto	0.0842	Avenged Sevenfold	0.0248	Stephen King	0.0306	Slipknot
0.0557	PlayStation	0.0827	Manga	0.0243	Nine Inch Nails	0.0302	Fight Club
0.0556	The Colbert Report	0.0704	Korn	0.0236	Nirvana	0.0290	Dimebag Darrell
0.0554	Linkin Park	0.0696	Bleach	0.0233	Shawshank Redemption	0.0290	Korn
...	...	...	...	...	...	...	...
-0.0517	Buffalo Wild Wings	-0.0628	Forever 21	-0.0277	Lance Armstrong	-0.0265	Gucci Mane
-0.0521	H&M	-0.0646	Alicia Keys	-0.0283	I love SLEEP	-0.0267	Movies
-0.0538	Rihanna	-0.0662	Wiz Khalifa	-0.0285	Basketball	-0.0271	Unlimited Texting
-0.0559	Nicki Minaj	-0.0685	T.I.	-0.0292	Social Interview	-0.0273	Running
-0.0572	Family Feud	-0.0715	Family Feud	-0.0300	Movies	-0.0283	I Hate Mosquitos
-0.0579	Chick-fil-A	-0.0765	Eminem	-0.0305	Volleyball	-0.0292	Softball
-0.0624	Basketball	-0.0816	Victoria's Secret	-0.0317	I Hate Mosquitos	-0.0325	The Bible
-0.0642	Victoria's Secret	-0.0828	Drake	-0.0354	Starbucks	-0.0332	Bible
-0.0686	Eminem	-0.0829	Nicki Minaj	-0.0360	Soccer	-0.0357	Sports
-0.0753	T.I.	-0.0928	Victoria's Secret Pink	-0.0380	The Bible	-0.0366	Soccer

Table 6.6: The LDA pages associated with the highest and lowest $SIB_i@100$ divided by case study and backbone recommender system.

recommendation algorithms to predict the top-$\{10, 25, 50, 75, 100\}$* items for each user. Respectively, SLIM (see Section 2.2.5.4), and Mult-VAE (see Section 2.2.5.6) were selected to investigate both linear and deep learning based approaches. We performed a vertical stratified 5-fold grid search cross-validation on the training set to find the best hyperparameters of the models by optimizing *NDCG*. The search space for SLIM was defined by exploring χ in $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ and ω in $\{10^{-2}, 10^{-3}\}$ as introduced in Eq. (2.21). As for Mult-VAE we set the batch size to 512, the annealing steps to 10,000, the ψ regularizer introduced in Eq. (2.23) to 0.2, and we train for 100 epochs, selecting the model with the best validation *NDCG* while searching for different neural architectures in $\{n\text{-}100\text{-}n, n\text{-}200\text{-}n, n\text{-}400\text{-}n, n\text{-}200\text{-}100\text{-}200\text{-}n, n\text{-}400\text{-}200\text{-}400\text{-}n\}$ with n as the total number of items. We then re-trained the models on the full training set with the hyperparameters found in the previous step and evaluated $NDCG@k$, $AIR@k$ and $SIB@k$ on the test set (see Section 6.3.2).

In Table 6.7 we report the results of various experiments divided by case study (i.e., depression, aggression), backbone recommender system (i.e., SLIM, Mult-VAE), algorithmic approach used (i.e., backbone, diversification, re-ranking and calibration), best selected hyperparameter, and number of k recommended items. In addition to the $NDCG@k$ (overall, sensitive and non-sensitive) and $AIR@k$ (sensitive and non-sensitive), we also report the absolute difference in $AIR@k$ between sensitive and non-sensitive user groups ($|\Delta_{AIR@k}|$) and the absolute percentage difference $|\Delta_{AIR@k}\%| = \frac{|\Delta_{AIR@k}|}{|\Delta_{AIR@k}|_{bkb}}$ calculated between the $|\Delta_{AIR@k}|$ and its corresponding backbone value $|\Delta_{AIR@k}|_{bkb}$. As can be observed, in both case studies, SLIM shows superior $NDCG@k$ performance compared to Mult-VAE both overall and measured in the sensitive and non-sensitive user groups for the different top-k settings. Moreover, $AIR@k$ results indicate a tendency for all algorithms to over-recommend controversial items to potentially aggressive users and to under-recommend favorable items to potentially depressed users compared to the non-sensitive groups.

*The choice to use top-$\{10, 25, 50, 75, 100\}$ setting for evaluation is compliant with other work proposing recommender systems based on myPersonality [315, 358] and with Steck's original work [399].

					NDCG@k			AIR@k			
Case Study	Backbone	k	Approach	Hyper	Overall	Sensitive	Non-Sensitive	Sensitive	Non-Sensitive	$\lvert\Delta_{AIR@k}\rvert$	$\lvert\Delta_{AIR@k}\%\rvert$
Depression	SLIM	10	Backbone	-	0.8558	0.8726	0.8556	0.0185	0.0318	0.0133	-
			Diversification	$\eta = 0.10$	0.8535	0.8705	0.8532	0.0362	0.0544	0.0044	33.08 %
			Re-ranking	$\alpha = 0.00$	0.8558	0.8726	0.8556	0.0185	0.0318	0.0133	100.00 %
			Calibration	$\lambda = 0.60$	0.8558	0.8694	0.8556	0.0316	0.0318	0.0002	1.50 %
		25	Backbone	-	0.7590	0.7742	0.7588	0.0290	0.0419	0.0129	-
			Diversification	$\eta = 0.10$	0.7584	0.7739	0.7581	0.0365	0.0509	0.0054	41.86 %
			Re-ranking	$\alpha = 0.04$	0.7590	0.7708	0.7588	0.0518	0.0419	0.0099	76.74 %
			Calibration	$\lambda = 0.60$	0.7590	0.7722	0.7588	0.0419	0.0419	0.0000	0.00 %
		50	Backbone	-	0.8290	0.8408	0.8288	0.0315	0.0454	0.0139	-
			Diversification	$\eta = 0.20$	0.8288	0.8407	0.8286	0.0383	0.0554	0.0071	51.08 %
			Re-ranking	$\alpha = 0.04$	0.8290	0.8399	0.8288	0.0510	0.0454	0.0056	40.29 %
			Calibration	$\lambda = 0.90$	0.8290	0.8399	0.8288	0.0413	0.0454	0.0041	29.50 %
		75	Backbone	-	0.8476	0.8583	0.8474	0.0325	0.0463	0.0138	-
			Diversification	$\eta = 0.30$	0.8475	0.8580	0.8473	0.0413	0.0567	0.0050	36.23 %
			Re-ranking	$\alpha = 0.02$	0.8476	0.8583	0.8474	0.0425	0.0463	0.0038	27.54 %
			Calibration	$\lambda = 0.99$	0.8476	0.8572	0.8474	0.0436	0.0463	0.0027	19.57 %
		100	Backbone	-	0.8572	0.8694	0.8570	0.0334	0.0457	0.0123	-
			Diversification	$\eta = 0.40$	0.8571	0.8693	0.8569	0.0419	0.0555	0.0038	30.89 %
			Re-ranking	$\alpha = 0.03$	0.8572	0.8692	0.8570	0.0440	0.0457	0.0017	13.82 %
			Calibration	$\lambda = 0.99$	0.8572	0.8689	0.8570	0.0446	0.0457	0.0011	8.94 %
	Mult-VAE	10	Backbone	-	0.6980	0.7028	0.6979	0.0219	0.0361	0.0142	-
			Diversification	$\eta = 0.20$	0.6905	0.6984	0.6903	0.0546	0.0832	0.0185	130.28 %
			Re-ranking	$\alpha = 0.00$	0.6980	0.7028	0.6979	0.0219	0.0361	0.0142	100.00 %
			Calibration	$\lambda = 0.80$	0.6979	0.6998	0.6979	0.0348	0.0361	0.0013	9.15 %
		25	Backbone	-	0.6127	0.6247	0.6125	0.0235	0.0425	0.0190	-
			Diversification	$\eta = 0.20$	0.6116	0.6246	0.6114	0.0350	0.0613	0.0075	39.47 %
			Re-ranking	$\alpha = 0.04$	0.6127	0.6226	0.6125	0.0522	0.0425	0.0097	51.05 %
			Calibration	$\lambda = 0.99$	0.6127	0.6214	0.6125	0.0399	0.0425	0.0026	13.68 %
		50	Backbone	-	0.6973	0.7040	0.6972	0.0271	0.0422	0.0151	-
			Diversification	$\eta = 0.30$	0.6969	0.7036	0.6968	0.0375	0.0569	0.0047	31.13 %
			Re-ranking	$\alpha = 0.02$	0.6973	0.7032	0.6972	0.0380	0.0422	0.0042	27.81 %
			Calibration	$\lambda = 0.99$	0.6973	0.7026	0.6972	0.0348	0.0422	0.0074	49.01 %
		75	Backbone	-	0.7334	0.7465	0.7332	0.0278	0.0432	0.0154	-
			Diversification	$\eta = 0.40$	0.7331	0.7465	0.7329	0.0388	0.0570	0.0044	28.57 %
			Re-ranking	$\alpha = 0.04$	0.7334	0.7462	0.7332	0.0417	0.0432	0.0015	9.74 %
			Calibration	$\lambda = 0.99$	0.7334	0.7459	0.7332	0.0359	0.0432	0.0073	47.40 %
		100	Backbone	-	0.7486	0.7626	0.7484	0.0272	0.0411	0.0139	-
			Diversification	$\eta = 0.50$	0.7484	0.7625	0.7482	0.0384	0.0550	0.0027	19.42 %
			Re-ranking	$\alpha = 0.03$	0.7486	0.7619	0.7484	0.0425	0.0411	0.0014	10.07 %
			Calibration	$\lambda = 0.99$	0.7486	0.7622	0.7484	0.0339	0.0411	0.0072	51.80 %
Aggression	SLIM	10	Backbone	-	0.8551	0.8547	0.8551	0.1008	0.0833	0.0175	-
			Diversification	$\eta = 0.00$	0.8551	0.8547	0.8551	0.1008	0.0833	0.0175	100.00 %
			Re-ranking	$\beta = 0.20$	0.8549	0.8525	0.8551	0.0859	0.0833	0.0026	14.86 %
			Calibration	$\lambda = 0.90$	0.8540	0.8381	0.8551	0.1000	0.0833	0.0167	95.43 %
		25	Backbone	-	0.7581	0.7603	0.7580	0.0953	0.0790	0.0163	-
			Diversification	$\eta = 0.00$	0.7581	0.7603	0.7580	0.0953	0.0790	0.0163	100.00 %
			Re-ranking	$\beta = 0.20$	0.7581	0.7592	0.7580	0.0887	0.0790	0.0097	59.51 %
			Calibration	$\lambda = 0.99$	0.7573	0.7469	0.7580	0.0800	0.0790	0.0010	6.13 %
		50	Backbone	-	0.8272	0.8303	0.8270	0.0809	0.0692	0.0118	-
			Diversification	$\eta = 0.00$	0.8272	0.8303	0.8270	0.0809	0.0692	0.0118	100.00 %
			Re-ranking	$\beta = 0.12$	0.8270	0.8283	0.8270	0.0703	0.0692	0.0011	9.32 %
			Calibration	$\lambda = 0.99$	0.8268	0.8242	0.8270	0.0694	0.0692	0.0002	1.69 %
		75	Backbone	-	0.8457	0.8497	0.8455	0.0773	0.0673	0.0100	-
			Diversification	$\eta = 0.00$	0.8457	0.8497	0.8455	0.0773	0.0673	0.0100	100.00 %
			Re-ranking	$\beta = 0.02$	0.8457	0.8488	0.8455	0.0681	0.0673	0.0008	8.00 %
			Calibration	$\lambda = 0.99$	0.8456	0.8471	0.8455	0.0693	0.0673	0.0020	20.00 %
		100	Backbone	-	0.8549	0.8557	0.8549	0.0751	0.0668	0.0083	-
			Diversification	$\eta = 0.00$	0.8549	0.8557	0.8549	0.0751	0.0668	0.0083	100.00 %
			Re-ranking	$\beta = 0.10$	0.8549	0.8553	0.8549	0.0671	0.0668	0.0003	3.61 %
			Calibration	$\lambda = 0.99$	0.8548	0.8544	0.8549	0.0697	0.0668	0.0029	34.94 %
	Mult-VAE	10	Backbone	-	0.6957	0.6938	0.6959	0.1088	0.0880	0.0208	-
			Diversification	$\eta = 0.00$	0.6957	0.6938	0.6959	0.1088	0.0880	0.0208	100.00 %
			Re-ranking	$\beta = 0.20$	0.6956	0.6916	0.6959	0.0864	0.0880	0.0016	7.69 %
			Calibration	$\lambda = 0.99$	0.6948	0.6800	0.6959	0.1000	0.0880	0.0120	57.69 %
		25	Backbone	-	0.6117	0.6152	0.6115	0.0982	0.0781	0.0202	-
			Diversification	$\eta = 0.00$	0.6117	0.6152	0.6115	0.0982	0.0781	0.0202	100.00 %
			Re-ranking	$\beta = 0.20$	0.6117	0.6147	0.6115	0.0864	0.0781	0.0083	41.09 %
			Calibration	$\lambda = 0.99$	0.6111	0.6060	0.6115	0.0798	0.0781	0.0017	8.42 %
		50	Backbone	-	0.6985	0.7035	0.6982	0.0947	0.0788	0.0158	-
			Diversification	$\eta = 0.00$	0.6985	0.7035	0.6982	0.0947	0.0788	0.0158	100.00 %
			Re-ranking	$\beta = 0.14$	0.6984	0.7023	0.6982	0.0772	0.0788	0.0016	10.13 %
			Calibration	$\lambda = 0.99$	0.6983	0.7000	0.6982	0.0801	0.0788	0.0013	8.23 %
		75	Backbone	-	0.7339	0.7367	0.7337	0.0922	0.0772	0.0149	-
			Diversification	$\eta = 0.00$	0.7339	0.7367	0.7337	0.0922	0.0772	0.0149	100.00 %
			Re-ranking	$\beta = 0.14$	0.7339	0.7364	0.7337	0.0772	0.0772	0.0000	0.00 %
			Calibration	$\lambda = 0.99$	0.7339	0.7355	0.7337	0.0801	0.0772	0.0029	19.46 %
		100	Backbone	-	0.7521	0.7548	0.7519	0.0878	0.0745	0.0133	-
			Diversification	$\eta = 0.00$	0.7521	0.7548	0.7519	0.0878	0.0745	0.0133	100.00 %
			Re-ranking	$\beta = 0.13$	0.7521	0.7542	0.7519	0.0751	0.0745	0.0006	4.51 %
			Calibration	$\lambda = 0.99$	0.7521	0.7545	0.7519	0.0794	0.0745	0.0049	36.84 %

Table 6.7: The overall, sensitive and non-sensitive $NDCG@k$, the sensitive and non-sensitive $AIR@k$, the absolute difference $|\Delta_{AIR@k}|$ and the percentage difference $|\Delta_{AIR@k}\%|$ divided by case study, backbone recommender, algorithmic approach, and number of k recommended items.

Considering a representative example with $k = 100$, in Table 6.6 we show the pages associated with the highest and lowest SIB_i@100 for the different case studies and recommendation algorithms. As can be observed, anime and manga pages such as *Naruto* or *Bleach* are recommended most frequently to the group of potentially depressed users. Sports such as basketball or singers such as *Alicia Keys* and *Eminem* are less frequently recommended. As for potentially aggressive users, heavy metal bands such as *Seether* and *Slipknot* are the most commonly recommended. Sports such as volleyball, soccer, or religious pages about *Bible* or *Jesus* are recommended less frequently.

6.4.3 Analysis of Algorithms to Balance Influential Items in Recommendation

As we observed in the previous experiment, SLIM and Mult-VAE show a tendency to over-recommend controversial items and under-recommend favorable items to sensitive users. Given this tendency, we exploited the methodologies introduced in Section 6.2, respectively, to promote the recommendation of favorable items to potentially depressed users and to discourage the recommendation of controversial items to potentially aggressive users. In this section, we study the variations in performance resulting from the application of these procedures.

The experiment proceeds as follows. We used the procedure presented in Section 6.2.1, referred to as *Calibration Approach*, to calibrate the recommendations for both groups of sensitive users based on the item distributions of the non-sensitive user groups. For the experiments, we varied the regularization coefficient λ defined in Eq. (6.5) in the range $[0, 0.99]$ to balance the distributions of the recommended items until they converge with the target distribution. We then compared the results with those obtained from the application of the procedure introduced in Section 6.2.2, referred to as *Re-ranking Approach*. In this case, for experiments involving potentially depressed users, we varied the coefficient α defined in Eq. (6.7) in the range $[0, 0.5]$ to promote the recommendation of favorable items. As for experiments with potentially aggressive users, instead, we varied the threshold β defined in Eq. (6.7) in the range $[0, 0.5]$ to discourage recommendations of controversial items. Both the α and β parameter ranges were selected to be able to analyze the performance of the algorithms at the point where the item distributions of sensitive users converged with those of non-sensitive users.

In Table 6.7 we present the results of experiments performed using the experimental setting discussed in Section 6.4.2. In the following we discuss the results obtained for *Calibration* and *Re-ranking* algorithms. Since we are not interested in a performance comparison but in a simple analysis of the results as the regularization parameters vary, we reported the results with parameters that minimize the absolute difference in $AIR@k$ between the sensitive and non-sensitive user groups ($|\Delta_{AIR@k}|$). However, in a real-world conditions a human operator (and possibly an additional validation set) would be required to set the hyperparameters of the post-processing approaches to obtain an acceptable tradeoff between diversity and the quality of recommendations for sensitive users. As can be observed, compared to backbones, the $AIR@k$ results indicate that the distributions of favorable items recommended to potentially depressed users and controversial items recommended to potentially aggressive users are more similar to those of non-sensitive users. The use of algorithmic approaches allowed in almost* all cases to optimize $|\Delta_{AIR@k}|$ without compromising $NDCG@k$ both overall and measured on the sensitive group.

*In the case of potentially depressed users with $k = 10$, the re-ranking algorithm do not to improve because, for numerical reasons, the % of influential items it adds to the sensitive users rank as α increases is always too high compared to that of non-sensitive users.

6.4.4 Comparison with Other Approaches from the Diversity Literature

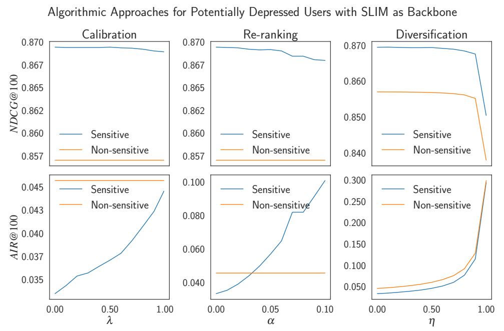
To address the problem discussed in Section 6.1 other methods from the diversity literature (Section 2.3.5.1) could also have been used. In this section we study pros and cons of these methodologies* and compare their performance with that of the approaches proposed in Section 6.2. As discussed in a recent survey [252], there are various approaches in the literature that can be used to introduce diversity into recommendations. It is well understood that the diversification of recommendations also results in the reduction of bias. Accordingly, some of the main diversification algorithms may be used to address the problem proposed in this article.

To understand whether some of the methodologies in the literature were applicable, we first analyzed the main references [1, 57, 61, 92, 197, 256, 260, 359, 396, 498] proposed in the work of Kunaver and Požrl [252]. We found that some of these algorithms could not be applied. Specifically, some algorithms are domain-specific [396] and can only be applied in the music domain. Other algorithms require additional information to be applied that is not available in our case, such as item meta-data descriptions or temporal information [1, 92, 256] or specific run-time user input [61]. Of the remaining approaches, some [57, 197, 260, 359, 498] could be applied to our problem. However, some of these [57, 197, 260] diversify the recommendations based on popularity criteria and do not take into consideration either the item category (i.e., influential vs non-influential) or the type of user (i.e., sensitive vs non-sensitive). Others [359, 498] diversify the recommendations by item category, but do not distinguish the type of user. Consequently, although the latters turn out to be applicable, since they make no distinction between sensitive and non-sensitive users the increase in diversity would also lead to a decrease in overall accuracy.

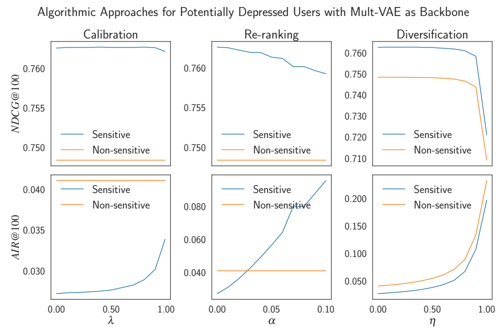
To compare the performance of diversification algorithms with those proposed in our paper, we implemented the algorithm proposed by Ziegler et al. [498] as it was considered the most meaningful for our context. The algorithm, referred to as *Diversification*, allows user recommendations to be diversified based on item category through a η diversification factor. The experimental methodology proceeds as follows. We used the algorithm [498] to diversify user recommendations according to the type of item: whether influential or not-influential. For the experiments, we varied the diversification factor η in the range $[0, 1]$. Tests were performed exploiting the experimental setting discussed in Section 6.4.2. In the following we discuss the results obtained.

In Table 6.7 we can observe the results of the *Diversification* algorithm [498] divided by case study, backbone recommender system and number of recommended items. As in the previous experiment, we report the results with the parameters that minimize the difference in $AIR@k$ between sensitive and non-sensitive users. To make a fair comparison in this case we calculated the $|\Delta_{AIR@k}|$ between the $AIR@k$ of sensitive users its corresponding non-sensitive backbone value. As can be seen by comparing the results with our methods: in the case of potentially depressed users, the diversification algorithm shows on average a higher $|\Delta_{AIR@k}\%|$ w.r.t. those obtained from Calibration and Re-ranking; in the case of potentially aggressive users, the algorithm does not obtain satisfactory performance, and the best results are obtained when no diversification is made ($\eta = 0.00$). The latter result on potentially aggressive users is due to the fact that, because of the design of the algorithm, the diversification always results in an upward shift of the distribution of influential and non-influential items, and thus the controversial

*Note that in addition to the diversity algorithms, other fairness algorithms (Section 2.3.5.2), if repurposed, could also have been used to address the problem. However, since the objective of the paper is to introduce the problem and propose two initial solutions, repurposing other fairness algorithms does not currently result in scope but would certainly be a promising future research direction.



(a) The $NDCG@100$ (upper corners) and $AIR@100$ (lower corners) obtained using SLIM as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).



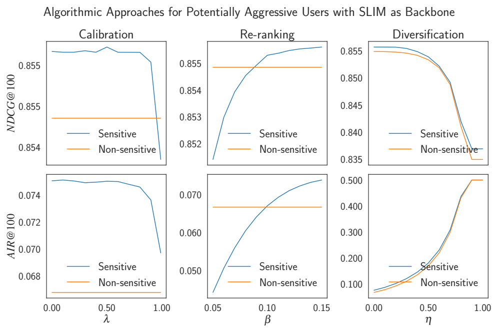
(b) The $NDCG@100$ (upper corners) and $AIR@100$ (lower corners) obtained using Mult-VAE as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).

Figure 6.6: The results of the experiments for potentially depressed users divided by backbone recommender system (SLIM left, Mult-VAE right) and algorithmic approach.

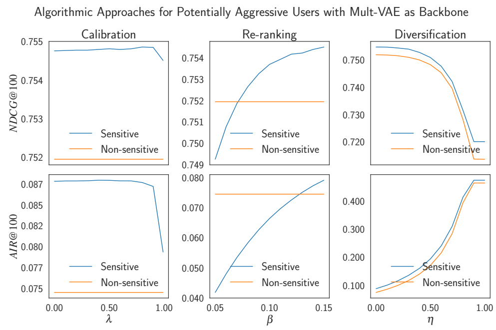
items of sensitive users never fall to the same level as those of non-sensitive users. In addition, the algorithm changes the item rankings also for non-sensitive users. This results in a slightly lower non-sensitive $NDCG@k$ performance when compared to our methods that leave the item rankings of non-sensitive users unaltered.

Below we show a representative example with $k = 100$ to observe how the metrics measured for the *Calibration*, *Re-ranking* and *Diversification* approaches vary as the hyperparameters vary. Other top-k settings show similar behavior. Figure 6.6 shows the results of the experiments that compare the different methodologies used to optimize the recommendations for potentially depressed users divided by backbone recommender system. As can be observed, using the calibration approach, as the coefficient λ increases, the $AIR@100$ tends to converge with the average percentage of influential items of non-sensitive users. Exploiting the re-ranking approach, on the other hand, as the α coefficient increases, the $AIR@100$ first converges with those of non-sensitive users and then exceeds them. As for the diversification approach, instead, by increasing the η parameter, the $AIR@100$ of influential items tends to increase and converge with that of non-influential items. All results were expected, as calibration aims to calibrate the item distributions of the recommendations for sensitive users until they converge with the target distribution, while re-ranking is used with the aim of promoting favorable item recommendations for potentially depressed users based on a percentage criterion and diversification is applied to both sensitive and non-sensitive users. In both re-ranking and calibration cases, the $NDCG@100$ of sensitive users decreases slightly without compromising overall performance. As for diversification, the decrease in $NDCG@100$ is much more important than calibration and re-ranking because it is applied also to non-sensitive users.

In Figure 6.7 we show the results of the experiments for potentially aggressive users. Overall outcomes are in line with previous ones with a small exception in the re-ranking approach. Since constraints tend to discourage rather than promote controversial items in sensitive users recommendations, as β decreases, the behavior we observe is the opposite of the depression case. In this case, as β decreases, the $AIR@100$ of the sensitive users first



(a) The $NDCG@100$ (upper corners) and $AIR@100$ (lower corners) obtained using SLIM as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).



(b) The $NDCG@100$ (upper corners) and $AIR@100$ (lower corners) obtained using Mult-VAE as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).

Figure 6.7: The results of the experiments for potentially aggressive users divided by backbone recommender system (SLIM left, Mult-VAE right) and algorithmic approach.

converges with that of the non-sensitive users and then diverges both in the case where the backbone is based on SLIM and Mult-VAE. Overall, as expected, sensitive users $NDCG@100$ decreases as λ and η increases and as β decreases. The $NDCG@k$ performance are not compromised in the case of re-ranking and calibration but there is a greater impact in the case of diversification since is applied also to non-sensitive users. Moreover, the diversification algorithm does not achieve satisfactory results in the case of potentially aggressive users because it never results in a reduction in the $AIR@100$ of sensitive users but always in its increase.

6.4.5 Analysis of Proposed Methodology to Combine Different Approaches

In this section we study the performance of a method that combines the output of different algorithms simultaneously with the goal of further improving the results and to relieve a practitioner from the need of selecting a-priori a single approach. The experiment proceeds as follows. We used the methodology presented in Section 6.2.3, referred to as the *Combination Approach* to combine the rankings obtained from *Calibration*, *Re-ranking* and *Diversification* algorithms discussed in previous sections. For each experimental scenario (i.e., the tuple consisting of case study, backbone recommender and number of recommender items), the algorithm was executed by combining the output of the techniques and hyper-parameters reported in Table 6.7.

In Table 6.8, we present the results of the *Combination* method divided by case study, backbone recommender system and number of recommended items, exploiting the experimental setting discussed in Section 6.4.2. The results are compared with those of the techniques that had obtained the best outcomes in previous experiments.

					NDCG@k			AIR@k			
Case Study	Backbone	k	Approach	Hyper	Overall	Sensitive	Non-Sensitive	Sensitive	Non-Sensitive	$\lvert\Delta_{AIR@k}\rvert$	$\lvert\Delta_{AIR@k}\%\rvert$
Depression	SLIM	10	Calibration	$\lambda = 0.60$	0.8558	0.8694	0.8556	0.0316	0.0318	0.0002	1.50 %
			Combination	-	0.8558	0.8700	0.8556	0.0310	0.0318	0.0008	6.02 %
		25	Calibration	$\lambda = 0.60$	0.7590	0.7722	0.7588	0.0419	0.0419	0.0000	0.00 %
			Combination	-	0.7589	0.7700	0.7588	0.0418	0.0419	0.0001	0.78 %
		50	Calibration	$\lambda = 0.90$	0.8290	0.8399	0.8288	0.0413	0.0454	0.0041	29.50 %
			Combination	-	0.8290	0.8395	0.8288	0.0452	0.0454	0.0002	1.44 %
		75	Calibration	$\lambda = 0.99$	0.8476	0.8572	0.8474	0.0436	0.0463	0.0027	19.57 %
			Combination	-	0.8476	0.8572	0.8474	0.0463	0.0463	0.0000	0.00 %
		100	Calibration	$\lambda = 0.99$	0.8572	0.8689	0.8570	0.0446	0.0457	0.0011	8.94 %
			Combination	-	0.8572	0.8689	0.8570	0.0457	0.0457	0.0000	0.00 %
	Mult-VAE	10	Calibration	$\lambda = 0.80$	0.6979	0.6998	0.6979	0.0348	0.0361	0.0013	9.15 %
			Combination	-	0.6979	0.7004	0.6979	0.0357	0.0361	0.0004	2.82 %
		25	Calibration	$\lambda = 0.99$	0.6127	0.6214	0.6125	0.0399	0.0425	0.0026	13.68 %
			Combination	-	0.6127	0.6216	0.6125	0.0424	0.0425	0.0001	0.53 %
		50	Re-ranking	$\alpha = 0.02$	0.6973	0.7032	0.6972	0.0380	0.0422	0.0042	27.81 %
			Combination	-	0.6973	0.7019	0.6972	0.0421	0.0422	0.0001	0.66 %
		75	Re-ranking	$\alpha = 0.04$	0.7334	0.7462	0.7332	0.0417	0.0432	0.0015	9.74 %
			Combination	-	0.7334	0.7455	0.7332	0.0431	0.0432	0.0001	0.65 %
		100	Re-ranking	$\alpha = 0.03$	0.7486	0.7619	0.7484	0.0425	0.0411	0.0014	10.07 %
			Combination	-	0.7486	0.7619	0.7484	0.0409	0.0411	0.0002	1.44 %
Aggression	SLIM	10	Re-ranking	$\beta = 0.20$	0.8549	0.8525	0.8551	0.0859	0.0833	0.0026	14.86 %
			Combination	-	0.8543	0.8426	0.8551	0.0833	0.0833	0.0000	0.00 %
		25	Calibration	$\lambda = 0.99$	0.7573	0.7469	0.7580	0.0800	0.0790	0.0010	6.13 %
			Combination	-	0.7573	0.7470	0.7580	0.0790	0.0790	0.0000	0.00 %
		50	Calibration	$\lambda = 0.99$	0.8268	0.8242	0.8270	0.0694	0.0692	0.0002	1.69 %
			Combination	-	0.8268	0.8246	0.8270	0.0691	0.0692	0.0001	0.85 %
		75	Re-ranking	$\beta = 0.02$	0.8457	0.8488	0.8455	0.0681	0.0673	0.0008	8.00 %
			Combination	-	0.8456	0.8471	0.8455	0.0673	0.0673	0.0000	0.00 %
		100	Re-ranking	$\beta = 0.10$	0.8549	0.8553	0.8549	0.0671	0.0668	0.0003	3.61 %
			Combination	-	0.8548	0.8544	0.8549	0.0668	0.0668	0.0000	0.00 %
	Mult-VAE	10	Re-ranking	$\beta = 0.20$	0.6956	0.6916	0.6959	0.0864	0.0880	0.0016	7.69 %
			Combination	-	0.6950	0.6822	0.6959	0.0880	0.0880	0.0000	0.00 %
		25	Calibration	$\lambda = 0.99$	0.6111	0.6060	0.6115	0.0798	0.0781	0.0017	8.42 %
			Combination	-	0.6112	0.6062	0.6115	0.0781	0.0781	0.0000	0.00 %
		50	Calibration	$\lambda = 0.99$	0.6983	0.7000	0.6982	0.0801	0.0788	0.0013	8.23 %
			Combination	-	0.6983	0.7005	0.6982	0.0788	0.0788	0.0000	0.00 %
		75	Re-ranking	$\beta = 0.14$	0.7339	0.7364	0.7337	0.0772	0.0772	0.0000	0.00 %
			Combination	-	0.7339	0.7356	0.7337	0.0772	0.0772	0.0000	0.00 %
		100	Re-ranking	$\beta = 0.13$	0.7521	0.7542	0.7519	0.0751	0.0745	0.0006	4.51 %
			Combination	-	0.7521	0.7543	0.7519	0.0745	0.0745	0.0000	0.00 %

Table 6.8: The overall, sensitive and non-sensitive $NDCG@k$, the sensitive and non-sensitive $AIR@k$, the absolute difference $|\Delta_{AIR@k}|$ and the percentage difference $|\Delta_{AIR@k}\%|$ divided by case study, backbone recommender, and number of k recommended items for the combination approach compared to the approaches that best performed in the previous experiments.

As can be observed, the combination approach achieved better results in almost* all cases, further decreasing the $|\Delta_{AIR@k}\%|$ and keeping the overall, sensitive and non-sensitive $NDCG@k$ high. The algorithm in particular seems to perform very well in the case of potentially aggressive users. This can be explained intuitively based on the design of the algorithm. In fact, as previously discussed in Section 6.2.3 since the algorithm is incremental and iterates on all sensitive users: for the first sensitive users processed the $|\Delta_{AIR@k}|$ tends to vary; as the number of users processed increases, the $|\Delta_{AIR@k}|$ tends to stabilize and decrease. Consequently, since the number of potentially aggressive users is higher than the number of potentially depressed users (Table 6.4), the algorithm obtains better results in the former case, because the initial instability phase is absorbed by the high number of users.

6.5 Discussion

In the previous experiments, we proposed two different cases to study the recommendation of favorable items (e.g. sports and hobbies pages) to potentially depressed users and the recommendation of controversial items (e.g. alcohol, weapons, and death metal pages) to potentially aggressive users. As we have seen (see Section 6.4.2), in some cases recommender systems tend to over-recommend controversial items and to under-recommend favorable items to sensitive users. The techniques introduced in Section 6.2 were used to promote the recommendation of favorable items to potentially depressed users and discourage the recommendation of controversial items to potentially aggressive users (Experiment 6.4.3). These techniques were also compared with one of the most significant recommendation diversification algorithms [498] for our context in Section 6.4.4. In Section 6.4.5 finally we also tried to combine the outputs from the various approaches used together to further improve the results. All the proposed techniques proved to be valid to address the problem, but with some differences. In this section, we discuss the pros and cons of the two approaches and the limitations of the present study.

The calibration algorithm proved to be a valid approach, but sometimes the distributions of influential items in sensitive and non-sensitive user recommendations did not converge. As λ in Eq. (6.5) increases it was expected that sensitive users $NDCG@k$ would decrease and the absolute difference in $AIR@k$ would decrease. If $\lambda = 0.99$ is set, the difference between the distributions should be minimized. However, the distributions often did not converge because the KL-based distance function tends to suffer from numerical instability when the difference between the distributions is small. Moreover, the computational complexity was higher compared to the constrained re-ranking approach and consequently also the execution time. Regarding the constraint-based approach, it proved to be effective and fast. Varying the parameters α and β in Eq. (6.7) allowed to easily balance the distribution of influential items. However, it was necessary to study the performance of the algorithm in the α and β ranges to obtain results that maximize $NDCG@k$ and minimize the absolute difference in $AIR@k$. In addition, analyzing the performance of the diversity algorithm that could best be exploited in our context [498] as comparison approach, although it allowed for the diversification of recommendations based on item category, it showed lower

*The only case where *Combination* does not get better results or very close to the baseline results is for potentially depressed users and $k = 10$. In this case, the number of sensitive users is not high enough (Table 6.4) to perfectly stabilize the $|\Delta_{AIR@k}|$. Consequently, as the algorithm iterates over all sensitive users by choosing the best performing rank at each iteration, it does not select only the ranks generated by calibration (that is the best performing method), but also the ones from re-ranking and diversification that did not perform well. As a result, it still gets good results, but suffers from the variability of the ranks selected in the first iterations.

overall performance than our methods. Furthermore, the algorithm could not be satisfactorily applied to the case of potentially aggressive users because, by algorithm design, as η increases, $AIR@k$ always tends to rise and that of sensitive users never falls to the level of that of non-sensitive ones. As for the combination approach, on the other hand, the algorithm presented some interesting features. Although the order in which the algorithm iterated over the sensitive users may have varied the final results, as different rankers were used to generate the candidate lists, the recommendations addressed to the single sensitive user still enjoyed the same properties as the original rankers used. Since rankers were selected based on the hyperparameters maximizing certain diversity indicators, the recommendations addressed to the single sensitive user always presented a certain degree of diversification of influential items. Moreover, iterating over the users, as the number of sensitive users increased, the overall $|\Delta_{AIR@k}|$ tended by design to stabilize and decrease.

6.6 Limitations of the Study and Future Work

We identified some limitations of the present work regarding the data behind the experiments, the way in which sensitive users and influential items were selected, and the algorithms used to address the problem.

Regarding the data, as far as the authors know, the current literature lacked datasets compatible with our experimental settings, i.e., containing information about the users that would have allowed the identification of sensitive ones and item descriptions that could have been used to identify influential items. In particular, although there were other datasets in the literature besides myPersonality that reported Big Five personality traits such as Personality2018 [323], ADS [377], and PsychoFlickr [173], these did not have the necessary characteristics to be used for experiments. In particular, Personality2018 [323] lacked contextual features to identify influential items and was based on a small set of only 1800 users. ADS [377], while being compatible with our experimental setting, exhibited an excessively small size to be used for experiments, i.e. 120 users and 300 items only. Assuming the same percentage of sensitive users as in our considered dataset, the identified number of sensitive users would have been too small (e.g., less than 10 users) to obtain meaningful results. Finally, PsychoFlickr [173], presented the dimensionality problems observed in previous datasets (i.e., 300 users), and was not suitable for use in collaborative filtering scenarios since each user was associated with a different set of items.

Considering instead the data used for experiments based on myPersonality, the pages that users liked were determined by breaking down the LDA topics of the most popular pages. Moreover, to train recommender systems, it was necessary to binarize the user-item interaction matrix considering only the top-k items for each user because the algorithms in the RecTorch library [357] did not accept real values as input. Furthermore, consistently with other work proposing recommendation systems based on myPersonality [315, 358] and the work of Steck [399] it was chosen to evaluate the results of the experiments using the top-$\{10, 25, 50, 75, 100\}$ settings. Together, these factors could have potentially influenced the experiments and interpretations of the results.

Some other points concerned the selection of sensitive users and influential items. As for sensitive users, in the experiments, it was proposed to select them based on certain correlations known in the literature with the Big Five personality traits. To define user groups, correlations with depressive disorders [247] and aggressive behavior [38] were exploited. Although initially relying on correlation might be a valid approach, this could lead to type 1 and 2 errors, selecting users who do not really suffer from the disorder or excluding others who do. Moreover, as for the influential items, these have been selected manually from an analysis of the topics of available pages. For both of the previous points, the selection procedure may have introduced errors and influenced the experiments.

As for the algorithms, the experiments were based exclusively on SLIM [328] and Mult-VAE [270] as backbones. Furthermore, optimization approaches were designed using exclusively post-processing methodologies taking inspiration from algorithmic fairness literature (Section 2.3.5.2). Since the purpose of this article was to introduce the problem and propose two solutions to address it, it was not studied how to adapt other fairness algorithms based on post-processing and in-processing approaches. Moreover, although we have also proposed an initial methodology to combine the outputs of different rankers simultaneously to further improve results, there are many studies in the literature addressing hybrid recommendation [63] and rank fusion [253] and new approaches could be developed or other existing algorithms repurposed. Together, these elements represent promising future research directions.

Furthermore, we identified several other future research directions. An initial research direction could focus on building more datasets to use for experiments. Furthermore, designing automatic methods to select sensitive users and influential items could also be a worthwhile research path. In fact, although the Big Five model provided a theoretical starting point to select sensitive users, asking platform users to fill out a questionnaire is not feasible in most real-world circumstances. In addition, even manually selecting the influential items would not be feasible. Another research direction might focus on diversification algorithms. The algorithms proposed in this paper took inspiration from approaches in the literature of algorithmic fairness applied to recommender systems. A promising research direction could aim to re-adapt other fairness algorithms to address the proposed problem as well. Moreover, the approaches proposed in this article were based on post-processing techniques and considered only one set of influential items for each sensitive user group. Research could investigate the use of in-processing methodologies and extend the problem to manage multiple influential item sets simultaneously. It could also be further investigated how to combine several algorithms simultaneously to achieve superior performance using methodologies from the literature on hybrid recommendation and rank fusion. Finally, another research direction could seek to study how user behavior is affected by the proposed techniques through the use of simulations.

6.7 Summary of Findings

In this chapter, we addressed the problem of recommending influential items to sensitive users as a way to optimize the value for the society as a whole. We defined as sensitive, users whose behavior can be influenced by specific types of items. Similarly we referred to influential, as those items that can influence sensitive users' behavior. In our study, we formalized the problem and proposed two techniques to maximize the performance of a recommender system that aimed to diversify the item distribution to positively affect sensitive users' behavior: mitigating potentially dangerous societal consequences and promoting healthier lifestyles. The first technique was a calibration approach that aimed to balance sensitive users' recommendations based on the distribution of non-sensitive users' influential items. The second technique was a re-ranking approach that aimed to optimize the performance of a recommender system under influential items' constraints. We also proposed a joint approach to combine the outputs of any technique together to achieve better results. We considered a real-world dataset to test the proposed techniques in two different case studies that involved potentially aggressive and depressive users. All techniques proved effective in allowing high performance to be maintained while diversifying influential items.

7
Conclusion and Future Work

This thesis aimed to study value-aware recommendation systems and investigate the benefits and potential harms of using these algorithms in practical applications. The problems we considered are relevant to a large number of business application domains. We can summarize the main contributions of the thesis as follows.

Value-aware recommendation systems are of great interest to industry, however, research is highly scattered nowadays and composed of many works proposed in isolated contexts. With this thesis, we helped academic researchers and industry practitioners to better understand the state-of-the-art. Specifically, in Chapter 3 and Chapter 4 we presented the first systematic reviews based on PRISMA guidelines of value-aware recommender systems and economic recommender systems. Both types of systems are well-suited to be used in real-world business applications such as e-commerce, media streaming sites, and advertising platforms since they offer various benefits to organizations to increase their business KPIs. Depending on the application domain, companies may prefer some value-aware algorithms over others, and some business value categories are more likely to be optimized. By analyzing a very large number of papers from different research streams, we identified various algorithmic approaches that we initially divided into in-processing and post-processing based on when the optimization of the business value occurs (i.e., at learning vs. prediction time). We also specialized the algorithms taxonomy based on the particular specificities of the underlying models. Furthermore, we studied various aspects concerning evaluation methodologies, discussing evaluation metrics, real-world studies, available datasets and reproducibility issues respectively. Moreover, although value-aware and economic recommender systems can bring great value to the business, the optimization of such value often brings new challenges. Correspondingly, our work has also helped to identify such challenges and propose several possible future research directions. In particular, in the future, it may be worth comparing different algorithmic approaches together and designing optimization methods able to consider multiple business value trade-offs (e.g., profitability, fairness and diversity). Overall, more in-depth research is required to improve the reliability of performance evaluation and to design higher-performing systems following trustworthy AI principles.

Other contributions in this thesis focus on some of these possible research paths. In particular, in Chapter 5 we investigated the use of modeling approaches with the goal of building profit-aware recommendation systems. In our work we extended, according to a value-aware perspective, four different families of state-of-the-art collaborative filtering algorithms widely used in industry, namely nearest neighbors, matrix factorization, learning-to-rank and neural models. The key idea is to integrate the objective function of such families of algorithms, embedding profit awareness at learning time and generating more profitable yet relevant recommendations at prediction time. However, our contribution is not limited to modeling aspects but also includes the comparison of proposed methods with the post-processing algorithms that are most commonly used in the literature to generate profit-aware recommendations. An experimental analysis of three different real-world datasets showed that the proposed models were consistently effective in generating more profitable yet relevant recommendations without requiring any computational overhead at prediction time. Hence they can be considered viable alternatives to post-processing approaches, which are currently more popular but have several limitations in practice, e.g., considering large-scale production systems with millions of active users and catalog items. Many extensions of this work are possible in the future and we hope that our study will foster further research in this area. In particular, future works may consider embedding profit awareness into other algorithmic classes or into other algorithms belonging to the same class. Moreover, in the future, it may be worth complementing the analyses with the study of alternative approaches to optimize the profitability of recommendations, which although theoretically possible have never been found in the literature (e.g., considering pre-processing methods). Future developments in this line of research could also include the study of temporal dynamics and long-term business value optimization aspects.

Finally in Chapter 6 we presented a study on the problem of recommending influential items to sensitive users, focusing on broader issues regarding value optimization for society as a whole. Indeed, just as it is important for the business to have a recommender system that can optimize business performance indicators, it is of great value to society to have a recommender system that can maintain high user well-being. Specifically, with our work we proposed three different algorithms that take inspiration from the diversity and fairness literature to calibrate the exposure of sensitive users to influential items while maintaining high performance of the recommender system. Considering a real-world dataset, we extensively study the performance of the proposed techniques in two different case studies involving potentially aggressive and depressive users. All proposed techniques proved effective in allowing high performance to be maintained while properly diversifying influential items: mitigating overexposure of potentially aggressive users to controversial items that may have a negative impact on their behavior and encouraging exposure of potentially depressed users to favorable items that may have a positive impact. Various research directions related to this topic may be interesting to pursue in the future. For example, it might be interesting to study in more detail how to automatically select influential items and sensitive users, i.e., without using procedures that are not feasible in practice such as requiring platform users to fill out questionnaires or requiring expert users to manually select influential items. In addition, the research could focus on the study of other post-processing methods or new in-processing techniques that can calibrate the diversity of recommendations while handling multiple influential item sets targeting different sensitive user groups simultaneously. Furthermore, it might be interesting to study in more detail how user behavior is impacted by the proposed techniques through the use of simulations. In particular, new techniques could be investigated to improve users' future lifestyles and bring value to society while still considering business interests.

References

[1] Zeinab Abbassi, Vahab S. Mirrokni, and Mayur Thakur. Diversity maximization under matroid constraints. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 32–40. ACM, 2013. doi: 10.1145/2487575.2487636.

[2] Himan Abdollahpouri. Popularity bias in ranking and recommendation. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 529–530, 2019.

[3] Himan Abdollahpouri and Robin Burke. Multistakeholder Recommender Systems. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 647–677. Springer US, 2022. ISBN 978-1-07-162196-7 978-1-07-162197-4. doi: 10.1007/978-1-0716-2197-4_17.

[4] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1):127–158, March 2020. ISSN 0924-1868, 1573-1391. doi: 10.1007/s11257-019-09256-1.

[5] Sibel Adali and Jennifer Golbeck. Predicting personality with social behavior. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 302–309. IEEE, 2012. doi: 10.1109/ASONAM.2012.58.

[6] Panagiotis Adamopoulos and Alexander Tuzhilin. The business value of recommendations: A privacy-preserving econometric analysis. In *Proceedings of the International Conference on Information Systems*, 2015.

[7] Panagiotis Adamopoulos and Alexander Tuzhilin. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology*, 5(4): 1–32, January 2015. ISSN 2157-6904, 2157-6912. doi: 10.1145/2559952.

[8] Hekmat Adelnia Najafabadi, Ahmadreza Shekarchizadeh, Akbar Nabiollahi, Naser Khani, and Hamid Rastgari. Dynamic pricing for information goods using revenue management and recommender systems. *Journal of Revenue and Pricing Management*, 21(2):153–163, April 2022. ISSN 1476-6930, 1477-657X. doi: 10.1057/s41272-020-00276-w.

[9] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, June 2005. ISSN 1041-4347. doi: 10.1109/TKDE.2005.99.

[10] Gediminas Adomavicius and YoungOk Kwon. Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE Transactions on Knowledge and Data Engineering*, 24(5):896–911, May 2012. ISSN 1041-4347. doi: 10.1109/TKDE.2011.15.

[11] Gediminas Adomavicius, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang. Do Recommender Systems Manipulate Consumer Preferences? A Study of Anchoring Effects. *Information Systems Research*, 24(4):956–975, December 2013. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.2013.0497.

[12] Gediminas Adomavicius, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang. Effects of Online Recommendations on Consumers' Willingness to Pay. *Information Systems Research*, 29(1):84–102, March 2018. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.2017.0703.

[13] Gediminas Adomavicius, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang. Recommender systems, ground truth, and preference pollution. *AI Magazine*, 43(2):177–189, June 2022. ISSN 0738-4602, 2371-9621. doi: 10.1002/aaai.12055.

[14] M. Mehdi Afsar, Trafford Crump, and Behrouz Far. Reinforcement Learning based Recommender Systems: A Survey. *ACM Computing Surveys*, 55(7):1–38, 2022. ISSN 0360-0300, 1557-7341. doi: 10.1145/3543846.

[15] Saloni Agarwal, Aparupa Das Gupta, and Amit Pande. Improving Bundles Recommendation Coverage in Sparse Product Graphs. In *Companion Proceedings of the Web Conference*, pages 1037–1045. ACM, April 2022. ISBN 978-1-4503-9130-6. doi: 10.1145/3487553.3524715.

[16] Leman Akoglu and Christos Faloutsos. ValuePick: Towards a Value-Oriented Dual-Goal Recommender System. In *Proceedings of the IEEE International Conference on Data Mining*, pages 1151–1158. IEEE, December 2010. ISBN 978-1-4244-9244-2. doi: 10.1109/ICDMW.2010.68.

[17] Sumaia Mohammed Al-Ghuribi and Shahrul Azman Mohd Noah. Multi-Criteria Review-Based Recommender System–The State of the Art. *IEEE Access*, 7:169446–169468, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2954861.

[18] Bushra Alhijawi, Arafat Awajan, and Salam Fraihat. Survey on the Objectives of Recommender Systems: Measures, Solutions, Evaluation Methodology, and New Perspectives. *ACM Computing Surveys*, 55(5): 1–38, May 2023. ISSN 0360-0300, 1557-7341. doi: 10.1145/3527449.

[19] Bayad Jamal Ali and Govand Anwar. Marketing Strategy: Pricing strategies and its influence on consumer purchasing decision. *International Journal of Rural Development, Environment and Health Research*, 5 (2):26–39, 2021. ISSN 24568678. doi: 10.22161/ijreh.5.2.4.

[20] Alaa Alslaity and Thomas Tran. Goal Modeling-based Evaluation for Personalized Recommendation Systems. In *Adjunct Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization*, pages 276–283. ACM, June 2021. ISBN 978-1-4503-8367-7. doi: 10.1145/3450614.3464619.

[21] Xavier Amatriain. Beyond data: from user information to business value through personalized recommendations and consumer science. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 2201–2208. ACM Press, 2013. ISBN 978-1-4503-2263-8. doi: 10.1145/2505515.2514701.

[22] Xavier Amatriain and Justin Basilico. Past, Present, and Future of Recommender Systems: An Industry Perspective. In *Proceedings of the ACM Conference on Recommender Systems*, pages 211–214. Association for Computing Machinery, 2016. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959144.

[23] Yair Amichai-Hamburger and Gideon Vinitzky. Social network use and personality. *Computers in Human Behavior*, 26(6):1289–1295, 2010. doi: 10.1016/j.chb.2010.03.018.

[24] James C. Anderson and James A. Narus. Business marketing: understand what customers value. *Harvard Business Review*, 76:53–67, 1998.

[25] James C. Anderson and James A. Narus. *Understanding, Creating, and Delivering Value. Business Market Management*. New Jersey: Prentice Hall, Upper Saddle River, 1999.

[26] Vito Walter Anelli, Alejandro Bellogín, Tommaso Di Noia, Dietmar Jannach, and Claudio Pomo. Top-n recommendation algorithms: A quest for the state-of-the-art. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization*, pages 121–131, 2022. doi: 10.1145/3503252.3531292.

[27] Arda Antikacioglu and R. Ravi. Post Processing Recommender Systems for Diversity. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 707–716, Halifax NS Canada, August 2017. ACM. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098173.

[28] Markus Appel, Caroline Marker, and Timo Gnambs. Are social media ruining our lives? a review of meta-analytic evidence. *Review of General Psychology*, 24(1):60–74, 2020. doi: 10.1177/1089268019880891.

[29] Guy Aridor and Duarte Gonçalves. Recommenders' originals: The welfare effects of the dual role of platforms as producers and recommender systems. *International Journal of Industrial Organization*, 83: 102845, July 2022. ISSN 0167-7187. doi: 10.1016/j.ijindorg.2022.102845.

[30] Kamelia Aryafar, Devin Guillory, and Liangjie Hong. An Ensemble-based Approach to Click-Through Rate Prediction for Promoted Listings at Etsy. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1–6, Halifax NS Canada, August 2017. ACM. ISBN 978-1-4503-5194-2. doi: 10.1145/3124749.3124758.

[31] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013. doi: 10.1176/appi.books.

[32] Tzoof Avny Brosh, Amit Livne, Oren Sar Shalom, Bracha Shapira, and Mark Last. BRUCE: Bundle Recommendation Using Contextualized item Embeddings. In *Proceedings of the ACM Conference on Recommender Systems*, pages 237–245. ACM, September 2022. ISBN 978-1-4503-9278-5. doi: 10.1145/3523227.3546754.

[33] Amos Azaria, Avinatan Hassidim, Sarit Kraus, Adi Eshkol, Ofer Weintraub, and Irit Netanely. Movie recommender system for profit maximization. In *Proceedings of the ACM Conference on Recommender Systems*, pages 121–128. ACM, October 2013. ISBN 978-1-4503-2409-0. doi: 10.1145/2507157.2507162.

[34] Klaus Backhaus, Joerg Becker, Daniel Beverungen, Margarethe Frohs, Oliver Mueller, and Matthias Weddeling. Incorporating Willingness-to-Pay Data into Online Recommendations for Value-Added Services. In *Proceedings of the European Conference on Information Systems*, page 16, 2010.

[35] Klaus Backhaus, Jörg Becker, Daniel Beverungen, Margarethe Frohs, Oliver Müller, Matthias Weddeling, Ralf Knackstedt, and Michael Steiner. Enabling individualized recommendations and dynamic pricing of value-added services through willingness-to-pay data. *Electronic Markets*, 20(2):131–146, June 2010. ISSN 1019-6781, 1422-8890. doi: 10.1007/s12525-010-0032-0.

[36] Jinze Bai, Chang Zhou, Junshuai Song, Xiaoru Qu, Weiting An, Zhao Li, and Jun Gao. Personalized Bundle List Recommendation. In *Proceedings of the International Conference on World Wide Web*, pages 60–71. ACM, May 2019. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313568.

[37] Linas Baltrunas, Bernd Ludwig, and Francesco Ricci. Matrix factorization techniques for context aware recommendation. In *Proceedings of the ACM Conference on Recommender Systems*, pages 301–304. ACM, October 2011. ISBN 978-1-4503-0683-6. doi: 10.1145/2043932.2043988.

[38] Christopher P. Barlett and Craig A. Anderson. Direct and indirect relations between the big 5 personality traits and aggressive and violent behavior. *Personality and Individual Differences*, 52(8):870–875, 2012. doi: 10.1016/j.paid.2012.01.029.

[39] Shreya Basu. Personalized product recommendations and firm performance. *Electronic Commerce Research and Applications*, 48:101074, July 2021. ISSN 1567-4223. doi: 10.1016/j.elerap.2021.101074.

[40] Joeran Beel, Corinna Breitinger, Stefan Langer, Andreas Lommatzsch, and Bela Gipp. Towards reproducibility in recommender-systems research. *User Modeling and User-Adapted Interaction*, 26(1):69–101, March 2016. ISSN 0924-1868, 1573-1391. doi: 10.1007/s11257-016-9174-x.

[41] Moran Beladev, Lior Rokach, and Bracha Shapira. Recommender systems for product bundling. *Knowledge-Based Systems*, 111:193–206, November 2016. ISSN 09507051. doi: 10.1016/j.knosys.2016.08.013.

[42] Alejandro Bellogín and Alan Said. Information Retrieval and Recommender Systems. In Alan Said and Vicenç Torra, editors, *Data Science in Practice*, volume 46, pages 79–96. Springer International Publishing, 2019. ISBN 978-3-319-97555-9 978-3-319-97556-6. doi: 10.1007/978-3-319-97556-6_5.

[43] Thiago Belluf, Leopoldo Xavier, and Ricardo Giglio. Case study on the business value impact of personalized recommendations on a large online retailer. In *Proceedings of the ACM Conference on Recommender Systems*, pages 277–280. ACM, September 2012. ISBN 978-1-4503-1270-7. doi: 10.1145/2365952.2366014.

[44] Izak Benbasat and Weiquan Wang. Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems*, 6(3):72–101, March 2005. ISSN 15369323. doi: 10.17705/1jais.00065.

[45] James Bennett, Stan Lanning, et al. The Netflix prize. In *Proceedings of the KDD Cup*, volume 2007, page 35, 2007.

[46] Dirk Bergemann and Deran Ozmen. Optimal pricing with recommender systems. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 43–51. Association for Computing Machinery, 2006. ISBN 978-1-59593-236-5. doi: 10.1145/1134707.1134713.

[47] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Optimization in online content recommendation services: Beyond click-through rates. *Manufacturing and Service Operations Management*, 18(1):15–33, 2016.

[48] Roger J. Best. *Market-Based Management: Strategies for Growing Customer Value and Profitability.* Pearson, 2013. ISBN 0-13-038775-4.

[49] Saumitra N. Bhaduri and David Fogarty. Customer Lifecycle Value—Past, Present, and Future. In *Advanced Business Analytics*, pages 139–156. Springer Singapore, 2016. ISBN 978-981-10-0726-2 978-981-10-0727-9. doi: 10.1007/978-981-10-0727-9_11.

[50] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 405–414, 2018. doi: 10.1145/3209978.3210063.

[51] Daniel Billsus and Michael J. Pazzani. Learning collaborative information filters. In *Proceedings of the International Conference on Machine Learning*, pages 46–54, 1998.

[52] Kunal Biswas, Palaiahnakote Shivakumara, Umapada Pal, Tapabrata Chakraborti, Tong Lu, and Mohamad Nizam Bin Ayub. Fuzzy and genetic algorithm based approach for classification of personality traits oriented social media images. *Knowledge-Based Systems*, 241:108024, 2022. doi: 10.1016/j.knosys.2021.108024.

[53] Robert C. Blattberg, Edward C. Malthouse, and Scott A. Neslin. Customer Lifetime Value: Empirical Generalizations and Some Conceptual Questions. *Journal of Interactive Marketing*, 23(2):157–168, May 2009. ISSN 1094-9968, 1520-6653. doi: 10.1016/j.intmar.2009.02.005.

[54] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[55] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013. doi: 10.1016/j.knosys.2013.03.012.

[56] Anand V Bodapati. Recommendation Systems with Purchase Data. *Journal of Marketing Research*, 45 (1):77–93, February 2008. ISSN 0022-2437. doi: 10.1509/jmkr.45.1.77.

[57] Rubi Boim, Tova Milo, and Slava Novgorodov. Diversification and refinement in collaborative filtering recommender. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 739–744, 2011. doi: 10.1145/2063576.2063684.

[58] Dirk Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark Graus. Understanding choice overload in recommender systems. In *Proceedings of the ACM Conference on Recommender Systems*, pages 63–70. ACM, September 2010. ISBN 978-1-60558-906-0. doi: 10.1145/1864708.1864724.

[59] Keith Bradley and Barry Smyth. Improving recommendation diversity. In *Proceedings of the Irish Conference on Artificial Intelligence and Cognitive Science*, volume 85, pages 141–152. Citeseer, 2001.

[60] Matthew Brand. A random walks perspective on maximizing satisfaction and profit. In *Proceedings of the SIAM International Conference on Data Mining*, pages 12–19. Society for Industrial and Applied Mathematics, April 2005. ISBN 978-0-89871-593-4 978-1-61197-275-7. doi: 10.1137/1.9781611972757.2.

[61] Derek Bridge and John Paul Kelly. Ways of computing diverse collaborative recommendations. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 41–50. Springer, 2006. doi: 10.1007/11768012_6.

[62] Björn Brodén, Mikael Hammar, Bengt J. Nilsson, and Dimitris Paraschakis. A Bandit-Based Ensemble Framework for Exploration/Exploitation of Diverse Recommendation Components: An Experimental Study within E-Commerce. *ACM Transactions on Interactive Intelligent Systems*, 10(1):1–32, January 2020. ISSN 2160-6455, 2160-6463. doi: 10.1145/3237187.

[63] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted IInteraction*, 12(4):331–370, 2002. doi: 10.1023/A:1021240730564.

[64] Robin Burke. Hybrid web recommender systems. In *The Adaptive Web*, pages 377–408. Springer, 2007.

[65] Robin Burke, Gediminas Adomavicius, Ido Guy, Jan Krasnodebski, Luiz Pizzato, Yi Zhang, and Himan Abdollahpouri. VAMS 2017: Workshop on Value-Aware and Multistakeholder Recommendation. In *Proceedings of the ACM Conference on Recommender Systems*, pages 378–379, Como Italy, August 2017. ACM. ISBN 978-1-4503-4652-8. doi: 10.1145/3109859.3109957.

[66] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on Fairness, Accountability and Transparency*, pages 202–214. PMLR, 2018. doi: 10.18122/b2gq53.

[67] Francis Buttle. *Customer Relationship Management Concepts and Tools, Edisi Pertama*. Bayumedia Publishing, 2007.

[68] Chun Hing Cai, Ada Wai-Chee Fu, Chun Hung Cheng, and Wang Wai Kwong. Mining association rules with weighted items. In *Proceedings of the International Symposium Database Engineering and Applications*, pages 68–77. IEEE Comput. Soc, 1998. ISBN 978-0-8186-8307-7. doi: 10.1109/IDEAS.1998.694360.

[69] Yuanfeng Cai and Dan Zhu. Trustworthy and profit: A new value-based neighbor selection method in recommender systems under shilling attacks. *Decision Support Systems*, 124:113112, September 2019. ISSN 01679236. doi: 10.1016/j.dss.2019.113112.

[70] Joana Carvalho and Pedro J. Nobre. Five-factor model of personality and sexual aggression. *International Journal of Offender Therapy and Comparative Criminology*, 63(5):797–814, 2019. doi: 10.1177/0306624X13481941.

[71] Carlos Castillo. Fairness and transparency in ranking. In *Proceedings of the ACM SIGIR Forum*, volume 52, pages 64–71. ACM New York, NY, USA, 2019. doi: 10.1145/3308774.3308783.

[72] Emanuele Cavenaghi, Lorenzo Camaione, Paolo Minasi, Gabriele Sottocornola, Fabio Stella, and Markus Zanker. An Online Experiment of a Price-Based Re-Rank Algorithm. In *Proceedings of the Workshop on Recommenders in Tourism*, 2022.

[73] Jianxin Chang, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Bundle Recommendation with Graph Convolutional Networks. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1673–1676. ACM, July 2020. ISBN 978-1-4503-8016-4. doi: 10.1145/3397271.3401198.

[74] Jianxin Chang, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Bundle Recommendation and Generation with Graph Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2023. ISSN 1041-4347, 1558-2191, 2326-3865. doi: 10.1109/TKDE.2021.3114586.

[75] Olivier Chapelle, Eren Manavoglu, and Romer Rosales. Simple and Scalable Response Prediction for Display Advertising. *ACM Transactions on Intelligent Systems and Technology*, 5(4):34, 2014.

[76] Patrick Y.K. Chau, Shuk Ying Ho, Kevin K.W. Ho, and Yihong Yao. Examining the effects of malfunctioning personalized services on online users' distrust and behaviors. *Decision Support Systems*, 56:180–191, December 2013. ISSN 01679236. doi: 10.1016/j.dss.2013.05.023.

[77] Jia Chen, Qin Jin, Shiwan Zhao, Shenghua Bao, Li Zhang, Zhong Su, and Yong Yu. Does product recommendation meet its Waterloo in unexplored categories? No, price comes to help. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 667–676. ACM, July 2014. ISBN 978-1-4503-2257-7. doi: 10.1145/2600428.2609608.

[78] Jia Chen, Qin Jin, Shiwan Zhao, Shenghua Bao, Li Zhang, Zhong Su, and Yong Yu. Boosting Recommendation in Unexplored Categories by User Price Preference. *ACM Transactions on Information Systems*, 35 (2):1–27, April 2017. ISSN 1046-8188, 1558-2868. doi: 10.1145/2978579.

[79] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Transactions on Information Systems*, 41 (3):1–39, July 2023. ISSN 1046-8188, 1558-2868. doi: 10.1145/3564284.

[80] Laming Chen, Guoxin Zhang, and Hanning Zhou. Improving the diversity of top-n recommendation via determinantal point process. In *Proceedings of the Large Scale Recommendation Systems Workshop*, 2017.

[81] Li Chen, Wen Wu, and Liang He. How personality influences users' needs for recommendation diversity? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 829–834, 2013. doi: 10.1145/2468356.2468505.

[82] Long-Sheng Chen, Fei-Hao Hsu, Mu-Chen Chen, and Yuan-Chia Hsu. Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences*, 178(4):1032–1048, February 2008. ISSN 00200255. doi: 10.1016/j.ins.2007.09.027.

[83] Mingang Chen. Performance Evaluation of Recommender Systems. *International Journal of Performability Engineering*, 2017. doi: 10.23940/ijpe.17.08.p7.12461256.

[84] Pei-Yu Chen, Shin-yi Wu, and Jungsun Yoon. The Impact of Online Recommendations and Consumer Feedback on Sales. In *Proceedings of the International Conference on Information Systems*, pages 711–724, 2004.

[85] Pei-Yu Chen, Yen-Chun Chou, and Robert J. Kauffman. Community-based recommender systems: Analyzing business models from a systems operator's perspective. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 1 – 10, 2009. doi: 10.1109/HICSS.2009.117.

[86] Shih-Fen S Chen, Kent B Monroe, and Yung-Chien Lou. The effects of framing price promotion messages on consumers' perceptions and purchase intentions. *Journal of Retailing*, 74(3):353–372, September 1998. ISSN 00224359. doi: 10.1016/S0022-4359(99)80100-6.

[87] Shiuann-Shuoh Chen, Bhaskar Choubey, and Vinay Singh. A neural network based price sensitive recommender model to predict customer choices based on price effect. *Journal of Retailing and Consumer Services*, 61:102573, July 2021. ISSN 09696989. doi: 10.1016/j.jretconser.2021.102573.

[88] Yaohua Chen, Yan Zhao, and Yiyu Yao. A Profit-Based Business Model for Evaluating Rule Interestingness. In Ziad Kobti and Dan Wu, editors, *Advances in Artificial Intelligence*, volume 4509, pages 296–307. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-72664-7 978-3-540-72665-4. doi: 10.1007/978-3-540-72665-4_26.

[89] Yujin Chen, Jing Wang, Zhihao Wu, and Youfang Lin. Integrating user-group relationships under interest similarity constraints for social recommendation. *Knowledge-Based Systems*, 249:108921, 2022. doi: 10.1016/j.knosys.2022.108921.

[90] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide & deep learning for recommender systems. In *Proceedings of the Workshop on Deep Learning for Recommender Systems*, pages 7–10, 2016.

[91] Young Sung Cho, Kyung Ah Kim, Song Chul Moon, Soo Ho Park, and Keun Ho Ryu. Effective Purchase Pattern Mining with Weight Based on FRAT Analysis for Recommender in e-Commerce. In James J. Park, Ivan Stojmenovic, Hwa Young Jeong, and Gangman Yi, editors, *Computer Science and its Applications*, volume 330, pages 443–454. Springer Berlin Heidelberg, 2015. ISBN 978-3-662-45401-5 978-3-662-45402-2. doi: 10.1007/978-3-662-45402-2_67.

[92] Sang-Min Choi and Yo-Sub Han. A content recommendation system based on category correlations. In *Proceedings of the International Multi-Conference on Computing in the Global Information Technology*, pages 66–70. IEEE, 2010. doi: 10.1109/ICCGI.2010.31.

[93] Chuan He, Cong Wang, Yi-Xin Zhong, and Rui-Fan Li. A survey on learning to rank. In *Proceedings of the International Conference on Machine Learning and Cybernetics*, pages 1734–1739. IEEE, July 2008. ISBN 978-1-4244-2095-7. doi: 10.1109/ICMLC.2008.4620685.

[94] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 2021. doi: 10.1073/pnas.2023301118.

[95] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 659–666, 2008. doi: 10.1145/1390334.1390446.

[96] José A. Concha-Carrasco, Miguel A. Vega-Rodríguez, and Carlos J. Pérez. A multi-objective artificial bee colony approach for profit-aware recommender systems. *Information Sciences*, 625:476–488, May 2023. ISSN 00200255. doi: 10.1016/j.ins.2023.01.050.

[97] Alan DJ Cooke, Harish Sujan, Mita Sujan, and Barton A Weitz. Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations. *Journal of marketing research*, 39(4):488–497, 2002. doi: 10.1509/jmkr.39.4.488.1912.

[98] Paul T Costa and Robert R McCrae. *The NEO personality inventory*. Psychological assessment resources Odessa, FL, 1985.

[99] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5):455–496, November 2008. ISSN 0924-1868, 1573-1391. doi: 10.1007/s11257-008-9051-3.

[100] Paolo Cremonesi and Dietmar Jannach. Progress in recommender systems research: Crisis? What crisis? *AI Magazine*, 2021. doi: 10.1609/aimag.v42i3.18145.

[101] Paolo Cremonesi, Franca Garzotto, and Roberto Turrin. Investigating the Persuasion Potential of Recommender Systems from a Quality Perspective: An Empirical Study. *ACM Transactions on Interactive Intelligent Systems*, 2(2):1–41, June 2012. ISSN 2160-6455, 2160-6463. doi: 10.1145/2209310.2209314.

[102] Xinyi Dai, Jiawei Hou, Qing Liu, Yunjia Xi, Ruiming Tang, Weinan Zhang, Xiuqiang He, Jun Wang, and Yong Yu. U-rank: Utility-oriented Learning to Rank with Implicit Feedback. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 2373–2380. ACM, October 2020. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3412756.

[103] Aparna Das, Claire Mathieu, and Daniel Ricketts. Maximizing profit using recommender systems. *arXiv*, August 2009. doi: 10.48550/arXiv.0908.3633.

[104] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The YouTube video recommendation system. In *Proceedings of the ACM Conference on Recommender Systems*, pages 293–296, 2010.

[105] Alvise De Biasio, Merylin Monaro, Luca Oneto, Lamberto Ballan, and Nicolò Navarin. On the problem of recommendation for sensitive users and influential items: Simultaneously maintaining interest and diversity. *Knowledge-Based Systems*, page 110699, 2023. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2023.110699. URL https://www.sciencedirect.com/science/article/pii/S0950705123004495.

[106] Alvise De Biasio, Andrea Montagna, Fabio Aiolli, and Nicolò Navarin. A systematic review of value-aware recommender systems. *Expert Systems with Applications*, page 120131, 2023. ISSN 0957-4174. doi: 10.1016/j.eswa.2023.120131. URL https://www.sciencedirect.com/science/article/pii/S0957417423006334.

[107] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. Fairness in recommender systems: Research landscape and future directions. *User Modeling and User-Adapted Interaction*, 2023. doi: 10.1007/s11257-023-09364-z.

[108] Emre M. Demirezen and Subodha Kumar. Optimization of Recommender Systems Based on Inventory. *Production and Operations Management*, 25(4):593–608, April 2016. ISSN 10591478. doi: 10.1111/poms.12540.

[109] Feng Deng. Utility-based Recommender Systems Using Implicit Utility and Genetic Algorithm. In *Proceedings of the International Conference on Mechatronics, Electronic, Industrial and Control Engineering*. Atlantis Press, 2015. ISBN 978-94-6252-062-2. doi: 10.2991/meic-15.2015.197.

[110] Qilin Deng, Kai Wang, Minghao Zhao, Zhene Zou, Runze Wu, Jianrong Tao, Changjie Fan, and Liang Chen. Personalized Bundle Recommendation in Online Games. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 2381–2388. ACM, October 2020. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3412734.

[111] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems*, 22(1):143–177, 2004.

[112] Gaddiel Desirena, Armando Diaz, Jalil Desirena, Ismael Moreno, and Daniel Garcia. Maximizing customer lifetime value using stacked neural networks: An insurance industry application. In *Proceedings of the IEEE International Conference On Machine Learning And Applications*, pages 541–544, 2019. doi: 10.1109/ICMLA.2019.00101.

[113] Sahraoui Dhelim, Nyothiri Aung, and Huansheng Ning. Mining user interest based on personality-aware hybrid filtering in social networks. *Knowledge-Based Systems*, 206:106227, 2020. doi: 10.1016/j.knosys.2020.106227.

[114] Sahraoui Dhelim, Nyothiri Aung, Mohammed Amine Bouras, Huansheng Ning, and Erik Cambria. A survey on personality-aware recommendation systems. *Artificial Intelligence Review*, pages 1–46, 2022. doi: 10.1007/s10462-021-10063-7.

[115] M. Benjamin Dias, Dominique Locher, Ming Li, Wael El-Deredy, and Paulo J.G. Lisboa. The Value of Personalised Recommender Systems to E-Business: A Case Study. In *Proceedings of the ACM Conference on Recommender Systems*, pages 291–294, 2008.

[116] Manqing Dong, Feng Yuan, Lina Yao, Xianzhi Wang, Xiwei Xu, and Liming Zhu. A survey for trust-aware recommender systems: A deep learning perspective. *Knowledge-Based Systems*, 249:108954, August 2022. ISSN 09507051. doi: 10.1016/j.knosys.2022.108954.

[117] Tim Donkers and Jürgen Ziegler. The dual echo chamber: Modeling social media polarization for interventional recommending. In *Proceedings of the ACM Conference on Recommender Systems*, pages 12–22, 2021. doi: 10.1145/3460231.3474261.

[118] M. Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, 18(2):192, 2006. doi: 10.1037/1040-3590.18.2.192.

[119] Navin Dookeram, Zahira Hosein, and Patrick Hosein. A Recommender System for the Upselling of Telecommunications Products. In *Proceedings of the International Conference on Advanced Communication Technology*, pages 66–72. IEEE, February 2022. ISBN 979-11-88428-08-3. doi: 10.23919/ICACT53585.2022.9728818.

[120] Verena Dorner and Michael Scholz. Predicting And Economically Exploiting Utility Thresholds With Utility-Based Recommendation Systems. In *Proceedings of the European Conference on Information Systems*, 2013.

[121] Verena Dorner, Olga Ivanova, and Michael Scholz. Think Twice Before You Buy! How Recommendations Affect Three-Stage Purchase Decision Processes. In *Proceedings of the International Conference on Information Systems*, 2013.

[122] Kevin S Douglas, Gina M Vincent, and John F Edens. *Risk for criminal recidivism: The role of psychopathy.* The Guilford Press, 2018.

[123] Peter Doyle. *Value-Based Marketing: Marketing Strategies for Corporate Growth and Shareholder Value*. John Wiley and Sons, 2000. ISBN 0 471 877271.

[124] Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. The Yahoo!Music dataset and KDD Cup '11. In *Proceedings of the KDD Cup*, pages 3–18, 2012.

[125] Azar Eftekhar, Chris Fullwood, and Neil Morris. Capturing personality from facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior*, 37:162–170, 2014. doi: 10.1016/j.chb.2014.04.048.

[126] Michael D. Ekstrand, F. Maxwell Harper, Martijn C. Willemsen, and Joseph A. Konstan. User perception of differences in recommender algorithms. In *Proceedings of the ACM Conference on Recommender Systems*, pages 161–168, 2014. doi: 10.1145/2645710.2645737.

[127] Michael D. Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D. Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on Fairness, Accountability and Transparency*, pages 172–186. PMLR, 2018.

[128] Mehdi Elahi, Matthias Braunhofer, Francesco Ricci, and Marko Tkalcic. Personality-based active learning for collaborative filtering recommender systems. In *Congress of the Italian Association for Artificial Intelligence*, pages 360–371. Springer, 2013. doi: 10.1007/978-3-319-03524-6_31.

[129] Levent Eraslan and Ahmet Kukuoglu. Social relations in virtual world and social media aggression. *World Journal on Educational Technology: Current Issues*, 11(2):1–11, 2019. doi: 10.18844/wjet.v11i2.4145.

[130] Joan Esteban and Gerald Schneider. Polarization and conflict: Theoretical and empirical issues, 2008.

[131] Markus Ettl, Pavithra Harsha, Anna Papush, and Georgia Perakis. A Data-Driven Approach to Personalized Bundle Pricing and Recommendation. *Manufacturing and Service Operations Management*, 22(3): 461–480, May 2020. ISSN 1523-4614, 1526-5498. doi: 10.1287/msom.2018.0756.

[132] Yan Fang, Xinyue Xiao, Xiaoyu Wang, and Huiqing Lan. Customized Bundle Recommendation by Association Rules of Product Categories for Online Supermarkets. In *Proceedings of the IEEE International Conference on Data Science in Cyberspace*, pages 472–475. IEEE, June 2018. ISBN 978-1-5386-4210-8. doi: 10.1109/DSC.2018.00076.

[133] Chong Feng, Muzammil Khan, Arif Ur Rahman, and Arshad Ahmad. News recommendation systems-accomplishments, challenges & future directions. *IEEE Access*, 8:16702–16725, 2020.

[134] Yufei Feng, Fuyu Lv, Weichen Shen, Menghan Wang, Fei Sun, Yu Zhu, and Keping Yang. Deep session interest network for click-through rate prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2301–2307, 2019.

[135] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the ACM Conference on Recommender Systems*, pages 101–109. ACM, September 2019. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3347058.

[136] Maurizio Ferrari Dacrema, Simone Boglio, Paolo Cremonesi, and Dietmar Jannach. A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research. *ACM Transactions on Information Systems*, 39(2):1–49, April 2021. ISSN 1046-8188, 1558-2868. doi: 10.1145/3434185.

[137] Gavan J. Fitzsimons and Donald R. Lehmann. Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses. *Marketing Science*, 23(1):82–94, February 2004. ISSN 0732-2399, 1526-548X. doi: 10.1287/mksc.1030.0033.

[138] Daniel Fleder and Kartik Hosanagar. Recommender systems and their impact on sales diversity. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 192–199. ACM, June 2007. ISBN 978-1-59593-653-0. doi: 10.1145/1250910.1250939.

[139] Daniel Fleder and Kartik Hosanagar. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science*, 55(5):697–712, 2009.

[140] Daniel Fleder, Kartik Hosanagar, and Andreas Buja. Recommender systems and their effects on consumers: the fragmentation debate. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 229–230. ACM, June 2010. ISBN 978-1-60558-822-3. doi: 10.1145/1807342.1807378.

[141] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Antonio Gomariz, Ted Gueniche, Azadeh Soltani, Zhihong Deng, and Hoang Thanh Lam. The SPMF open-source data mining library version 2. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 36–40, 2016.

[142] Danilo Franco, Nicolò Navarin, Michele Donini, Davide Anguita, and Luca Oneto. Deep fair models for complex data: Graphs labeling and explainable face recognition. *Neurocomputing*, 470:318–334, 2022. doi: 10.1016/j.neucom.2021.05.109.

[143] Chen Gao, Yu Zheng, Nian Li, Yinfeng Li, Yingrong Qin, Jinghua Piao, Yuhan Quan, Jianxin Chang, Depeng Jin, Xiangnan He, and Yong Li. A Survey of Graph Neural Networks for Recommender Systems: Challenges, Methods, and Directions. *ACM Transactions on Recommender Systems*, 1(1):1–51, March 2023. ISSN 2770-6699, 2770-6699. doi: 10.1145/3568022.

[144] Danilo Garcia and Sverker Sikström. The dark side of facebook: Semantic representations of status updates predict the dark triad of personality. *Personality and Individual Differences*, 67:92–96, 2014. doi: 10.1016/j.paid.2013.10.001.

[145] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin, and Amr Huber. Offline and online evaluation of news recommender systems at swissinfo.ch. In *Proceedings of the ACM Conference on Recommender Systems*, pages 169–176. ACM, October 2014. ISBN 978-1-4503-2668-1. doi: 10.1145/2645710.2645745.

[146] Robert Garfinkel, Ram Gopal, Arvind Tripathi, and Fang Yin. Design of a shopbot and recommender system for bundle purchases. *Decision Support Systems*, 42(3):1974–1986, December 2006. ISSN 01679236. doi: 10.1016/j.dss.2006.05.005.

[147] Robert Garfinkel, Ram Gopal, Bhavik Pathak, Rajkumar Venkatesan, and Fang Yin. Empirical Analysis of the Business Value of Recommender Systems. *Social Science Research Network*, 2007. doi: 10.2139/SSRN.958770.

[148] Robert Garfinkel, Ram Gopal, Bhavik Pathak, and Fang Yin. Shopbot 2.0: Integrating recommendations and promotions with comparison shopping. *Decision Support Systems*, 46(1):61–69, December 2008. ISSN 01679236. doi: 10.1016/j.dss.2008.05.006.

[149] Mouzhi Ge, Fatih Gedikli, and Dietmar Jannach. Placing high-diversity items in top-n recommendation lists. In *Proceedings of the Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*, 2011.

[150] Xinyu Ge, Yousha Zhang, Yu Qian, and Hua Yuan. Effects of product characteristics on the bundling strategy implemented by recommendation systems. In *Proceedings of the International Conference on Service Systems and Service Management*, pages 1–6. IEEE, June 2017. ISBN 978-1-5090-6370-3. doi: 10.1109/ICSSSM.2017.7996297.

[151] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Shijie Geng, Zuohui Fu, and Yongfeng Zhang. Maximizing Marginal Utility per Dollar for Economic Recommendation. In *Proceedings of the International Conference on World Wide Web*, pages 2757–2763. ACM Press, 2019. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313725.

[152] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Zuohui Fu, Fei Sun, and Yongfeng Zhang. Learning Personalized Risk Preferences for Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 409–418. ACM, July 2020. ISBN 978-1-4503-8016-4. doi: 10.1145/3397271.3401056.

[153] Yong Ge, Hui Xiong, Alexander Tuzhilin, and Qi Liu. Cost-Aware Collaborative Filtering for Travel Tour Recommendations. *ACM Transactions on Information Systems*, 32(1):1–31, January 2014. ISSN 1046-8188, 1558-2868. doi: 10.1145/2559169.

[154] Paul Geroski, Steve Machin, and John Van Reenen. The Profitability of Innovating Firms. *The RAND Journal of Economics*, 24(2):198, 1993. ISSN 07416261. doi: 10.2307/2555757.

[155] Nada Ghanem, Stephan Leitner, and Dietmar Jannach. Balancing consumer and business value of recommender systems: A simulation-based analysis. *Electronic Commerce Research and Applications*, 55:101195, September 2022. ISSN 15674223. doi: 10.1016/j.elerap.2022.101195.

[156] Abhijeet Ghoshal, Subodha Kumar, and Vijay Mookerjee. Impact of Recommender System on Competition Between Personalizing and Non-Personalizing Firms. *Journal of Management Information Systems*, 31(4):243–277, January 2015. ISSN 0742-1222, 1557-928X. doi: 10.1080/07421222.2014.1001276.

[157] Abhijeet Ghoshal, Vijay S. Mookerjee, and Sumit Sarkar. Recommendations and Cross-selling: Pricing Strategies when Personalizing Firms Cross-sell. *Journal of Management Information Systems*, 38(2):430–456, April 2021. ISSN 0742-1222, 1557-928X. doi: 10.1080/07421222.2021.1912930.

[158] Katie A Gleason, Lauri A Jensen-Campbell, and Deborah South Richardson. Agreeableness as a predictor of aggression in adolescence. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 30(1):43–61, 2004. doi: 10.1002/ab.20002.

[159] Chai Lee Goi. A Review of Marketing Mix: 4Ps or More? *International Journal of Marketing Studies*, 1 (1), 2009.

[160] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

[161] Lewis R Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993. doi: 10.1037/0003-066x.48.1.26.

[162] Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1):84–96, 2006. doi: 10.1016/j.jrp.2005.08.007.

[163] Carlos A. Gomez-Uribe and Neil Hunt. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems*, 6(4):1–19, January 2016. ISSN 2158-656X, 2158-6578. doi: 10.1145/2843948.

[164] Michele Gorgoglione, Umberto Panniello, and Alexander Tuzhilin. Recommendation strategies in personalization applications. *Information and Management*, 56(6):103143, September 2019. ISSN 03787206. doi: 10.1016/j.im.2019.01.005.

[165] Amit Goyal and Laks V.S. Lakshmanan. RecMax: exploiting recommender systems for fun and profit. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1294–1302. ACM, August 2012. ISBN 978-1-4503-1462-6. doi: 10.1145/2339530.2339731.

[166] Thore Graepel, Joaquin Quinonero Candela, Thomas Borchert, and Ralf Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proceedings of the International Conference on International Conference on Machine Learning*. Omnipress, 2010.

[167] Asnat Greenstein-Messica and Lior Rokach. Personal price aware multi-seller recommender system: Evidence from eBay. *Knowledge-Based Systems*, 150:14–26, June 2018. ISSN 09507051. doi: 10.1016/j.knosys.2018.02.026.

[168] Asnat Greenstein-Messica, Lior Rokach, and Asaf Shabtai. Personal-discount sensitivity prediction for mobile coupon conversion optimization. *Journal of the Association for Information Science and Technology*, 68(8):1940–1952, August 2017. ISSN 23301635. doi: 10/gbqjdn.

[169] Ulrike Gretzel and Daniel R. Fesenmaier. Persuasion in Recommender Systems. *International Journal of Electronic Commerce*, 11(2):81–100, December 2006. ISSN 1086-4415, 1557-9301. doi: 10.2753/JEC1086-4415110204.

[170] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, Lixin Zou, Yiding Liu, and Dawei Yin. Deep Multifaceted Transformers for Multi-objective Ranking in Large-Scale E-commerce Recommender Systems. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 2493–2500, Virtual Event Ireland, October 2020. ACM. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3412697. URL `https://dl.acm.org/doi/10.1145/3340531.3412697`.

[171] Asela Gunawardana, Guy Shani, and Sivan Yogev. Evaluating Recommender Systems. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 547–601. Springer US, 2022. ISBN 978-1-07-162196-7 978-1-07-162197-4. doi: 10.1007/978-1-0716-2197-4_15.

[172] Odd Erik Gundersen and Sigbjørn Kjensmo. State of the Art: Reproducibility in Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), April 2018. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v32i1.11503.

[173] Sharath Chandra Guntuku, Sujoy Roy, and Lin Weisi. Personality modeling based image recommendation. In *Proceedings of the International Conference on MultiMedia Modeling*, pages 171–182. Springer, 2015. doi: 10.1007/978-3-319-14442-9_15.

[174] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: a factorization-machine based neural network for ctr prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1725–1731, 2017.

[175] Junpeng Guo, Zeng Gao, Na Liu, and Yi Wu. Recommend products with consideration of multi-category inter-purchase time and price. *Future Generation Computer Systems*, 78:451–461, January 2018. ISSN 0167739X. doi: 10.1016/j.future.2017.02.031.

[176] Liyi Guo, Junqi Jin, Haoqi Zhang, Zhenzhe Zheng, Zhiye Yang, Zhizhuang Xing, Fei Pan, Lvyin Niu, Fan Wu, Haiyang Xu, Chuan Yu, Yuning Jiang, and Xiaoqiang Zhu. We Know What You Want: An Advertising Strategy Recommender System for Online Advertising. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2919–2927, Virtual Event Singapore, August 2021. ACM. ISBN 978-1-4503-8332-5. doi: 10.1145/3447548.3467175.

[177] Yan Guo, Xiaonan Hu, Yu Zou, Shouwei Li, Wenchuan Cheng, and Zhiyong Li. Maximizing E-Tailers' Sales Volume through the Shipping-Fee Discount and Product Recommendation System. *Discrete Dynamics in Nature and Society*, 2020:1–14, March 2020. ISSN 1026-0226, 1607-887X. doi: 10.1155/2020/7349162.

[178] Benjamin Haibe-Kains, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors, Thakkar Shraddha, Rebecca Kusko, Susanna-Assunta Sansone, Weida Tong, Russ D. Wolfinger, Christopher E. Mason, Wendell Jones, Joaquin Dopazo,

Cesare Furlanello, Levi Waldron, Bo Wang, Chris McIntosh, Anna Goldenberg, Anshul Kundaje, Casey S. Greene, Tamara Broderick, Michael M. Hoffman, Jeffrey T. Leek, Keegan Korthauer, Wolfgang Huber, Alvis Brazma, Joelle Pineau, Robert Tibshirani, Trevor Hastie, John P. A. Ioannidis, John Quackenbush, and Hugo J. W. L. Aerts. Transparency and reproducibility in artificial intelligence. *Nature*, 586(7829): E14–E16, October 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2766-y.

[179] Christian Hakulinen, Marko Elovainio, Laura Pulkki-Råback, Marianna Virtanen, Mika Kivimäki, and Markus Jokela. Personality and depressive symptoms: Individual participant meta-analysis of 10 cohort studies. *Depression and Anxiety*, 32(7):461–470, 2015. doi: 10.1002/da.22376.

[180] Elaheh Malekzadeh Hamedani and Marjan Kaedi. Recommending the long tail items through personalized diversification. *Knowledge-Based Systems*, 164:348–357, 2019.

[181] Mikael Hammar, Robin Karlsson, and Bengt J. Nilsson. Using maximum coverage to optimize recommendation systems in e-commerce. In *Proceedings of the ACM Conference on Recommender Systems*, pages 265–272, Hong Kong China, October 2013. ACM. ISBN 978-1-4503-2409-0. doi: 10.1145/2507157.2507169.

[182] Jianhua Han, Yong Yu, Feng Liu, Ruiming Tang, and Yuzhou Zhang. Optimizing Ranking Algorithm in Recommender System via Deep Reinforcement Learning. In *Proceedings of the International Conference on Artificial Intelligence and Advanced Manufacturing*, pages 22–26, Dublin, Ireland, October 2019. IEEE. ISBN 978-1-72814-691-1. doi: 10.1109/AIAM48774.2019.00011.

[183] Songqiao Han, Hailiang Huang, and Yuqing Tang. Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems*, 194:105550, 2020. doi: 10.1016/j.knosys.2020.105550.

[184] Payam Hanafizadeh, Mahdi Barkhordari Firouzabadi, and Khuong Minh Vu. Insight monetization intermediary platform using recommender systems. *Electronic Markets*, 31(2):269–293, June 2021. ISSN 1019-6781, 1422-8890. doi: 10.1007/s12525-020-00449-w.

[185] Bari A. Harlam, Aradhna Krishna, Donald R. Lehmann, and Carl Mela. Impact of bundle type, price framing and familiarity on purchase intention for the bundle. *Journal of Business Research*, 33(1):57–66, May 1995. ISSN 01482963. doi: 10.1016/0148-2963(94)00014-6.

[186] F. Maxwell Harper and Joseph A. Konstan. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):1–19, January 2016. ISSN 2160-6455, 2160-6463. doi: 10.1145/2827872.

[187] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, 2009. doi: 10.1007/978-0-387-21606-5.

[188] Naieme Hazrati. Impact of Recommender Systems on the Dynamics of Users' Choices. In *Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems*, pages 1811–1813, 2021.

[189] Pengfei He, Haochen Liu, Xiangyu Zhao, Hui Liu, and Jiliang Tang. PROPN: Personalized Probabilistic Strategic Parameter Optimization in Recommendations. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 3152–3161. ACM, October 2022. ISBN 978-1-4503-9236-5. doi: 10.1145/3511808.3557130.

[190] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the International Conference on World Wide Web*, pages 173–182, 2017.

[191] Xinran He, Stuart Bowers, Joaquin Quiñonero Candela, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, and Ralf Herbrich. Practical Lessons from Predicting Clicks on Ads at Facebook. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1–9, New York, NY, USA, 2014. ACM Press. ISBN 978-1-4503-2999-6. doi: 10.1145/2648584.2648589.

[192] Patrick C.L. Heaven. Personality and self-reported delinquency: Analysis of the "big five" personality dimensions. *Personality and Individual Differences*, 20(1):47–54, 1996. doi: 10.1016/0191-8869(95)00136-T.

[193] Dale M Heien. Markup pricing in a dynamic model of the food industry. *American Journal of Agricultural Economics*, 62(1):10–18, 1980.

[194] Eelco Herder. The Need for Identifying Ways to Monetize Personalization and Recommendation. In *Adjunct Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization*, pages 291–294. ACM, June 2019. ISBN 978-1-4503-6711-0. doi: 10.1145/3314183.3323844.

[195] Oliver Hinz and Jochen Eckert. The Impact of Search and Recommendation Systems on Sales in Electronic Commerce. *Business and Information Systems Engineering*, 2(2):67–77, April 2010. ISSN 1867-0202. doi: 10.1007/s12599-010-0092-x.

[196] Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. Algorithms for association rule mining – a general survey and comparison. *ACM SIGKDD Explorations Newsletter*, 2(1):58–64, June 2000. ISSN 1931-0145, 1931-0153. doi: 10.1145/360402.360421.

[197] Yu-Chieh Ho, Yi-Ting Chiang, and Jane Yung-Jen Hsu. Who likes it more?: mining worth-recommending items from long tails by modeling relative preference. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 253–262, New York New York USA, February 2014. ACM. ISBN 978-1-4503-2351-2. doi: 10.1145/2556195.2566589.

[198] Donna L Hoffman and Thomas P Novak. A Conceptual Framework for Considering Web-Based Business Models and Potential Revenue Streams. *International Journal of Marketing Education*, 1(1):7–34, 2005.

[199] Tomáš Horváth and André C. P. L. F. de Carvalho. Evolutionary computing in recommender systems: a review of recent research. *Natural Computing*, 16(3):441–462, September 2017. ISSN 1567-7818, 1572-9796. doi: 10.1007/s11047-016-9540-y.

[200] Kartik Hosanagar, Ramayya Krishnan, and Liye Ma. Recommended for you: the impact of profit incentives on the relevance of online recommendations. In *Proceedings of the International Conference on Information Systems*, page 31, 2008.

[201] Kartik Hosanagar, Daniel Fleder, Dokyun Lee, and Andreas Buja. Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science*, 60 (4):805–823, April 2014. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.2013.1808.

[202] Patrick Hosein, Inzamam Rahaman, Keanu Nichols, and Kiran Maharaj. Recommendations for Long-Term Profit Optimization. In *Proceedings of the Workshop on the Impact of Recommender Systems*, 2019.

[203] Chun-Nan Hsu, Hao-Hsiang Chung, and Han-Shen Huang. Mining skewed and sparse transaction data for personalized shopping recommendation. *Machine Learning*, 57:35–59, 2004.

[204] Rong Hu and Pearl Pu. Using personality information in collaborative filtering for new users. *Recommender Systems and the Social Web*, 17, 2010.

[205] Chao Huang, Yi Ding, Weihao Hu, Yi Jiang, and Yongzhen Li. Cost-Based attraction recommendation for tour operators under stochastic demand. *Omega*, 102:102314, July 2021. ISSN 03050483. doi: 10.1016/j.omega.2020.102314.

[206] Cheng-Lung Huang, Mu-Chen Chen, Wen-Chen Huang, and Sheng-Huang Huang. Incorporating frequency, recency and profit in sequential pattern based recommender systems. *Intelligent Data Analysis*, 17(5):899–916, September 2013. ISSN 15714128, 1088467X. doi: 10.3233/IDA-130611.

[207] Shiu-li Huang. Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods. *Electronic Commerce Research and Applications*, 10(4):398–407, July 2011. ISSN 15674223. doi: 10.1016/j.elerap.2010.11.003.

[208] Neil Hurley and Mi Zhang. Novelty and diversity in top-n recommendation–analysis and evaluation. *ACM Transactions on Internet Technology*, 10(4):1–30, 2011. doi: 10.1145/1944339.1944341.

[209] Gerald Häubl and Valerie Trifts. Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, 19(1):4–21, February 2000. ISSN 0732-2399, 1526-548X. doi: 10.1287/mksc.19.1.4.15178.

[210] Eugene Ie, Chih-wei Hsu, Martin Mladenov, Vihan Jain, Sanmit Narvekar, Jing Wang, Rui Wu, and Craig Boutilier. RecSim: A Configurable Simulation Platform for Recommender Systems, September 2019.

[211] Tomoharu Iwata, Kazumi Saito, and Takeshi Yamada. Recommendation method for extending subscription periods. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 574. ACM Press, 2006. ISBN 978-1-59593-339-3. doi: 10.1145/1150402.1150470.

[212] Tomoharu Iwata, Kazumi Saito, and Takeshi Yamada. Recommendation Method for Improving Customer Lifetime Value. *IEEE Transactions on Knowledge and Data Engineering*, 20(9):1254–1263, September 2008. ISSN 1041-4347. doi: 10.1109/TKDE.2008.55.

[213] Dietmar Jannach and Gediminas Adomavicius. Recommendations with a Purpose. In *Proceedings of the ACM Conference on Recommender Systems*, pages 7–10. ACM, September 2016. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959186.

[214] Dietmar Jannach and Gediminas Adomavicius. Price and Profit Awareness in Recommender Systems. In *Proceedings of the Workshop on Value-Aware and Multistakeholder Recommendation*, July 2017.

[215] Dietmar Jannach and Kolja Hegelich. A case study on the effectiveness of recommendations in the mobile internet. In *Proceedings of the ACM Conference on Recommender Systems*, pages 205–208. ACM, October 2009. ISBN 978-1-60558-435-5. doi: 10.1145/1639714.1639749.

[216] Dietmar Jannach and Michael Jugovac. Measuring the Business Value of Recommender Systems. *ACM Transactions on Management Information Systems*, 10(4):1–23, December 2019. ISSN 2158-656X, 2158-6578. doi: 10.1145/3370082.

[217] Dietmar Jannach and Malte Ludewig. Determining characteristics of successful recommendations from log data: a case study. In *Proceedings of the Symposium on Applied Computing*, pages 1643–1648. ACM, April 2017. ISBN 978-1-4503-4486-9. doi: 10.1145/3019612.3019757.

[218] Dietmar Jannach and Markus Zanker. Value and impact of recommender systems. In *Recommender Systems Handbook*, pages 519–546. Springer, 2022.

[219] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY, USA, 2010. ISBN 0521493366, 9780521493369.

[220] Kalervo Järvelin and Jaana Kekäläinen. IR evaluation methods for retrieving highly relevant documents. In *Proceedings of the ACM SIGIR Forum*, volume 51, pages 243–250. ACM New York, NY, USA, 2017.

[221] Luo Ji, Qi Qin, Bingqing Han, and Hongxia Yang. Reinforcement Learning to Optimize Lifetime Value in Cold-Start Recommendation. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 782–791. ACM, October 2021. ISBN 978-1-4503-8446-9. doi: 10.1145/3459637.3482292.

[222] Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 383–390, 2019. doi: 10.1145/3306618.3314288.

[223] Yuanchun Jiang and Yezheng Liu. Optimization of online promotion: a profit-maximizing model integrating price discount and product recommendation. *International Journal of Information Technology and Decision Making*, 11(05):961–982, September 2012. ISSN 0219-6220, 1793-6845. doi: 10.1142/S0219622012500289.

[224] Yuanchun Jiang, Jennifer Shang, Chris F. Kemerer, and Yezheng Liu. Optimizing E-tailer Profits and Customer Savings: Pricing Multistage Customized Online Bundles. *Marketing Science*, 30(4):737–752, July 2011. ISSN 0732-2399, 1526-548X. doi: 10.1287/mksc.1100.0631.

[225] Yuanchun Jiang, Jennifer Shang, Yezheng Liu, and Jerrold May. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. *International Journal of Production Economics*, 167:257–270, September 2015. ISSN 09255273. doi: 10.1016/j.ijpe.2015.02.028.

[226] Zhimeng Jiang, Xiaotian Han, Chao Fan, Fan Yang, Ali Mostafavi, and Xia Hu. Generalized demographic parity for group fairness. In *International Conference on Learning Representations*, 2021.

[227] Jianxin Jiao and Yiyang Zhang. Product portfolio identification based on association rule mining. *Computer-Aided Design*, 37(2):149–172, February 2005. ISSN 00104485. doi: 10.1016/j.cad.2004.05.006.

[228] Oliver P John, Sanjay Srivastava, et al. *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives*, volume 2. University of California Berkeley, 1999.

[229] Jo Yong Ju, Il Young Choi, Hyun Sil Moon, and Jae Kyeong Kim. Reinforcement Learning for Profit Maximization of Recommender Systems. In *Proceedings of the Pre-ICIS SIGDSA Symposium*, 2017.

[230] Toshihiro Kamishima and Shotaro Akaho. Personalized pricing recommender system: multi-stage epsilon-greedy approach. In *Proceedings of the International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, pages 57–64. ACM Press, 2011. ISBN 978-1-4503-1027-7. doi: 10.1145/2039320.2039329.

[231] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. News recommender systems–survey and roads ahead. *Information Processing & Management*, 54(6):1203–1227, 2018.

[232] Nooria Karimi, Adyasha Dash, Siddharth Swarup Rautaray, and Manjusha Pandey. Customer Profiling and Retention Using Recommendation System and Factor Identification to Predict Customer Churn in Telecom Industry. In Manjusha Pandey and Siddharth Swarup Rautaray, editors, *Machine Learning: Theoretical Foundations and Practical Applications*, volume 87, pages 155–172. Springer Singapore, 2021. ISBN 978-981-336-517-9 978-981-336-518-6. doi: 10.1007/978-981-33-6518-6_9.

[233] Davinder Kaur, Suleyman Uslu, Kaley J Rittichier, and Arjan Durresi. Trustworthy artificial intelligence: a review. *ACM Computing Surveys*, 55(2):1–38, 2022.

[234] Ralph L. Keeney and Howard Raiffa. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge University Press, 1993.

[235] Venkataramana Kini and Ashwin Manjunatha. Revenue Maximization using Multitask Learning for Promotion Recommendation. In *Proceedings of the International Conference on Data Mining Workshops*, pages 144–150. IEEE, November 2020. ISBN 978-1-72819-012-9. doi: 10.1109/ICDMW51313.2020.00029.

[236] Evan Kirshenbaum, George Forman, and Michael Dugan. A Live Comparison of Methods for Personalized Article Recommendation at Forbes.com. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Peter A. Flach, Tijl De Bie, and Nello Cristianini, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 7524, pages 51–66. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33485-6 978-3-642-33486-3. doi: 10.1007/978-3-642-33486-3_4.

[237] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. In *AEA Papers and Proceedings*, volume 108, pages 22–27, 2018. doi: 10.1257/pandp.20181018.

[238] Mei Kobayashi and Koichi Takeda. Information retrieval on the web. *ACM Computing Surveys*, 32(2), 2000. doi: 10.1145/358923.358934.

[239] Komiak and Benbasat. The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly*, 30(4):941, 2006. ISSN 02767783. doi: 10.2307/25148760.

[240] Michal Kompan, Peter Gaspar, Jakub Macina, Matus Cimerman, and Maria Bielikova. Exploring Customer Price Preference and Product Profit Role in Recommender Systems. *IEEE Intelligent Systems*, 37 (1):89–98, January 2022. ISSN 1541-1672, 1941-1294. doi: 10.1109/MIS.2021.3092768.

[241] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 426–434, 2008.

[242] Yehuda Koren and Robert Bell. Advances in Collaborative Filtering. In *Recommender Systems Handbook*, pages 91–142. Springer, 2022.

[243] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8):30–37, August 2009. ISSN 0018-9162. doi: 10.1109/MC.2009.263.

[244] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013. doi: 10.1073/pnas.1218772110.

[245] Philip Kotler. *Marketing Management: The Millennium Edition*, volume 10. Prentice Hall Upper Saddle River, NJ, 2000.

[246] Philip Kotler, Gary Armstrong, Michael Harker, and Ross Brennan. *Marketing: an introduction*, volume 1. Prentice-Hall Englewood Cliffs, NJ, 1990.

[247] Roman Kotov, Wakiza Gamez, Frank Schmidt, and David Watson. Linking “big” personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological Bulletin*, 136(5):768, 2010. doi: 10.1037/a0020327.

[248] Pigi Kouki, Ilias Fountalis, Nikolaos Vasiloglou, Nian Yan, Unaiza Ahsan, Khalifeh Al Jadda, and Huiming Qu. Product collection recommendation in online retail. In *Proceedings of the ACM Conference on Recommender Systems*, pages 486–490. ACM, September 2019. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3347003.

[249] Tobias Kowatsch, Wolfgang Maass, Andreas Filler, and Sabine Janzen. Knowledge-Based Bundling of Smart Products on a Mobile Recomendation Agent. In *Proceedings of the International Conference on Mobile Business*, pages 181–190. IEEE, July 2008. doi: 10.1109/ICMB.2008.34.

[250] Jan Krasnodebski and John Dines. Considering Supplier Relations and Monetization in Designing Recommendation Systems. In *Proceedings of the ACM Conference on Recommender Systems*, pages 381–382. ACM, September 2016. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959124.

[251] Saurabh Kulkarni and Sunil F. Rodd. Context Aware Recommendation Systems: A review of the state of the art techniques. *Computer Science Review*, 37:100255, August 2020. ISSN 15740137. doi: 10.1016/j.cosrev.2020.100255.

[252] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems – A survey. *Knowledge-Based Systems*, 123:154–162, May 2017. ISSN 09507051. doi: 10.1016/j.knosys.2017.02.009.

[253] Oren Kurland and J Shane Culpepper. Fusion in information retrieval: Sigir 2018 half-day tutorial. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1383–1386, 2018. doi: 10.1145/3209978.3210186.

[254] Sören Köcher, Michael Jugovac, Dietmar Jannach, and Hartmut H. Holzmüller. New Hidden Persuaders: An Investigation of Attribute-Level Anchoring Effects of Product Recommendations. *Journal of Retailing*, 95(1):24–41, March 2019. ISSN 00224359. doi: 10.1016/j.jretai.2018.10.004.

[255] Preethi Lahoti, Krishna P Gummadi, and Gerhard Weikum. ifair: Learning individually fair data representations for algorithmic decision making. In *Proceedings of the International Conference on Data Engineering*, pages 1334–1345. IEEE, 2019.

[256] Neal Lathia, Stephen Hailes, Licia Capra, and Xavier Amatriain. Temporal diversity in recommender systems. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 210–217, 2010. doi: 10.1145/1835449.1835486.

[257] Dokyun Lee and Kartik Hosanagar. Impact of Recommender Systems on Sales Volume and Diversity. In *Proceedings of the International Conference on Information Systems*, page 15, 2014.

[258] Dokyun Lee and Kartik Hosanagar. How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment. *Information Systems Research*, 30(1):239–259, 2019.

[259] Dongwon Lee, Kihwan Nam, Ingoo Han, and Kanghyun Cho. From free to fee: Monetizing digital content through expected utility-based recommender systems. *Information & Management*, 59(6):103681, 2022.

[260] Kibeom Lee and Kyogu Lee. Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications*, 42(10):4851–4858, 2015. doi: 10.1016/j.eswa.2014.07.024.

[261] Jun Li, Chaochao Chen, Huiling Chen, and Changfei Tong. Towards context-aware social recommendation via individual trust. *Knowledge-Based Systems*, 127:58–66, 2017. doi: 10.1016/j.knosys.2017.02.032.

[262] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the International Conference on World Wide Web*, page 661, Raleigh, North Carolina, USA, 2010. ACM Press. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772758.

[263] Lusi Li, Jianqing Chen, and Srinivasan Raghunathan. Recommender System Rethink: Implications for an Electronic Marketplace with Competing Manufacturers. *Information Systems Research*, 29(4):1003–1023, December 2018. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.2017.0765.

[264] Lusi Li, Jianqing Chen, and Srinivasan Raghunathan. Informative Role of Recommender Systems in Electronic Marketplaces: A Boon or a Bane for Competing Sellers. *MIS Quarterly*, 44(4):1957–1985, December 2020. ISSN 02767783, 21629730. doi: 10.25300/MISQ/2020/14614.

[265] Miao Li, Xuguang Bao, Liang Chang, Zhoubo Xu, and Long Li. A Survey of Researches on Personalized Bundle Recommendation Techniques. In Xiaofeng Chen, Hongyang Yan, Qiben Yan, and Xiangliang Zhang, editors, *Machine Learning for Cyber Security*, volume 12487, pages 290–304. Springer International Publishing, 2020. ISBN 978-3-030-62459-0 978-3-030-62460-6. doi: 10.1007/978-3-030-62460-6_26.

[266] Wei Lu Shanshan Chen Keqian Li and Laks VS Lakshmanan. Show me the money: Dynamic recommendations for revenue maximization. *Proceedings of the VLDB Endowment*, 7(14), 2014.

[267] Xinyu Li, Chun-Hsien Chen, Pai Zheng, Zuhua Jiang, and Linke Wang. A context-aware diversity-oriented knowledge recommendation approach for smart engineering solution design. *Knowledge-Based Systems*, 215:106739, 2021. doi: 10.1016/j.knosys.2021.106739.

[268] Yu Li, Yi Zhang, Lu Gan, Gengwei Hong, Zimu Zhou, and Qiang Li. RevMan: Revenue-aware Multi-task Online Insurance Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35 (1):303–310, May 2021. ISSN 2374-3468.

[269] Zhi Li, Daichi Amagata, Yihong Zhang, Takuya Maekawa, Takahiro Hara, Kei Yonekawa, and Mori Kurokawa. HML4Rec: Hierarchical meta-learning for cold-start recommendation in flash sale e-commerce. *Knowledge-Based Systems*, 255:109674, 2022. doi: 10.1016/j.knosys.2022.109674.

[270] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the International Conference on World Wide Web*, pages 689–698, 2018. doi: 10.1145/3178876.3186150.

[271] Guanzhong Liang, Junhao Wen, and Wei Zhou. Individual diversity preference aware neural collaborative filtering. *Knowledge-Based Systems*, 258:109730, 2022. doi: 10.1016/j.knosys.2022.109730.

[272] Mengqi Liao, S. Shyam Sundar, and Joseph B. Walther. User Trust in Recommendation Systems: A comparison of Content-Based, Collaborative and Demographic Filtering. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–14. ACM, April 2022. ISBN 978-1-4503-9157-3. doi: 10.1145/3491102.3501936.

[273] Donald R. Lichtenstein, Nancy M. Ridgway, and Richard G. Netemeyer. Price Perceptions and Consumer Shopping Behavior: A Field Study. *Journal of Marketing Research*, 30(2):234, May 1993. ISSN 00222437. doi: 10.2307/3172830.

[274] Junjie Lin, Wenji Mao, and Daniel D Zeng. Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems*, 132:204–214, 2017. doi: 10.1016/j.knosys.2017.06.031.

[275] Xiao Lin, Hongjie Chen, Changhua Pei, Fei Sun, Xuanji Xiao, Hanxiao Sun, Yongfeng Zhang, Wenwu Ou, and Peng Jiang. A pareto-efficient algorithm for multiple objective optimization in e-commerce recommendation. In *Proceedings of the ACM Conference on Recommender Systems*, pages 20–28. Association for Computing Machinery, September 2019. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3346998.

[276] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003. doi: 10.1109/MIC.2003.1167344.

[277] Adam Lindgreen and Finn Wynstra. Value in business markets: What do we know? where are we going? *Industrial Marketing Management*, 34(7):732–748, 2005.

[278] Duen-Ren Liu and Ya-Yueh Shih. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software*, 77(2):181–191, August 2005. ISSN 01641212. doi: 10.1016/j.jss.2004.08.031.

[279] Duen-Ren Liu and Ya-Yueh Shih. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*, 42(3):387–400, March 2005. ISSN 03787206. doi: 10.1016/j.im.2004.01.008.

[280] Guannan Liu, Yanjie Fu, Guoqing Chen, Hui Xiong, and Can Chen. Modeling Buying Motives for Personalized Product Bundle Recommendation. *ACM Transactions on Knowledge Discovery from Data*, 11 (3):1–26, August 2017. ISSN 1556-4681, 1556-472X. doi: 10.1145/3022185.

[281] Qinghua Liu, Andrew Henry Reiner, Arnoldo Frigessi, and Ida Scheel. Diverse personalized recommendations with uncertainty from implicit preference data with the bayesian mallows model. *Knowledge-Based Systems*, 186:104960, 2019. doi: 10.1016/j.knosys.2019.104960.

[282] Yezheng Liu, Jiajia Wang, and Yuanchun Jiang. Pt-lda: A latent variable model to predict personality traits of social network users. *Neurocomputing*, 210:155–163, 2016. doi: 10.1016/j.neucom.2015.10.144.

[283] Cheng Long, Raymond Chi-Wing Wong, and Victor Junqiu Wei. Profit Maximization with Sufficient Customer Satisfactions. *ACM Transactions on Knowledge Discovery from Data*, 12(2):1–34, April 2018. ISSN 1556-4681, 1556-472X. doi: 10.1145/3110216. URL `https://dl.acm.org/doi/10.1145/3110216`.

[284] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. *Recommender Systems Handbook*, pages 73–105, 2011.

[285] Raphael Louca, Moumita Bhattacharya, Diane Hu, and Liangjie Hong. Joint Optimization of Profit and Relevance for Recommendation Systems in E-commerce. In *Proceedings of the Workshop on Recommendation in Multistakeholder Environments*, 2019.

[286] Feng Lu, Anca Dumitrache, and David Graus. Beyond Optimizing for Clicks: Incorporating Editorial Values in News Recommendation. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization*, pages 145–153, Genoa Italy, July 2020. ACM. ISBN 978-1-4503-6861-2. doi: 10.1145/3340631.3394864.

[287] Wei Lu, Shanshan Chen, Keqian Li, and Laks V. S. Lakshmanan. Show me the money: dynamic recommendations for revenue maximization. *Proceedings of the VLDB Endowment Conference*, 7(14):1785–1796, October 2014. ISSN 2150-8097. doi: 10.14778/2733085.2733086.

[288] G.T. Lumpkin, Keith H. Brigham, and Todd W. Moss. Long-term orientation: Implications for the entrepreneurial orientation and performance of family businesses. *Entrepreneurship and Regional Development*, 22(3-4):241–264, May 2010. ISSN 0898-5626, 1464-5114. doi: 10.1080/08985621003726218.

[289] Runyu Ma, Hantian Li, Jin Cen, and Amrinder Arora. Placement-and-Profit-Aware Association Rules Mining:. In *Proceedings of the International Conference on Agents and Artificial Intelligence*, pages 639–646. SCITEPRESS - Science and Technology Publications, 2019. ISBN 978-989-758-350-6. doi: 10.5220/0007380606390646.

[290] Edward C. Malthouse, Yasaman Kamyab Hessary, Khadija Ali Vakeel, Robin Burke, and Morana Fudurić. An Algorithm for Allocating Sponsored Recommendations and Content: Unifying Programmatic Advertising and Recommender Systems. *Journal of Advertising*, 48(4):366–379, August 2019. ISSN 0091-3367, 1557-7805. doi: 10.1080/00913367.2019.1652123.

[291] Edward C. Malthouse, Khadija Ali Vakeel, Yasaman Kamyab Hessary, Robin Burke, and Morana Fudurić. A Multistakeholder Recommender Systems Algorithm for Allocating Sponsored Recommendations. In *Proceedings of the Workshop on Recommendation in Multistakeholder Environments*, 2019.

[292] Christopher D. Manning. *Introduction to Information Retrieval*. Syngress Publishing, 2008.

[293] Reza Yousefi Maragheh, Ramin Giahi, Jianpeng Xu, Lalitesh Morishetti, Shanu Vashishtha, Kaushiki Nag, Jason Cho, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Prospect-Net: Top-K Retrieval Problem Using Prospect Theory. In *Proceedings of the IEEE International Conference on Big Data*, pages 3945–3951. IEEE, December 2022. ISBN 978-1-66548-045-1. doi: 10.1109/BigData55660.2022.10020542.

[294] Tara C Marshall, Katharina Lefringhausen, and Nelli Ferenczi. The big five, self-esteem, and narcissism as predictors of the topics people write about in facebook status updates. *Personality and Individual Differences*, 85:35–40, 2015. doi: 10.1016/j.paid.2015.04.039.

[295] Ewa Maslowska, Edward C. Malthouse, and Linda D. Hollebeek. The role of recommender systems in fostering consumers' long-term platform engagement. *Journal of Service Management*, 33(4/5):721–732, July 2022. ISSN 1757-5818. doi: 10.1108/JOSM-12-2021-0487.

[296] Mohamed Massoud and Mohamed Abo-Rizka. A Conceptual Model of Personalized Pricing Recommender System Based on Customer Online Behavior. *International Journal of Computer Science and Network Security*, 12(6):129, 2012.

[297] Christian Matt, Thomas Hess, and Christian Weiß. The Differences between Recommender Technologies in their Impact on Sales Diversity. In *Proceedings of the International Conference on Information Systems*, 2013.

[298] Julian McAuley, Rahul Pandey, and Jure Leskovec. Inferring Networks of Substitutable and Complementary Products. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, Sydney NSW Australia, August 2015. ACM. ISBN 978-1-4503-3664-2. doi: 10.1145/2783258.2783381.

[299] Campbell R McConnell and Stanley L Brue. *Macroeconomics: principles, problems, and policies*. McGraw-Hill, 2020.

[300] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[301] Brian McFee, Thierry Bertin-Mahieux, Daniel PW Ellis, and Gert RG Lanckriet. The million song dataset challenge. In *Proceedings of the International Conference on World Wide Web*, pages 909–916, 2012.

[302] H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: a view from the trenches. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1222–1230, Chicago Illinois USA, August 2013. ACM. ISBN 978-1-4503-2174-7. doi: 10.1145/2487575.2488200.

[303] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6):1–35, 2021. doi: 10.1145/3457607.

[304] Rishabh Mehrotra and Benjamin Carterette. Recommendations in a marketplace. In *Proceedings of the ACM Conference on Recommender Systems*, pages 580–581. ACM, September 2019. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3346952.

[305] Yash Mehta, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, pages 1–27, 2019. doi: 10.1007/s10462-019-09770-z.

[306] Edward Melomed, Irina Gorbach, Alexander Berger, and Py Bateman. *Microsoft SQL Server 2005 Analysis Services*. Sams Publishing, 2006.

[307] Zaiqiao Meng, Richard McCreadie, Craig Macdonald, and Iadh Ounis. Exploring Data Splitting Strategies for the Evaluation of Recommendation Models. In *Proceedings of the ACM Conference on Recommender Systems*, pages 681–686. ACM, September 2020. ISBN 978-1-4503-7583-2. doi: 10.1145/3383313.3418479.

[308] Vafopoulos Michalis and Oikonomou Michael. Recommendation Systems: Bridging Technical Aspects with Marketing Implications. In Ioannis E. Anagnostopoulos, Mária Bieliková, Phivos Mylonas, and Nicolas Tsapatsoulis, editors, *Semantic Hyper/Multimedia Adaptation*, volume 418, pages 155–180. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-28976-7 978-3-642-28977-4. doi: 10.1007/978-3-642-28977-4_5.

[309] Lawrence D. Miles. *Techniques of Value Analysis and Engineering*. New York: McGraw-Hill Book Company., 1961.

[310] Faye Mishna, Cheryl Regehr, Ashley Lacombe-Duncan, Joanne Daciuk, Gwendolyn Fearing, and Melissa Van Wert. Social media, cyber-aggression and student mental health on a university campus. *Journal of Mental Health*, 27(3):222–229, 2018. doi: 10.1080/09638237.2018.1437607.

[311] Effrosyni Mitsopoulou and Theodoros Giovazolias. Personality traits, empathy and bullying behavior: A meta-analytic approach. *Aggression and Violent Behavior*, 21:61–72, 2015. doi: 10.1016/j.avb.2015.01.007.

[312] Merylin Monaro, Andrea Toncini, Stefano Ferracuti, Gianmarco Tessari, Maria G Vaccaro, Pasquale De Fazio, Giorgio Pigato, Tiziano Meneghel, Cristina Scarpazza, and Giuseppe Sartori. The detection of malingering: a new tool to identify made-up depression. *Frontiers in Psychiatry*, 9:249, 2018. doi: 10.3389/fpsyt.2018.00249.

[313] Andrea Montagna, Alvise De Biasio, Nicolo' Navarin, Fabio Aiolli, et al. Graph-based explainable recommendation systems: Are we rigorously evaluating explanations? In *CEUR Workshop Proceedings*, volume 3502. CEUR-WS, 2023.

[314] Neil A. Morgan, Kimberly A. Whitler, Hui Feng, and Simos Chari. Research in marketing strategy. *Journal of the Academy of Marketing Science*, 47(1):4–29, January 2019. ISSN 0092-0703, 1552-7824. doi: 10.1007/s11747-018-0598-1.

[315] Vincenzo Moscato, Antonio Picariello, and Giancarlo Sperli. An emotional recommender system for music. *IEEE Intelligent Systems*, 36(5):57–68, 2020. doi: 10.1109/MIS.2020.3026000.

[316] Mu-Chen Chen, Long-Sheng Chen, Fei-Hao Hsu, Yuanjia Hsu, and Hsiao-Ying Chou. HPRS: A profitability based recommender system. In *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, pages 219–223. IEEE, December 2007. ISBN 978-1-4244-1528-1 978-1-4244-1529-8. doi: 10.1109/IEEM.2007.4419183.

[317] Cataldo Musto, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Context-aware graph-based recommendations exploiting personalized pagerank. *Knowledge-Based Systems*, 216:106806, 2021. doi: 10.1016/j.knosys.2021.106806.

[318] Hekmat Adelnia Najafabadi, Ahmadreza Shekarchizadeh, Akbar Nabiollahi, Naser Khani, and Hamid Rastgari. Dynamic pricing for information goods using revenue management and recommender systems. *Journal of Revenue and Pricing Management*, 2021. doi: 10.1057/S41272-020-00276-W.

[319] Halil Shevket Neap and Tahir Celik. Value of a product: A definition. *International Journal of Value-Based Management*, 12(2):181–191, 1999.

[320] Yaser Nemati and Hossein Khademolhosseini. Devising a profit-aware recommender system using multi-objective GA. *Journal of Advances in Computer Research*, 11(3):109–120, 2020.

[321] Andreas C Neves, Ignacio González, John Leander, and Raid Karoumi. A new approach to damage detection in bridges using machine learning. In *Experimental Vibration Analysis for Civil Structures: Testing, Sensing, Monitoring, and Control 7*, pages 73–84. Springer, 2018.

[322] Phong Nguyen, John Dines, and Jan Krasnodebski. A Multi-Objective Learning to re-Rank Approach to Optimize Online Marketplaces for Multiple Stakeholders. *arXiv*, August 2017.

[323] Tien T. Nguyen, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. User personality and user satisfaction with recommender systems. *Information Systems Frontiers*, 20:1173–1189, 2018. doi: 10.1007/s10796-017-9782-y.

[324] Tuan (Kellan) Nguyen and Pei-Fang Hsu. More Personalized, More Useful? Reinvestigating Recommendation Mechanisms in E-Commerce. *International Journal of Electronic Commerce*, 26(1):90–122, January 2022. ISSN 1086-4415, 1557-9301. doi: 10.1080/10864415.2021.2010006.

[325] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, pages 188–197. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1018.

[326] Athanasios N. Nikolakopoulos, Xia Ning, Christian Desrosiers, and George Karypis. Trust Your Neighbors: A Comprehensive Survey of Neighborhood-Based Methods for Recommender Systems. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 39–89. Springer US, 2022. ISBN 978-1-07-162196-7 978-1-07-162197-4. doi: 10.1007/978-1-0716-2197-4_2.

[327] Mehrbakhsh Nilashi, Dietmar Jannach, Othman bin Ibrahim, Mohammad Dalvi Esfahani, and Hossein Ahmadi. Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications*, 19:70–84, September 2016. ISSN 15674223. doi: 10.1016/j.elerap.2016.09.003.

[328] Xia Ning and George Karypis. Slim: Sparse linear methods for top-n recommender systems. In *Proceedings of the IEEE International Conference on Data Mining*, pages 497–506. IEEE, 2011.

[329] [no authors]. Example Database FoodMart of Microsoft Analysis Services, 1998. URL `https://github.com/julianhyde/foodmart-data-hsqldb`.

[330] [no authors]. Dunnhumby - the complete journey, 2007. URL `https://www.dunnhumby.com/source-files/`.

[331] [no authors]. Track 2: Personalized E-Commerce Search Challenge, 2016. URL `https://competitions.codalab.org/competitions/11161`.

[332] [no authors]. NetEase Music, BX Book and Douban Movie Recommendation Datasets and Model Codes, 2018. URL `https://doi.org/10.7910/DVN/T2GWX4`.

[333] [no authors]. eCommerce Events History in Cosmetics Shop, 2019. URL `https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-cosmetics-shop`.

[334] Oliver Oechslein. On the search for new revenue models: An empirical investigation of personalized news aggregators. *Proceedings of the BLED Conference*, pages 38–52, 2014.

[335] Gal Oestreicher-Singer and Arun Sundararajan. The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets. *Management Science*, 58(11):1963–1981, November 2012. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.1120.1536.

[336] Rodrigo Oliveira, Ramon CF Araújo, Fabrício JB Barros, Adriano Paranhos Segundo, Ronaldo F Zampolo, Wellington Fonseca, Victor Dmitriev, Fernando S Brasil, et al. A system based on artificial neural networks for automatic classification of hydro-generator stator windings partial discharges. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, 16:628–645, 2017.

[337] Natalia Ortiz-de Mandojana and Pratima Bansal. The long-term benefits of organizational resilience through sustainable business practices: The Long-Term Benefits of Sustainable Business Practices. *Strategic Management Journal*, 37(8):1615–1631, August 2016. ISSN 01432095. doi: 10.1002/smj.2410.

[338] Timur Osadchiy, Ivan Poliakov, Patrick Olivier, Maisie Rowland, and Emma Foster. Recommender system based on pairwise association rules. *Expert Systems with Applications*, 115:535–542, January 2019. ISSN 09574174. doi: 10.1016/j.eswa.2018.07.077.

[339] Wentao Ouyang, Xiuwu Zhang, Li Li, Heng Zou, Xin Xing, Zhaojie Liu, and Yanlong Du. Deep Spatio-Temporal Neural Networks for Click-Through Rate Prediction. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2078–2086, Anchorage AK USA, July 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330655.

[340] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88:105906, April 2021. ISSN 17439191. doi: 10.1016/j.ijsu.2021.105906.

[341] Umberto Panniello. How to use recommender systems in e-business domains. *Webology*, 11(2), 2014.

[342] Umberto Panniello, Michele Gorgoglione, and Alexander Tuzhilin. In CARSs we trust: How context-aware recommendations affect customers' trust and other business performance measures of recommender systems. *Information Systems Research*, 27(1):182–196, 2016. doi: 10.1287/isre.2015.0610.

[343] Umberto Panniello, Shawndra Hill, and Michele Gorgoglione. The impact of profit incentives on the relevance of online recommendations. *Electronic Commerce Research and Applications*, 20:87–104, November 2016. ISSN 15674223. doi: 10.1016/j.elerap.2016.10.003.

[344] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072, September 2012. ISSN 09574174. doi: 10.1016/j.eswa.2012.02.038.

[345] Yoon-Joo Park. The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, 25(8):1904–1915, August 2013. ISSN 1041-4347. doi: 10.1109/TKDE.2012.119.

[346] Yoon-Joo Park and Alexander Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proceedings of the ACM conference on Recommender Systems*, page 11, Lausanne, Switzerland, 2008. ACM Press. ISBN 978-1-60558-093-7. doi: 10.1145/1454008.1454012. URL `http://portal.acm.org/citation.cfm?doid=1454008.1454012`.

[347] Jeannette Paschen, Matthew Wilson, and João J. Ferreira. Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, 63(3):403–414, May 2020. ISSN 00076813. doi: 10.1016/j.bushor.2020.01.003.

[348] Apurva Pathak, Kshitiz Gupta, and Julian McAuley. Generating and personalizing bundle recommendations on steam. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1073–1076, 2017.

[349] Bhavik Pathak, Robert Garfinkel, Ram D Gopal, Rajkumar Venkatesan, and Fang Yin. Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2):159–188, 2010. doi: 10.2753/MIS0742-1222270205.

[350] Priyadarsini Patnaik and Biswajeet Patnaik. The Influential Role of Trust in Understanding Online Product Recommendation System. In Rabi Narayan Subudhi, Sumita Mishra, Abu Saleh, and Dariush Khezrimotlagh, editors, *Future of Work and Business in Covid-19 Era*, pages 155–170. Springer Nature Singapore, 2022. ISBN 978-981-19035-6-4 978-981-19035-7-1. doi: 10.1007/978-981-19-0357-1_14.

[351] Gourab K. Patro, Lorenzo Porcaro, Laura Mitchell, Qiuyue Zhang, Meike Zehlike, and Nikhil Garg. Fair ranking: a critical review, challenges, and future directions. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, pages 1929–1942. ACM, June 2022. ISBN 978-1-4503-9352-2. doi: 10.1145/3531146.3533238.

[352] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[353] Changhua Pei, Xinru Yang, Qing Cui, Xiao Lin, Fei Sun, Peng Jiang, Wenwu Ou, and Yongfeng Zhang. Value-aware Recommendation based on Reinforcement Profit Maximization. In *Proceedings of the International Conference on World Wide Web*, pages 3123–3129. ACM, May 2019. ISBN 978-1-4503-6674-8.

[354] J. Pisharath, Y. Liu, B. Ozisikyilmaz, R. Narayanan, W.K. Liao, A. Choudhary, and G. Memik. NU-MineBench version 2.0 dataset and technical report, 2005. URL `http://cucis.ece.northwestern.edu/projects/DMS/MineBench`.

[355] Thomas Piton, Julien Blanchard, and Fabrice Guillet. CAPRE: A New Methodology for Product Recommendation Based on Customer Actionability and Profitability. In *Proceedings of the IEEE International Conference on Data Mining Workshops*, pages 466–473. IEEE, December 2011. ISBN 978-1-4673-0005-6 978-0-7695-4409-0. doi: 10.1109/ICDMW.2011.42.

[356] Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. Fairness in Rankings and Recommendations: An Overview. *The VLDB Journal*, pages 1–28, 2022. doi: 10.1007/s00778-021-00697-y.

[357] Mirko Polato. rectorch: pytorch-based framework for top-N recommendation, sep 2020. URL `https://doi.org/10.5281/zenodo.153841898991`.

[358] Marco Polignano, Fedelucio Narducci, Marco de Gemmis, and Giovanni Semeraro. Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Systems with Applications*, 170:114382, 2021. doi: 10.1016/j.eswa.2020.114382.

[359] Wichian Premchaiswadi, Pitaya Poompuang, Nipat Jongswat, and Nucharee Premchaiswadi. Enhancing diversity-accuracy technique on user-based top-n recommendation algorithms. In *Proceedings of the IEEE Computer Software and Applications Conference Workshops*, pages 403–408. IEEE, 2013. doi: 10.1109/COMPSACW.2013.68.

[360] Nan Qiu, BoYu Gao, Huawei Tu, Feiran Huang, Quanlong Guan, and Weiqi Luo. Ldgc-sr: Integrating long-range dependencies and global context information for session-based recommendation. *Knowledge-Based Systems*, 248:108894, 2022. doi: 10.1016/j.knosys.2022.108894.

[361] Meng Qu, Hengshu Zhu, Junming Liu, Guannan Liu, and Hui Xiong. A cost-effective recommender system for taxi drivers. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 45–54. ACM, August 2014. ISBN 978-1-4503-2956-9. doi: 10.1145/2623330.2623668.

[362] Ulrich Rabanser and Francesco Ricci. Recommender Systems: Do They Have a Viable Business Model in e-Tourism? In Andrew J. Frew, editor, *Information and Communication Technologies in Tourism 2005*, pages 160–171. Springer-Verlag, 2005. ISBN 978-3-211-24148-6. doi: 10.1007/3-211-27283-6_15.

[363] Andreas Rauch, Michael Frese, and Andreas Utsch. Effects of Human Capital and Long–Term Human Resources Development and Utilization on Employment Growth of Small–Scale Businesses: A Causal Analysis. *Entrepreneurship Theory and Practice*, 2005.

[364] Kathleen W. Reardon, Jennifer L. Tackett, and Don Lynam. The personality context of relational aggression: A five-factor model profile analysis. *Personality Disorders: Theory, Research, and Treatment*, 9(3): 228, 2018. doi: 10.1037/per0000231.

[365] Caixia Ren, Ping Zhu, and Hua Zhang. A new collaborative filtering technique to improve recommendation diversity. In *Proceedings of the IEEE International Conference on Computer and Communications*, pages 1279–1282. IEEE, 2016. doi: 10.1109/CompComm.2016.7924910.

[366] Jing Ren and Yimiao Zhang. Is All-Win Possible? A Novel Smart Contracts-based Music Recommendation System. In *Proceedings of the International Conference on Information Communication and Management*, pages 18–23. ACM, August 2021. ISBN 978-1-4503-9019-4. doi: 10.1145/3484399.3484402.

[367] Jing Ren, Robert J Kauffman, and Dave King. Two-Sided Value-Based Music Artist Recommendation in Streaming Music Services. In *Proceedings of the Hawaii International Conference on System Sciences*, 2019.

[368] Steffen Rendle. Item Recommendation from Implicit Feedback. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 143–171. Springer US, 2022. ISBN 978-1-07-162196-7 978-1-07-162197-4. doi: 10.1007/978-1-0716-2197-4_4.

[369] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 452–461, 2009.

[370] Steffen Rendle, Walid Krichene, Li Zhang, and John Anderson. Neural Collaborative Filtering vs. Matrix Factorization Revisited. In *Proceedings of the ACM Conference on Recommender Systems*, pages 240–248. ACM, September 2020. ISBN 978-1-4503-7583-2. doi: 10.1145/3383313.3412488.

[371] Steffen Rendle, Walid Krichene, Li Zhang, and Yehuda Koren. Revisiting the performance of ials on item recommendation benchmarks. In *Proceedings of the ACM Conference on Recommender Systems*, pages 427–435, 2022. doi: 10.1145/3523227.3548486.

[372] Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3), 1997.

[373] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, pages 175–186, 1994.

[374] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender Systems Handbook*, pages 1–35. Springer, 2011. doi: 10.1007/978-0-387-85820-3_1.

[375] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: Techniques, applications, and challenges. *Recommender Systems Handbook*, pages 1–35, 2022.

[376] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 61–70, 2002.

[377] Giorgio Roffo and Alessandro Vinciarelli. Personality in computational advertising: A benchmark, 2016.

[378] David Rohde, Stephen Bonner, Travis Dunlop, Flavian Vasile, and Alexandros Karatzoglou. RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising. *arXiv*, September 2018.

[379] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[380] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[381] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the International Conference on Machine Learning*, page 791–798, 2007. doi: 10.1145/1273496.1273596.

[382] Tanya Sammut-Bonnici and Derek F Channon. Pricing strategy. *Wiley Encyclopedia of Management*, pages 1–3, 2015.

[383] Masahiro Sato, Hidetaka Izumo, and Takashi Sonoda. Discount Sensitive Recommender System for Retail Business. In *Proceedings of the Workshop on Emotions and Personality in Personalized Systems*, pages 33–40. ACM, September 2015. ISBN 978-1-4503-3615-4. doi: 10.1145/2809643.2809646.

[384] Lisa M Saulsman and Andrew C Page. The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, 23(8):1055–1085, 2004. doi: 10.1016/j.cpr.2002.09.001.

[385] J. Ben Schafer, Joseph Konstan, and John Riedl. Recommender Systems in E-Commerce. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 158–166, 1999. doi: 10.1145/336992.337035.

[386] J Ben Schafer, Joseph A Konstan, and John Riedl. E-commerce recommendation applications. *Data mining and knowledge discovery*, 5:115–153, 2001. doi: 10.1023/A:1009804230409.

[387] Michael Scholz, Verena Dorner, Markus Franz, and Oliver Hinz. Measuring consumers' willingness to pay with utility-based recommendation systems. *Decision Support Systems*, 72:60–71, April 2015. ISSN 01679236. doi: 10.1016/j.dss.2015.02.006.

[388] John Scott. Rational choice theory. *Understanding contemporary society: Theories of the present*, 129:126–138, 2000.

[389] Sylvain Senecal and Jacques Nantel. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2):159–169, January 2004. ISSN 00224359. doi: 10.1016/j.jretai.2004.04.001.

[390] Sinan Seymen, Anna-Lena Sachs, and Edward C Malthouse. Making smart recommendations for perishable and stockout products. In *Proceedings of the Workshop on Multi-Objective Recommender Systems*, 2022.

[391] J.P. Sharpe and S. Desai. The revised neo personality inventory and the mmpi-2 psychopathology five in the prediction of aggression. *Personality and Individual Differences*, 31(4):505–518, 2001. doi: 10.1016/S0191-8869(00)00155-0.

[392] Ya-Yueh Shih and Duen-Ren Liu. Hybrid Recommendation Approaches: Collaborative Filtering via Valuable Content Information. In *Proceedings of the Hawaii International Conference on System Sciences*, 2005.

[393] Ya-Yueh Shih and Duen-Ren Liu. Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands. *Expert Systems with Applications*, 35(1-2):350–360, July 2008. ISSN 09574174. doi: 10.1016/j.eswa.2007.07.055.

[394] Yingwai Shiu, Cheng Guo, Min Zhang, Yiqun Liu, and Shaoping Ma. Identifying Price Sensitive Customers in E-commerce Platforms for Recommender Systems. In Shichao Zhang, Tie-Yan Liu, Xianxian Li, Jiafeng Guo, and Chenliang Li, editors, *Information Retrieval*, volume 11168, pages 225–236. Springer International Publishing, 2018. ISBN 978-3-030-01011-9 978-3-030-01012-6. doi: 10.1007/978-3-030-01012-6_18.

[395] Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2219–2228, 2018. doi: 10.1145/3219819.3220088.

[396] Malcolm Slaney and William White. Measuring playlist diversity for recommendation systems. In *Proceedings of the ACM Workshop on Audio and Music Computing Multimedia*, pages 77–82, 2006. doi: 10.1145/1178723.1178735.

[397] Ryosuke Sonoda. A pre-processing method for fairness in ranking. *arXiv*, 2021.

[398] Paschalia Spyridou, Constantinos Djouvas, and Dimitra Milioni. Modeling and validating a news recommender algorithm in a mainstream medium-sized news organization: An experimental approach. *Future Internet*, 14(10):284, 2022.

[399] Harald Steck. Calibrated recommendations. In *Proceedings of the ACM Conference on Recommender Systems*, pages 154–162, 2018. doi: 10.1145/3240323.3240372.

[400] Harald Steck, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, and Justin Basilico. Deep learning for recommender systems: A netflix case study. *AI Magazine*, 42(3):7–18, 2021. doi: 10.1609/aimag.v42i3.18140.

[401] Xiaoyuan Su and Taghi M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009:1–19, October 2009. ISSN 1687-7470, 1687-7489. doi: 10.1155/2009/421425.

[402] Chanchal Suman, Sriparna Saha, Aditya Gupta, Saurabh Kumar Pandey, and Pushpak Bhattacharyya. A multi-modal personality prediction system. *Knowledge-Based Systems*, 236:107715, 2022. doi: 10.1016/j.knosys.2021.107715.

[403] Chris Sumner, Alison Byers, Rachel Boochever, and Gregory J. Park. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *Proceedings of the International Conference on Machine Learning and Applications*, volume 2, pages 386–393. IEEE, 2012. doi: 10.1109/ICMLA.2012.218.

[404] Zhu Sun, Qing Guo, Jie Yang, Hui Fang, Guibing Guo, Jie Zhang, and Robin Burke. Research commentary on recommendations with side information: A survey and research directions. *Electronic Commerce Research and Applications*, 37, 2019. doi: 10.1016/j.elerap.2019.100879.

[405] Zhu Sun, Di Yu, Hui Fang, Jie Yang, Xinghua Qu, Jie Zhang, and Cong Geng. Are We Evaluating Rigorously? Benchmarking Recommendation for Reproducible Evaluation and Fair Comparison. In *Proceedings of the ACM Conference on Recommender Systems*, pages 23–32. ACM, September 2020. ISBN 978-1-4503-7583-2. doi: 10.1145/3383313.3412489.

[406] Zhu Sun, Jie Yang, Kaidong Feng, Hui Fang, Xinghua Qu, and Yew Soon Ong. Revisiting Bundle Recommendation: Datasets, Tasks, Challenges and Opportunities for Intent-aware Product Bundling. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2900–2911. ACM, July 2022. ISBN 978-1-4503-8732-3. doi: 10.1145/3477495.3531904.

[407] Cass R. Sunstein. The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, (91), 1999. doi: 10.7551/mitpress/11974.003.0005.

[408] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[409] Zahra Tabaei and Mohammad Fathian. Using Customer lifetime Value Model for Product Recommendation: An Electronic Retailing Case Study. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(1), 2012.

[410] David J. Teece. Internal Organization and Economic Performance: An Empirical Analysis of the Profitability of Principal Firms. *The Journal of Industrial Economics*, 30(2):173, December 1981. ISSN 00221821. doi: 10.2307/2098201.

[411] Georgios Theocharous, Philip S. Thomas, and Mohammad Ghavamzadeh. Personalized Ad Recommendation Systems for Life-Time Value Optimization with Guarantees. In *Proceedings of the International Conference on World Wide Web*, page 7, 2015. doi: 10.1145/2740908.2741998.

[412] Ting-Kai Hwang and Yung-Ming Li. Optimal Recommendation and Long-Tail Provision Strategies for Content Monetization. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 1316–1323. IEEE, January 2014. ISBN 978-1-4799-2504-9. doi: 10.1109/HICSS.2014.169.

[413] Nava Tintarev and Judith Masthoff. A Survey of Explanations in Recommender Systems. In *Proceedings of the IEEE International Conference on Data Engineering Workshop*, pages 801–810. IEEE, April 2007. ISBN 978-1-4244-0831-3 978-1-4244-0832-0. doi: 10.1109/ICDEW.2007.4401070.

[414] Nava Tintarev, Matt Dennis, and Judith Masthoff. Adapting recommendation diversity to openness to experience: a study of human behaviour. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 190–202. Springer, 2013. doi: 10.1007/978-3-642-38844-6_16.

[415] Marko Tkalcic and Li Chen. Personality and recommender systems. In *Recommender Systems Handbook*, pages 715–739. Springer, 2015. doi: 10.1007/978-1-4899-7637-6_21.

[416] Marko Tkalcic, Matevz Kunaver, Andrej Košir, and Jurij Tasic. Addressing the new user problem with a personality based user similarity measure. In *Proceedings of the International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems*, volume 106, pages 106–123, 2011.

[417] Ching-Chih Tsao, Cheng-Yi Tang, Yu-Wen Chang, Yin-Hsuan Sung, Shih-Yi Chien, and Szu-Yin Lin. The Influence of a Robot Recommender System on Impulse Buying Tendency. In *Companion Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 672–676. ACM, March 2023. ISBN 978-1-4503-9970-8. doi: 10.1145/3568294.3580171.

[418] Ernest C. Tupes and Raymond E. Christal. Recurrent personality factors based on trait ratings. *Journal ofPpersonality*, 60(2):225–251, 1992. doi: 10.1111/j.1467-6494.1992.tb00973.x.

[419] Panniello Umberto. Developing a price-sensitive recommender system to improve accuracy and business performance of ecommerce applications. *International Journal of Electronic Commerce Studies*, 6(1):1–18, 2015.

[420] Hirofumi Uzawa. Production functions with constant elasticities of substitution. *The Review of Economic Studies*, 29(4):291–299, 1962.

[421] Guido van Capelleveen, Chintan Amrit, Devrim Murat Yazan, and Henk Zijm. The recommender canvas: A model for developing and documenting recommender system design. *Expert Systems with Applications*, 129:97–117, September 2019. ISSN 09574174. doi: 10.1016/j.eswa.2019.04.001.

[422] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the ACM Conference on Recommender Systems*, pages 209–216, 2014. doi: 10.1145/2645710.2645743.

[423] Saúl Vargas and Pablo Castells. Improving sales diversity by recommending users to items. In *Proceedings of the ACM Conference on Recommender Systems*, pages 145–152, Foster City, Silicon Valley, California, USA, 2014. ACM Press. ISBN 978-1-4503-2668-1. doi: 10.1145/2645710.2645744.

[424] Súul Vargas. New approaches to diversity and novelty in recommender systems. In *Proceedings of the BCS-IRSG Symposium on Future Directions in Information Access*, pages 8–13, 2011. doi: 10.14236/ewic/FDIA2011.2.

[425] Alina Venermo, Jukka Rantala, and Timo Holopainen. From Sales Funnel to Customer Journey. In Jussi Ilari Kantola, Salman Nazir, and Vesa Salminen, editors, *Advances in Human Factors, Business Management and Leadership*, volume 1209, pages 200–206. Springer International Publishing, 2020. ISBN 978-3-030-50790-9 978-3-030-50791-6. doi: 10.1007/978-3-030-50791-6_25.

[426] R. Venkatesh and Vijay Mahajan. The Design and Pricing of Bundles: A Review of Normative Guidelines and Practical Approaches. In *Handbook of Pricing Research in Marketing*, pages 232–257. Edward Elgar Publishing, 2009. ISBN 978-1-84844-744-8. doi: 10.4337/9781848447448.00020.

[427] Norha M Villegas, Cristian Sánchez, Javier Díaz-Cely, and Gabriel Tamura. Characterizing context-aware recommender systems: A systematic literature review. *Knowledge-Based Systems*, 140:173–200, 2018. doi: 10.1016/j.knosys.2017.11.003.

[428] Alexandra Vultureanu-Albişi and Costin Bădică. A survey on effects of adding explanations to recommender systems. *Concurrency and Computation: Practice and Experience*, page e6834, 2022.

[429] Soumya Wadhwa, Ashish Ranjan, Selene Xu, Jason H D Cho, Sushant Kumar, and Kannan Achan. Personalizing Item Recommendation via Price Understanding. In *Proceedings of the Workshop on the Impact of Recommender Systems*, 2020.

[430] Cong Wang, Guoqing Chen, Qiang Wei, Guannan Liu, and Xunhua Guo. Personalized Promotion Recommendation Through Consumer Experience Evolution Modeling. In Ralph Baker Kearfott, Ildar Batyrshin, Marek Reformat, Martine Ceberio, and Vladik Kreinovich, editors, *Fuzzy Techniques: Theory and Applications*, volume 1000, pages 692–703. Springer International Publishing, 2019. ISBN 978-3-030-21919-2 978-3-030-21920-8. doi: 10.1007/978-3-030-21920-8_61.

[431] Cong Wang, Xunhua Guo, Guannan Liu, and Guoqing Chen. Personalized Promotion Recommendation: A Dynamic Adaptation Modeling Approach. In *Proceedings of the International Conference on Information Systems*, 2020.

[432] Hanfei Wang, Yuan Zuo, Hong Li, and Junjie Wu. Cross-domain recommendation with user personality. *Knowledge-Based Systems*, 213:106664, 2021. doi: 10.1016/j.knosys.2020.106664.

[433] Hongpeng Wang, Rong Du, Shizhong Ai, and Zhe Chi. The Moderating Role of Perceived Effectiveness of Provider Recommendations on Consumers' Satisfaction, Trust, and Online Repurchase Intention. In Fiona Fui-Hoon Nah and Chuan-Hoo Tan, editors, *HCI in Business*, volume 9191, pages 382–391. Springer International Publishing, 2015. ISBN 978-3-319-20894-7 978-3-319-20895-4. doi: 10.1007/978-3-319-20895-4_35.

[434] Hsiao-Fan Wang and Cheng-Ting Wu. A mathematical model for product selection strategies in a recommender system. *Expert Systems with Applications*, 36(3):7299–7308, April 2009. ISSN 09574174. doi: 10.1016/j.eswa.2008.09.006.

[435] Hsiao-Fan Wang and Cheng-Ting Wu. A strategy-oriented operation module for recommender systems in E-commerce. *Computers and Operations Research*, 39(8):1837–1849, August 2012. ISSN 03050548. doi: 10.1016/j.cor.2010.03.011.

[436] Jian Wang and Yi Zhang. Utilizing marginal net utility for recommendation in e-commerce. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1003–1012. ACM, July 2011. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2010050.

[437] Ke Wang and Senqiang Zhou. Mining Customer Value: From Association Rules to Direct Marketing. *Data Mining and Knowledge Discovery*, 2005. doi: 10.1007/s10618-005-1355-x.

[438] Ke Wang, Simon FraserUniversity, and Ming-Yen Thomas Su. Item Selection By "Hub-Authority" Profit Ranking. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 652–657, 2002. doi: 10.1145/775047.775144.

[439] Ke Wang, Senqiang Zhou, and Jiawei Han. Profit Mining: From Patterns to Actions. In G. Goos, J. Hartmanis, J. van Leeuwen, Christian S. Jensen, Simonas Šaltenis, Keith G. Jeffery, Jaroslav Pokorny, Elisa Bertino, Klemens Böhn, and Matthias Jarke, editors, *Advances in Database Technology*, volume 2287, pages 70–87. Springer Berlin Heidelberg, 2002. ISBN 978-3-540-43324-8 978-3-540-45876-0. doi: 10.1007/3-540-45876-X_7.

[440] Lina Wang, Xuyun Zhang, Ruili Wang, Chao Yan, Huaizhen Kou, and Lianyong Qi. Diversified service recommendation with high accuracy and efficiency. *Knowledge-Based Systems*, 204:106196, 2020. doi: 10.1016/j.knosys.2020.106196.

[441] Shanfeng Wang, Maoguo Gong, Haoliang Li, and Junwei Yang. Multi-objective optimization for long tail recommendation. *Knowledge-Based Systems*, 104:145–155, July 2016. ISSN 09507051. doi: 10.1016/j.knosys.2016.04.018.

[442] Weiquan Wang and May Wang. Effects of Sponsorship Disclosure on Perceived Integrity of Biased Recommendation Agents: Psychological Contract Violation and Knowledge-Based Trust Perspectives. *Information Systems Research*, 30(2):507–522, June 2019. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.2018.0811.

[443] Weiquan Wang, Lingyun Qiu, Dongmin Kim, and Izak Benbasat. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, 86:48–60, June 2016. ISSN 01679236. doi: 10.1016/j.dss.2016.03.007.

[444] Yi-Fan Wang, Ding-An Chiang, Mei-Hua Hsu, Cheng-Jung Lin, and I-Long Lin. A recommender system to avoid customer churn: A case study. *Expert Systems with Applications*, 36(4):8071–8075, May 2009. ISSN 09574174. doi: 10.1016/j.eswa.2008.10.089.

[445] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems*, 41(3):1–43, 2023.

[446] Stephan Winter, German Neubaum, Sabrina C Eimler, Vanessa Gordon, Jenna Theil, Jonathan Herrmann, Judith Meinert, and Nicole C Krämer. Another brick in the facebook wall–how personality traits relate to the content of status updates. *Computers in Human Behavior*, 34:194–202, 2014. doi: 10.1016/j.chb.2014.01.048.

[447] Han Wu, Hongzhe Zhang, Liangyue Li, Zulong Chen, Fanwei Zhu, and Xiao Fang. Cheaper Is Better: Exploring Price Competitiveness for Online Purchase Prediction. In *Proceedings of the IEEE International Conference on Data Engineering*, pages 3399–3412. IEEE, May 2022. ISBN 978-1-66540-883-7. doi: 10.1109/ICDE53745.2022.00320.

[448] Liang Wu, Diane Hu, Liangjie Hong, and Huan Liu. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 365–374. ACM, June 2018. ISBN 978-1-4503-5657-2. doi: 10.1145/3209978.3209993.

[449] Qingyun Wu, Hongning Wang, Liangjie Hong, and Yue Shi. Returning is Believing: Optimizing Long-term User Engagement in Recommender Systems. In *Proceedings of the ACM on Conference on Information and Knowledge Management*, pages 1927–1936. ACM, November 2017. ISBN 978-1-4503-4918-5. doi: 10.1145/3132847.3133025.

[450] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5):1–37, 2022.

[451] Wen Wu, Li Chen, and Liang He. Using personality to adjust diversity in recommender systems. In *Proceedings of the ACM Conference on Hypertext and Social Media*, pages 225–229, 2013. doi: 10.1145/2481492.2481521.

[452] Yi-Jing Wu and Wei-Guang Teng. An enhanced recommendation scheme for online grocery shopping. In *Proceedings of the IEEE International Symposium on Consumer Electronics*, pages 410–415. IEEE, June 2011. ISBN 978-1-61284-843-3. doi: 10.1109/ISCE.2011.5973860.

[453] Bin Xia, Junjie Yin, Jian Xu, and Yun Li. We-rec: A fairness-aware reciprocal recommendation based on walrasian equilibrium. *Knowledge-Based Systems*, 182:104857, 2019. doi: 10.1016/j.knosys.2019.07.028.

[454] Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS quarterly*, pages 137–209, 2007. doi: 10.2307/25148784.

[455] Bo Xiao and Izak Benbasat. An empirical examination of the influence of biased personalized product recommendations on consumers' decision making outcomes. *Decision Support Systems*, 110:46–57, June 2018. ISSN 01679236. doi: 10.1016/j.dss.2018.03.005.

[456] Zhichao Xu, Yi Han, Yongfeng Zhang, and Qingyao Ai. E-commerce Recommendation with Weighted Expected Utility. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1695–1704, October 2020. doi: 10.1145/3340531.3411993.

[457] Ruiliang Yan and Subir Bandyopadhyay. The profit benefits of bundle pricing of complementary products. *Journal of Retailing and Consumer Services*, 18(4):355–361, July 2011. ISSN 09696989. doi: 10.1016/j.jretconser.2011.04.001.

[458] Ke Yang and Julia Stoyanovich. Measuring fairness in ranked outputs. In *Proceedings of the International Conference on Scientific and Statistical Database Management*, pages 1–6, 2017. doi: 10.1145/3085504.3085526.

[459] Ruixin Yang, Mingyang Xu, Paul Jones, and Nagiza Samatova. Real time utility-based recommendation for revenue optimization via an adaptive online Top-K high utility itemsets mining model. In *Proceedings of the International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pages 1859–1866. IEEE, July 2017. ISBN 978-1-5386-2165-3. doi: 10.1109/FSKD.2017.8393050.

[460] Tzyy-Ching Yang and Hsiangchu Lai. Comparison of product bundling strategies on different online shopping behaviors. *Electronic Commerce Research and Applications*, 5(4):295–304, December 2006. ISSN 15674223. doi: 10.1016/j.elerap.2006.04.006.

[461] Wang Yanping and Cheng Yan. Psychology reactance to online recommendations: The influence of time pressure. In *Proceedings of the International Conference on System Science, Engineering Design and Manufacturing Informatization*, pages 347–350. IEEE, October 2012. ISBN 978-1-4673-0915-8 978-1-4673-0914-1. doi: 10.1109/ICSSEM.2012.6340745.

[462] Sirui Yao and Bert Huang. Beyond parity: Fairness objectives for collaborative filtering. *Advances in Neural Information Processing Systems*, 30, 2017.

[463] Tal Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44(3):363–373, 2010. doi: 10.1016/j.jrp.2010.04.001.

[464] Sangyoon Yi, Dongyeon Kim, and Jaehyeon Ju. Recommendation technologies and consumption diversity: An experimental study on product recommendations, consumer search, and sales diversity. *Technological Forecasting and Social Change*, 178:121486, May 2022. ISSN 00401625. doi: 10.1016/j.techfore.2022.121486.

[465] Wu Youyou, Michal Kosinski, and David Stillwell. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015. doi: 10.1073/pnas.1418680112.

[466] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the ACM Conference on Information and Knowledge Management*, pages 1569–1578, 2017. doi: 10.1145/3132847.3132938.

[467] Meike Zehlike, Ke Yang, and Julia Stoyanovich. Fairness in Ranking, Part I: Score-Based Ranking. *ACM Computing Surveys*, 55(6):1–36, July 2023. ISSN 0360-0300, 1557-7341. doi: 10.1145/3533379.

[468] Meike Zehlike, Ke Yang, and Julia Stoyanovich. Fairness in Ranking, Part II: Learning-to-Rank and Recommender Systems. *ACM Computing Surveys*, 55(6):1–41, July 2023. ISSN 0360-0300, 1557-7341. doi: 10.1145/3533380.

[469] Valarie A Zeithaml. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of marketing*, 52(3):2–22, 1988.

[470] Ruohan Zhan, Konstantina Christakopoulou, Ya Le, Jayden Ooi, Martin Mladenov, Alex Beutel, Craig Boutilier, Ed Chi, and Minmin Chen. Towards Content Provider Aware Recommender Systems: A Simulation Study on the Interplay between User and Provider Utilities. In *Proceedings of the Web Conference*, pages 3872–3883. ACM, April 2021. ISBN 978-1-4503-8312-7. doi: 10.1145/3442381.3449889.

[471] Huiwang Zhang, Pengpeng Zhao, Xuefeng Xian, Victor S. Sheng, Yongjing Hao, and Zhiming Cui. Click is Not Equal to Purchase: Multi-task Reinforcement Learning for Multi-behavior Recommendation. In Richard Chbeir, Helen Huang, Fabrizio Silvestri, Yannis Manolopoulos, and Yanchun Zhang, editors, *Proceedings of the Conference on Web Information Systems Engineering*, volume 13724, pages 443–459. Springer International Publishing, 2022. ISBN 978-3-031-20890-4 978-3-031-20891-1. doi: 10.1007/978-3-031-20891-1_32.

[472] Jingjing Zhang and Shawn P. Curley. Exploring Explanation Effects on Consumers' Trust in Online Recommender Agents. *International Journal of Human–Computer Interaction*, 34(5):421–432, May 2018. ISSN 1044-7318, 1532-7590. doi: 10.1080/10447318.2017.1357904.

[473] Weinan Zhang, Jun Wang, Bowei Chen, and Xiaoxue Zhao. To personalize or not: a risk management perspective. In *Proceedings of the ACM Conference on Recommender Systems*, pages 229–236. ACM, October 2013. ISBN 978-1-4503-2409-0. doi: 10.1145/2507157.2507167.

[474] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems*, 28, 2015.

[475] Xiaochen Zhang, Pedro Ferreira, Miguel Godinho de Matos, and Rodrigo Belo. Welfare Properties of Profit Maximizing Recommender Systems: Theory and Results from a Randomized Experiment. *MIS Quarterly*, 45(1):1–43, March 2021. ISSN 02767783, 21629730. doi: 10.25300/MISQ/2021/14971.

[476] Xiaokun Zhang, Bo Xu, Liang Yang, Chenliang Li, Fenglong Ma, Haifeng Liu, and Hongfei Lin. Price DOES Matter!: Modeling Price and Interest Preferences in Session-based Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1684–1693. ACM, July 2022. ISBN 978-1-4503-8732-3. doi: 10.1145/3477495.3532043.

[477] Yongfeng Zhang and Xu Chen. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends in Information Retrieval*, 14(1):1–101, 2020. doi: 10.1561/1500000066.

[478] Yongfeng Zhang, Qi Zhao, Yi Zhang, Daniel Friedman, Min Zhang, Yiqun Liu, and Shaoping Ma. Economic Recommendation with Surplus Maximization. In *Proceedings of the International Conference on World Wide Web*, pages 73–83. International World Wide Web Conferences Steering Committee, April 2016. ISBN 978-1-4503-4143-1. doi: 10.1145/2872427.2882973.

[479] Yongfeng Zhang, Yi Zhang, and Daniel Friedman. Economic recommendation based on pareto efficient resource allocation. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, 2017.

[480] Zhiwei Zhang, Ning Chen, Jun Wang, and Luo Si. SMART: Sponsored mobile app recommendation by balancing app downloads and appstore profit. In *Proceedings of the IEEE International Conference on Big Data*, pages 1600–1609. IEEE, December 2017. ISBN 978-1-5386-2715-0. doi: 10.1109/BigData.2017.8258094.

[481] Qi Zhao, Yi Zhang, Daniel Friedman, and Fangfang Tan. E-commerce Recommendation with Personalized Promotion. In *Proceedings of the ACM Conference on Recommender Systems*, pages 219–226. ACM, September 2015. ISBN 978-1-4503-3692-5. doi: 10.1145/2792838.2800178.

[482] Qi Zhao, Yongfeng Zhang, Yi Zhang, and Daniel Friedman. Multi-Product Utility Maximization for Economic Recommendation. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 435–443. ACM, February 2017. ISBN 978-1-4503-4675-7. doi: 10.1145/3018661.3018674.

[483] Wayne Xin Zhao, Zihan Lin, Zhichao Feng, Pengfei Wang, and Ji-Rong Wen. A Revisiting Study of Appropriate Offline Evaluation for Top- *N* Recommendation Algorithms. *ACM Transactions on Information Systems*, 41(2):1–41, April 2023. ISSN 1046-8188, 1558-2868. doi: 10.1145/3545796.

[484] Yifei Zhao, Yu-Hang Zhou, Mingdong Ou, Huan Xu, and Nan Li. Maximizing Cumulative User Engagement in Sequential Recommendation: An Online Optimization Perspective. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2784–2792. ACM, August 2020. ISBN 978-1-4503-7998-4. doi: 10.1145/3394486.3403329.

[485] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In *Proceedings of the Conference on World Wide Web*, pages 167–176, Lyon, France, 2018. ACM Press. ISBN 978-1-4503-5639-8. doi: 10.1145/3178876.3185994. URL `http://dl.acm.org/citation.cfm?doid=3178876.3185994`.

[486] Yong Zheng. Utility-based multi-criteria recommender systems. In *Proceedings of the SIGAPP Symposium on Applied Computing*, pages 2529–2531. ACM, April 2019. ISBN 978-1-4503-5933-7. doi: 10.1145/3297280.3297641.

[487] Yong Zheng and Aviana Pu. Utility-Based Multi-Stakeholder Recommendations by Multi-Objective Optimization. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 128–135. IEEE, December 2018. ISBN 978-1-5386-7325-6. doi: 10.1109/WI.2018.00-98.

[488] Yong Zheng and David Xuejun Wang. A survey of recommender systems with multi-objective optimization. *Neurocomputing*, 474:141–153, February 2022. ISSN 09252312. doi: 10.1016/j.neucom.2021.11.041.

[489] Yu Zheng, Chen Gao, Xiangnan He, Yong Li, and Depeng Jin. Price-aware Recommendation with Graph Convolutional Networks. In *Proceedings of the IEEE International Conference on Data Engineering*, pages 133–144. IEEE, April 2020. ISBN 978-1-72812-903-7. doi: 10.1109/ICDE48307.2020.00019.

[490] Yu Zheng, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Incorporating Price into Recommendation with Graph Convolutional Networks. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2021. ISSN 1041-4347, 1558-2191, 2326-3865. doi: 10.1109/TKDE.2021.3091160.

[491] Bo Zhou and Tianxin Zou. Competing for Recommendations: The Strategic Impact of Personalized Product Recommendations in Online Marketplaces. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3761350.

[492] Chi Zhou, Mingming Leng, Zhibing Liu, Xin Cui, and Jing Yu. The impact of recommender systems and pricing strategies on brand competition and consumer search. *Electronic Commerce Research and Applications*, 53:101144, 2022.

[493] Chi Zhou, Mingming Leng, Zhibing Liu, Xin Cui, and Jing Yu. The impact of recommender systems and pricing strategies on brand competition and consumer search. *Electronic Commerce Research and Applications*, 53:101144, May 2022. ISSN 1567-4223. doi: 10.1016/j.elerap.2022.101144.

[494] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep Interest Network for Click-Through Rate Prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1059–1068, London United Kingdom, July 2018. ACM. ISBN 978-1-4503-5552-0. doi: 10.1145/3219819.3219823. URL https://dl.acm.org/doi/10.1145/3219819.3219823.

[495] Yayu Zhou, Bobby J Calder, Edward C Malthouse, and Yasaman Kamyab Hessary. Not all clicks are equal: detecting engagement with digital content. *Journal of Media Business Studies*, pages 1–18, 2021.

[496] Han Zhu, Xiang Li, Pengye Zhang, Guozheng Li, Jie He, Han Li, and Kun Gai. Learning tree-based deep model for recommender systems. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1079–1088, 2018.

[497] Tao Zhu, Patrick Harrington, Junjun Li, and Lei Tang. Bundle recommendation in ecommerce. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 657–666. ACM, July 2014. ISBN 978-1-4503-2257-7. doi: 10.1145/2600428.2609603.

[498] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the International Conference on World Wide Web*. ACM Press, 2005. ISBN 978-1-59593-046-0. doi: 10.1145/1060745.1060754.

[499] Morteza Zihayat, Anteneh Ayanso, Xing Zhao, Heidar Davoudi, and Aijun An. A utility-based news recommendation system. *Decision Support Systems*, 117:14–27, 2019.

[500] Lixin Zou, Long Xia, Zhuoye Ding, Jiaxing Song, Weidong Liu, and Dawei Yin. Reinforcement Learning to Optimize Long-term User Engagement in Recommender Systems. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2810–2818. ACM, July 2019. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330668.