Università degli Studi di Padova

Head Office: University of Padova

Department: Department of Mathematics

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

Curriculum: Computer Science and Innovation for Societal Challenges

Series: XXXVI

Integration of Explainability in Recommender Systems to Enhance Enterprise Value Strategies

Thesis written with the financial contribution of Estilos srl

Coordinator: Prof. Anna Spagnolli
Supervisor: Prof. Fabio Aiolli
Co-Supervisor: Prof. Anna Spagnolli

Ph.D. student: Andrea Montagna

“Alea iacta est”
— Gaius Iulius Caesar

Abstract

Companies are often looking for strategies to achieve business goals in the most efficient way. In their journey, they rely on systems and algorithms to support their decisions. One family of algorithms, that helps companies in choosing which actions to take, is Recommender Systems (RSs). RSs are a family of algorithms that generate suggestions of the item-user type. They are widely used in social networks, e-commerce, news and advertising, and online streaming applications. However, one area that has not yet been widely addressed in the literature is related to the business effects of recommendations. The business effect can be interpreted as the generation of value for the company, which can take many shapes and involve many users. For example, the business value may consist not only of a direct measure of company metrics, such as revenue or profit, but also of changing the sales distribution, increasing Clicks-Through Rates for an advertisement, or even keeping a user's interest in an item high.

This topic in the literature is called Value-Aware Recommender Systems (VARSs). VARSs are a particular class of RSs that aim to maximize one or more business indicators to achieve a well-established business objective. Through VARSs, a company can adopt strategies to increase efficiency and answer business questions by driving the market dynamically. However, end-users must understand a recommendation to get the best tradeoff between demand and response. In the first part of this thesis, we explore the state-of-the-art of VARSs, collecting and classifying all VARSs available in literature in a first-of-its-topic systematic review. Additionally, we propose a more specific taxonomic categorization for the different models provided, highlighting the model outcome and the economic perspective.

Furthermore, in order to improve the comprehension of the suggestions provided by these algorithms, we focus on Explainable Artificial Intelligence (XAI) that aims to promote transparency in RSs and thus incentivize user adoption. XAI focuses on different perspectives, such as industrial, social, and end-user, giving explanations based on the context. In the second part of the thesis, we suggest a novel, explainable, value-aware recommender system that aims to balance XAI and VARS perspectives. Applying XAI to VARSs is still exploratory and has several potential evolutions and academic-industrial interests.

Scientific research has shown that there are many advantages to complementing a recommendation with a convincing explanation. For example, such explanations often lead to improved user trust, which in turn improves the effectiveness of recommendations and system adoption. In particular, for this reason, many research works are studying explainable recommendation algorithms based on graphs, i.e., exploiting Knowledge Graph (KG) or Graph Neural Networks (GNNs) methods. The use of graphs is very promising since algorithms can, in principle, combine the benefits of personalization and graph reasoning, thus potentially improving the effectiveness of both recommendations and explanations. However, although graph-based al-

gorithms have been repeatedly shown to bring improvements in terms of ranking quality, not much literature has yet studied how to properly evaluate the quality of the corresponding explanations. In the final part of this thesis, we discuss a problem that affects explainability features applied on KG and GNN models, examining how they are currently assessed and suggesting the direction for a future more quantitative and comparable evaluation.

Acknowledgments

I would like to express my gratitude to the many persons who have supported and inspired me throughout my doctoral journey. This achievement would not have been possible without their unwavering encouragement, guidance, and understanding.

First of all, I would like to sincerely thank Estilos, the organization that allowed me to conduct my research during this important moment in my career, and all my colleagues. In particular, I thank Suleima Citron, Marco Colusso, and Paolo Trevisan for handling projects in the best possible way so that I could concentrate on my Ph.D. program. Finally, I would like to thank Andrea Bergamo, CEO of Estilos srl, who has been my human, intellectual, and professional mentor throughout this journey. Your commitment to innovation and excellence has provided me with a rich and stimulating environment to explore the complexities of artificial intelligence systems. I am truly grateful for your support.

I owe a debt of gratitude to my colleague and friend, Alvise De Biasio. Our shared experiences, challenges, and triumphs have made this journey not only intellectually rewarding but also deeply and humanly meaningful. Thank you for being an incredible companion on this path.

A warm greeting to the HCAI team in Magdeburg, Germany, who welcomed me during my abroad period. Your hospitality and competence have enriched my academic experience and broadened my horizons.

To my family, who have stood by me from afar, offering me their unwavering support and encouragement, especially during the difficult times of the past year, I am deeply grateful. Your trust in me has been a constant source of motivation.

Last but certainly not least, I want to express my love and gratitude to my partner, Silvia Balzan. Your boundless faith in my abilities and your constant support have been my rock. Your presence in my life has made this journey even more meaningful. You are my companion on this and future journeys.

To all my mentors, friends, and loved ones who have played a role in my academic and personal growth, I extend my warmest thanks. Your contributions, whether big or small, have shaped my path.

Thank you, one and all, for being an integral part of this incredible journey.

Contents

Listing of figures

Listing of tables

Listing of acronyms

CLV . . . . Customer Lifetime Value

CTR . . . . Click-through rate

EMF . . . . Explainable Matrix Factorization

E-NDCG . . . . . . Explainable Normalized Discounted Cumulative Gain

GNN . . . . Graph Neural Network

GNNxRS . . . . . . GNN-based Explainable Recommender System

GRS . . . . Graph-based Recommender System

GxRS . . . . Graph-Based Explainable Recommender System

HR . . . . Hit Ratio

KG . . . . Knowledge Graph

MDP . . . . Markov decision process

MF . . . . Matrix Factorization

NDCG . . . . Normalized Discounted Cumulative Gain

NDCV . . . . Normalized Discounted Cumulative Value

RL . . . . Reinforcement Learning

RLRS . . . . Reinforcement Learning Recommender Systems

RS . . . . Recommender System

VARS . . . . Value-Aware Recommender System

xRS . . . . Explainable Recommender System

XAI . . . . Explainable Artificial Intelligence

xVARS . . . . Explainable Value-Aware Recommender System

XVMF . . . . Explainable Value-aware Matrix Factorization

1 Introduction

In the contemporary landscape of information abundance, the pivotal role of recommender systems has emerged as a key to society's interaction with digital content and services. These systems, driven by advanced algorithms and artificial intelligence, play a critical role in assisting users to navigate the abundant volume of choices and enable efficient and tailored access to relevant information, products, and experiences. This introductory chapter represents a prelude to an in-depth exploration of recommender systems in the context of business applications, delving into the growing user needs to comprehend better the suggestion of a model.

Over the past decade, recommender systems have gained significant attention due to their transformative impact on modern user experiences. This increasing interest can be attributed to the convergence of several factors, such as the proliferation of online platforms, the exponential growth of data available and collected, and the advancement of machine learning techniques. As a concrete example of this paradigm shift, consider the rise of streaming services. These platforms leverage recommender systems to propose personalized playlists, exploiting user preferences and behaviors to create tailored music (or movies) recommendations. This not only amplifies user engagement but also exemplifies the efficacy of these systems in delivering highly relevant content. Moreover, the business world has demonstrated an increasing awareness of the potential of recommender systems and this trend can be observed, for instance, in the spreading of the e-commerce sector. E-commerce heavily relies on recommender systems to enhance customer engagement and augment revenue. Enterprise clients, in turn, increasingly seek the integration of these systems into their business strategies, recognizing their potential to drive sales, customer satisfaction, and brand loyalty. The complex interaction between cutting-edge technologies and real-world profitability demonstrates the need to improve and advance the recommendation system.

In an attempt to exploit the full potential of recommender systems, the need for transparency and comprehensibility in their decision-making processes is a key challenge. As these systems evolve in complexity and sophistication, the resulting recommendations can often appear as *black box* results to end users and stakeholders. This opacity prevents the creation of trust and limits the overall usefulness of the system, generating a demand for tools to clarify the underlying mechanisms of recommendation systems and make their results understandable to both users and companies.

This thesis includes an examination of the state of the art of recommender systems with the aim of generating business value in enterprise applications. The overall objective is two-fold: firstly, to survey and summarise existing advances in the field, and secondly, to develop explainability strategies that improve the applicability of these systems to real-world business scenarios. A focal point of investigation concerns the optimization of recommendation models with respect to the delicate balance between their generated value and their explanatory capabilities. The aspiration is to propose a direction in the debated context between the accuracy of recommendations and their comprehensibility, ensuring that recommendations can be interpreted within a business context and thus support actions and decisions.

Over the past few years, a variety of families of models have emerged in the recommendation system landscape, each with unique strengths and limitations. Value-aware recommender systems stand out in this context because they take into account not only user preferences but also the intrinsic value of the recommended items, thus allowing for more precise recommendations tailored to the enterprise context. One of the most widely used approaches in this scenario is called *matrix factorization* techniques. This methodology exploits the latent structures of user-object interaction matrices to reveal latent patterns for accurate predictions. In contrast, a novel approach is the use of graph networks that natively present capabilities to exploit intricate user-item relationships, infusing recommendations with rich contextual information, even if not yet used concretely to generate business value and optimized explanations to business stakeholders.

In this thesis, these families of models will be meticulously explored and analyzed, evaluating mechanisms, limitations, and trade-offs. Leveraging the insights gained from this analysis, the research aspires to develop explainable recommendation strategies that contribute to business users' understanding of the results as an effective support for their daily work.

1.1 Research Motivation and Contribution

This research was initiated with the desire to establish a collaboration between academia and industry in order to bring the two worlds closer together and capture aspects that are still little discussed. The initiative is promoted by an IT consulting company (Estilos srl) that has been working for years to understand the needs of its customers and drive an AI-driven transition, recognizing the value and unexpressed potential of corporate data. The pursuit of this research aligns perfectly with the academic commitment to research and the company's effort towards innovation, realizing the theoretical advances that belong to the context of recommender systems. This thesis embarks on a rigorous journey through the realm of recommender systems and explanation capabilities, beginning with an exploration of their economic significance and culminating in a search for more effective and interpretable business-oriented models. The following chapters delve into the complexities of these systems, assessing the mechanisms and analyzing their impact on the business landscape, providing directions for future solutions and challenges that characterize the domain of recommender systems applied to the contemporary enterprise. Four distinct interrelated contributions have been developed to rationalize and advance the state of the art in this domain:

1. **Analysis and taxonomy of Value-Aware Recommender Systems:** the first contribution of this thesis consists of a comprehensive analysis of the state of the art of recommender systems, focusing exclusively on the identification and categorization of the family of systems designed to generate value for companies, known as Value-Aware Recommender Systems (VARSs). Through a systematic literature review based on PRISMA constraints, the various categories and technical approaches within this family are grouped into a novel taxonomy to facilitate the adoption and understanding of existing studies. Furthermore, the research investigates real-life applications of VARSs proposed in the literature, analyzing the availability of different datasets that embed business value information. Finally, by rigorously examining the strengths, limitations, and implications of these systems, this contribution aims to provide a holistic view of their applicability in practical business scenarios;

2. **Exploring the connection of explainability and business significance:** the second contribution explores the critical aspect of explainability in the realm of recommendation systems, particularly in a business context. Taking from industry experiences and existing literature, this contribution explores the importance of assessing explainability in decision-making processes. The industrial perspective exposes how a lack of explainability can hinder the acceptance and adoption of recommender systems in real-world business contexts. Combining practical insights and theoretical perspectives, this con-

tribution emphasizes the need for recommendations that are not only accurate but also understandable and transparent to stakeholders;

3. **Bridging business value and explainability:** the third contribution of this research aims to establish a connection between the commercial value generated by recommendation models and the demand for explainability. The search for a balance between the accuracy of recommendations and the ability to interpret results is a central challenge in the field of recommendation systems. This contribution focuses on the development of a novel model that aims to balance business value and recommendation explainability, investigating innovative approaches that reconcile these requirements. The objective is to find a balance between the quantitative benefits and qualitative insights that recommender systems bring to the business landscape;

4. **Explainability features in graph networks, a novel challenge:** the fourth contribution of this research concerns the emerging application of explainable graph networks in recommendation systems. Graph networks, with their ability to capture complex relationships between items and users, have been promised to improve the accuracy of recommendations. This contribution takes this a step further by focusing on the evaluation of explainability integrated into graph-based recommendations, aiming to improve transparency and user acceptance of graph network-based recommendations. Through an empirical investigation, this contribution studies the state-of-the-art of how Graph-Based Explainable Recommender Systems (GxRSs) is proposed and how they are evaluated, providing a direction and an improvement of recommendation outcomes in business contexts.

In conclusion, this research explores the multifaceted landscape of business-applied recommender systems, enhancing the comprehension, effectiveness, and transparency of these systems. Through rigorous analysis, empirical investigation, and a strategic fusion of theoretical and practical perspectives, this thesis aims to guide the field of recommender systems toward a business-oriented direction.

1.2 Structure of the thesis

This thesis is organized as follows: Chapter 2 delves into a systematic literature review of Value Aware Recommender Systems (VARS), which constitute a main aspect of the research's original contributions. It explores categories and technical methodologies within the business-oriented recommender system, culminating in the groping into VARS family. By examining their real-world applications and limitations, this chapter aims to provide a comprehensive overview of VARS within a business context, and a novel proposed taxonomy that organizes these systems for enhancing comprehension.

Subsequently, Chapter 3 explores the complex relationship between the explainability of recommender systems and their ability to deliver business value. Based on practical industrial experience and theoretical frameworks, it emphasizes the crucial need to make recommendation outputs transparent and understandable. Furthermore, it examines the difficulty of balancing the value-driven goals of these systems with the requirement for clear and feasible explanations. Finally, in order to facilitate the development of more efficient and valuable enterprise recommendation models, this study seeks to establish a clear link between the importance of business and the ability to provide adequate explanations, providing a novel model that aims to evaluate the trade-off between the explainability feature and the value generated.

Then, in Chapter 4, we explore the fast-growing field of graph networks, which are capable of improving the comprehensibility of recommender outcomes. Recognizing the inherent challenges, we examine how explainability features can be incorporated into graph-based recommendations. Furthermore, this chapter highlights a significant gap in the research: the lack of a uniform evaluation method for comparing explainability-enhanced graph networks. Finally, we aim to help establish a foundation for the evaluation and benchmarking of graph-based recommender systems in the context of explainability by examining the advantages and disadvantages of these new techniques.

Finally, in Chapter 5, we conclude with a discussion of the contributions of this thesis and potential avenues for future research in both academic and industrial settings.

1.3 Publications

Part of the contribution of this thesis produced a systematic review under PRISMA constraints [1] published in the journal *Expert Systems with Applications*, and a position paper on the quality evaluation of graph-based explainable recommendation systems published in the conference CHItaly'23 and presented in the HCAI4U'23 workshop.

- **A systematic review of value-aware recommender systems**. [2]
  Alvise De Biasio, Andrea Montagna, Fabio Aiolli, and Nicolò Navarin.
  Expert Systems with Applications, September 2023.
  ISSN 0957-4174
  doi: 10.1016/j.eswa.2023.120131

- **Graph-based Explainable Recommendation Systems: Are We Rigorously Evaluating Explanations?**. [3]
  Andrea Montagna, Alvise De Biasio, Nicolò Navarin, and Fabio Aiolli.
  Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter.
  CHItaly'23, September 2023.

1.4 Work in progress

The concepts, approaches, and model proposed in Chapter 3 are going to be formalized in a paper and submitted to a conference or journal.

- **XVMF: a recommender system designed to balance business profit and explainability features**.
  Andrea Montagna, Alvise De Biasio, Nicolò Navarin, and Fabio Aiolli.
  [Conference or journal for submission to be defined.]

2
Value-Aware Recommender System

Recommender Systems (RSs) help users make daily decisions [4] and they are used in various applications, including e-commerce systems [5], news [6] and online streaming [7] services, and advertising platforms [8]. To alleviate the problem of information overload [9], RSs recommend items of greatest interest for users to focus their attention on. Indeed, traditional recommendation models are designed to provide personalized recommendations relevant to the user [4]. In order to match customer preferences, an RS typically addresses a specific objective [10], namely, minimizing the prediction error or maximizing the ranking quality.

However, although providing products and services to satisfy customers is a fundamental requirement for the sustainability of any business, an organization often decides to adopt a recommender to improve business performance [11, 12]. For these reasons, in the past few years, there has been increased interest in Value-Aware Recommender Systems (VARSs) [13]. These systems are designed to optimize the *economic value* of recommendations by balancing the interests of multiple stakeholders [14], i.e., consumers, providers, and organizations. Some examples of VARS include recommenders designed to maximize profits [15], increase user engagement [16], and improve customer lifetime value [17].

In this chapter, we present a systematic literature review on value-aware RSs based on the PRISMA guidelines [1]. Most surveys in the RS field have investigated related domains, such as multi-objective RS [18], multi-stakeholder RS [19], multi-criteria RS [20], context-aware RS [21], and attribute-aware RS [22]. However, to the best of the authors' knowledge, no survey or review has focused specifically on VARSs. Therefore, we aim to help academic researchers and industry stakeholders understand how VARS can be used to optimize value, the principal application domains, open challenges to be addressed, and future research directions. Finally, this chapter represents a necessary step in the research because it provides an industrial perspective that is enhanced in the following chapter 3 through the integration of explainability capabilities into recommender systems and the proposal of a model that aims to balance explainability performance with business value generation.

The main contributions of this chapter can be summarised as follows:

- we provide the first systematic literature review focused on Value-Aware Recommender System, based on PRISMA guidelines, by discussing articles collected from different research streams and industrial perspectives;
- we describe the main value categories that are traditionally optimized by VARS and the technical approaches used in the design of these value-generation algorithms;
- we discuss the various application domains, the most commonly used datasets, the main challenges, and possible future research directions.

2.1 Background and Concepts

In this section, we introduce the basic facts of RSs to provide background elements necessary for better understanding the related work section that reports VARS similarities and differences from other general recommender classes.

2.1.1 Recommendation Algorithms

RSs are algorithms designed to offer suggestions of items of interest for users [4]. Various service providers have deployed RSs in different domains, including e-commerce [5], online streaming [7] and news services [6]. Users interact with these systems through various online sites whenever they are looking for a product to purchase, news to read, or a movie to watch. RSs help users evaluate a large number of alternatives [9] by suggesting items that might be of the greatest interest. These suggestions are offered to the user in the form of a ranking of items [4]. The ranking is generated by algorithms that exploit information collected explicitly (e.g., item ratings) or implicitly (e.g., browsing behaviour, product reviews) from the user's interaction with the platform hosting the service [4].

The Recommendation Problem

The recommendation problem can be formulated primarily in two ways, i.e., determining the degree of user interest in a particular item (*prediction problem*) or identifying a set of k items of interest to the user (*t*op-k recommendation problem) [10]. Formally, in both cases, given a set $\mathcal{U} = \{u_1, \ldots, u_m\}$ of users and a set $\mathcal{I} = \{i_1, \ldots, i_n\}$ of items, a RS is designed to predict a matrix of scores $\hat{\mathbf{R}} \in \mathbb{R}^{m \times n}$ from a matrix $\mathbf{R} \in \mathbb{R}^{m \times n}$ of ground-truth preferences [4, 23]. Although it is always possible to order the predicted ratings to obtain a rank of k items for the user, algorithms are developed specifically for the prediction problem (e.g., matrix factorization [24]) or for the top-k recommendation problem (e.g., sparse linear method [25]). There are two main modes by which recommender systems are evaluated [11, 12], i.e., online A/B tests or offline evaluation. Online evaluation based on conversion rates (e.g., measuring how often a user chooses a recommended item) is the most direct method for evaluating the effectiveness of an RS [11, 12]. However, these types of studies are often difficult in practice because they require access to existing systems with large user groups, where potentially risky testing can be performed that could impact the economic performance of service providers. Therefore, the evaluation of offline performance based on historical data is generally preferred to online testing [11, 12]. For the prediction problem, where the algorithm tries to infer the rating for a given user and item, performance is traditionally evaluated by calculating the prediction error (e.g., mean absolute error and root mean square error) [26]. On the other hand, for the top-k recommendation problem, since the algorithm outputs a list of k items for the user, performance is typically evaluated using relevance or ranking metrics (e.g., precision, recall, and normalized discounted cumulative gain) [26]. Therefore, offline evaluation provides an indirect performance measurement [11, 12], potentially correlated with online metrics, which is used as a proxy for the latter.

Main Classes of Recommender Systems

To suggest the most relevant items, RSs personalize recommendations [4]. Different users receive different recommendations according to their interests. Based on the type of personalization, recommender systems are often divided into different classes. One of the best-known taxonomies of recommender systems [4] divides algorithms into *content-based filtering (CB) [27], collaborative filtering (CF) [28] and hybrid systems (HS) [29]. CB systems [27] suggest items with characteristics similar to those with which the user interacted in the past. CF systems [28] recommend items that other users with similar tastes have engaged with* in the past. By contrast, HS systems [29] rely on a combination of the previous techniques. In addition, further distinctions can be made within this taxonomy. CF systems are traditionally divided into *neighborhood* [23] and *model-based* [30] approaches. The former [23] recommend new items using user/item similarity criteria (e.g., user-based neighborhood, item-based neighborhood). The latter [31, 32, 30] learn a predictive model from historical ratings to make new recommendations (e.g., rule-based collaborative filtering, latent factors models).

Typical Challenges in Recommender Systems

There are various advantages and disadvantages depending on the type of algorithm chosen [4]. For example, RSs may suffer to a greater or lesser degree from the cold-start problem [33], which occurs whenever the amount of information available is insufficient to produce recommendations that effectively reflect the interests of new users and recommendations of new items that have not yet been engaged with by users. In particular, [34] CB methods are generally more robust to new item cold-start than are CF methods. However, CBs often produce recommendations of items that are highly similar to those engaged thus far, preventing the user from discovering surprisingly relevant items. By contrast, HS methods can work well in cold-start settings; however, the computational cost is often very high, and it is difficult to produce an explanation of the rationale behind the recommendations.

2.2 Related Work

2.2.1 Value-Aware Recommender Systems

In this section, we introduce the economic concept of *v*alue. We also note the typical types of business value generated by recommendations. Next, we provide a chronological overview of *v*alue-aware recommender systems. These algorithms are designed to directly optimize various types of business value of recommendations for organizations.

An Economic Perspective on the Concept of Value

From early academic definitions in the mid-1950s, the term value has had multiple meanings, closely related to the application scenario considered. In early studies, Miles [35] defines the concept of value by distinguishing use value, estimated value, cost, and exchange value. As reported in the author's research, use value is the ability to perform a certain function, i.e., considering a mobile phone, its use value is the ability to make a phone call. On the other hand, the estimated value is related to the sphere of attractiveness and desirability, e.g., a mobile phone with a color display is more desirable than one with a black-and-white display. Cost value is related to the economic quantity used to produce an item, e.g., the cost to produce every component and assemble a mobile phone. Finally, exchange value is related to the increase in value over time, i.e., the mobile phone after ten years.

On the basis of these theoretical concepts, authors have proposed alternative definitions that focus on different factors. In some work [36, 37], the concept of the value of a product or service is related to the expected benefit that the buyer receives as a function of the price paid. For example, if the purchase of a product produces certain savings, the value lies precisely in the delta between the savings and the price paid. On the other hand, other works [38, 39, 40] define value according to customer perception. According to this interpretation, the value of a product or service is highly dependent on factors related to the emotional and sensory sphere of the customer. For two distinct customers, the same product might have a different value depending on the emotions/feelings it generates.

As can easily be inferred from the above considerations, the definition of value is not unique and may differ depending on the perspective considered. The value for the user/customer is often related to the concepts of quality and personalization, experience and trust, features, and benefits [41, 42]. Moreover, the value for the producer/business is often linked to the loyalty relationship established with the customer and the economic results of sales [43, 44, 45]. Therefore, when business value is referred to in the literature, it represents the impact on the company's economic indicators (e.g., revenues, costs, margins, profits, and losses).

On the Business Value of Recommendations

As discussed in the previous section, the concept of value has multiple definitions in the literature and is largely context dependent. In the field of RSs, a recent study by Jannach and Zanker [11, 12] proposes a heterogeneous taxonomy based on five distinct definitions of business value that recommendations may generate:

- *Click-Through Rate (CTR)*: according to which the business value of recommendations is defined according to the number of user clicks;
- *Adoption and Conversion Rate*: according to which the business value depends on the degree of customer adoption of the system;

- *Sales and Revenue*: where business value is defined as a function of total sales of products and services;
- *Effects on Sales Distributions*: according to which the value depends on the effects of recommendations on the distribution of items sold;
- *User Engagement and Behaviour*: according to which the value depends on the customer's overall engagement with the platform.

As can easily be observed, the business value of recommendations depends on the application context (e.g., product recommendation, news, ads) and the company's business model (e.g., direct sales, rental, subscription). As a result, the value of recommendations could differ depending on, for example, whether the company sells physical products through e-commerce or sells a subscription service by streaming video content.

Some studies [46, 47, 48, 49, 50] provide quantitative evidence by relating recommendations to specific types of value (e.g., sales and revenue, effects on sales and distribution). For example, in some research [46, 47, 49], the effect of recommendations on the diversity of products sold is measured. According to the authors, a recommendation system would individually lead the user to increase or decrease the diversity of items purchased. However, on average, recommendations lead to an overall decrease in diversity in favour of the most popular items. On the other hand, with regard to the effect of recommendations on the overall sales volume, in the literature [48], it is found that depending on the type of design (i.e., collaborative filtering, content-based), one algorithm could show higher performance than another. Furthermore, as found in more recent studies [50], these two factors, namely, diversity and sales volume, are correlated. In particular, greater diversity correlates with higher purchase rates, average purchase amounts, and cross-purchase rates.

Introduction to Value-Aware Recommender Systems

Personalization has traditionally led RSs to focus on the user [4]. 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 vary [11, 12]. As recently argued in research on *multistakeholder recommender systems* (*MSRS*) [14], RSs should consider the interests of multiple parties, known as stakeholders, to generate recommendations. e.g., consumers who receive the recommendations, providers who supply items behind the recommendations, and organizations that manage the recommendation service. Depending on the perspective from which the MSRS is designed, recommendations will be generated to optimize the utility of one or more stakeholders. Within the multi-stakeholder taxonomy, a particular class of algorithms known as Value-Aware Recommender Systems (VARSs) can be distinguished. VARS are systems that aim to directly maximize the economic value of recommendations. These include systems designed to increase sales, improve customer lifetime value, and optimize profitability.

The strategic goal of optimizing the value of recommendations emerged with the growth of e-business. The first studies in the VARS field [51] date to 2007. These works propose several methodologies to generate more profitable recommendations to increase the business value of an organization. However, the first explicit reference to the term *value-aware* [52]. In their study, the authors introduce VARS as a future research direction for industrial applications. Research on VARS was subsequently brought to the attention of the academic community in *Workshop on Value-Aware and Multi-S*takeholder Recommendation (*VAMS 2017*) [13]. The workshop encouraged researchers to formulate a common vision on this emerging research area by inviting the submission of papers on various topics, including value-aware recommendations and multi-stakeholder recommendations. After VAMS 2017, there has been an increase in the number of specialised articles on VARS. Some studies have investigated how to design VARS using specific methodologies including post-processing approaches [53] and reinforcement learning algorithms [15]. Other articles have proposed methodologies that aim to optimisw value in certain application contexts, i.e., e-commerce [15], advertising [54], news [16] and others. Furthermore, other studies [55, 56, 53] have investigated the main benefits and risks of using VARS in real-world circumstances, considering customer pricing preferences, the trade-off between profitability and accuracy, and the short- and long-term consequences for organizations.

2.2.2 Other Classes of Recommender Systems and Related Works

Research on VARS is an emerging topic. However, other RSs have been proposed in the literature to solve related problems. The latter include the following:

- *Multi-Objective Recommender Systems* [57]: in which the system aims to produce recommendations that optimize several objectives (e.g. accuracy, novelty, diversity) simultaneously;

- *Multi-Criteria Recommender Systems* [58]: in which the system exploits a user's preferences on different item criteria (e.g., room cleanliness, location, safety) to provide better suggestions;

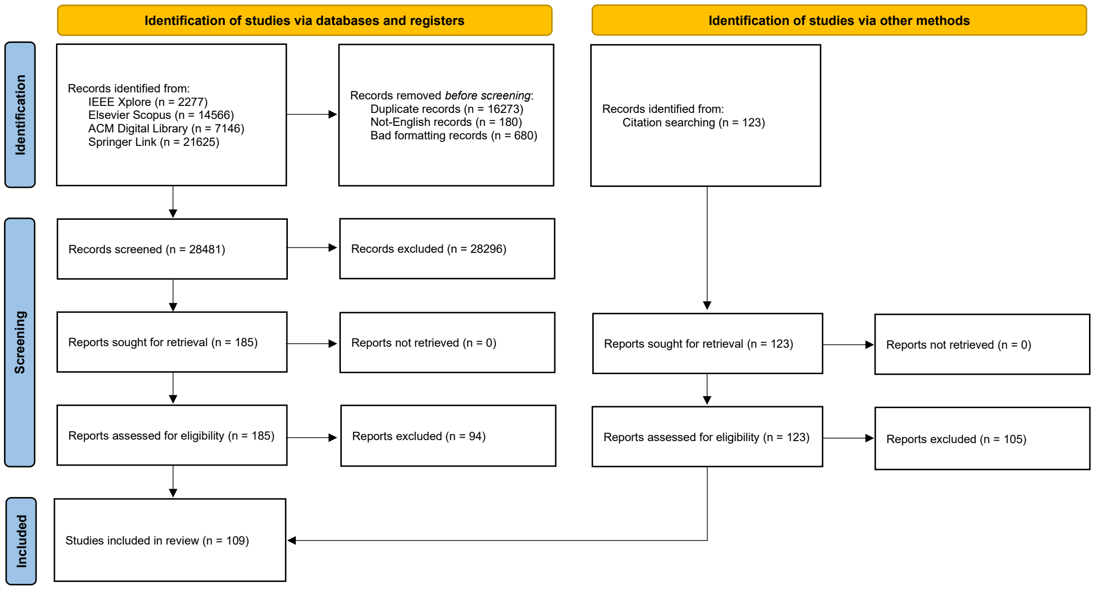


Figure 2.1: PRISMA flow diagram representing the systematic literature review process.

- *Multi-Stakeholder Recommender Systems* [59]: in which the system considers the interests of multiple stakeholders (e.g., consumers, suppliers, organizations) to generate recommendations;
- *Context-Aware Recommender Systems* [60]: in which the system uses contextual information (e.g., location, time) to provide personalized recommendations to the user;
- *Attribute-Aware Recommender Systems* [22]: in which the system exploits additional user (e.g., gender), item (e.g., category), and rating (e.g., time) information to provide more personalized recommendations;
- *Price-Aware Recommender Systems* [61]: in which the system exploits the user's price preferences and sensitivity to increase the accuracy of recommendations.

There are various surveys and reviews on RSs since the research field has been studied in the past several decades. Some of these works [62] approach the problem from a general perspective. Others specialise in certain topics, such as recommendations based on deep learning [63, 64] or reinforcement learning [65]. Many surveys focus on different types of RSs, such as multi-objective RS [18], multi-criteria [20], multi-stakeholder [19, 59], context-aware [21, 66], attribute-aware [22] and fairness [67, 68]. As introduced earlier, VARS differ from the previously mentioned categories since they aim to directly maximize economic value. To the authors' knowledge, although there is growing interest in the literature [12, 11] on the topic of RSs' value creation for business stakeholders, no surveys or reviews based on PRISMA guidelines [1] focused on VARS have been conducted.

2.3 Proposed Approach

To select studies for inclusion, we adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [1] guidelines. The rigor and coverage of the PRISMA selection process is recognised throughout the scientific community as an indication of high reliability and quality. 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.

2.3.1 Research Questions

The systematic review aims to answer the following research questions (RQ):

- 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?

2.3.2 Eligibility Criteria

Only articles that met the following eligibility criteria (EC) were included:

- 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.

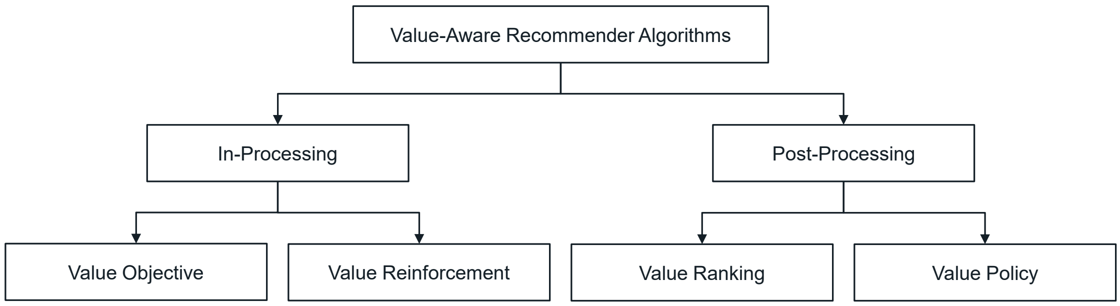


Figure 2.2: Value-aware recommender algorithm taxonomy

.

- 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.

2.3.3 Search Strategy

We identified all articles in various online journal databases from 2006 to 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”)).*

To stay below the maximum number of items that could be extracted from 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.

2.3.4 Selection Process

As shown in the PRISMA flow diagram in Figure 2.1, a total of 2,277 articles from IEEE Xplore, 14,566 articles from Elsevier Scopus, 7,146 articles from ACM Digital Library, and

21,625 articles from Springer Link were identified in this 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. In the screening stage, the titles and abstracts of 28,481 articles were analysed, and 28,296 records were excluded 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. 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. At the end of this overall process, a total of 109 studies were included in the review.

2.3.5 Study Limitations

The main limitations of the present study are as follows:

- 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.
- Unpublished articles, non-English articles, articles whose content was not accessible, graduate theses, doctoral dissertations, commercial products, and demos were not included.

2.4 Results

In this section, we present the results of the systematic review. First, we classify and describe VARS algorithms. Then, we review some of the work that has studied applications of VARS in the past few years. Finally, we present the most commonly used datasets.

2.4.1 Value-Aware Recommender Algorithm Taxonomy

In this section, we introduce the main algorithms in the literature on VARS. 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 2.2.1), we provide a classification of VARS according to the technical approaches used to highlight the different mechanisms underlying the various algorithms. As indicated in Figure 2.2, VARS 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.

Value-Aware Post-Processing Algorithms

Post-processing algorithms can be applied to any recommendation algorithm (treated as a black box) to optimize the value of recommendations

In traditional scenarios, a recommender system suggests to user u a rank $\mathcal{Y}^*_{u,k}$ of k items that maximizes the expected interest:

$$\underset{\mathcal{Y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathcal{Y}_{u,k}} \hat{r}_{u,i} \tag{2.1}$$

by sorting the predicted scores $\hat{r}_{u,i}$ of the unrated items in descending order and selecting the first k. Post-processing methods rely on predicted scores and other economic information to rerank the output of the original algorithm.

Value Ranking

This class of methods extends the approach in Eq. (2.1) by incorporating economic value information into the objective function to rerank the output of the original algorithm.

Given a value $v_i \in \mathbb{R}$ associated with item i (e.g., product profit), a strategy commonly used by these systems [51, 69, 70, 71, 72, 73, 55, 74, 75, 6, 53] is to recommend the set $\mathcal{Y}^*_{u,k}$ of items that maximize the weighted expected interest:

$$\underset{\mathcal{Y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathcal{Y}_{u,k}} \hat{r}_{u,i} \cdot v_i \tag{2.2}$$

selecting the first k items with the highest $\hat{r}_{u,i} \cdot v_i$. As noted in some studies [51, 69], 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 noted in various works [55, 75, 53], 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 [72, 74, 6, 53, 76, 77, 78] have proposed simple extensions of Eq. (2.2) to account for the perspectives of different stakeholders and determine the best trade-off between economic value for the organization and customer interests:

$$\underset{\mathcal{Y}_{u,k}}{\operatorname{argmax}} \sum_{i \in \mathcal{Y}_{u,k}} (1-\alpha) \cdot \hat{r}_{u,i} + \alpha \cdot v_i \tag{2.3}$$

using a regularization parameter $\alpha \in [0, 1]$ to control the equation. Some variants [70, 71, 75, 73] of the approach in Eq. (2.3) 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 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.

Various studies [79, 80, 81] have proposed multiple-step process-based approaches to optimize economic value. For example, one study [79] 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 [80] by incorporating various factors such as price, profitability, product competition, and saturation effects to improve profitability over a finite time horizon. Recent work [81] has proposed 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.

Other works [5, 82, 7, 83] have proposed methodologies optimizing the value of recommendations by integrating dynamic pricing algorithms. For example, some works [83] have proposed to optimize the discount of recommended items by exploiting multi-armed bandits. By contrast, a more recent work [7] has proposed personalizing the price of recommended products based on customer willingness to pay to simultaneously optimize service provider profit and customer surplus.

Application Domain	Most Frequently Used Technique	Typically optimized Value	Details
Product Recommendation	Value Objective	Sales and Revenue	Table 2.2
Advertising Recommendation	Value Reinforcement	Sales and Revenue	Table 2.3
News Recommendation	Value Reinforcement	CTR, User Engagement	Table 2.4
Media Recommendation	Value Objective	Distribution	Table 2.5

Table 2.1: Application domains of value-aware recommender systems.

Value-Aware In-Processing Algorithms

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

Value Objective

This class of methods contains algorithms that integrate the objective function of known or domain-specific algorithms to generate more valuable recommendations.

For example, some work [84, 85, 86, 87] has proposed modifying the well-known *n*eighbourhood recommender system [23]. The original algorithm computes the similarity between users (*user-based approach*) or items (*item-based approach*) belonging to a given neighborhood to predict the scores. A neighborhood refers to a set of users who share similar interests or a set of items that have been engaged with by similar users. For example, in the user-based approach, the algorithm computes the similarity $sim(u, v)$ between a user $u \in \mathcal{U}$ who did not express a preference for item i and a user v belonging to a set $\mathcal{P}(u, i)$ of users who expressed preferences similar to u and rated i.

(2.4)

based on the ratings of the neighbours. Some value objective algorithms have been proposed by partially modifying the function in Eq. (2.4) to optimize some types of economic value. For example, some work [87] has proposed a neighbour selection algorithm to increase the overall profitability of recommended products while maintaining accuracy under shilling attacks, i.e., identifying malicious users who generate biased ratings to influence recommendations for their own interests. Moreover, other studies [86] have proposed increasing sales diversity by recommending users to items by reversing the original neighbour computation algorithm.

Furthermore, other research [88, 89, 90, 91, 92] extends the well-known *m*atrix factorization algorithm [24, 30] by incorporating value information into the objective function. In the traditional matrix factorization algorithm, the user-item interaction matrix **R** is decomposed into the product of two rectangular lower-dimensional latent space matrices representing users and items. Decomposition is often performed using a dimensionality reduction algorithm known as singular value decomposition [93].

$$\hat{r}_{u,i} = \mathbf{p}_u^\intercal \mathbf{q}_i \tag{2.5}$$

by computing the dot product between the l-dimensional latent feature vector $\mathbf{p}_u \in \mathbb{R}^l$ of user u and $\mathbf{q}_i \in \mathbb{R}^l$ of item i. Some value objective algorithms have been proposed by incorporating other factors into the calculation of predicted scores in Eq. (2.5) to optimize certain types of economic value. For example, some works [88, 90, 91, 92] have used predicted scores determined via matrix factorization and other economic information to improve the utility of recommended products to the user.

Value Reinforcement

Recent studies have proposed value-aware recommendation algorithms by exploiting Reinforcement Learning (RL) [94] techniques, 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. An RL environment can be formalised through a Markov decision process (MDP) in the tuple $(\mathcal{S}, \mathcal{A}, \Omega, \mathbb{P}, \gamma)$, where $\mathcal{S}$ is a set of possible states, $\mathcal{A}$ is a set of possible actions, Ω is a reward function, $\mathbb{P}$ is the probability of transition from one state to another following an action and $\gamma \in [0, 1]$ is a discount factor. Typically, in RL algorithms, an agent interacts with the environment to maximize the expected discounted cumulative reward:

$$\begin{aligned} &\max \qquad \mathbb{E}[\Omega(\tau)] \\ &\Omega(\tau) = \sum_{t=0}^{\tau} \gamma^t \cdot \omega(a_t, s_t) \end{aligned} \tag{2.6}$$

with $\omega(a_t, s_t)$ as the reward for taking action $a \in \mathcal{A}$ in state $s \in \mathcal{S}$ at time t. The objective of the algorithm is to determine an optimal policy $\pi(a|s)$ that involves taking an action in a given state to maximize the reward.

Given the sequential nature of user interaction with an RS, Reinforcement Learning Recommender Systems (RLRS) [65] have emerged as alternative approaches based on RL techniques to generate recommendations. Much of the literature on VARS [95, 96, 16, 97, 17, 15, 98, 54, 8, 99, 100] exploits this methodology to maximize the long-term value of recommendations, implementing the agent reward function $\Omega(\tau)$ in Eq. (2.6) to take into account the value $\mathbf{v}$ of the recommended items. For example, in one study [95], an algorithm was designed to maximize the customer lifetime value (CLV), i.e., the total value generated by the customer throughout his or her history. By contrast, in another study [15], the reward function was modified via the concept of click-conversion rate (CVR) to generate recommendations that maximize the economic value from each user action (e.g., click, add-to-cart, pay).

2.4.2 Value-Aware Recommender Systems Applications

Recent years have witnessed growing interest in VARS. Since algorithms are often designed based on domain-dependent characteristics, in this section, we review the literature on VARS in various application domains. As indicated in Table 2.1, 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.

Product Recommendation

Many VARS have been developed to optimize product sales. 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, the influence of price on the willingness to pay, and real-world studies.

Table 2.2 summarises the literature on value-aware product recommendation systems.

Accuracy-Profitability Trade-Off

Business interest in leveraging recommender systems to increase revenue or other key performance indicators of global e-tailers existed since the 2000s. In early work, Chen et al. [51, 69] proposed a methodology to weight the recommendations of a collaborative filtering algorithm with product profitability factors (i.e., revenues minus costs). 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. While some techniques based on constrained optimization [70, 71, 101, 80, 102] or multi-objective algorithms [76, 78, 103, 87, 104, 105, 106, 107, 108, 109] have been proposed (see Section 2.4.1 to balance the potentially conflicting interests of multiple stakeholders simultaneously, other studies [55, 53, 110, 108] have investigated the accuracy-profitability trade-off through offline simulations. As argued by Jannach and Adomavicius [55], the items that are most important for the user may not be those that produce the highest business value for the service provider. Biasing algorithms in the direction of higher profitability could actually increase marginality while maintaining the relevance of recommended products. However, above a certain threshold, the probability of purchase drops dramatically, and the business value generated as a result is reduced. Taking this reasoning to its logical consequence, Zhou and Zou [110] argue in a theoretical study that a profit-based recommender system could influence a marketplace by leading sellers to strategically increase product prices to compete in recommendations, leading to a decrease in overall profitability.

On the Usefulness of Customer Recommendations

In contrast to previous studies, a different research perspective [88, 111, 90, 91, 92] finds that the usefulness of customer recommendations is directly proportional to the sales performance of the recommendation system. In fact, according to leading economic theories, a rational customer would choose products that maximize their utility. Based on this perspective, Wang and Zhang [88] develop a recommendation algorithm that maximizes the net marginal utility of recommended products for the customer by exploiting the economic principle of diminishing marginal utility. Yang et al. [111] propose an adaptive association rule mining algorithm to recommend the highest utility products. By contrast, Zhang et al. [90] design a recommendation system that jointly optimizes the interests of customers and sellers in an online marketplace. The system optimizes customer surplus, defined as the amount of utility that a customer obtains with respect to the price that he or she pays, and producer surplus, defined as the amount of revenue that a producer obtains after costs. Further developing previous approaches, Zhao et al. [91] propose maximizing the usefulness of recommendations based on the concept of the marginal rate of substitution. The algorithm considers the complementarity and substitutability of the products to be recommended to the customer compared to those already purchased. Finally, Ge et al. [92] aim to optimize the utility of recommended products by maximizing the marginal utility per dollar (MUD) under 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 [55] and Ghanem et al. [108], the performance of an RS also depends to a large extent on the long-term effects of recommendations on customers. 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 more stable levels of customer satisfaction. Hosanagar et al. [79] considered this factor by arguing 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. Additional work by Liu and Shih [120, 121, 119], Shih and Liu [122] and Tabaei and Fathian [123] proposed methodologies based on customer lifetime value (CLV), a popular metric in the marketing and management literature that measures the overall value that a customer generates for the organization throughout his or her history. However, previous studies focused on the use of CLV to improve the quality of recommendations rather than to optimize the long-term value for the organization. More recent works [114, 96, 15, 115, 100] have proposed directly optimizing the long-term performance of recommender systems by exploiting probabilistic approaches [114] or reinforcement learning [96, 15, 115, 100] algorithms (see Section 2.4.1). The latter have been used, for example, to maximize the cumulative value obtained via all user actions (i.e., click, add-to-cart, pay) [15] or to optimize customer lifetime value in cold-start scenarios [99].

Static vs. Dynamic Pricing

The majority of research on VARS is based on algorithms that keep prices static. However, a pioneering alternative approach is represented by systems that integrate recommendations with dynamic pricing algorithms [83, 112, 5, 82, 106]. According to this philosophy, Kamishima

Reference	Technique Used	optimized Value	Dataset
[51]	Value Ranking	Sales and Revenue	Foodmart
[69]	Value Ranking	Sales and Revenue	Foodmart
[79]	Value Policy	Sales and Revenue	N/A
[70]	Value Ranking	Sales and Revenue	Self-collected
[71]	Value Ranking	Sales and Revenue	N/A
[102]	Value Objective	Sales and Revenue	N/A
[83]	Value Policy	Sales and Revenue	MovieLens
[88]	Value Objective	User Engagement	Self-collected
[112]	Value Objective	Sales and Revenue	Self-collected
[106]	Value Objective	Sales and Revenue	Foodmart
[101]	Value Objective	Sales and Revenue	Self-collected
[80]	Value Policy	Sales and Revenue	Self-collected
[5]	Value Policy	Sales and Revenue	Self-collected
[113]	Value Policy	Sales and Revenue	Self-collected
[82]	Value Policy	Sales and Revenue	Self-collected
[56]	Value Policy	Sales and Revenue	Self-collected
[90]	Value Objective	User Engagement	Self-collected
[91]	Value Objective	User Engagement	Self-collected
[55]	Value Ranking	Sales and Revenue	MovieLens
[111]	Value Objective	User Engagement	Foodmart, Chain-Store, Amazon Review
[96]	Value Reinforcement	Sales and Revenue	Dunnhumby
[87]	Value Objective	Sales and Revenue	Book-Crossing
[114]	Value Objective	Sales and Revenue	MovieLens
[104]	Value Objective	Sales and Revenue	Self-collected
[105]	Value Objective	Sales and Revenue	SPMF/Retail
[92]	Value Objective	User Engagement	Amazon Review
[15]	Value Reinforcement	Sales and Revenue	REC-RL
[107]	Value Ranking	Sales and Revenue	Self-collected
[78]	Value Objective	User Engagement	EC-REC
[103]	Value Objective	User Engagement	JD
[115]	Value Reinforcement	Sales and Revenue	Self-collected
[53]	Value Ranking	Sales and Revenue	Self-collected
[116]	Value Policy	Sales and Revenue	Self-collected
[100]	Value Reinforcement	Sales and Revenue	Self-collected
[99]	Value Reinforcement	User Engagement	Self-collected
[109]	Value Policy	Sales and Revenue	N/A
[117]	Value Ranking	CTR	Self-Collected
[108]	Value Ranking	Sales and Revenue	MovieLens
[118]	Value Objective	Sales and Revenue	Self-Collected
[110]	Value Policy	Sales and Revenue	N/A

Table 2.2: Product value-aware recommender systems.

and Akaho propose a system that strategically adjusts the price of items recommended to customers through a discount 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 et al. [113], who designed a system that recommends products and simultaneously optimizes associated promotional discounts to maximize the total profit gain for the company. Instead, Jiang and Liu [112] optimize the discount of promotional products to increase the overall profitability of non-promotional products. The authors propose exploiting intra/cross-category effects of products purchased at a discounted price to stimulate customers to purchase non-discounted products. Additionally, regarding personalized promotions, Zhao et al., [5] propose customizing the discount of recommended products based on customer willingness to pay predictions, while Beladev et al. [82] propose recommending product bundles by pricing them to maximize the organization's revenue.

Real-World Studies

Some research has studied the performance of VARS in real-world environments. In particular, the model designed by Hosanagar et al. [79] has been used in many research works. The algorithm was designed according to the following assumptions: when a customer trusts an RS, the system biases the recommendations to increase profitability; when customer trust is below a certain threshold, the system recommends the most relevant products to restore trust at the expense of profitability. Some online studies [56, 116] used this algorithm to study the sales performance of a profit-based recommender system. In particular, Panniello et al. [56], in a randomised field experiment, showed that the Hosanagar et al. algorithm achieved higher revenue than that of a content-based algorithm without affecting the customer's trust in the organization. In another experiment, Basu [116] found that the relevance of recommendations and customer trust in the organization were positively correlated with the revenue generated from recommendations.

Similar results were reported by Kompan et al. [53] in a study based on a real-world e-commerce dataset in the fashion domain. Integrating product profit factors and customer price preferences into the algorithms could actually increase the profitability and, in some cases, even the accuracy of the recommendations. However, an excessive bias could lead to opposite effects. Moreover, as argued by Cavenaghi et al. [117], the price and rank position of a recommended product are two key factors that can influence CTR and other business value indicators.

Reference	Technique Used	optimized Value	Dataset
[95]	Value Reinforcement	User Engagement	Self-collected
[74]	Value Ranking	Sales and Revenue	Self-collected
[124]	Value Policy	Sales and Revenue	Package, NBA
[75]	Value Ranking	Sales and Revenue	Self-collected
[17]	Value Reinforcement	Sales and Revenue	MovieLens
[54]	Value Reinforcement	User Engagement	Self-collected
[8]	Value Reinforcement	Adoption	Self-collected
[81]	Value Policy	All Values	Amazon Review

Table 2.3: Advertising value-aware recommender systems.

Advertising Recommendation

Several value-aware systems have been proposed to optimize the value of advertising. In the following, we provide an overview of traditional systems in this field and recent perspectives that aim to optimize customer lifetime value.

Table 2.3 summarises the literature on value-aware advertising recommendation systems.

Traditional Advertising Strategies

In advertising systems, sponsored space is traditionally sold through auctions, where different advertisers compete for customers' attention. The systems often work as follows [125, 126, 8]: 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 [127, 128, 129, 130, 125, 131, 126, 132] investigates algorithms to predict the previous metrics as accurately as possible from the characteristics of the recommended items. Early work [127] by Microsoft proposed a Bayesian algorithm based on a probit regression model to predict CTR in a Microsoft Bing sponsored search. Subsequent work describes the ad systems of Google [128], Facebook [130] and Yahoo [129], as well as the algorithms used to estimate CTR. More recent approaches proposed by Etsy [125] and Alibaba [131, 126, 132] leverage ensemble learning models and neural networks, respectively, to make predictions by exploiting features associated with items (e.g., text and images), and customer purchase behaviour.

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 [124, 74, 75]. Indiscriminately promoting high-profit items that do not match users' interests could push users away from the system. Thus, to consider both the interests of the organization and the users, Zhang et al. [74] proposed a methodology to balance the revenue generated from the ads of an app store and the overall quality of the recommendations. The system optimizes, in the same objective function, the profit from downloading financially sponsored apps and the number of downloads of non-sponsored apps relevant to the user. Adopting a similar perspective: Long et al. [124] developed an algorithm that optimizes the overall profitability of a promotional campaign while maintaining a certain number of satisfied customers; Malthouse et al. [75] proposed a multi-stakeholder advertising system that jointly optimizes ad revenue and user utility. Considering the user's interests in recommendations would increase customer lifetime value and improve other drivers of business value. However, as noted by He et al. [81], maximizing multiple strategic parameters in the same objective function (e.g., click-through rate, user engagement) could lead to suboptimal results in individual indicators. Instead, considering that many indicators are interrelated, by adopting a probabilistic optimization methodology, optimizing one parameter at a time could have positive effects on other business values as well.

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 com-

Reference	Technique Used	optimized Value	Dataset
[133]	Value Reinforcement	CTR, User Engagement	Self-collected
[134]	Value Ranking	CTR, User Engagement	Self-collected
[16]	Value Reinforcement	User Engagement	Self-collected
[97]	Value Reinforcement	CTR	Self-collected
[135]	Value Ranking	User Engagement	Self-Collected
[98]	Value Reinforcement	CTR, User Engagement	Self-collected
[6]	Value Ranking	User Engagement	Self-collected
[136]	Value Ranking	CTR, User Engagement	Self-Collected

Table 2.4: News value-aware recommender systems.

pany may lose valuable sources of revenue. To address this problem, some works [95, 17, 54, 8] have studied how to optimize the performance of a long-term advertising system. Instead of recommending to customers ads that have the highest probability of being clicked, Theocharous et al. [95] and Han et al. [17] proposed leveraging reinforcement learning techniques to optimize customer lifetime value and, more generally, cumulative reward for the platform. Zhao et al. [54] further adaptedd 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, from a multi-stakeholder perspective, 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. [8] 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.

News Recommendation

Some value-aware recommenders have been proposed to optimize the value of news systems. 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.

Table 2.4 summarises the literature on value-aware news recommendation systems.

Conventional News Recommendation Strategies

The reputation of a news company is directly related to the impact of the information it provides on society (i.e., breaking news) [137]. The business model may be subscription-based, advertising-based, or both. Conventionally the number of clicks or views a given news item obtains during its overall lifespan is directly related to the organization's returns. As a result, traditionally, companies whose core business is sharing information in the form of news may be interested in generating higher profits by optimizing user interaction. Since the click-through rate (CTR) is directly related to an organization's revenue, a common goal is to maximize the number of clicks. Therefore, traditional news RSs [138, 139] use CTR as a primary indicator to feed probabilistic techniques 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-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 reading it. Consequently, a growing body of work [134, 135, 6] has considered the relationship between CTR and user engagement by proposing to optimize the latter. Besbes et al. [134] formulated a heuristic methodology that examines the probability of clicking on a news item and the engagement effect that it triggers. Specifically, they express the relationship between clicks (the likelihood of clicking on an article when recommended) and engagement (the probability of clicking on an article when it hosts a recommendation). Through this formulation, the news is proposed also considering the future navigation paths of the contents. Instead, Zihayat et al. [135] proposed a probabilistic methodology that simultaneously considers an article's recentness and user-article interaction (i.e., dwell time) to recommend news based on user

Reference	Technique Used	optimized Value	Dataset
[140]	Value Objective	Sales and Revenue	Self-collected
[141]	Value Objective	Sales and Revenue	Self-collected
[84]	Value Objective	Distribution	MovieLens
[72]	Value Ranking	Sales and Revenue	Self-collected
[85]	Value Objective	Distribution	MovieLens, Book-Crossing
[89]	Value Objective	Distribution	MovieLens
[86]	Value Objective	Distribution	Netflix Prize, Million Song
[142]	Value Ranking	Distribution	MovieLens, Netflix Prize, Jester
[143]	Value Ranking	Distribution	MovieLens, Netflix Prize
[144]	Value Ranking	Distribution	MovieLens, Netflix Prize
[7]	Value Policy	Sales and Revenue	Self-collected
[145]	Value Policy	Sales and Revenue	Self-collected

Table 2.5: Media value-aware recommender systems.

utility criteria. Moreover, as observed by Lu et al. [6] and Spyridou et al. [136], 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 to the user. Taking this into consideration, several works [16, 133, 97, 98] have proposed methodologies to optimize long-term metrics. For example, Wu et al. [16] 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. [133] to maximize the total number of user clicks. More advanced approaches based on reinforcement learning have been proposed by Zheng et al. and Zou et al. [97, 98] to optimize both CTR and long-term user engagement while considering the user's return pattern on the platform in addition to click information.

Media Recommendation

Some value-aware recommender systems have been designed to optimize the value of multimedia services. Below, we provide an overview of the main topics in the literature concerning the optimization of user engagement, the effects on the distribution of items with which the user interacts, and the resulting increase in sales.

Table 2.5 summarises the literature on value-aware media recommendation systems.

On The Effects of Optimizing User Engagement on Item 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 [7]. As with news systems, the main business 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 [146].

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 [47, 48, 49]. To keep users engaged, one of the main techniques is to optimize the distribution of recommended items (recall effects on distribution are part of the value taxonomy in Section 2.2.1) with the goal of helping the user discover surprisingly new and relevant items. This can be done, for example, by increasing the diversity [147] of recommendations [86, 143, 144] or promoting long-tail items [84, 85, 89] that tend to be proposed less by RSs because of popularity bias.

Optimizing Sales Revenue According to the Business Model

In addition to user engagement, research on media VARSs propose approaches to optimize other value indicators. Some works [72, 140, 141] 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 business model*) to have the opportunity to watch several movies in a given time frame [140, 72] or a fixed price (*on-demand business model*) for individual movies [141, 72].

The importance of the value-aware approach on the overall revenues of a movie provider based on an on-demand business model has also been studied in detail in two recent papers [145, 7]. In particular, according to Zhang et al. [145], recommendation systems that aim solely at profit optimization could produce negative effects on customer surplus (i.e., the price paid by the customer minus willingness to pay) and risk driving customers away from the company. Instead, according to Najafabadi et al. [7], personalizing pricing would allow the offer to be more tailored to the customer's willingness to pay and simultaneously create more profit for the sellers and surplus for the customers.

2.4.3 Datasets

In many studies, VARS have been trained and evaluated on public datasets. Unlike traditional datasets, the majority of the latter contain economic value information. Below, we present the main datasets used in the literature.

As shown in Table 2.6, most datasets are used to design product or media value-aware recommender systems. Some studies that proposed product VARS [51, 69, 111] have exploited the FoodMart dataset [148]. This is a Microsoft SQL Server 2000 sample database of a supermarket. The dataset contains 5,581 customers, 1,559 products, and 20,522 purchase transactions. In addition, master data about customers (e.g., country) and products (e.g., brand) are presented together with sales data (e.g., price, cost, profitability). Other studies on product VARS [80, 87, 111, 81] have exploited different datasets crawled from Amazon [159, 149] and Epinions [153]. 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). Furthermore, other works [96, 105, 111] have leveraged supermarket transaction datasets such as Dunnhumby [150], SPMF/Retail [151], and Chainstore [152], which contain customer, product, and purchase transaction information. Finally, other works [15, 78, 76, 103, 98] used e-commerce datasets such as EC-REC [78], REC-RL [15] and JD [103] that, together with customer, product, and pricing information, contain user-item interaction data (e.g., click, add-to-cart).

Other studies that investigated media VARS [55, 114, 81, 84, 85, 89, 142, 143, 144, 17] re-

Dataset	Domain	Content	Availability
FoodMart [148]	Product	Contains transaction data, product metadata and customer demographics of a supermarket chain	https://github.com/julianhyde/foodmart-data-hsqldb
Amazon Review [149]	Product	Contains product review data and metadata crawled from Amazon e-commerce site	https://nijianmo.github.io/amazon/index.html
JD [103]	Product	Contains data collected from the recommender systems logs of the JD Chinese e-commerce site	https://github.com/guyulongcs/CIKM2020_DMT
Dunnhumby [150]	Product	Contains transaction data from a subset of households that make frequent purchases from a retailer	https://www.dunnhumby.com/sourcefiles
SPMF/Retail [151]	Product	Contains customer transaction data from a Belgian retail store	https://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php
ChainStore [152]	Product	Contains transaction data and product metadata from a supermarket chain in California	http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html
EC-REC [78]	Product	Contains records of impressions, clicks and purchases from a large-scale e-commerce platform	https://drive.google.com/open?id=1rbidQksa_mLQz-V1d2X43WuUQQVa7P8H
REC-RL [15]	Product	Contains user interaction data collected from a real-world e-commerce platform	https://github.com/rec-agent/rec-rl
Epinions [153]	Product	Contains who-trust-whom online social network data from the Epinions consumer review site	https://snap.stanford.edu/data/soc-Epinions1.html
MovieLens [154]	Media	Contains movie ratings collected over various time periods from the MovieLens web site	https://grouplens.org/datasets/movielens/
Netflix Prize [155]	Media	Contains anonymous movie ratings from subscribers to the Netflix online movie rental service	https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data
Book-Crossing [156]	Media	Contains anonymised data of implicit/explicit book ratings from the Book-Crossing community	http://www2.informatik.uni-freiburg.de/~cziegler/BX/
Million Song [157]	Media	Contains audio features and metadata for over a million contemporary popular music tracks	http://millionsongdataset.com/
Jester [158]	Media	Contains anonymous ratings of jokes by users of the Jester Joke Recommender System	https://eigentaste.berkeley.edu/dataset/

Table 2.6: Datasets used in VARS literature.

lied on the well-known MovieLens dataset [154]. This is a very popular dataset that is used extensively in research on RSs. The dataset exists in several versions and contains movie rating data from the homonymous website. The largest version is MovieLens 25 M which contains 162,000 users, 62,000 movies, and 25 million ratings. Unlike the previous datasets, MovieLens does not explicitly contain product pricing data. Therefore, in several studies [55, 114, 81, 17], some methodologies based on probability calculations have been proposed to add this information to design VARS. Furthermore, other studies [84, 85, 89, 142, 143, 144] have used the dataset to design algorithms capable of optimizing product distributions without the need to add pricing information. Different research works on media VARS [84, 85, 86, 142, 143, 144] have adopted a similar philosophy and are based on famous datasets that do not contain pricing information, such as Netflix Prize [155], Book-Crossing [156], Million Song [157], and Jester [158].

2.5 Discussion

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.

2.5.1 Balancing Accuracy and Profitability

Early studies in the literature [51, 69] focused on optimizing a particular value driver of interest (e.g., CTR, sales, conversion rate). However, although biasing recommendations can in many cases improve some key business indicators, a system that always recommends irrelevant high-profit items could hurt the company's reputation by driving customers away [55, 53, 110]. To address this issue, many studies [79, 70, 71, 80, 134, 74, 124, 87, 75, 104, 105, 108] propose algorithms to determine the best trade-off between recommendation accuracy and value maximization. In fact, to avoid losing valuable sources of revenue for the organization, the RS must appropriately balance the interests of all key stakeholder groups. Although several studies have addressed this issue, the proposed algorithms are often based on assumptions about a particular type of industry (e.g., retail, entertainment, insurance) [7, 100] or business model (e.g., direct sales, subscriptions) [140, 141], and the methodologies are not always applicable in different contexts. Future research should study this problem in more detail by generalizing methodologies to propose algorithms suitable for each application context. Furthermore, optimizing a certain type of value (e.g., click-through rate) often affects other related indicators (e.g., sales) [81]. Studying more in-depth the correlations between the various indicators and their relationships with other business KPIs such as cash flows or inventory levels [73, 160], to optimize other types of value for organizations as well (e.g., reducing logistics delays, cost-to-serve or interest rates), could be another interesting research direction for the future.

2.5.2 On the Long-Term Perspective of Value Creation

To balance the interests of different parties, many algorithms have been proposed based on constrained optimization techniques. However, these algorithms often perform post-processing operations to optimize short-term performance without considering the long-term effects of recommendations [79, 55, 108]. Although widely used in the literature, this approach is risky because if a potential client notices that the recommendations are biased, they may lose trust in the organization and decide to purchase from competitors. To address this issue, reinforcement learning techniques have recently been proposed [95, 96, 16, 97, 17, 15, 98, 54, 8, 99, 100]. In this way, the recommendation can be modeled as a sequential decision problem in which an agent interacts with customers to maximize a cumulative reward for the organization. In general, we think that the use of reinforcement learning algorithms to optimize long-term recommendation performance will be highly valued in the next generation of VARS.

2.5.3 Dynamic Pricing for Value Optimization

Another important point to consider is the variability of certain information (e.g., price) that, together with recommendations, helps influence a customer's decision to purchase from an online service. To date, the literature on VARS has primarily studied how to optimize recommendations while keeping prices stable. Some specialized works [83, 112, 5, 82, 7, 117] have instead proposed further optimizing the sales process by integrating dynamic pricing algorithms into recommendations. In fact, the price of a product is one of the most important purchase drivers for a customer [117, 161]. Therefore, it is possible to act on this information as well to increase revenue and overall profitability for the organization. The study of the integration of dynamic pricing algorithms in value-aware systems is currently still in its infancy but could be a valuable future research direction.

2.5.4 Value-Aware Performance Evaluation

To evaluate the performance of VARS [12, 11, 162], some studies [56, 116] have performed online A/B tests. Specifically, given the non-deterministic nature of customer purchase choices, randomised controlled field tests are typically considered one of the few reliable performance evaluation methods. However, performing these tests is costly in terms of both time and money on the part of organizations: often, an A/B test can last several months if long-term aspects are to be evaluated and unexpected effects can sometimes occur, for example, due to particular world events that affect the results, making it necessary to rerun the test. 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 to evaluate performance, most studies on VARS [51, 69, 80, 76, 55, 111, 131, 87, 114, 78, 105, 15, 81] exploit offline approaches based on public datasets. However, the most popular public datasets [155, 154] often do not contain business information (e.g., prices, profits), making it difficult to measure the potential value generated by the recommender. Another important limitation is that it is often unclear under what circumstances the data were sampled. The results obtained by the algorithms could therefore be affected by bias, e.g., due to a particular promotion or certain population characteristics leading to erroneous conclusions. In addition, the results of studies in the literature are often not comparable because the authors measure offline value using ad hoc metrics or simulation systems that attempt to emulate the user's purchase choice. As a result, future research should address this issue to provide more reliable and sustainable performance evaluation methods.

2.5.5 Trustworthy Value-Aware Recommender Systems

Finally, like other AI-based systems, value-aware recommenders should be designed to respect important principles of AI trustworthiness [163], including alignment with human values, robustness and safety, privacy preservation, fairness [68], explainability [164] and transparency, reproducibility, and accountability. Studying each of these aspects in detail could be a profitable research direction. Investigating how to explain VARS recommendations without degrading business value or studying the reproducibility of major algorithms in the literature could provide interesting hints for future contributions.

2.6 Contribution

This chapter presented a systematic review of the latest developments in Value-Aware Recommender Systems. These systems have been strategically designed for seamless integration into commercial applications, aiming to enhance the economic value derived from recommendations. Value-Aware Recommender Systems find applications in diverse domains, including e-commerce for profit optimization, advertising platforms to enhance customer lifetime value, and online news services to maximize user engagement. The main algorithms within the literature of value-aware systems are based on post-processing techniques, the incorporation of objective functions from established recommendation algorithms, and the application of reinforcement learning methodologies.

However, if on the one hand, these systems provide key benefits for companies, on the other hand, they introduce new challenges such as appropriately balancing at the same time the interests of consumers, producers, and organizations while maintaining high recommendation performance in the short and long term. More in-depth, research is required to design higher-performing systems following recent trustworthy AI principles, effectively manage pricing information to optimize value, and improve offline and online performance evaluation methodologies.

However, while these systems offer substantial advantages to enterprises, they also introduce new challenges, such as appropriately balancing at the same time the interests of consumers, producers, and organizations while maintaining high recommendation performance in the short and long term. Furthermore, research is required to design higher-performing systems following recent trustworthy AI principles, effectively manage pricing information to optimize value, and improve offline and online performance evaluation methodologies.

Finally, these systems could be enhanced with the integration of explainability features that provide users insights into the suggestions of the models to make better and informed decisions. In the next chapter, we present explainability concepts and applications, proposing a novel RS model that aims to suggest items to business users, providing a balancing of value generated and explainability indices.

3

Explainable Recommender Systems and Business Value

A Recommender System (RS) is a type of information filtering algorithm that provides suggestions for objects related to a particular user. However, not all recommendation systems are transparent about how to generate their recommendations, which can affect multiple user criteria, such as trust and satisfaction. To address this problem, during the last decade researchers have defined a novel branch of study named Explainable Artificial Intelligence (XAI), in particular, they aim to address the problem of user comprehension. Explanations applied to RSs aim to address this problem by providing the reason, criteria, or evidence behind the model results. These explanations can help users understand, assess, and accept recommendations and improve decision-making processes.

From an industrial perspective, the potential for explanation is increasingly in demand, helping companies understand how AI models provide information and providing more insight to make decisions. The knowledge of the system that suggests which actions you should take is essential for the buildup of trust between clients, employees, and regulators. Indeed, companies that adopt best practices to allow explanation could understand the results and lead the profits, and it can also increase productivity by quickly revealing errors and areas for improvement.

Although the potential impact of these complex systems is challenging for enterprises, it has been affected by some limitations. One of the limitations that emerges from the current state-of-the-art is that the more complex an AI system becomes, the harder it is to precisely know how it obtained a particular insight, especially if the model aims to provide explanations targeted to specific users. In particular, when different kinds of users (consumers, business clients, modelers) interact with these models, they expect to solve a specific problem based on multiple explanation requirements. For example, a consumer rejected by a bank for a mortgage would probably want to understand the reason for this decision. Or, a marketing user wants to be assisted by an Explainable Recommender System while preparing a marketing campaign for a specific business target, motivating why he makes these decisions (i.e., compose the proper marketing proposal for a campaign). Due to the significant potential of these models, the balancing between business value and explainability will lead to future industrial research, helping companies for their target. To the best of our knowledge, no studies try to apply explanations to Value-Aware Recommender System (VARS).

In this chapter, we introduce the concept of the Explainable Recommender System (xRS), review the main techniques for creating and evaluating explanations, and present literature examples of applications for explaining recommendations (see section 3.1 and 3.2). Finally, we propose a novel model called Explainable Value-aware Matrix Factorization (XVMF) that is, to the best of our knowledge, the first tentative to create a model targeted to define a balance between explainability performance and business value generation, applying XAI concepts to VARSs.

3.1 Background and Concepts

Recommender Systems requires a proper evaluation analysis to be comparable and acceptable, defining metrics to facilitate the selection by stakeholders [165]. To enhance the motivations provided during the best-model selection phase, explainability can provide some useful insights and perspectives, simplifying the comprehension of the results and the acceptance from users. In this section, we report the main concepts that describe RSs evaluation metrics and, in particular, the explainability models state-of-the-art, focusing on techniques, performance indicators, and human aspects. This information represents the basis for the comprehension of the approach proposed in the chapter, and it extends the knowledge reported in the chapter 2.

3.1.1 Recommendation Evaluation Metrics

Recommender Systems operate in various contexts, ranging from e-commerce platforms to content recommendations in media streaming services. Each domain introduces distinct challenges and requirements. The evaluation of recommender system models is a critical aspect in assessing their effectiveness and suitability for diverse applications [166]. Due to the different nature of recommendation algorithms and the requirements of the specific domain of applications, an exhaustive evaluation process is essential to accurately determine the model's performance and empower stakeholders to make informed decisions, thus favoring the continuous advancement of recommendation system techniques. Hence, the choice of evaluation metrics becomes pivotal in reflecting the system's performance within its specific application. An evaluation process tailored to the application domain ensures that the recommender system aligns with user expectations, business goals, and the underlying data distribution. The comprehension of what metrics evaluate is crucial to understand better the need for a proper evaluation of the model developed, distinguishing between measuring both recommendation results and explainability performance [167].

Properties of Recommendation System

According to Herlocker et al. [168] and Gunawardana and Shani [169], several properties could define a recommender system model. These important characteristics are crucial during the selection of a solution to be implemented in an application or in a productive system. In the last decade, most of the researchers' effort has been on creating models that improve one or more properties at the same time, trying to find the best trade-off and avoiding degradation of the system performance. Variations of application, experiments (online and offline), dataset improvement, and user feedback, are some of the solutions provided by a lot of scientific works to overcome numerous challenges. In every implementation, the developers use metrics to compare the new models to the baseline. In table 3.1, we report the most used performance metrics among recommendation system models.

Accuracy Prediction accuracy is a fundamental aspect of recommender systems that has garnered significant attention within the recommendation system literature. Indeed, most recommender systems rely on a prediction engine responsible for forecasting user preferences or usage probabilities over items. The underlying assumption is that users will favour systems that provide more accurate predictions. This property comprises three primary categories: the accuracy of rating predictions, the accuracy of usage predictions, and the accuracy of rankings of items.

Evaluating the accuracy of rating prediction is the main objective in scenarios interested in user ratings (i.e., movie ratings). The main metrics used in this context are Root Mean Squared

Property	Focus	Metric	Formula
Accuracy	Rating	RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{u,i}^{m,n} \lvert\hat{y}_{ui} - y_{ui}\rvert^2}$
		MAE	$MAE = \frac{1}{N}\sum_{u,i}^{m,n} \lvert\hat{y}_{ui} - y_{ui}\rvert$
	Usage	Precision	$Precision = \frac{TP}{TP+FP}$
		Recall	$Recall = \frac{TP}{TP+FN}$
		F-measure	$F\text{-}measure = \frac{2(Precision*Recall}{Precision+Recall}$
	Ranking	NDCG	$NDCG = \frac{\sum_{i=1}^{n} \frac{2^{rel_i}-1}{\log_2(i+1)}}{\sum_{i=1}^{n_{max}} \frac{2^{rel_i}-1}{\log_2(i+1)}}$
		Spearman's Correlation	$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$
		NDPM	$NDPM = \frac{DP}{IDP}$
		Pearson Correlation	$\rho(x,y) = \frac{E[x,y]}{\sigma_x\sigma_y}$
Coverage	Item space	Gini Index	$G = 1 - \frac{2}{n(n-1)}\sum_{i=1}^{n}(n-i)\cdot R_i$
		Shannon Entropy	$H = -\sum_{i=1}^{n} p_i \log_2(p_i)$

Table 3.1: Most used evaluation metrics for Recommender Systems.

Error (RMSE) and Mean Absolute Error (MAE)[168]. RMSE quantifies the square root of the average squared difference between predicted and actual ratings, while MAE calculates the average absolute difference. Additionally, two metrics named Normalized RMSE (NRMSE) and Normalized MAE (NMAE) are the normalization version of, respectively, RMSE and MAE by the range of ratings.

On the contrary, scenarios where is important to predict usage behavior (i.e., item selection by user) are supported by metrics like Precision, Recall, and F-measure [169]. These metrics help the evaluation of the accuracy of predicting interactions in recommender systems. In particular, precision assesses the proportion of correctly predicted selections among all predicted ones, while Recall measures the proportion of correctly predicted selections among all actual ones. In order to summarize the performance of the two measures, F-measure can be used to evaluate the balance between precision and recall providing a single score.

Recommendation system output is commonly presented as a list of items in a ranked order that must be evaluated by their position to asses the final accuracy of the model. Ranking metrics include metrics like Normalized Cumulative Discounted Gain (NDCG), Normalized Distance-based Performance Measure (NDPM), Spearman's correlation, and Pearson Correlation [169]. NDCG evaluates ranking by assessing the cumulative discounted gain of the recommended items compared to an ideal ranking. It is calculated using Discounted Cumulative Gain (DCG) normalized by the ideal DCG. On the other hand, NDPM quantifies the performance of ranking items by considering pairs of items and their ranking orders.

Coverage Coverage focuses on the proportion of the items that the system can analyze and produce a recommendation, giving the stakeholders insights into the diversity and completeness of the results [4, 170]. A high coverage value indicates that the system can suggest the majority of the items, covers multiple user preferences, and prevents the risk of underestimating the real user behavior avoiding limitation on the subset of items considered. At the same time, keeping attention to the coverage level could introduce users to novel and undiscovered content.

For evaluating coverage the two main metrics, Gini Index and Shannon Entropy, offer different perspectives to quantify the distribution of recommendations across items [169]. The Gini Index, commonly employed in the economic domain to measure wealth distribution, can also be applied to recommender systems to quantify the distribution of recommended items. A low Gini Index means a more equally distributed set of recommendations, and it corresponds to a higher coverage score. On the contrary, Shannon Entropy is a metric borrowed from information theory and it assesses an uncertainty or unpredictability distribution in the recommendation model, where a higher value of Shannon Entropy indicates a more diverse recommendation (better coverage). Both metrics help assess the diversity and distribution of recommended elements, ensuring that a wide range of options is presented to users, and allowing evaluators to evaluate the system's ability to introduce users to a wide range of options.

Finally, coverage evaluation guarantees not only to respect the diversity during the recommendation process but also to ensure the discovery capabilities of the system which is essential in real-world scenarios, especially in enterprise strategies [101].

Others Properties In addition to the previous main properties of accuracy and coverage, the evaluation of recommender systems is based on a range of other dimensions that are equally useful in shaping the user experience and enhancing the utility of the recommendations provided. These properties provide valuable performance insights across various perspectives, enabling researchers and stakeholders to make informed decisions and improvements. In the following list, we report the properties that complete the evaluation framework for designing and optimizing recommender systems [168]:

- *Confidence* concerns the system's ability to quantify the certainty of predictions, offering insight into the reliability of its recommendations;

- *Trust* involves the user's perspective in the recommendations, ensuring that users feel secure and informed in their decision-making phase;
- *Novelty* and *Serendipity* emphasize the system's capability to introduce users to unexplored and unexpected content, enriching their experience of discovering novel products;
- *Diversity* measures the system's ability to suggest a diverse range of items, responding to various user preferences and avoiding emphasizing the most popular items;
- *Utility* scores the relevance and value of recommendations according to users' needs and preferences, enhancing user satisfaction and engagement;
- *Risk* assesses the potential drawbacks and consequences of recommendations, ensuring that system suggestions do not inadvertently expose users to undesirable results;
- *Robustness* examines the system's stability and performance under multiple conditions, maintaining its reliability across different scenarios;
- *Privacy* aims to assess the safeguarding of user information while allowing for personalized recommendations and without compromising user confidentiality;
- *Adaptivity* focuses on the system's capability to evolve and adjust recommendations, based on user interactions and ensure the relevance of suggestions over time;
- *Scalability* targets to the system's efficiency and capability to handle complexity (larger datasets and user bases). This property is essential in enterprise adoption of AI strategies where the amount of data constantly increases.

3.1.2 Explainability on RS

Explainable Recommender Systems (xRSs) have become an important area of research in recent years due to their potential to improve user trust and satisfaction with recommendation systems [171], facilitate system designers for better system debugging [166], and helps users facilitate the access to resources or relevant items in an information-overloaded environment [172]. There are several approaches to improving explainability in recommender systems. One approach is to use post-hoc explainability methods that provide explanations after a recommendation has been made. Another approach is to use model-based explainability methods that provide explanations based on the internal working methods of the recommendation model [166].

xRSs attempts to develop models that generate high-quality recommendations and intuitive explanations. The explanations may either be post-hoc or directly come from an explainable model (also called interpretable or transparent model in some contexts) [166]. There are several types of models used in explainable recommendations. One main type is *rule-based models*, which use a set of rules to make recommendations and provide explanations based on those rules. Another most-used family is *matrix factorization models*, which use techniques to learn latent factors representing user preferences and item attributes. These models can provide explanations based on the learned factors [173]. Another type of model is the *decision tree-based model*, which uses decision trees to make recommendations and provide explanations based on the decision paths taken by the tree. These models can be used for both classification and regression tasks [174]. Finally, there are *neural network-based models*, which use neural networks to learn representations of users and items and can provide explanations based on these learned entities [175].

With the diffusion of the explainable model, it was fundamental to define a way to make them comparable and based on similar evaluation criteria. One of the most significant works aimed to solve this need is from Tintarev and Masthoff [165], where the authors formalize the guideline that every Explainable Recommender System should be based on. In the work, the first guideline concerns the requirement of the target definition for the xRS, choosing from transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, or satisfaction. The authors continue reminding developers and modelers not to confuse the *evaluation methods* of the explainability with those of RS, which are close but different (this is the focal point of the chapter 4 in this thesis). In addition, *presentation* and *interaction* with explanations are factors that affect the achievement degree of the goal. Finally, explanation *styles* classify how the explanations are generated and whether they are useful for achieving the model goal despite the presence or absence of a recommendation model under the hood.

Numerous other works have refined these concepts by highlighting perspectives and applications: for instance, Mohseni et al. [176], Adadi and Berrada [177] released surveys that collect evaluation methods for other algorithms family of intelligent systems, but not for RSs. Moreover, Zhang and Chen [166] collects and summarizes all state-of-the-art applications of Explainable Recommender Systems. Finally, Vultureanu-Albişi and Bădică [178], Chen et al. [172] provided novel and updated perspective through their surveys on xRS, specializing the primary work of Tintarev and Masthoff [165].

Style	**Example**
User	Customers who bought item A also bought items B, C.
Item	Item A is recommended because you highly rated or bought item B, C.
Feature	You might be interested in this item because you are looking for feature A (word based).
Popularity	We suggest items A, B, since they are very popular among people similar to you.
Social	A and B also like the item C.
Pairwise personalized	We guess you would like item A more than B because you may prefer C over D.
Visual	Explanation provided by an image

Table 3.2: List of the explanation styles

Explanation styles One of the aspects concerning the implementation of an explainable model regards the presentation of the results. Explanations could be shown using different styles to maximize the desired effect. In the literature, several styles are proposed in order to define and enhance a specific perspective. Tintarev and Masthoff [165], for instance, collect and define six types of style: case-based, collaborative-based [179], content-based [180], conversational [181], demographic-based, and knowledge-based [182].

Another perspective that aims to improve these style categories was provided by Papadimitriou et al. [183] and summarized by Zhang and Chen [166]. Different from the Tintarev and Masthoff approach that distinguishes between styles and evaluation goals, the proposal from Papadimitriou et al. could give a second perspective on the styles used for an explanation. The authors create an entity-relation classification that suggests eight kinds of groups: user, item, feature, popularity, social, pairwise personalized, sentence-level, and visual (see table 3.2). The *user-based* model provides explanations based on the user's preferences and interests, and it is useful for building personalized explanations that are tailored to the consumer's needs. *Item-based* approach, on the contrary, gives explanations based on the attributes of the recommended item, focusing on the item's features and characteristics. The third main style is *feature-based*, which focuses on how the model internally works and why it suggests an item to a user, exposing internal features of the recommendation model. Finally, *hybrid* explanations aim to combine two or more styles to provide more comprehensive and informative explanations.

Presentation, Display Incorporating presentation consideration when an explanation model provides insight to users is essential to enhance the reader experience and ensure the effectiveness of the suggestions. Moreover, there are two approaches that model developers could follow to influence user comprehension and acceptance: defining the appearance and the presentation structure.

Firstly, considering the appearance of explanations is essential as it directly impacts how users perceive the information. It could be performed as natural language text (a recent promising research branch), graphical visualization [179], or based on a presentation template [184] such as a structured form to be filled with the information.

Moreover, the design of the structure is followed by companies with primary interest because business users should interact with the explanations during the productive phase, leaving important feedback (implicit or explicit) in the system. It regards the design of a user interface and user experience that aims for consistency and empowers users to seek clarification or explore alternatives based on preferences. The implementation of a user interface represents the solution to integrate the results of these AI models into the selection process, especially in the enterprise context where insights lead to better decisions for the business.

Personalization Personalization of explanations defines who is the target user. When a recommendation proposes an item to a different user, it could be explained in different ways. Herlocker et al. [179] affirm that non-personalized ratings are more persuasive than personalized from users' neighbors in preference space. On the contrary, Vig et al. [180] design tag-based explanations to improve the personalized user experience but imply a negative effect on different designs. In addition, Guesmi et al. [185] with his work developed a transparent model called *Recommendation and Interest Modeling Application* (RIMA) that aims to personalize explanations with different detail levels targeting the results on specific kinds of end-users. Finally, Tintarev and Masthoff [184] assert that the positive or negative effects of an explanation depend on the design chosen for the model.

Explanation focus Explanations can also be classified based on the part of the recommendation process they are trying to explain: recommendation input, recommendation process, recommendation output [186, 187]. Explanations applied to the input part of the recommendation system expose the system's hypothesis about the user's interests, preferences, or needs, analyzing the input information to drive awareness of personal tastes or detect misunderstandings made by the system. Conversely, focusing on output explanations provides a justification for a particular recommendation without revealing internal elements and logic (black-box mode). Finally, the most challenging approach is to understand better what internal processes ensure the production of a specific suggestion by the recommendation model. The explainability applied to the internal logic provides the decisions the algorithm made to produce a specific item list. However, due to the complexity of the algorithm, motivating the internal process is a challenging task.

3.1.3 Explainability Evaluation Process

Evaluating the explanations produced by a model is one of the main challenges in the literature. An explanation model must first address the task of evaluating the results obtained, not only as recommendation items but also by assessing the intrinsic explanatory capacity of the model[165]. Although some studies aim at objectivity, the evaluation must recognize the intrinsic relationship between articles and users, Measuring an explainability model, especially an xRS model, numerically, requires time, assumptions, and applications. Evaluation of recommendation explanations is particularly challenging, according to Chen et al. [172] because obtaining a ground truth is difficult, and it is complex to incorporate objective human emotions into a measurement algorithm.

The evaluation aims to define quantitative metrics that quantify the system performance [188], such as accuracy or coverage. Basically, it uses the same measures for evaluating conventional RS adapting the definition of the relative metric to the new context. Based on the type of experiment chosen, the evaluation could be grouped into three sets: user studies, online evaluations, and offline evaluations. *User studies* require explicit feedback on the explanation quality but it is difficult to gather due to the large amount of users and their specific differences. In this scenario, users are conscious of the study they are subjected to. On the contrary, *online evaluation* helps the model to understand the implicit feedback that users give during the gathering phase. Realistically, it should be implemented in a system with a large number of users (registered to the system). An example of this kind of evaluation could use the Click-through rate (CTR) to collect user feedback. Finally, the *offline evaluation* doesn't require direct interaction with the users and provides a score of the number of recommendations that could be explained by the model (regardless of the quality of the recommendation). Although this method lacks direct user feedback, it has the advantage that the model could directly evaluate the quality of the outcome results, depending on the format and type, and could be compared to other models.

The idea based on the role of humans in the explanations evaluation is argued by Ehsan and Riedl [189] defining the concept of Human-Centered XAI, such as a mixture of human interaction, cognitive psychology, and interpretable outcome, offered by the intelligent system. Furthermore, the quality of the outcome can only be interpreted by users involving usability studies on human subjects [188]. To perform a proper user evaluation and prepare the interaction between humans and systems, some stakeholders should participate in the primary definition phase of the model as end-users (humans that are AI and data novices), and experts (humans that are aware of data and AI) [190]. The feedback collected has the same importance value as the rating for the traditional RS [191].

The work of Mohseni et al. [176] expands the Human-Centered evaluation perspective, exploring how they are perceived by users by cognition, interaction, satisfaction, and trust. They examine the cognitive aspects of the explainability model, describing how users comprehend the system's capabilities and interact with the model results via output and failure prediction ratings from a cognitive psychology perspective. Moreover, Tsai and Brusilovsky [192] contribute with a work about the extraction of the feedback from the text provided by users during the approval of the recommendation. In addition, Dominguez et al. [193] consider the RS interactivity to better suit users' needs. Furthermore, Purificato et al. [194] combine different explainable interfaces to provide proper explanations to users during the interaction phase, aiming to asses different displaying results in the most comprehensible structure. Most of the studies focus on user satisfaction, extracting the level of understandability, fairness, usefulness, and sufficiency of details through interviews, self-reports, and questionnaires [195, 196, 197]. Finally, trust is an attribute that influences the system in a positive or negative way, and it requires time to be gained by continuous interaction with users. It is measured by asking users' opinions through interviews or questionnaires, and it could observed indirectly by examining personality traits [198].

3.1.4 Human aspects on the importance of explanations

The concept of explainability can be analyzed under different motivations, as proposed by Adadi and Berrada [177] in a recent study. In their study, the authors explore the motivations behind the study of Explainable Artificial Intelligence by defining four different needs: to justify, to control, to improve, and to discover. When users need an explanation of the model's results to make an informed decision, the system should be capable of *justifying* how a particular result was achieved rather than merely demonstrating the methods or reasoning steps taken to complete the process. Users of an explanation-capable recommendation system desire to receive a justification of the suggested proposal with regard to the process that led to its development. Providing explanations of this nature not only enhances the user's confidence in the system but also improves the system's reliability. This enables the user to acquire a more profound understanding of the system. *Control* is particularly useful for developers because they can easily identify and solve errors, gain more visibility into unknown vulnerabilities and problems, and learn more about system behavior. RSs assist in more effective interaction between users and systems by enabling better future action selection based on recommendations and influencing user behavior and system outcomes. The explanations have both the objective of orienting the administration of system behavior as well as informing the user. The *improvement* of intelligent systems is constantly sought, and the need to constantly improve recommendation systems is one of the reasons why, in this thesis, we study the integration of explainability. If users understand why certain recommendations were made, they will be able to make better decisions. Even if people cannot explain their decisions, the system must be able to support the new information it has provided to those affected. According to this understanding, explanation becomes a useful tool for *discovering* new elements provided by AI.

In addition, Miller [199] discusses the social significance of Explainable Artificial Intelligence and identifies the *sender* and *receiver* as the two main profiles in the explanation interaction. Despite lacking agreement on the definitions and explanations in scientific research, Miller presents three processes involved in explanation scenarios: cognitive, product, and social. The first process aims to identify a possible cause of an event and subsequently provide an explanation based on it. The last focuses on the human context required to interpret the explanation in terms of expectations, skills, and tasks, aiming to transmit the least amount of information from the explainer to the explained.

Furthermore, van Engers and de Vries [200] propose a framework to evaluate the explanation produced by a XAI model based on six quality criteria: *external coherence* (the probability of acceptance of the decision increases when explanations are compatible with the user knowledge of a particular context), *internal coherence* (sentences must be logically connected to improve the explanation and the understanding), *simplicity* (few and simple causes are easier to understand rather than complex sentences), *articulation* (quality of explanations is influenced by the length or words number contained in a statement), *contrastiveness* (a property that expresses the clarity provided by two consecutive events or in certain conditions), and *interaction* (related to human-computer interaction concepts, it expresses how an explanation could be customized interactively by the user).

Moreover, Sovrano et al. [201] propose SAGE model as a discovery criteria to determine a good path in the explanatory space. It is based on the principles of *sourced* (the references that lead the explanation are the input of the system), *adaptation* (the ability to create an argumentative framework during the exploration), *grounding* (showing the source in the original format, used for counterfactuals), and *expansion* (improving the explanation with the proper information retrieved).

Finally, Barredo Arrieta et al. [202] aims to clarify the relationship between interpretability (oriented to generate meaning for human beings) and explainability, in terms of an interface between people and the decision-maker. Explanations are classified as global or local, where global offers complete transparent explanations and local explains aspects of the reasoning process [203, 204].

3.2 Related Work

The idea of explaining a model is powerful, but it leads to a complexity that numerous researchers solved by providing different approaches and models. In this section, we report the criteria, the methods, and the limitations regarding the evaluation of Explainable Recommender Systems. Since explainability is important for the business in order to give the best solutions as an answer to a specific business requirement, we refer to chapter 2 for business value concepts applied to RSs.

3.2.1 Criteria for Explainable RSs evaluation

Among the literature, multiple approaches are proposed to evaluate the explainability of a system. The evaluation elements can be found under different titles as evaluation characteristics, benefits, goals, metrics, or perspective [184, 172]. However, according to the guidelines proposed by Tintarev and Masthoff [165], the evaluation could be properly defined as a set of *criteria* and targeted one or more stakeholders (*model designers*, *users*, and *providers*). These criteria aim to standardize the method based on the objective of the algorithms: transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction.

- **Transparency** is one of the most significant goals for an Explainable Recommender System, which directly affects some related aspects of user satisfaction: accessibility, usability, information, understanding, and auditability. Since the lack of comprehension could cause a major misunderstanding for users regarding the objectives of recommender systems [205], a model should provide transparent explanations to reveal the main advantages and disadvantages provided by the systems (embedding explanations) [206]. *Model designers* are the target stakeholders since explanations can reveal the internal working principles of the recommender models.

- **Scrutability** is closely linked to transparency. Although transparency pays more attention to details inside the model in order to reveal the working mechanisms, on the other hand, scrutability is more interested in the relationship between explanations and outputs, and it can be applied properly to an agnostic model. Moreover, explanations based on scrutability criteria can support both the user's accurate argumentation and make the system valid. Indeed, in a scrutable system (enabled by explanations), users can comprehend the motivation on the basis of algorithmic results, and then they can control the generated recommendation results by correcting the system's assumptions. Finally, the target of scrutability criteria is *model designers*.

- **Trust** increases the user confidence in the system and is rigidly related to transparency. If users understand why a particular recommendation was made, they will have more confidence in the recommendation system. *Users* are the criteria target.

- **Effectiveness** targets the *users* helping them to make quality decisions. Effective explanations will allow the user to assess the quality of the proposed items more accurately on the basis of their preferences. Tintarev and Masthoff [184] provide an overview of evaluating the effectiveness and the metrics used in particular in the recommendation domain.

- **Persuasiveness** aims to convince users to buy or try one or more items. In order to persuade users to buy more goods (i.e., to convince them), the explanations cannot accurately reveal the operational mechanism of the model (i.e., transparency), but must highlight the features that can promote sales. The final objective is to gain benefits for the system rather than for the user. Explanations may increase the user acceptance of the system, as well as of the provided recommendations. The target of this criteria is the *provider* of the system that receives an improvement from the application of the model.

- Finally, there are two more criteria, both targeted to *users*. **Efficiency** helps the user to decide faster, giving them the most suitable suggestions for the item recommended. Moreover, the explanation can improve user **satisfaction** and make the use of the recommendation system more pleasant. When assessing the satisfaction of explanations, it is important to distinguish between the satisfaction of the recommendation process and the satisfaction of the recommended items.

3.2.2 Methods for Explainable RSs evaluation

After the definition of criteria that Explainable Recommender Systems should be targeted to, we need to define how to evaluate the system and the objective to measure. These evaluation methods are grouped into four chunks: *qualitative*, *quantitative*, *crowdsourcing*, and *online experiments*.

Qualitative evaluations consist of case studies designed by model designers and provide comprehension of intuitive aspects. Although the objective is improving effectiveness and transparency criteria, they are affected by bias and by the absence of not comparable features [172]. We discuss this limitation deeply in chapter 4, providing symptoms of most of the actual Graph-Based Explainable Recommender System, and possible solutions.

Quantitative evaluations aim to evaluate offline models through the use of specific metrics, ensuring effectiveness and scrutability criteria [172]. Using metrics as a source of evaluation offers model developers advantages during the phase of comparing the developed solution with those belonging to the state-of-the-art. At the same time, the use of objective metrics facilitates the selection phase by corporate suppliers interested in implementing these solutions in a production environment. The advantages of these methods are efficiency at the comparison stage (i.e., ensuring that model performance is above a threshold), quantitative aspects, and ease of benchmarking. In contrast, measuring explanations without appropriate metrics (i.e., using an unsuitable metric proposed to solve a non-similar application) could deviate from the primary goals of the explanation and produce less effective results due to approximation. We collect the numerous metrics defined in the state-of-the-art in table 3.3 (see section 3.2.3).

Different from qualitative evaluation based on the use cases, **crowdsourcing** implies the active presence of a heterogeneous study population that extracts real human feeling. Since this method would put the right attention to the user interaction with the system, studying deeply causes and effects on humans, it demands a significant cost for the experimentation set. The issue of cost is significant in an industrial environment, where the difference between accepting or rejecting a project lies precisely in the time and cost involved versus the return on investment that can be achieved. The criteria addressed are effectiveness and scrutability (as in quantitative evaluation), transparency (as in qualitative evaluation), and persuasiveness. It can also be performed on public datasets, constructed datasets or annotation injections into public datasets [172].

Similarly to crowdsourcing, **online experiments** aim to collect user feedback in real-time, producing a highly reliable result at a significant cost of the experiment, although the crite-

Metrics	Evaluation Objective	References
BLEU	user review words	[209]
ROUGE	user review words	[210]
Unique Sentence Ratio (USR)	user review diversity	[211]
Feature Coverage Ratio (FCR)	user review diversity	[211]
Feature Diversity (FD)	user review diversity	[211]
PN, PS	necessity, sufficiency of a feature	[208]
MEP, MER, EP	user-item explainability	[173]
Model Fidelity	user-item explainability	[212]
E-NDCG	user-item explainability	[213]
NR^2	user-item explainability	[214]
LIR, LID, SEP, SED, PTD, PTC	recency, diversity, popularity	[215, 216]

Table 3.3: Most significant quantitative metrics to evaluate the quality of explanations.

ria aim to enhance effectiveness and persuasiveness. The explorations of the user acceptance affected by the explanations (i.e., using A/B tests, evaluating the Click-through rate and the conversion rate), are explored by Zhang et al. [207].

3.2.3 Metrics for Explainable RSs evaluation

In literature, some works proposed metrics to evaluate different quantitative perspectives of the model's explainability capabilities. One point to consider is distinguishing between assessing the model-generated explanation performance and the recommendation outcomes (see chapter 4). Regarding explainable evaluation, authors defined metrics to count words (i.e., BLEU and ROUGE index evaluate natural language explanations comparing the results with user reviews), to evaluate diversity (i.e., Unique Sentence Ratio (USR), Feature Coverage Ratio (FCR), Feature Diversity (FD)), and to predict features (i.e., Feature-level Precision (FP), Recall (FR), and F1 score (FF)). Furthermore, Tan et al. [208] suggest two indexes named Probability of Necessity (PN) and Probability of Sufficiency (PS), that estimate the necessity and sufficiency of a feature if the item is, respectively, removed or still in the suggestions list.

Explainable Matrix Factorization Researchers studied different recommendation system algorithm classes (i.e., Content-Based, Collaborative Filtering, Knowledge-Based, Hybrid, and Neural), identifying and studying advantages, methods, and challenges. Alongside the classic approaches built on the decision tree, rule-based, or Bayesian algorithms, Matrix Factorization (MF) is a family of model-based methods applied to collaborative filtering RS that primarily has the advantage of producing latent features that discern the connection between article entities and users.

One of the most interesting works that aims to embed explainable capabilities into a matrix factorization-based recommender system is proposed by Abdollahi and Nasraoui [173], where the authors optimize the performance of the models from two perspectives: recommendations system results and the explainable capabilities to the users. Indeed, Abdollahi and Nasraoui propose to measure the performance of the explanations generated by the algorithm using a quantitative approach, enhancing the evaluation of the RS model with a new user-centric perspective. The two proposed indices, Mean Explainability Precision (MEP) and Mean Explainability Recall (MER), are based on the capability of the system to explain a suggestion using the neighborhood of a user (under specific constraints). Both measures are the mean across all testing users, respectively, of Explainability Precision (EP) and Explainability Recall (ER).

The precision EP is calculated as the ratio between the number of explainable items in the top-n recommendation list and the total recommended items for each user:

$$MEP = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i : i \in L_u, E_{ui} > 0\}|}{N} \tag{3.1}$$

where U is the set of all users, i is an item of the recommendation list L_u (provided to user u), and N is the number of recommended items.

Additionally, to calculate ER, it takes the ratio between the number of explainable items in the top-n recommendation list and the total number of explainable items for given users [173]:

$$MER = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i : i \in L_u, E_{ui} > 0\}|}{W} \tag{3.2}$$

where U is the set of all users, i is an item of the recommendation list L_u (provided to user u), and W is the number of explainable items.

To implement this concept into the basic matrix factorization objective function, the authors add an additional regularization term named Explainability Power E_{ij} as a measure of how many nearest neighbors for a user (with a similar profile) highly rated an item, constructing a matrix E that expose the explainability power calculated with the following formula:

$$E_{u,i} = \sum_{\forall r \in R, r \geq P_\tau} r \times |NN^k(u)_{i,r}| \tag{3.3}$$

where i is an item for a specific user u, r is the rating given by user u to item i, and $NN^k(u)_{i,r}$ is the k entity nearest neighbours of user u [213].

Moreover, the standard matrix factorization objective function is enhanced with the new regularization term, as formulated in:

$$G_{expl} = \sum_{u,i \in R} (r_{ui} - a_u b_i^T)^2 + \frac{\beta}{2}(\|a_u\|^2 + \|b_i\|^2) + \lambda \|a_u - b_i\|^2 E_{ui} \tag{3.4}$$

where i is the item rated by user u, R is the set of ratings r_{ui}, a_u and b_i are the two terms to regularize the contribution of user u and item i to the objective function, and E_{ij} contains the information of the nearest neighbors (with similar profile) for a user u.

Then, to minimize the objective function G_{expl} (see Eq.3.4), authors used the stochastic gradient descent approach applied to the two following rules:

$$a'_u = a_u + \eta \left(2(r_{ui} - a_u b_i^T) b_i - \beta a_u - \lambda (a_u - b_i) E_{ui}\right) \tag{3.5}$$

$$b'_i = b_i + \eta \left(2(r_{ui} - a_u b_i^T) a_u - \beta b_i + \lambda (a_u - b_i) E_{ui}\right) \tag{3.6}$$

where η is the learning rate retrieved from the minimization computation of the gradient descendant, β controls the magnitude of the latent vectors (user and item), and λ aims to control the explanation matrix impact.

Finally, the model proposed by Abdollahi and Nasraoui was generalized by Peake and Wang [212] proposing the model *Fidelity score* as the percentage of explainable items in the recommendation list.

Explainable-NDCG Relying on the work by Abdollahi and Nasraoui, Coba et al. [213] introduce an enhanced model named Novel EMF (NEMF) that aims to find the trade-off between the novelty of an item and the explainability capability. In addition, they noticed that the MEP metric assigns inaccurate ranking positions of the explainable and recommendation items, proposing a new metric named Explainable Normalized Discounted Cumulative Gain (E-NDCG), based on NDCG and able to precisely identify explainable items.

Starting from the NDCG formula 3.7, the authors consider the position of the recommended item L_u (see Eq. 3.1) during the evaluation of the explainability power for user u and item i.

If NDCG is formalized with:

$$NDCG = \frac{DCG_u}{IDCG} \tag{3.7}$$

the novel E-NDCG proposed by Coba et al. is defined by:

$$E\text{–}NDCG = \frac{E\text{–}DCG_u}{E\text{–}IDCG} \tag{3.8}$$

The two terms of the ratio are defined by the following equations:

$$E\text{–}DCG_u = E_{u,l_1} + \sum_{i=1}^{N} \frac{E_{u,l_i}}{\log_2(i)} \tag{3.9}$$

$$E\text{–}IDCG = E_{max} + \sum_{i=1}^{N} \frac{E_{max}}{\log_2(i)} \tag{3.10}$$

where l is an item of the list L_u, and $E_{u,l}$ is the explainability power, while E_{max} is the explainability power considering all recommended items in the list L_u.

Considering the two distinct objective functions aimed to maximize explanation and novel:

$$G_{expl} = \sum_{u,i \in R} (r_{ui} - a_u b_i^T)^2 + \frac{\beta}{2}(\|a_u\|^2 + \|b_i\|^2) + \lambda \|a_u - b_i\| E_{ui} \tag{3.11}$$

$$G_{novel} = \sum_{u,i \in R} (r_{ui} - a_u b_i^T)^2 + \frac{\beta}{2}(\|a_u\|^2 + \|b_i\|^2) + \theta \|a_u - b_i\| N_{ui} \tag{3.12}$$

Coba et al. formalized the objective function that describes the Novel Explainable Matrix Factorization (NEMF) model as follows:

$$G_{expl\&novel} = \sum_{u,i \in R} (r_{ui} - a_u b_i^T)^2 + \frac{\beta}{2}(\|a_u\|^2 + \|b_i\|^2) + \|a_u - b_i\|(\lambda E_{ui} + \theta N_{ui}) \tag{3.13}$$

where the final results depend on both explainable and novel elements traded off of the two hyperparameters (λ and θ), and a_u and b_i components are formulated as:

$$a'_u = a_u + \eta \left(2(r_{ui} - a_u b_i^T) b_i - \beta a_u - sgn(a_u - b_i)(\lambda E_{ui} + \theta N_{ui})\right) \tag{3.14}$$

$$b'_i = b_i + \eta \left(2(r_{ui} - a_u b_i^T) a_u - \beta b_i + sgn(a_u - b_i)(\lambda E_{ui} + \theta N_{ui})\right) \tag{3.15}$$

3.3 PROPOSED APPROACH

Companies increasingly rely on AI tools to support their decisions and maximise business value (see Chapter 2,). However, understanding by non-expert users is a limitation and a current challenge in order to enhance the adoption of these systems. In particular, with our work, we want to support this business phase through the creation of a model that, to the best of our knowledge, is the first to simultaneously consider business value and model explanation capabilities. This section presents our model XVMF that aims to define a quantitative, evaluated, explainable, and value-aware recommender system.

In analogy with the approach of NEMF [213], we study how properly integrate a novel value regularization term to the original object function of EMF proposed by Abdollahi and Nasraoui [173] (see Eq. 3.4) to control the business value of the generated recommendations. In particular, for the value-aware term, we define v_i as the business value of the item i. As asserted in chapter 2, the business value of an item is most considered equal for all users and it represents the control key for the enterprise user to answer business questions and adopt an AI-based strategy in the decision-making phase.

The following equation formalizes the objective function of XVMF:

$$\mathcal{L} = \sum_{(u,i)\in\mathcal{S}} (r_{u,i} - \mathbf{p}_u \cdot \mathbf{q}_i^T)^2 + \frac{\beta}{2}(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2) + \; + \|\mathbf{p}_u - \mathbf{q}_i\|^2(\lambda \mathbf{W}_{u,i} + \delta v_i) \tag{3.16}$$

where,

- $\mathcal{L}$ is the loss function;
- u is a single user from the set $\mathcal{U} = \{1, \ldots, m\}$ of all m users;
- i is a single item from the set $\mathcal{I} = \{1, \ldots, n\}$ of all n items
- $\mathcal{S}$ is the set of combinations (user, item)
- $r_{u,i} \in \mathbf{R}$ is the rating for item i by user u in value range $[1, 5]$
- $\mathbf{p}_u \in \mathbb{R}^{1\times f}$ is the latent factor f-dimensional of user u (where f is the embedding dimension, hyper-parameter of the algorithm)
- $\mathbf{q}_i \in \mathbb{R}^{f\times 1}$ is the latent factor f-dimensional of item i (where f is the embedding dimension, hyper-parameter of the algorithm)
- β represents the loss hyper-parameter that controls the bias $(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2)$

- λ represents the loss hyper-parameter that controls the recommendation explainability
- $\mathbf{W} \in \mathbb{R}^{m \times n}$ is the explainability power weighted matrix
- δ represents the loss hyper-parameter that controls the business value
- v_i is the business value related to item i

As can be shown by Eq. 3.16, we set a trade-off between the explainability and the value-aware component, respectively, through the use of the two hyperparameters λ and δ on the following rules:

$$p'_u = p_u + \eta \left(2(r_{ui} - p_u q_i^T)q_i - \beta p_u - sgn(p_u - q_i)(\lambda W_{ui} + \delta v_i)\right) \tag{3.17}$$
$$q'_i = q_i + \eta \left(2(r_{ui} - p_u q_i^T)p_u - \beta q_i + sgn(p_u - q_i)(\lambda W_{ui} + \delta v_i)\right) \tag{3.18}$$

where η is the learning rate.

3.4 Experiments and Results

In the following section, we experimentally compare our approach with other explainable recommendation models. We compare our novel model XVMF with Matrix Factorization (MF)[217] and Explainable Matrix Factorization (EMF)[173]. Since in our model we evaluate both business value and explainability capabilities, the calculation of the explainability matrix is based on the code by Coba et al., while the value is originally assigned to each item in the selected datasets.

3.4.1 Dataset

To test our approach, the experiment is prepared with two different datasets *Yelp* [1] and *Amazon Tools and Home Improvement*[2] [218] (see table 3.4 for numerical information about the datasets). In both datasets, the rating dimension is scored in the range [1,5], where 1 is the lowest user rate while 5 is the highest one. The item business value is native in the datasets selected, user-independent, and shows the enterprise return generated by purchasing a specific item (profit).

[1]https://www.yelp.com/dataset/challenge/
[2]https://jmcauley.ucsd.edu/data/amazon_v2/index.html

Property	Yelp	Amazon
User number	589379	4699605
Item number	15290	559628
Interaction number	1191794	9437704
Filter on user interaction	20	25
Filter on item interaction	0	25
User filtered number	3876	5252
Item filtered number	11893	26927
Interaction filtered number	124999	134323

Table 3.4: Information about datasets used in the experiment and filter applied to avoid cold start issue.

Since in the literature the number of available datasets that expose the business value for an item is limited, the selection of these datasets (Yelp and Amazon) is based on the information they present in the form of user ratings for an item, and the corresponding business value. Although it is possible to generate synthetically the value for an item by using specific methodologies, we discourage this approach that could be affected by bias and generates results that do not represent by definition the real-world behaviors (where value could be assigned due to specific business requirements).

From the original datasets, a data refinement was applied to avoid cold start problems [154, 219, 52, 220]. In particular, for the Yelp dataset, we experimentally selected the threshold of 20 as the minimum number of interactions per user. Finally, to be compliant with the original information source, we verified that the correlation and distribution of the data were maintained after the filtering operation. The same approach was applied to the Amazon dataset but in this case, to avoid memory overflow on GPU, we applied a user interaction threshold of 25 and in addition an item interaction threshold of 25, ensuring the compliance of correlation and distribution to the original dataset.

3.4.2 Experimental Setup

The experiment consists of generating item recommendations for users evaluating the performance of explainability and business value capabilities of the model. All experiments are performed with a train-validation-test split of 60%-20%-20%. We consider the first 20 elements recommended by the algorithms for each user.

Regarding the assessment phase, to evaluate the accuracy of the model we use NDCG metric.

To evaluate explainability, instead of using MEP metric, suggested by Abdollahi and Nasraoui, we adopt E-NDCG calculation, as argued by Coba et al. [213]. Finally, to evaluate the business impact, we define a novel evaluation metric named the Normalized Discounted Cumulative Value (NDCV). NDCV is a metric that combines 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 [104], i.e., a variant of the previously defined NDCG@k where the gain is given by the item's price. In our context, instead of explicitly considering the price, we consider a generic business value (e.g., short-term profit) that the company may aim to optimize in accordance with the purposes of value-aware RSs.

We systematically assess the performance MF, EMF, and XVMF. Our evaluation methodology involves the manipulation of explainability and value regularization parameters while holding all other variables constant at specified values. For the Yelp dataset, the fixed parameters include a latent factor count of 128, a learning rate η set at 0.001, and a threshold of 20 users to provide explanations for each recommendation. Moreover, referring to the Amazon dataset, we maintain parameter uniformity with 128 latent factors, an η value of 0.001, and the same 20-user threshold for explanation provision. In addition, all the experiments are conducted with an upper bound of 1000 epochs of learning, also implementing a control that early-stops the learning when the validation error starts to increase or stops improving.

Finally, the experiments were conducted on a Ubuntu 22.04 machine with AMD Ryzen™ 5 2600X processor, 16GB DDR4 3000MHz ram, Nvidia GeForce RTX 2080 Ti graphic card.

3.4.3 Results

In this sections, we discuss the results of our proposed model when we set variation on the two hyperparameters λ and δ, respectively the explainability and value regularization terms.

We report the variation of the three considered evaluation metrics in two different charts, repeating the procedure for the two datasets. In the charts, we aim to show the trend of curves when we vary the regularization term and calculate the NDCG, NDCV, and E-NDCG indexes.

Performance evidence on the Yelp dataset

In chart 3.1a, we report the trend of E-NDCG compared to the NDCG value, increasing the explainability term while fixing the δ at 0.001. The curves show a similar trend with a higher level of E-NDCG for λ less than or equal to 0.05, keeping the peak at 0.001. On the other hand, 3.1b exposes the results for NDCG and NDCV calculation at $\lambda = 0.001$ while the

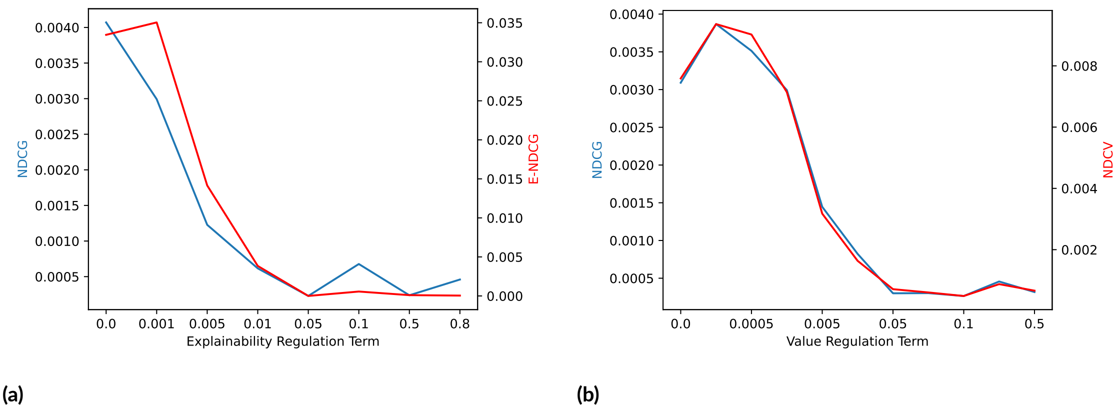


Figure 3.1: Performance analysis for **Yelp** dataset: metrics calculation for the individual variation of each regularisation term.

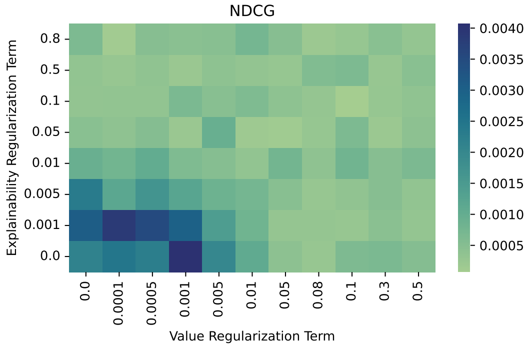
model varies the value regularization term. The two curves design a similar path that degrades at $\delta >= 0.005$.

Considering, instead, the simultaneous variation of both explainability and value regularization term, we can notice how the NDCG, NDCV, and E-NDCG change. We represent the evaluation of the three metrics using three different heatmaps (see Figure 3.2), where each cell represents the metric score for the considered couple of regularization terms. The higher the value of the metric is, the darker blue the cell is colored. Evaluating the obtained charts, they show an existing value interval for both of the regularization terms where we can calculate the maximum value for each metric. These intervals overlap approximately for each heatmap, and each metric value degrades quickly by increasing the regularisers above a certain threshold (0.01 for both explainability and value regularization term).

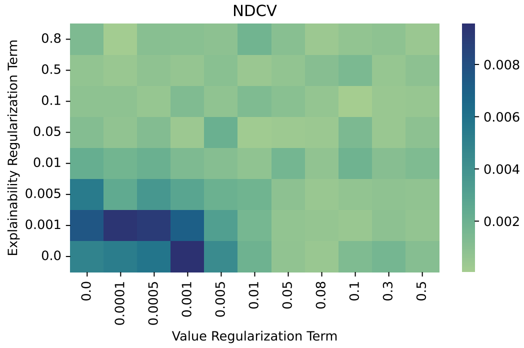
Performance evidence on the Amazon dataset

Considering the Amazon dataset, in the chart 3.3a we could highlight the E-NDCG compared to the NDCG, while increasing the explainability term and keeping the hyperparameter for value at $\delta = 0.01$. We obtain the best explainability result for $\lambda = 0.01$ and simultaneously a decrease of NDCG, while on other values the explainability performance degrades to the benefit of NDCG. On the other hand, 3.3b exposes a similar trend for the NDCG and NDCV curves, varying the value regularization term and keeping the best λ value at 0.01.

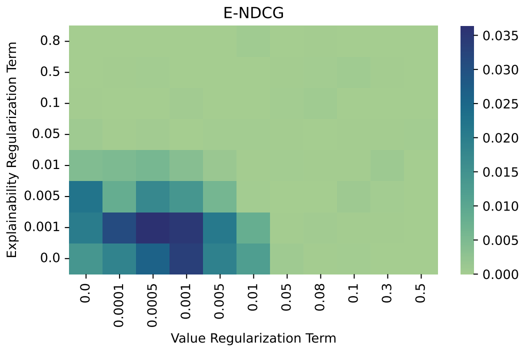
In Figure 3.4, we integrate a simultaneous variation in the weight of the two regularization terms and show the relative metric value in each cell (darker blue color for higher metric score,



(a)

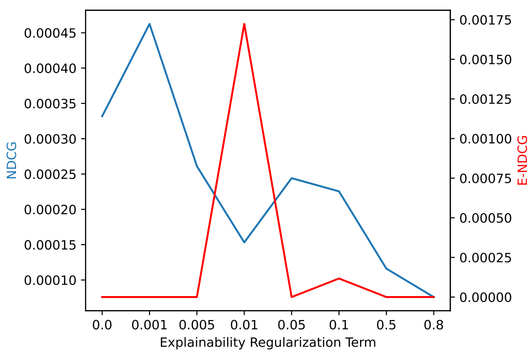


(b)

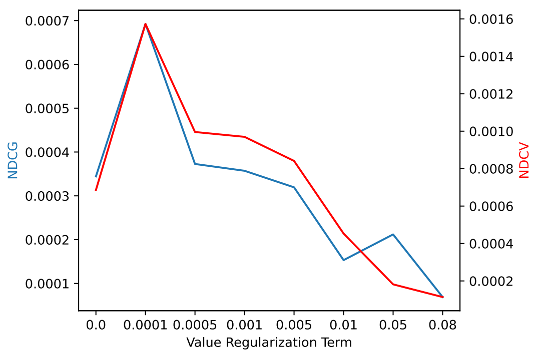


(c)

Figure 3.2: Performance analysis for **Yelp** dataset: metrics calculation for simultaneous variation of both regularization terms.



(a)



(b)

Figure 3.3: Performance analysis for **Amazon** dataset: metrics calculation for the individual variation of each regularisation term.

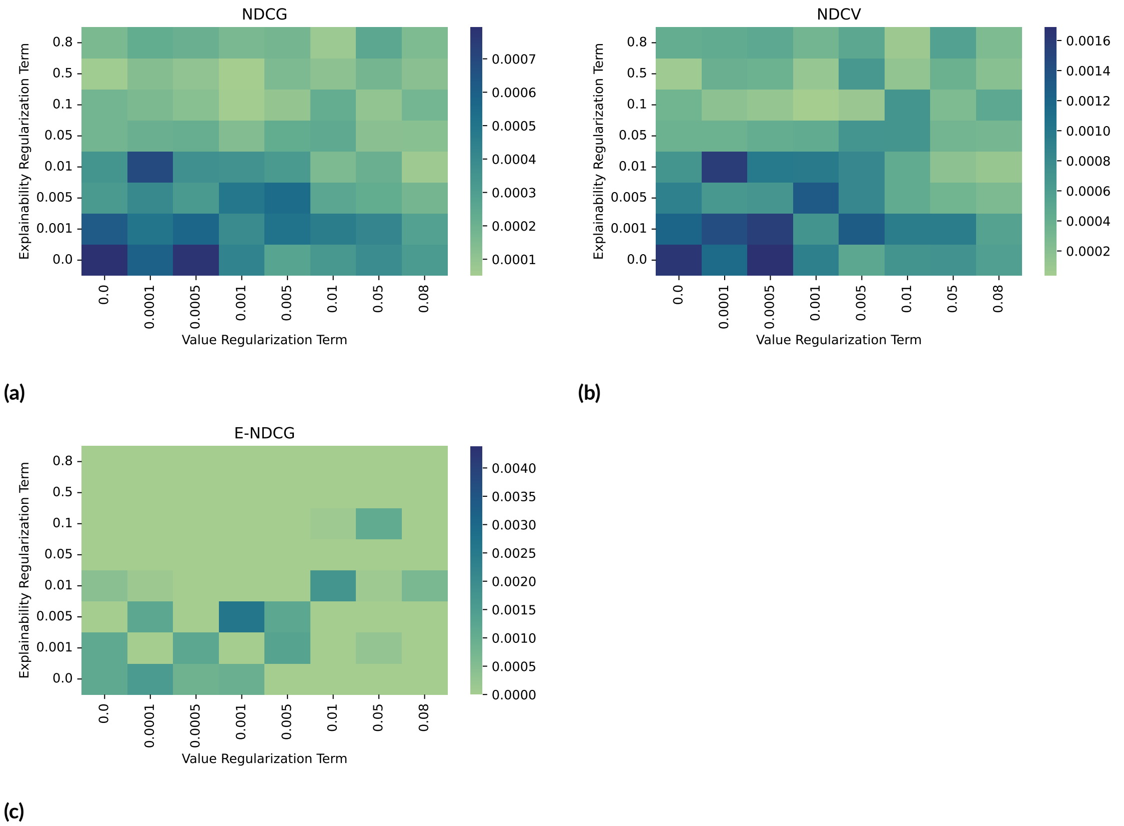


Figure 3.4: Performance analysis for **Amazon** dataset: metrics calculation for simultaneous variation of both regularization terms.

lighter green color for lower one). We can point out that in 3.4a and 3.4b, the NDCG and NDCV values follow a similar behavior with a good approximation, while the degradation of the metrics is present but not defined by a precise threshold, presenting a wider area with low-level score for higher value of explainability and value terms. Conversely, the behavior of E-NDCG metric exposes a tradeoff under a threshold for both the regularization term ($\lambda = 0.05$ and $\delta = 0.005$). Over these values, the explainability metric degrades.

3.4.4 Comparison with other methods

Our model XVMF is matrix factorization-based, and for the comparison experimentation, we adopt MF as the baseline, and EMF [173] to compare the performance with an explainabile model. Finally, In Table 3.5 and 3.6 are reported the results of the comparison analysis between the models selected, respectively for the outcome obtained with Yelp and Amazon datasets.

Since the balance obtained with the variation of the hyperparameters in XVMF could shift the performance of the results, we report:

- **XVMF-e** as the best solution for explainability score considering the maximum E-NDCG obtained;
- **XVMF-v** as the best solution for business value considering the maximum NDCV obtained;
- **XVMF-ev** as the best solution at the intersection of the previous two and considering both of the previous indexes simultaneously.

Considering the results of the Yelp dataset (see Table 3.5), our model XVMF has the highest score for all metrics, and the solution that maximizes the explainability and business value perspective, alone or simultaneously, is the same. Conversely, considering the Amazon dataset, the **XVMF-ev** is not the top model in terms of NDCV and E-NDCG, but, the solutions that consider individual value and explainability get the highest score.

Algorithm	NDCG	NDCV	E-NDCG
MF	0.1281%	0.2069%	0.8799%
EMF	0.1003%	0.2067%	0.8101%
XVMF-e	**0.2493%**	**0.5042%**	**1.9954%**
XVMF-v	**0.2493%**	**0.5042%**	**1.9954%**
XVMF-ev	**0.2493%**	**0.5042%**	**1.9954%**

Table 3.5: Evaluation metrics MF, EMF and XVMF for **Yelp** dataset. The highest value per index is indicated in **bold**.

Algorithm	NDCG	NDCV	E-NDCG
MF	**0.0525%**	0.0792%	0.0836%
EMF	0.0121%	0.0253%	0.0000%
XVMF-e	0.0125%	0.0275%	**0.1723%**
XVMF-v	0.0514%	**0.1154%**	0.0123%
XVMF-ev	0.0365%	0.1117%	0.0184%

Table 3.6: Evaluation metrics MF, EMF and XVMF for **Amazon** dataset. The highest value per index is indicated in **bold**.

3.5 Discussion

Upon analyzing the preceding section's outcomes, it is evident that depending on the nature of the interaction saved into the database, we can find slightly different behaviors that the mathematical model tries to learn. Overall, based on the metrics presented in Tables 3.5 and 3.6, it can be noticed that, although the values are low, they are promising since they generally exceed the baselines used, particularly for the Yelp dataset. Moreover, the experiments show that the balance between the two regularization terms is achieved for low value, and performance degrades quickly after a certain threshold range that is dependent on the application context. The experimental model could be enhanced in the future by considering alternative perspectives or evaluation metrics, either previously existing in the literature or newly proposed. Adopting qualitative measurement indicators, already present in the state-of-the-art, is desirable to enhance this new field of research with more comparable models. Finally, we used the profit as business value, but other companies could integrate another kind of value (see chapter 2) according to the business strategy. Different kinds of values could require variation on the model to adapt the mathematical interpretation of the problem to the real context.

The model is focused on the definition of a way to balance the explainability term and the business value. We had to deal with the classic recommendation problems (i.e., cold start, long tail) while creating a model that introduced new complexity to handle and respond to a specific requirement of enterprises. Indeed, one of the problems that this model aims to solve is enabling the company to understand better their internal processes enhanced by AI instruments. In particular, integrating a black-box AI model into a decision process can lead to unpleasant effects in the adoption of system results (i.e., lack of trust and understanding). Balancing business value and enhancing the recommendations with a quantitative score of explainability could reduce the distance between the user and the model, enabling the adoption and leading the use.

In future works, the research on this topic could be directed to improve XVMF or create a new model, considering other perspectives or using different metrics and datasets. Industries need to lead and understand the complexity beyond their process, yet with the support of these kinds of systems. Another future direction that we could suggest is the development of an interactive user interface to handle the balance between the two perspectives and collect feedback from the users, which is essential for empowering the models.

3.5.1 Real World utility

Explainability and value awareness are two pivotal aspects in the implementation of recommender systems for business decisions. Explainability addresses the need for transparency and comprehension in the recommendations provided to users and stakeholders. In a business context, understanding the rationale behind recommendations is paramount, as it instills trust and confidence in the system's suggestions. Explainable recommender systems allow users and decision-makers to grasp not only what is recommended, but also why it is recommended, enabling them to make informed choices aligned with their preferences and needs.

On the other hand, value-awareness adds a strategic dimension to recommender systems by focusing on the generation of tangible business value. Traditional recommender systems might solely emphasize user preferences, but in a business setting, it is crucial to align recommendations with organizational objectives. Value-aware recommender systems take into account not only user preferences but also the potential impact on key business metrics, such as sales, customer retention, or engagement. By optimizing recommendations to simultaneously satisfy user preferences and drive business outcomes, value awareness becomes a key tool for enhancing the overall efficiency and effectiveness of a company's decision-making processes.

Incorporating explainability and value-awareness into recommender systems can thus empower businesses to create a harmonious balance between user satisfaction and organizational success. It not only enhances user experiences but also positions the organization for improved competitiveness and growth through recommendations that are not only accurate but also understandable and aligned with strategic goals. Real-world datasets describe different contexts of application and the adaptation of the model proposed should be taken into account. In addition, by providing an interface to business decision-makers, it could create an interactive way to access the information provided by the model proposed, guiding the system to produce more refined explanations or more valuable outputs, depending on the preferred level of balance and business goals.

3.6 Contribution

In this chapter, we provided the novel recommendation model XVMF, which is, to the best of our knowledge, the first model to balance explainability capabilities with the business value generated. Evaluating three different metrics, NDCG, NDCV, and E-NDCG, we aimed to understand how the model affects the different perspective of explainability and value business. Finally, changing the hyperparameters simultaneously, we derived the cross-dependence performance of the model.

Although the benefits of the system could be different for the companies, it appears challenging to create a general system that could be applied to multiple sources (i.e., datasets of multiple companies) keeping a high and accurate level of performance. The application of the model in a real-world scenario must require a refinement of the parameters trained and a deep understanding of the business decision process to encapsulate the right enterprise dynamics within the system. Measuring model performance in the right way is essential for enterprises in the design and evaluation phase because it provides a comparable way to assess multiple models, selecting the best one.

Moreover, we discussed how explainability affects the RSs field, focusing on the importance of the evaluation phase. Among the evaluation metrics reported, we selected the most used and promising ones to evaluate our model.

In the next chapter of this thesis, we conclude our research journey with an analysis of a promising recommender systems family based on graphs. In particular, we focus on the reason for the importance of defining common evaluation criteria for explainability, highlighting a common symptom that affects the Graph-Based Explainable Recommender Systems, and suggesting a direction for future research works in this field to support the enterprise model design process.

4

Explainability applied on Graph-based Recommendation Systems

In recent years, we have registered an increase in the amount of published research in the field of Explainable Recommender Systems. As reported in the chapter 3, these systems are designed to help users find the items of the most interest by providing not only suggestions but also the reasons behind those recommendations. Research has shown that there are many advantages to complementing a recommendation with a convincing explanation. For example, such an explanation can often lead to an increase in user trust, which in turn can improve recommendation effectiveness and system adoption. In particular, many research works focus on explainable recommendation algorithms based on graphs due to the natural graph structure presentation capabilities (i.e., exploiting knowledge graphs or graph neural networks-based methods). The use of graphs is very promising since algorithms can, in principle, combine the benefits of personalization and graph reasoning, thus potentially improving the effectiveness of both recommendations and explanations. However, although graph-based algorithms have been repeatedly shown to bring improvements in terms of ranking quality, not much literature has yet studied how to properly evaluate the quality of the corresponding explanations. As reported in chapter 2 and 3, from an industrial perspective is essential for a model to be comparable and quantitative evaluable, not only in the selection phase of the best performer but also during project setup, establishing expectations and return on investment. In this chapter, we focus on this problem, examining in detail how the explainable recommender systems based on graphs are currently evaluated, discussing how they could be improved, and suggesting a future direction in a more quantitative and comparable way.

As users have increasingly demanded explanations for recommendations in recent years, it becomes crucial for model developers to provide insights into how and why those recommendations are made. Unlike traditional recommender systems, this kind of approach allows any recommendation to be generated by simultaneously integrating it with a corresponding explanation. The explanations provided can be very persuasive [221, 222, 223] in that they often exploit graph reasoning logic that allows the explanation to be represented as a path on the graph. For example, an explanation for a "Back to the Future" recommendation, may be intuitively represented with the path "user $\xrightarrow{\text{watched}}$ Forrest Gump $\xrightarrow{\text{directed by}}$ Robert Zemeckis $\xrightarrow{\text{that also directed}}$ Back to the Future" [224]. However, although graph-based methods have repeatedly shown remarkable performance in modeling complex user-item relational dependencies and generating relevant recommendations, the transparency and interpretability of the underlying reasoning process still remain a significant concern. In particular, the main problem lies in establishing a clear understanding of what a Graph-Based Explainable Recommender System (GxRS) should provide as an explanation, how it could be useful for the final users, and especially how accurate the outcome is.

To clarify the challenges mentioned above, the remainder of the chapter is organized as follows. In Section 4.1 we offer a general introduction to recent graph-based explainable recommendation algorithms. Subsequently, in Section 4.2 we focus on the current methodologies used in the literature to evaluate these algorithms. Then, in Section 4.3 we bring to the community's attention the importance of the integration of quantitative explainable performance evaluation, while also discussing potential ways to use these guidelines in future research works.

4.1 Background and concepts

Graph-based algorithms have attracted the interest of many researchers because of the capabilities they offer to represent the world of interactions, particularly those related to humans. They are a promising area because the learning process can be based directly on graphs that not only represent user-object interactions but can also include contextual information such as user demographics, product categories, and other attributes. In particular with the objective of capturing these connections and exploiting these potentials through user suggestions, Knowledge Graphs (KGs) and Graph Neural Networks (GNNs) have gained significant attention in recent years, especially in the Recommender Systems (RSs) field to ensure fairness

Table 4.1: Graph-Based Explainable Recommender Systems surveyed from literature, and grouped by algorithmic method and explainability evaluation approach.

Algorithmic Method	No Explainability Evaluation	Qualitative Explainability Evaluation	Quantitative Explainability Evaluation
Embedding-based	[225, 226, 227, 228, 229]	[230, 231, 232]	
Path-based		[233, 234, 235, 236, 237, 238, 239, 224, 240, 241, 242, 243, 244]	[245, 246, 247, 214, 248, 249]

[250], improve business value [2], or generate relevant, yet explainable, recommendations to users from graphs [251, 194].

Analyzing the most recent surveys [252, 253, 254, 255, 256, 257] in the field of GNN- and KG-based recommender systems, from 2018 to 2023 (see Table 4.1), what emerges is that many existing studies have focused on the use of graphs to improve the performance of recommender systems in terms of recommendation quality, diversity, and other conventional measures, while less rigorously assessing the quality of the corresponding explanations. Studies often offer only qualitative case-based examples where a particular explanation is represented graphically as a path on the graph. However, what is generally lacking is a quantitative and comparable evaluation of the quality of explanations according to the widely known guidelines of explainable recommendation systems [165]. While graphical explanations are useful for getting an intuitive idea of the underlying reasoning process, they also severely limit the comparison of algorithms and, more generally, the progress in this particular field. Instead, it may be beneficial for research to focus more in the future on the evaluation aspects by designing metrics to provide quantitative insights into the complete decision-making process to ensure that the algorithms' explanations are useful in practice from a human perspective.

In the literature, a number of different studies proposed the use of GNN- or KG-based algorithms to generate relevant, yet explainable recommendations from graphs. These graphs often provide additional information in addition to the most commonly used user-item interactions, including demographic characteristics of the user (e.g., age, gender), various attributes of the item (e.g., product category, price range), and contextual features (e.g., time, location) interconnected in a graph. A graph can be classified into homogeneous (if the edge connects only two nodes and there is only one type of nodes and edges), heterogeneous (if the edge connects only two nodes and there are multiple types of nodes and edges), or hypergraph (where each edge joins more than two nodes). While edges represent a relation (interaction or property) of the node, each node represents an entity of the dataset that could have one or more associated properties and could interact with other entities. There are various techniques for using such graph-based information for recommendations and/or explanation purposes.

Depending on how the graph is handled in the learning process, we can distinguish different graph-based explainable recommendation techniques in the literature. For example, in certain cases, neural networks can be exploited to decompose the graph in the form of embeddings or paths (see Table 4.1). In particular, *embedding-based methods* typically aim to learn embedding representations of users, items, and other entities from the graph that can be used to generate recommendations or explanations. However, embedding-based methods generally lack the ability to discover multi-hop relational paths from the graph to generate explanations. Therefore, the explanations provided are generated by exploiting empirical criteria of similarity matching among the various embeddings in the graph to motivate post-hoc a given recommendation (*weak explainability*). Instead, *path-based methods* first identify connectivity paths between users and items and then feed those paths into the recommendation algorithms to generate recommendations and explanations. However, although the explanations provided by these models often appear quite convincing, as they are based on complex multi-hop reasoning, considering all possible paths between a given user-item pair may involve irrelevant ones that can lead to mismatches with real user preferences (*error propagation*). Given the current limitations of *pure* path-based and embedding-based algorithms, other hybrid [224, 240, 241] algorithms have also been studied in the literature. These methodologies should, in principle, improve recommendations and explanations by alleviating the weak explainability and error propagation problems. However, as we discuss in the next section, explanations are not always rigorously evaluated.

4.2 Evaluating the graph-based explainable recommenders

In this section, we explore the evolution of the explainability applied on Graph-Based Explainable Recommender Systems. For details on the explainability evaluation process, metrics, criteria, and limitations (on RSs in general), see sections 3.1 and 3.2.

In the above literature, a variety of methods and metrics are used for evaluation purposes.

In particular, all the surveyed studies employ well-known offline metrics (e.g., Precision, Recall, NDCG, AUC) from the RSs literature to evaluate the relevance of recommendations[1]. These metrics are typically used to evaluate the performance of a RS in recommending items of most interest to users. Proposed graph-based algorithms are often able to beat baselines in terms of relevance or other well-known quality factors such as diversity and coverage because, since graphs are often used as additional contextual information to the user-item interaction matrix, they allow, in principle, more accurate recommendations to be generated. However, especially for an Explainable Recommender Systems, while it is important to assess the relevance of recommendations, it should also be equally important to assess the quality of the corresponding explanations. Indeed, the recommendation algorithm could, in principle, produce good-quality recommendations but weak explanations.

Unfortunately, when analyzing the above literature in detail, it emerges that only a few works [245, 246, 247, 214, 248, 249] have evaluated the quality of explanations as rigorously as they have assessed recommendations relevance. Indeed, in terms of explanation quality, almost all studies (see Table 4.1) provide some qualitative case-based analysis to intuitively evaluate the quality of the algorithmic reasoning process. Typically, a specific recommendation of a certain item is selected for a given user, and a *graphical representation* of the explanation provided by the algorithm is proposed. Supplementing the graphical representation, some empirical observations are often provided to state that, at least intuitively, the considered explanation seems realistic. However, what is generally missing is a quantitative and comparable assessment of the quality of the system's overall explanations, i.e., a goal-oriented evaluation based on different factors of the explanations that the system should provide for each recommendation to every user, as is typically done instead when assessing recommendations relevance.

4.2.1 Current evaluation of explanations in graph-based recommender systems

Besides the typically employed qualitative case analyses, only a few articles, among the ones listed in Table 4.1, proposed to evaluate the explanations of the proposed Explainable Recommender Systems in a more quantitative way. For example, Lyu et al. [247] used *ROUGE* to evaluate explanations offline. The metric is typically used for the evaluation of text summarization tasks and measures the number of overlapping words between the generated text and the

[1] We refer readers to some recent surveys [170, 258] for further insights on well-known offline evaluation metrics that are widely used in the RSs literature to assess the relevance of recommendations.

ground truth. Since in the work, the explanations generated by the algorithm are expressed in natural language, the authors can use the metric to assess how close these explanations are to the ground truth user reviews. However, the metric can only be used to evaluate the natural language-based explanation style, which is a recent area of research in the literature. Therefore, the proposed evaluation is not suitable for evaluating the more widely adopted path-based explanation style, i.e., where the logical reasoning of the algorithm is represented as a path on the graph. To overcome this limitation, a similar methodology recently employed by Tai et al. [248] and Zhao et al. [249] consists of evaluating the ability of an algorithm to provide explanation paths that contain entities also present in the form of words in user reviews, exploiting well-known relevance-based metrics such as NDCG and Recall. However, as in the previous case, if user reviews are not present, this kind of evaluation methodology is not applicable. Exploiting a different methodology, Ma et al. [246] proposed to evaluate the quality of the explanations (in terms of relevance and diversity) from a *human* perspective. In particular, the authors randomly selected 100 user-item recommendation pairs and the corresponding path-based explanations generated by the recommender system. Then they selected 10 human raters who have machine learning experience to manually evaluate the quality of explanations. However, as is also known in other areas in the field of recommender systems, this particular online evaluation methodology can be very expensive to perform on a large scale and subject to user bias. Hence, the overall validity of the final results may be compromised if the human raters are not carefully selected. Another methodology has been proposed recently by Wang et al. [245]. In particular, given a certain explanation for a user recommendation, proposed to evaluate the degree to which the explanation path conforms to the particular user profile. Specifically, for a given user, the authors first construct the user profile containing his/her interactions with the entities of the graph. Then they measured the number of entities in the explanation path that are also present in the user profile. Moreover, since an explanation path can be based on multiple hops between different graph entities, very long reasoning paths would be able to match more user profile entities. Correspondingly, the authors' proposed evaluation is based on a hyper-parameter that considers only a certain number of entities in the explanation path for evaluation purposes. However, as noted by the authors, this evaluation methodology is very inefficient. Hence, they sampled only 100 test set users and evaluated the explanations of the top 20 recommendations for each of them. Finally, in another recent work by Geng et al. [214] it has been proposed to measure through the New Reach Ratio (NR^2) metric in which terms a graph-based explainable model is able to mitigate the recall bias.

4.3 Towards a more standardized evaluation

In the previous section, we highlighted that most of the works in the GxRS literature used a qualitative case-based analysis rather than a quantitative approach to intuitively evaluate the explanations provided by the model. Unfortunately, the lack of adoption of a quantitative and comparable framework for styling, presenting, personalizing, and evaluating such explanations, does not allow to compare the different models in terms of explainability results. The adoption of shared guidelines that employ quantitative metrics would allow this issue to be solved.

In 2015 Tintarev and Masthoff [165], and more recently Chen et al. [172] and Mohseni et al. [176] released well-known guidelines to create a common evaluation framework for Explainable Recommender Systems (of which GxRS are part). The guidelines provide a formal process for assessing the explainability of a model (see section 3.1). Following this process, the developer can define the goal of the model, the target user, and the evaluation metrics to determine how much the model performs in terms of explainability, considering style, presentation, personalization, and evaluation aspects. For example, in terms of *presentation*, the recommendations provider may desire to give explanations that are structured in a certain way. Currently, most of the explanations are provided to users by using a template-based structure (e.g., indicating how many similar users have the same tastes of the current user), a graphical representation (e.g., considering a path on a graph) or natural language [259, 260]. Instead, in terms of *evaluation*, the quality of explanations is typically assessed considering certain goals, as transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction. The ability of explanations to achieve such goals can be evaluated by calculating certain metrics, considering specific case studies (e.g., for a qualitative assessment), or through online experiments (e.g., user studies, A/B tests) [172]. Each of these methods has certain advantages and disadvantages. For example, qualitative case-based analyses can be used to intuitively assess whether explanations are realistic or not. However, the evaluation may be affected by bias, and the outcome of different models is not comparable. Instead, analyses based on quantitative metrics (e.g., *Probability of Necessity*, *Mean Explainability Precision*) [208, 173] could be less intuitive, but they are easy to benchmark, comparable, and more efficient.

For example, future works may exploit existing metrics proposed in the literature to evaluate the explanations quality (e.g., *Mean Explainability Precision* [173], or *Explainability Power* [213]). A representative list of explainability metrics from the literature [172, 256, 166, 261] is presented in Table 3.3. In particular, researchers could focus on balancing multiple perspectives at the same time, such as fairness and explainability measuring the Path and Diversity Scores proposed by Fu et al. [262]. Moreover, in the future, it could be worth to investigate the relation between the recommendation list and the explanation properties as proposed by Balloccu et al.. In particular, in recent works, e.g., [215, 216], Balloccu et al. propose six novel metrics to evaluate the quality of explanations, i.e., Linking Interaction Recency (LIR), Linking Interaction Diversity(LID), Shared Entity Popularity (SEP), Shared Entity Diversity (SED), Explanation Path Type Diversity (PTD), and Explanation Path Type Concentration (PTC). The use of such quantitative evaluation metrics can enhance the interpretability and transparency of recommendations, empowering users to make informed decisions based on comprehensible explanations. Overall the integration of such a quantitative evaluation framework may not only provide concrete evidence of model effectiveness but it may also contribute to the advancement and adoption of explainable recommender systems in real-world applications.

Besides exploiting existing metrics, other research directions may also be interesting for the future. In particular, while existing metrics can be used to measure certain aspects of explanations, future research may study other quantitative and comparable methods that can be used to evaluate the quality of explanations in compliance with the explainability guidelines, e.g., assessing explanations in terms of style, personalization and presentation aspects. Moreover, given the limitations of available datasets in reflecting real user preferences in terms of the explanations provided by the models, further research may focus on collecting datasets that have such information. Furthermore, another important aspect to consider for the future may be the inclusion of human-centered evaluation methods. Indeed, evaluating the performance and effectiveness of explainability from a human perspective is essential to gain valuable insights into the usability and impact of the explanations provided by an algorithm [188]. Finally, another effective enhancement for evaluating explanations could be providing explanations to users through an interface designed to facilitate access, increase comprehension, and collect users' feedback during the usage experience. Through this interface, providers could measure the adoption of a system that may seek to achieve one or more objectives at the same time, e.g., considering effectiveness [194], persuasiveness, scrutability aspects.

4.4 Contribution and future direction

In this chapter, we discussed how the explanations provided by graph-based explainable recommendation systems are currently evaluated, pointing out open challenges and future research directions in this area concerning evaluation methods. What emerges from our analysis is that most papers evaluated the quality of explanations through qualitative case-based analysis, while only a few articles proposed metrics for a more quantitative evaluation. Moreover, the current metrics are not sufficient to comprehensively evaluate all the different types of explanations and are only partially compliant with the well-known explainable recommender system guidelines. Future research will need to address current limitations by providing new guidelines-compliant evaluation methodologies. With this chapter, we encourage researchers to adopt a more quantitative and comparable approach when evaluating the quality of the explanations. We hope that our efforts will inspire further research in this field and lead to the creation of more comprehensive and guideline-compliant methods for assessing and comparing the quality of explanations of graph-based explainable recommendation algorithms.

5

Conclusion and future works

5.1 Conclusion

Approaching the research world with an enterprise perspective is the widest objective of the industrial doctoral program at the basis of this thesis. Companies must operate under most constraints, i.e., regarding cost, time, manpower, and expertise, and most of the time they follow only one path in the multitude of solutions offered by the academic world. While they find a concrete application scenario of the model available in the literature, the researchers should enhance the proposed model with explanations of what the black-box system generates as output to the non-expert users. This type of approach is doubly effective. While, on the one hand, business users understand the AI system and increase its adoption, on the other hand, academia is forced to express the results of the model not only in terms of how accurate it is but also what degree of understanding provides the system.

In this context, we prepared our journey path by collecting numerous contributions from the literature, developed our research based on industrial experience, and summarized the findings on how the AI models, in particular the Recommender Systems, can support enterprise strategies of selling a product or proposing an item. Firstly, we identified the different kinds of business values defined in the literature and proposed a taxonomy that aims to simplify comprehension and lead to adoption. Secondly, we analyzed how explainability approaches can be implemented in recommender systems from an industrial point and we proposed a model that aims to balance explainability and business value. Finally, we argued a symptom that affects many recent works in the literature of Graph-Based Explainable Recommender Systems, suggesting the implementation of quantitative metrics and the adoption of definition guidelines to create comparable models and, thus, facilitating the use by enterprises.

The main contribution of this thesis is the proposal of a model that aims to simultaneously analyze both the explainability and the business value, mathematically finding a tradeoff between two perspectives that are studied separately in the literature. In designing the model XVMF, we deal with some challenges. Primarily, we decided which family of algorithms we should use, and matrix factorization optimally supported our perspective. Then, we reviewed the literature to choose the used and promising metrics to evaluate the model created. Moreover, we decided on the datasets for the experiment based on the fundamental characteristics we wanted to study, such as the presence of a dimension representing business value. Finally, we defined the experiment with the mathematical model concepts and set the hyperparameters for the simulation execution.

On the other hand, we have also systematically investigated the state-of-the-art of Value-Aware Recommender Systems to collect all the resources published regarding this topic, analyzing the different approaches proposed and suggesting a proper taxonomy and datasets for future models. This contribution became the first systematic literature review on VARSs, creating a baseline for researchers who are interested in this topic.

Finally, due to the increased companies' interest in Graph-Based Explainable Recommender Systems (GxRSs), we decided to point out how these models could be beneficial for enterprises from an explainability perspective. In particular, since they are able to simplify the complexity of the information stored in enterprise datasets, we focused on the evaluation phase which is essential for a company. Indeed, during the decision phase, decision-makers strategically manage a wide number of information at the same time and require a solution to comprehend this process. Moreover, after highlighting the absence of quantitative evaluation metrics for the proposed explainable graph model, we suggested the adoption of a standard way to evaluate the explainability through the adoption of existing quantitative metrics and the definition of the property of the model based on released guidelines that aim to classify the model in terms of style, personalization, presentation, and evaluation.

5.2 Future works

After examining the limitations and promising areas of expansion within this thesis, we can highlight potential avenues for future research:

- improve explainable recommender systems with enhanced explainability capabilities through the integration of model definition, proper evaluation metrics, and adoption of guidelines;
- focus on the development of an explainable model targeted to business, using a proper dataset and developing an effective approach to solving real-world problems, such as profit or revenue maximizing;
- define new metrics that are designed to explain the black-box model to non-expert stakeholders to enhance the decision phase;
- design and integrate more Human-Centered interfaces able to present the information properly (already existing approaches are available; maybe it would be beneficial to focus on the industrial user), and that are able to collect user feedback, balancing the explainability engine to human business questions.

We hope that our work could add an additional element in a still unexplored area, due to the distance between the academic and enterprise world, but with a high level of interest and potential improvement.

References

[1] 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, et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, 10(1):1–11, 2021.

[2] Alvise De Biasio, Andrea Montagna, Fabio Aiolli, and Nicolò Navarin. A systematic review of value-aware recommender systems. *Expert Systems with Applications*, 226:120131, September 2023. ISSN 0957-4174. doi: 10.1016/j.eswa.2023.120131. URL https://www.sciencedirect.com/science/article/pii/S0957417423006334.

[3] Andrea Montagna, Alvise De Biasio, Nicolò Navarin, and Fabio Aiolli. Graph-based explainable recommendation systems: Are we rigorously evaluating explanations? In *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter*, CHItaly '23, New York, NY, USA, 2023. Association for Computing Machinery.

[4] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: Techniques, applications, and challenges. *Recommender Systems Handbook*, pages 1–35, 2022.

[5] Qi Zhao, Yi Zhang, Daniel Friedman, and Fangfang Tan. E-commerce Recommendation with Personalized Promotion. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 219–226, Vienna Austria, September 2015. ACM. ISBN 978-1-4503-3692-5. doi: 10.1145/2792838.2800178. URL https://dl.acm.org/doi/10.1145/2792838.2800178.

[6] Feng Lu, Anca Dumitrache, and David Graus. Beyond Optimizing for Clicks: Incorporating Editorial Values in News Recommendation. In *Proceedings of the 28th 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. URL https://dl.acm.org/doi/10.1145/3340631.3394864.

[7] 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.

[8] 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 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2919–2927, Virtual Event Singapore, August 2021. ACM. ISBN 978-1-4503-8332-5. doi: 10.1145/3447548.3467175. URL https://dl.acm.org/doi/10.1145/3447548.3467175.

[9] Dirk Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark Graus. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, page 63, Barcelona, Spain, 2010. ACM Press. ISBN 978-1-60558-906-0. doi: 10.1145/1864708.1864724. URL http://portal.acm.org/citation.cfm?doid=1864708.1864724.

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

[11] 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. URL https://dl.acm.org/doi/10.1145/3370082.

[12] Dietmar Jannach and Markus Zanker. Value and impact of recommender systems. In *Recommender Systems Handbook*, pages 519–546. Springer, 2022.

[13] 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 Eleventh ACM Conference on Recommender Systems*, pages 378–379, Como Italy, August 2017. ACM. ISBN 978-1-4503-4652-8. doi: 10.1145/3109859.3109957. URL https://dl.acm.org/doi/10.1145/3109859.3109957.

[14] Robin Burke, Himan Abdollahpouri, Edward C. Malthouse, Kp Thai, and Yongfeng Zhang. Recommendation in multistakeholder environments. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 566–567, Copenhagen Denmark, September 2019. ACM. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3346973. URL https://dl.acm.org/doi/10.1145/3298689.3346973.

[15] 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 *The World Wide Web Conference on - WWW '19*, pages 3123–3129, San Francisco, CA, USA, 2019. ACM Press. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313404. URL http://dl.acm.org/citation.cfm?doid=3308558.3313404. 17 citations (Semantic Scholar/DOI) [2022-02-04].

[16] Qingyun Wu, Hongning Wang, Liangjie Hong, and Yue Shi. Returning is Believing: Optimizing Long-term User Engagement in Recommender Systems. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1927–1936, Singapore Singapore, November 2017. ACM. ISBN 978-1-4503-4918-5. doi: 10.1145/3132847.3133025. URL https://dl.acm.org/doi/10.1145/3132847.3133025.

[17] Jianhua Han, Yong Yu, Feng Liu, Ruiming Tang, and Yuzhou Zhang. Optimizing Ranking Algorithm in Recommender System via Deep Reinforcement Learning. In *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, pages 22–26, Dublin, Ireland, October 2019. IEEE. ISBN 978-1-72814-691-1. doi: 10.1109/AIAM48774.2019.00011. URL https://ieeexplore.ieee.org/document/8950897/.

[18] Yong Zheng and David (Xuejun) Wang. A survey of recommender systems with multi-objective optimization. *Neurocomputing*, 474:141–153, February 2022. ISSN 0925-2312. doi: 10.1016/j.neucom.2021.11.041. URL https://www.sciencedirect.com/science/article/pii/S0925231221017185.

[19] 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 1573-1391. doi: 10.1007/s11257-019-09256-1. URL https://doi.org/10.1007/s11257-019-09256-1.

[20] Diego Monti, Giuseppe Rizzo, and Maurizio Morisio. A systematic literature review of multicriteria recommender systems. *Artificial Intelligence Review*, 54(1):427–468, January 2021. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-020-09851-4. URL https://link.springer.com/10.1007/s10462-020-09851-4.

[21] Shaina Raza and Chen Ding. Progress in context-aware recommender systems — An overview. *Computer Science Review*, 31:84–97, February 2019. ISSN 15740137. doi: 10.1016/j.cosrev.2019.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1574013718302120.

[22] Wen-Hao Chen, Chin-Chi Hsu, Yi-An Lai, Vincent Liu, Mi-Yen Yeh, and Shou-De Lin. Attribute-Aware Recommender System Based on Collaborative Filtering: Survey and Classification. *Frontiers in Big Data*, 2:49, January 2020. ISSN 2624-909X. doi: 10.3389/fdata.2019.00049. URL https://www.frontiersin.org/article/10.3389/fdata.2019.00049/full.

[23] Athanasios N Nikolakopoulos, Xia Ning, Christian Desrosiers, and George Karypis. Trust your neighbors: a comprehensive survey of neighborhood-based methods for recommender systems. *Recommender Systems Handbook*, pages 39–89, 2022.

[24] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[25] Xia Ning and George Karypis. Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th international conference on data mining*, pages 497–506. IEEE, 2011.

[26] Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.

[27] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Content-based Recommender Systems: State of the Art and Trends. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 73–105. Springer US, Boston, MA, 2011. ISBN 978-0-387-85819-7 978-0-387-85820-3. doi: 10.1007/978-0-387-85820-3_3. URL http://link.springer.com/10.1007/978-0-387-85820-3_3.

[28] 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. URL https://www.hindawi.com/journals/aai/2009/421425/.

[29] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[30] Yehuda Koren, Steffen Rendle, and Robert Bell. Advances in collaborative filtering. *Recommender systems handbook*, pages 91–142, 2022.

[31] Charu C Aggarwal. Model-based collaborative filtering. In *Recommender systems*, pages 71–138. Springer, 2016.

[32] Xavier Amatriain, Nuria Oliver, Josep M Pujol, et al. Data mining methods for recommender systems. In *Recommender systems handbook*, pages 39–71. Springer, 2011.

[33] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073, 2014.

[34] Marwa Hussien Mohamed, Mohamed Helmy Khafagy, and Mohamed Hasan Ibrahim. Recommender systems challenges and solutions survey. In *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)*, pages 149–155. IEEE, 2019.

[35] Lawrence D. Miles. *Techniques of Value Analysis and Engineering*. New York: McGraw-Hill Book Company., 1961.

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

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

[38] 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.

[39] Halil Shevket Neap and Tahir Celik. Value of a product: A definition. *International Journal of Value-Based Management*, 12(2):181–191, 1999.

[40] Philip Kotler. *Marketing Management: The Millennium Edition*, volume 10. Prentice Hall Upper Saddle River, NJ, 2000.

[41] 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.

[42] Milton Rokeach. *The Nature of Human Values*. Free press, 1973. ISBN 0-02-926750-1.

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

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

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

[46] Daniel M. Fleder and Kartik Hosanagar. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce - EC '07*, page 192, San Diego, California, USA, 2007. ACM Press. ISBN 978-1-59593-653-0. doi: 10.1145/1250910.1250939. URL http://portal.acm.org/citation.cfm?doid=1250910.1250939.

[47] 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, May 2009. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.1080.0974. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0974.

[48] Dokyun Lee and Kartik Hosanagar. Impact of recommender systems on sales volume and diversity. *ICIS 2014*, 2014.

[49] 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.

[50] Hyokmin Kwon, Jaeho Han, and Kyungsik Han. ART (Attractive Recommendation Tailor): How the Diversity of Product Recommendations Affects Customer Purchase Preference in Fashion Industry? In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2573–2580, Virtual Event Ireland, October 2020. ACM. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3412687. URL https://dl.acm.org/doi/10.1145/3340531.3412687.

[51] Mu-Chen Chen, Long-Sheng Chen, Fei-Hao Hsu, Yuanjia Hsu, and Hsiao-Ying Chou. HPRS: A profitability based recommender system. In *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 219–223, Singapore, December 2007. IEEE. ISBN 978-1-4244-1528-1 978-1-4244-1529-8. doi: 10.1109/IEEM.2007.4419183. URL http://ieeexplore.ieee.org/document/4419183/.

[52] Xavier Amatriain and Justin Basilico. Past, Present, and Future of Recommender Systems: An Industry Perspective. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, pages 211–214, New York, NY, USA, September 2016. Association for Computing Machinery. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959144. URL https://doi.org/10.1145/2959100.2959144.

[53] 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*, pages 1–1, 2021. ISSN 1541-1672, 1941-1294. doi: 10.1109/MIS.2021.3092768. URL https://ieeexplore.ieee.org/document/9466459/.

[54] 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 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2784–2792, Virtual Event CA USA, August 2020. ACM. ISBN 978-1-4503-7998-4. doi: 10.1145/3394486.3403329. URL https://dl.acm.org/doi/10.1145/3394486.3403329.

[55] Dietmar Jannach and Gediminas Adomavicius. Price and Profit Awareness in Recommender Systems. *Proceedings VAMS Workshop at ACM RecSys 2017*, July 2017. URL http://arxiv.org/abs/1707.08029. 25 citations (Semantic Scholar/arXiv) [2022-02-04] arXiv: 1707.08029.

[56] 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. URL https://linkinghub.elsevier.com/retrieve/pii/S1567422316300588.

[57] Mario Rodriguez, Christian Posse, and Ethan Zhang. Multiple objective optimization in recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems - RecSys '12*, page 11, Dublin, Ireland, 2012. ACM Press. ISBN 978-1-4503-1270-7. doi: 10.1145/2365952.2365961. URL http://dl.acm.org/citation.cfm?doid=2365952.2365961.

[58] Gediminas Adomavicius, Nikos Manouselis, and YoungOk Kwon. Multi-criteria recommender systems. In *Recommender systems handbook*, pages 769–803. Springer, 2011.

[59] Himan Abdollahpouri and Robin Burke. Multistakeholder recommender systems. In *Recommender systems handbook*, pages 647–677. Springer, 2022.

[60] Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

[61] Yu Zheng, Chen Gao, Xiangnan He, Yong Li, and Depeng Jin. Price-aware Recommendation with Graph Convolutional Networks. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 133–144, April 2020. doi: 10.1109/ICDE48307.2020.00019. 23 citations (Semantic Scholar/DOI) [2022-03-11] ISSN: 2375-026X.

[62] Hyeyoung Ko, Suyeon Lee, Yoonseo Park, and Anna Choi. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*, 11(1):141, January 2022. ISSN 2079-9292. doi: 10.3390/electronics11010141. URL https://www.mdpi.com/2079-9292/11/1/141.

[63] Aminu Da'u and Naomie Salim. Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53(4):2709–2748, April 2020. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-019-09744-1. URL http://link.springer.com/10.1007/s10462-019-09744-1.

[64] 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.

[65] M Mehdi Afsar, Trafford Crump, and Behrouz Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys (CSUR)*, 2021.

[66] Gediminas Adomavicius, Konstantin Bauman, Alexander Tuzhilin, and Moshe Unger. Context-aware recommender systems: From foundations to recent developments. In *Recommender Systems Handbook*, pages 211–250. Springer, 2022.

[67] Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. Fairness in rankings and recommendations: an overview. *The VLDB Journal*, October 2021. ISSN 1066-8888, 0949-877X. doi: 10.1007/s00778-021-00697-y. URL https://link.springer.com/10.1007/s00778-021-00697-y.

[68] Yifan Wang, Weizhi Ma, Min Zhang*, Yiqun Liu, and Shaoping Ma. A survey on the fairness of recommender systems. *ACM Journal of the ACM (JACM)*, 2022.

[69] L Chen, F Hsu, M Chen, and Y 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0020025507004641.

[70] 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0957417408006532.

[71] Aparna Das, Claire Mathieu, and Daniel Ricketts. Maximizing profit using recommender systems. *arXiv:0908.3633 [cs]*, August 2009. URL http://arxiv.org/abs/0908.3633.

[72] Amos Azaria, Avinatan Hassidim, Sarit Kraus, Adi Eshkol, Ofer Weintraub, and Irit Netanely. Movie recommender system for profit maximization. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 121–128, Hong Kong China, October 2013. ACM. ISBN 978-1-4503-2409-0. doi: 10.1145/2507157.2507162. URL https://dl.acm.org/doi/10.1145/2507157.2507162.

[73] 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. URL https://onlinelibrary.wiley.com/doi/10.1111/poms.12540.

[74] Zhiwei Zhang, Ning Chen, Jun Wang, and Luo Si. SMART: Sponsored mobile app recommendation by balancing app downloads and appstore profit. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1600–1609, Boston, MA, December 2017. IEEE. ISBN 978-1-5386-2715-0. doi: 10.1109/BigData.2017.8258094. URL http://ieeexplore.ieee.org/document/8258094/.

[75] Edward C Malthouse, Khadija Ali Vakeel, Yasaman Kamyab Hessary, Robin Burke, and Morana Fuduric. A multistakeholder recommender systems algorithm for allocating sponsored recommendations. In *RMSE@ RecSys*, 2019.

[76] Phong Nguyen, John Dines, and Jan Krasnodebski. A Multi-Objective Learning to re-Rank Approach to Optimize Online Marketplaces for Multiple Stakeholders. *arXiv:1708.00651 [cs]*, August 2017. URL http://arxiv.org/abs/1708.00651.

[77] Liang Wu, Diane Hu, Liangjie Hong, and Huan Liu. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pages 365–374, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-5657-2. doi: 10.1145/3209978.3209993. URL https://doi.org/10.1145/3209978.3209993.

[78] 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 13th ACM Conference on Recommender Systems*, RecSys '19, pages 20–28, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3346998. URL https://doi.org/10.1145/3298689.3346998. 37 citations (Semantic Scholar/DOI) [2022-03-11].

[79] Kartik Hosanagar, Ramayya Krishnan, and Liye Ma. Recomended for you: The impact of profit incentives on the relevance of online recommendations. *ICIS 2008 Proceedings*, page 31, 2008.

[80] 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.

[81] Pengfei He, Haochen Liu, Xiangyu Zhao, Hui Liu, and Jiliang Tang. Propn: Personalized probabilistic strategic parameter optimization in recommendations. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3152–3161, 2022.

[82] Moran Beladev, Lior Rokach, and Bracha Shapira. Recommender systems for product bundling. *Knowledge-Based Systems*, 111:193–206, November 2016. ISSN 0950-7051. doi: 10.1016/j.knosys.2016.08.013. URL https://www.sciencedirect.com/science/article/pii/S0950705116302751.

[83] Toshihiro Kamishima and Shotaro Akaho. Personalized pricing recommender system: multi-stage epsilon-greedy approach. In *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems - HetRec '11*, pages 57–64, Chicago, Illinois, 2011. ACM Press. ISBN 978-1-4503-1027-7. doi: 10.1145/2039320.2039329. URL http://dl.acm.org/citation.cfm?doid=2039320.2039329.

[84] Yoon-Joo Park and Alexander Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08*, 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.

[85] 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. URL http://ieeexplore.ieee.org/document/6226399/.

[86] Saúl Vargas and Pablo Castells. Improving sales diversity by recommending users to items. In *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*, pages 145–152, Foster City, Silicon Valley, California, USA, 2014. ACM Press. ISBN 978-1-4503-2668-1. doi: 10.1145/2645710.2645744. URL http://dl.acm.org/citation.cfm?doid=2645710.2645744.

[87] 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0167923619301411.

[88] Jian Wang and Yi Zhang. Utilizing marginal net utility for recommendation in e-commerce. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11*, page 1003, Beijing, China, 2011. ACM Press. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2010050. URL http://portal.acm.org/citation.cfm?doid=2009916.2010050.

[89] 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 7th 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. URL https://dl.acm.org/doi/10.1145/2556195.2566589.

[90] Yongfeng Zhang, Qi Zhao, Yi Zhang, Daniel Friedman, Min Zhang, Yiqun Liu, and Shaoping Ma. Economic Recommendation with Surplus Maximization. In *Proceedings of the 25th International Conference on World Wide Web*, pages 73–83, Montréal Québec Canada, April 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4143-1. doi: 10.1145/2872427.2882973. URL https://dl.acm.org/doi/10.1145/2872427.2882973.

[91] Qi Zhao, Yongfeng Zhang, Yi Zhang, and Daniel Friedman. Multi-Product Utility Maximization for Economic Recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 435–443, Cambridge United Kingdom, February 2017. ACM. ISBN 978-1-4503-4675-7. doi: 10.1145/3018661.3018674. URL https://dl.acm.org/doi/10.1145/3018661.3018674.

[92] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Shijie Geng, Zuohui Fu, and Yongfeng Zhang. Maximizing Marginal Utility per Dollar for Economic Recommendation. In *The World Wide Web Conference on - WWW '19*, pages 2757–2763, San Francisco, CA, USA, 2019. ACM Press. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313725. URL http://dl.acm.org/citation.cfm?doid=3308558.3313725.

[93] Michael E Wall, Andreas Rechtsteiner, and Luis M Rocha. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer, 2003.

[94] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[95] Georgios Theocharous, Philip S Thomas, and Mohammad Ghavamzadeh. Personalized ad recommendation systems for life-time value optimization with guarantees. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[96] Jo Yong Ju. Reinforcement Learning for Profit Maximization of Recommender Systems. *Reinforcement Learning*, page 9, 2017.

[97] 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 2018 World Wide Web Conference on World Wide Web - WWW '18*, 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.

[98] 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 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2810–2818, Anchorage AK USA, July 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330668. URL https://dl.acm.org/doi/10.1145/3292500.3330668.

[99] Luo Ji, Qi Qin, Bingqing Han, and Hongxia Yang. Reinforcement Learning to Optimize Lifetime Value in Cold-Start Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 782–791, Virtual Event Queensland Australia, October 2021. ACM. ISBN 978-1-4503-8446-9. doi: 10.1145/3459637.3482292. URL https://dl.acm.org/doi/10.1145/3459637.3482292.

[100] 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. URL https://ojs.aaai.org/index.php/AAAI/article/view/16105.

[101] Mikael Hammar, Robin Karlsson, and Bengt J. Nilsson. Using maximum coverage to optimize recommendation systems in e-commerce. In *Proceedings of the 7th 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. URL https://dl.acm.org/doi/10.1145/2507157.2507169.

[102] Leman Akoglu and Christos Faloutsos. ValuePick: Towards a Value-Oriented Dual-Goal Recommender System. In *2010 IEEE International Conference on Data Mining Workshops*, pages 1151–1158, Sydney, TBD, Australia, December 2010. IEEE. ISBN 978-1-4244-9244-2. doi: 10.1109/ICDMW.2010.68. URL http://ieeexplore.ieee.org/document/5693424/.

[103] 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 29th ACM International Conference on Information & 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.

[104] Raphael Louca, Moumita Bhattacharya, Diane Hu, and Liangjie Hong. Joint optimization of profit and relevance for recommendation systems in e-commerce. In *RMSE@ RecSys*, 2019.

[105] Runyu Ma, Hantian Li, Jin Cen, and Amrinder Arora. Placement-and-Profit-Aware Association Rules Mining:. In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, pages 639–646, Prague, Czech Republic, 2019. SCITEPRESS - Science and Technology Publications. ISBN 978-989-758-350-6. doi: 10.5220/0007380606390646. URL http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0007380606390646.

[106] 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, 2013.

[107] Gaddiel Desirena, Armando Diaz, Jalil Desirena, Ismael Moreno, and Daniel Garcia. Maximizing customer lifetime value using stacked neural networks: An insurance industry application. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 541–544, 2019. doi: 10.1109/ICMLA.2019.00101.

[108] 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, 2022. ISSN 1567-4223. doi: https://doi.org/10.1016/j.elerap.2022.101195. URL https://www.sciencedirect.com/science/article/pii/S1567422322000783.

[109] 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, 2022.

[110] Bo Zhou and Tianxin Zou. Competing for recommendations: The strategic impact of personalized product recommendations in online marketplaces. *Marketing Science*, 2022.

[111] 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 *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 1859–1866, Guilin, July 2017. IEEE. ISBN 978-1-5386-2165-3. doi: 10.1109/FSKD.2017.8393050. URL https://ieeexplore.ieee.org/document/8393050/.

[112] Yuanchun Jiang and Yezheng Liu. Optimization of online promotion: a profit-maximizing model integrating price discount and product recommendation. *International Journal of Information Technology & Decision Making*, 11(05):961–982, September 2012. ISSN 0219-6220, 1793-6845. doi: 10.1142/S0219622012500289. URL https://www.worldscientific.com/doi/abs/10.1142/S0219622012500289.

[113] 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, 2015.

[114] Patrick Hosein, Inzamam Rahaman, Keanu Nichols, and Kiran Maharaj. Recommendations for long-term profit optimization. In *ImpactRS@RecSys*, 2019.

[115] 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. URL https://dl.acm.org/doi/10.1145/3237187.

[116] 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. URL https://www.sciencedirect.com/science/article/pii/S1567422321000466.

[117] Emanuele Cavenaghi, Lorenzo Camaione, Paolo Minasi, Gabriele Sottocornola, Fabio Stella, and Markus Zanker. An online experiment of a price-based re-rank algorithm. In *RecSys Workshop on Recommenders in Tourism (RecTour 2022)*, 2022.

[118] 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.

[119] 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0164121204001426.

[120] Duen-Ren Liu and Ya-Yueh Shih. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3):387–400, March 2005. ISSN 03787206. doi: 10.1016/j.im.2004.01.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0378720604000394.

[121] Ya-Yueh Shih and Duen-Ren Liu. Hybrid Recommendation Approaches: Collaborative Filtering via Valuable Content Information. *th Hawaii International Conference on System Sciences*, page 7, 2005.

[122] Y Shih and D 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0957417407002539.

[123] 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):77, 2012.

[124] 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.

[125] Kamelia Aryafar, Devin Guillory, and Liangjie Hong. An Ensemble-based Approach to Click-Through Rate Prediction for Promoted Listings at Etsy. In *Proceedings of the ADKDD'17*, pages 1–6, Halifax NS Canada, August 2017. ACM. ISBN 978-1-4503-5194-2. doi: 10.1145/3124749.3124758. URL https://dl.acm.org/doi/10.1145/3124749.3124758.

[126] 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 28th International Joint Conference on Artificial Intelligence*, pages 2301–2307, 2019.

[127] 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 27th International Conference on International Conference on Machine Learning*. Omnipress, 2010.

[128] 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 19th 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. URL https://dl.acm.org/doi/10.1145/2487575.2488200.

[129] 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.

[130] 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 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining - ADKDD'14*, pages 1–9, New York, NY, USA, 2014. ACM Press. ISBN 978-1-4503-2999-6. doi: 10.1145/2648584.2648589. URL http://dl.acm.org/citation.cfm?doid=2648584.2648589.

[131] 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 & 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.

[132] 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 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2078–2086, Anchorage AK USA, July 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330655. URL https://dl.acm.org/doi/10.1145/3292500.3330655.

[133] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web - WWW '10*, page 661, Raleigh, North Carolina, USA, 2010. ACM Press. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772758. URL http://portal.acm.org/citation.cfm?doid=1772690.1772758.

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

[135] Morteza Zihayat, Anteneh Ayanso, Xing Zhao, Heidar Davoudi, and Aijun An. A utility-based news recommendation system. *Decision Support Systems*, 117:14–27, 2019.

[136] 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.

[137] 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.

[138] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. News recommender systems–survey and roads ahead. *Information Processing & Management*, 54(6):1203–1227, 2018.

[139] Chong Feng, Muzammil Khan, Arif Ur Rahman, and Arshad Ahmad. News recommendation systems-accomplishments, challenges & future directions. *IEEE Access*, 8: 16702–16725, 2020.

[140] Tomoharu Iwata, Kazumi Saito, and Takeshi Yamada. Recommendation method for extending subscription periods. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 574, Philadelphia, PA, USA, 2006. ACM Press. ISBN 978-1-59593-339-3. doi: 10.1145/1150402.1150470. URL http://portal.acm.org/citation.cfm?doid=1150402.1150470.

[141] T. Iwata, K. Saito, and T. 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. URL http://ieeexplore.ieee.org/document/4479464/.

[142] 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. URL https://linkinghub.elsevier.com/retrieve/pii/S0950705116300600.

[143] Arda Antikacioglu and R. Ravi. Post Processing Recommender Systems for Diversity. In *Proceedings of the 23rd 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. URL https://dl.acm.org/doi/10.1145/3097983.3098173.

[144] Elaheh Malekzadeh Hamedani and Marjan Kaedi. Recommending the long tail items through personalized diversification. *Knowledge-Based Systems*, 164:348–357, 2019.

[145] 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. URL https://misq.org/welfare-properties-of-profit-maximizing-recommender-systems-theory-and-results-html.

[146] Carlos A Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015.

[147] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems–a survey. *Knowledge-based systems*, 123:154–162, 2017.

[148] Microsoft Corporation. Example Database FoodMart of Microsoft Analysis Services, 1998. URL https://github.com/julianhyde/foodmart-data-hsqldb.

[149] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1018. URL https://aclanthology.org/D19-1018.

[150] Ventatesan. Dunnhumby - the complete journey, 2007. URL https://www.dunnhumby.com/source-files/.

[151] 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 *Joint European conference on machine learning and knowledge discovery in databases*, pages 36–40. Springer, 2016.

[152] Pisharath. NU-MineBench version 2.0 dataset and technical report. *URI: http://cucis. ece. northwestern. edu/projects/DMS/MineBench. html (last access on 2 March 2015)*, 2005.

[153] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 61–70, 2002.

[154] 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. URL https://dl.acm.org/doi/10.1145/2827872.

[155] James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. Citeseer, 2007.

[156] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32, 2005.

[157] Brian McFee, Thierry Bertin-Mahieux, Daniel PW Ellis, and Gert RG Lanckriet. The million song dataset challenge. In *Proceedings of the 21st International Conference on World Wide Web*, pages 909–916, 2012.

[158] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *information retrieval*, 4(2):133–151, 2001.

[159] Julian McAuley, Rahul Pandey, and Jure Leskovec. Inferring Networks of Substitutable and Complementary Products. In *Proceedings of the 21th 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. URL https://dl.acm.org/doi/10.1145/2783258.2783381.

[160] Xingyue Zhang, James A Dearden, and Yuliang Yao. Let them stay or let them go? online retailer pricing strategy for managing stockouts. *Production and Operations Management*, 2022.

[161] 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.

[162] Asela Gunawardana, Guy Shani, and Sivan Yogev. Evaluating recommender systems. In *Recommender systems handbook*, pages 547–601. Springer, 2022.

[163] Davinder Kaur, Suleyman Uslu, Kaley J Rittichier, and Arjan Durresi. Trustworthy artificial intelligence: a review. *ACM Computing Surveys (CSUR)*, 55(2):1–38, 2022.

[164] 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.

[165] Nava Tintarev and Judith Masthoff. Explaining Recommendations: Design and Evaluation. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 353–382. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7637-6. doi: 10.1007/978-1-4899-7637-6_10. URL https://doi.org/10.1007/978-1-4899-7637-6_10.

[166] Yongfeng Zhang and Xu Chen. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval*, 14(1):1–101, 2020. ISSN 1554-0669, 1554-0677. doi: 10.1561/1500000066. URL http://arxiv.org/abs/1804.11192. 297 citations (Semantic Scholar/arXiv) [2022-02-14] 297 citations (Semantic Scholar/DOI) [2022-02-14] arXiv: 1804.11192.

[167] Giacomo Balloccu, Ludovico Boratto, Gianni Fenu, and Mirko Marras. XRecSys: A framework for path reasoning quality in explainable recommendation. *Software Impacts*, 14:100404, December 2022. ISSN 26659638. doi: 10.1016/j.simpa.2022.100404. URL https://linkinghub.elsevier.com/retrieve/pii/S266596382200094X.

[168] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, January 2004. ISSN 1046-8188, 1558-2868. doi: 10.1145/963770.963772. URL https://dl.acm.org/doi/10.1145/963770.963772.

[169] Asela Gunawardana and Guy Shani. Evaluating Recommender Systems. In Francesco Ricci, Lior Rokach, and Bracha Shapira, editors, *Recommender Systems Handbook*, pages 265–308. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7637-6. doi: 10.1007/978-1-4899-7637-6_8. URL https://doi.org/10.1007/978-1-4899-7637-6_8.

[170] Thiago Silveira, Min Zhang, Xiao Lin, Yiqun Liu, and Shaoping Ma. How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5):813–831, May 2019. ISSN 1868-808X. doi: 10.1007/s13042-017-0762-9. URL https://doi.org/10.1007/s13042-017-0762-9.

[171] Janneth Chicaiza and Priscila Valdiviezo-Diaz. Explainable Recommender Systems: From Theory to Practice. In Atulya K. Nagar, Dharm Singh Jat, Durgesh Kumar Mishra, and Amit Joshi, editors, *Intelligent Sustainable Systems*, Lecture Notes in Networks and Systems, pages 449–459, Singapore, 2023. Springer Nature. ISBN 978-981-19766-3-6. doi: 10.1007/978-981-19-7663-6_42.

[172] Xu Chen, Yongfeng Zhang, and Ji-Rong Wen. Measuring "Why" in Recommender Systems: a Comprehensive Survey on the Evaluation of Explainable Recommendation, February 2022. URL http://arxiv.org/abs/2202.06466. Number: arXiv:2202.06466 arXiv:2202.06466 [cs].

[173] Behnoush Abdollahi and Olfa Nasraoui. Explainable Matrix Factorization for Collaborative Filtering. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 5–6, Republic and Canton of Geneva, CHE, April 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4144-8. doi: 10.1145/2872518.2889405. URL https://doi.org/10.1145/2872518.2889405.

[174] Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. TEM: Tree-enhanced Embedding Model for Explainable Recommendation. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1543–1552, Republic and Canton of Geneva, CHE, April 2018. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-5639-8. doi: 10.1145/3178876.3186066. URL https://doi.org/10.1145/3178876.3186066.

[175] Yu-Yao Liu, Bo Yang, Hong-Bin Pei, and Jing Huang. Neural Explainable Recommender Model Based on Attributes and Reviews. *Journal of Computer Science and Technology*, 35(6):1446–1460, November 2020. ISSN 1860-4749. doi: 10.1007/s11390-020-0152-8. URL https://doi.org/10.1007/s11390-020-0152-8.

[176] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems*, 11(3-4):24:1–24:45, 2021. ISSN 2160-6455. doi: 10.1145/3387166. URL https://doi.org/10.1145/3387166.

[177] Amina Adadi and Mohammed Berrada. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2870052. Conference Name: IEEE Access.

[178] Alexandra Vultureanu-Albişi and Costin Bădică. A survey on effects of adding explanations to recommender systems. *Concurrency and Computation: Practice and Experience*, 34(20):e6834, 2022. ISSN 1532-0634. doi: 10.1002/cpe.6834. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6834. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.6834.

[179] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, CSCW '00, pages 241–250, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 978-1-58113-222-9. doi: 10.1145/358916.358995. URL https://doi.org/10.1145/358916.358995.

[180] Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*, IUI '09, pages 47–56, New York, NY, USA, February 2009. Association for Computing Machinery. ISBN 978-1-60558-168-2. doi: 10.1145/1502650.1502661. URL https://doi.org/10.1145/1502650.1502661.

[181] Florian Pecune, Shruti Murali, Vivian Tsai, Yoichi Matsuyama, and Justine Cassell. A Model of Social Explanations for a Conversational Movie Recommendation System. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, HAI '19, pages 135–143, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6922-0. doi: 10.1145/3349537.3351899. URL https://doi.org/10.1145/3349537.3351899. 23 citations (Semantic Scholar/DOI) [2022-02-14].

[182] Weiquan Wang and Izak Benbasat. Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs. *Journal of Manage-

ment Information Systems, 23(4):217–246, 2007. ISSN 0742-1222. doi: 10.2753/MIS0742-1222230410. URL https://doi.org/10.2753/MIS0742-1222230410.

[183] Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, 24(3):555–583, May 2012. ISSN 1573-756X. doi: 10.1007/s10618-011-0215-0. URL https://doi.org/10.1007/s10618-011-0215-0.

[184] Nava Tintarev and Judith Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4):399–439, October 2012. ISSN 1573-1391. doi: 10.1007/s11257-011-9117-5. URL https://doi.org/10.1007/s11257-011-9117-5.

[185] Mouadh Guesmi, Mohamed Amine Chatti, Laura Vorgerd, Shoeb Joarder, Shadi Zumor, Yiqi Sun, Fangzheng Ji, and Arham Muslim. On-demand Personalized Explanation for Transparent Recommendation. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 246–252. Association for Computing Machinery, New York, NY, USA, 2021. ISBN 978-1-4503-8367-7. URL https://doi.org/10.1145/3450614.3464479.

[186] Ruijing Zhao, Izak Benbasat, and Hasan Cavusoglu. Transparency in Advice-Giving Systems: A Framework and a Research Model for Transparency Provision. volume 2327 of *CEUR Workshop Proceedings*, 2019. URL https://ceur-ws.org/Vol-2327/IUI19WS-IUIATEC-2.pdf. ISSN: 1613-0073.

[187] Mouadh Guesmi, Mohamed Amine Chatti, Laura Vorgerd, Shoeb Joarder, Qurat Ul Ain, Thao Ngo, Shadi Zumor, Yiqi Sun, Fangzheng Ji, and Arham Muslim. Input or Output: Effects of Explanation Focus on the Perception of Explainable Recommendation with Varying Level of Details. In *Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*, volume 2948 of *CEUR Workshop Proceedings*, pages 55–72, Online Event, September, September 2021. CEUR. URL https://ceur-ws.org/Vol-2948/#paper4. ISSN: 1613-0073.

[188] Giulia Vilone and Luca Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76(C):89–106, 2021. ISSN

1566-2535. doi: 10.1016/j.inffus.2021.05.009. URL https://doi.org/10.1016/j.inffus.2021.05.009.

[189] Upol Ehsan and Mark O. Riedl. Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In Constantine Stephanidis, Masaaki Kurosu, Helmut Degen, and Lauren Reinerman-Jones, editors, *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, Lecture Notes in Computer Science, pages 449–466, Cham, 2020. Springer International Publishing. ISBN 978-3-030-60117-1. doi: 10.1007/978-3-030-60117-1_33.

[190] Lauren Shluzas, Joel A. Sadler, Rebecca Currano, Martin Steinert, Riitta Katila, and Taryn Sanks. Comparing Novice and Expert User Inputs in Early Stage Product Design, December 2012. URL https://papers.ssrn.com/abstract=2372677.

[191] Donghee Shin. How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109:106344, August 2020. ISSN 0747-5632. doi: 10.1016/j.chb.2020.106344. URL https://www.sciencedirect.com/science/article/pii/S0747563220300984.

[192] Chun-Hua Tsai and Peter Brusilovsky. Explaining Social Recommendations to Casual Users: Design Principles and Opportunities. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, pages 1–2, New York, NY, USA, March 2018. Association for Computing Machinery. ISBN 978-1-4503-5571-1. doi: 10.1145/3180308.3180368. URL https://doi.org/10.1145/3180308.3180368.

[193] Vicente Dominguez, Pablo Messina, Ivania Donoso-Guzmán, and Denis Parra. The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 408–416, New York, NY, USA, March 2019. Association for Computing Machinery. ISBN 978-1-4503-6272-6. doi: 10.1145/3301275.3302274. URL https://doi.org/10.1145/3301275.3302274.

[194] Erasmo Purificato, Baalakrishnan Aiyer Manikandan, Prasanth Vaidya Karanam, Mahantesh Vishvanath Pattadkal, and Ernesto William De Luca. Evaluating Explainable Interfaces for a Knowledge Graph-Based Recommender System. In Peter Brusilovsky,

Marco de Gemmis, Alexander Felfernig, Elisabeth Lex, Pasquale Lops, Giovanni Semeraro, and Martijn C. Willemsen, editors, *Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*, volume 2948 of *CEUR Workshop Proceedings*, pages 73–88, Online Event, September, September 2021. CEUR. URL https://ceur-ws.org/Vol-2948/#paper5. ISSN: 1613-0073.

[195] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2119–2128, New York, NY, USA, April 2009. Association for Computing Machinery. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1519023. URL https://doi.org/10.1145/1518701.1519023.

[196] Yongfeng Zhang, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservables. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1189–1198, New York, NY, USA, November 2014. Association for Computing Machinery. ISBN 978-1-4503-2598-1. doi: 10.1145/2661829.2661976. URL https://doi.org/10.1145/2661829.2661976.

[197] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, 10(4):31:1–31:40, November 2020. ISSN 2160-6455. doi: 10.1145/3365843. URL https://doi.org/10.1145/3365843.

[198] Shlomo Berkovsky, Ronnie Taib, and Dan Conway. How to Recommend? User Trust Factors in Movie Recommender Systems. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, pages 287–300, New York, NY, USA, March 2017. Association for Computing Machinery. ISBN 978-1-4503-4348-0. doi: 10.1145/3025171.3025209. URL https://doi.org/10.1145/3025171.3025209.

[199] Tim Miller. Explanation in Artificial Intelligence: Insights from the Social Sciences. *arXiv:1706.07269 [cs]*, August 2018. URL http://arxiv.org/abs/1706.07269. arXiv: 1706.07269.

[200] Tom M. van Engers and Dennis M. de Vries. Governmental Transparency in the Era of Artificial Intelligence. *Legal Knowledge and Information Systems*, pages 33–42, 2019. doi: 10.3233/FAIA190304. URL https://ebooks.iospress.nl/doi/10.3233/FAIA190304. Publisher: IOS Press.

[201] Francesco Sovrano, Fabio Vitali, and Monica Palmirani. Modelling GDPR-Compliant Explanations for Trustworthy AI. In Andrea Kő, Enrico Francesconi, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, editors, *Electronic Government and the Information Systems Perspective*, Lecture Notes in Computer Science, pages 219–233, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58957-8. doi: 10.1007/978-3-030-58957-8_16.

[202] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, June 2020. ISSN 1566-2535. doi: 10.1016/j.inffus.2019.12.012. URL https://www.sciencedirect.com/science/article/pii/S1566253519308103.

[203] Alexandra Vultureanu-Albişi and Costin Bădică. Improving Students' Performance by Interpretable Explanations using Ensemble Tree-Based Approaches. In *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 215–220, May 2021. doi: 10.1109/SACI51354.2021.9465558.

[204] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939778. URL https://doi.org/10.1145/2939672.2939778.

[205] Nasim Sonboli, Jessie J. Smith, Florencia Cabral Berenfus, Robin Burke, and Casey Fiesler. Fairness and Transparency in Recommendation: The Users' Perspective. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '21, pages 274–279, New York, NY, USA, 2021. Association for Com-

puting Machinery. ISBN 978-1-4503-8366-0. doi: 10.1145/3450613.3456835. URL https://doi.org/10.1145/3450613.3456835.

[206] Larissa Chazette and Kurt Schneider. Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering*, 25(4):493–514, December 2020. ISSN 1432-010X. doi: 10.1007/s00766-020-00333-1. URL https://doi.org/10.1007/s00766-020-00333-1.

[207] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, SIGIR '14, pages 83–92, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 978-1-4503-2257-7. doi: 10.1145/2600428.2609579. URL https://doi.org/10.1145/2600428.2609579.

[208] Juntao Tan, Shuyuan Xu, Yingqiang Ge, Yunqi Li, Xu Chen, and Yongfeng Zhang. Counterfactual Explainable Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, pages 1784–1793, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 978-1-4503-8446-9. doi: 10.1145/3459637.3482420. URL https://doi.org/10.1145/3459637.3482420.

[209] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 311, Philadelphia, Pennsylvania, 2001. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL http://portal.acm.org/citation.cfm?doid=1073083.1073135.

[210] Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013.

[211] Lei Li, Yongfeng Zhang, and Li Chen. Generate Neural Template Explanations for Recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, pages 755–764, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-6859-9. doi:

10.1145/3340531.3411992. URL https://doi.org/10.1145/3340531.3411992. 26 citations (Semantic Scholar/DOI) [2022-02-14].

[212] Georgina Peake and Jun Wang. Explanation Mining: Post Hoc Interpretability of Latent Factor Models for Recommendation Systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 2060–2069, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-5552-0. doi: 10.1145/3219819.3220072. URL https://doi.org/10.1145/3219819.3220072.

[213] Ludovik Coba, Panagiotis Symeonidis, and Markus Zanker. Personalised novel and explainable matrix factorisation. *Data & Knowledge Engineering*, 122:142–158, July 2019. ISSN 0169-023X. doi: 10.1016/j.datak.2019.06.003. URL https://www.sciencedirect.com/science/article/pii/S0169023X1830332X.

[214] Shijie Geng, Zuohui Fu, Juntao Tan, Yingqiang Ge, Gerard de Melo, and Yongfeng Zhang. Path Language Modeling over Knowledge Graphsfor Explainable Recommendation. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pages 946–955, New York, NY, USA, April 2022. Association for Computing Machinery. ISBN 978-1-4503-9096-5. doi: 10.1145/3485447.3511937. URL https://dl.acm.org/doi/10.1145/3485447.3511937.

[215] Giacomo Balloccu, Ludovico Boratto, Gianni Fenu, and Mirko Marras. Post Processing Recommender Systems with Knowledge Graphs for Recency, Popularity, and Diversity of Explanations. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, pages 646–656, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 978-1-4503-8732-3. doi: 10.1145/3477495.3532041. URL https://dl.acm.org/doi/10.1145/3477495.3532041.

[216] Giacomo Balloccu, Ludovico Boratto, Gianni Fenu, and Mirko Marras. Reinforcement recommendation reasoning through knowledge graphs for explanation path quality. *Knowledge-Based Systems*, 260:110098, January 2023. ISSN 0950-7051. doi: 10.1016/j.knosys.2022.110098. URL https://www.sciencedirect.com/science/article/pii/S0950705122011947.

[217] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. doi: 10.1109/MC.2009.263.

[218] 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*, 35(2):1609–1623, February 2023. ISSN 1558-2191. doi: 10.1109/TKDE.2021.3091160. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[219] 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 1567-4223. doi: 10.1016/j.elerap.2022.101195. URL https://www.sciencedirect.com/science/article/pii/S1567422322000783.

[220] Robert Garfinkel, Ram Gopal, Bhavik Pathak, Rajkumar Venkatesan, and Fang Yin. Empirical Analysis of the Business Value of Recommender Systems. doi: 10.2139/SSRN.958770.

[221] Sofia Gkika and George Lekakos. The Persuasive Role of Explanations in Recommender Systems. In Anssi Öörni, Saskia Kelders, Lisette van Gemert-Pijnen, and Harri Oinas-Kukkonen, editors, *Proceedings of the Second International Workshop on Behavior Change Support Systems*, volume 1153 of *CEUR Workshop Proceedings*, pages 59–68, Padua, Italy, May 2014. CEUR. URL https://ceur-ws.org/Vol-1153/#Paper_6. ISSN: 1613-0073.

[222] 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):11:1–11:41, 2012. ISSN 2160-6455. doi: 10.1145/2209310.2209314. URL https://doi.org/10.1145/2209310.2209314.

[223] 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*, 275:110699, September 2023. ISSN 0950-7051. doi: 10.1016/j.knosys.2023.110699. URL https://www.sciencedirect.com/science/article/pii/S0950705123004495.

[224] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 417–426, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-6014-2. doi: 10.1145/3269206.3271739. URL https://dl.acm.org/doi/10.1145/3269206.3271739.

[225] Qianyu Li, Xiaoli Tang, Tengyun Wang, Haizhi Yang, and Hengjie Song. Unifying Task-Oriented Knowledge Graph Learning and Recommendation. *IEEE Access*, 7:115816–115828, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2932466. Conference Name: IEEE Access.

[226] Xiao Sha, Zhu Sun, and Jie Zhang. Hierarchical attentive knowledge graph embedding for personalized recommendation. *Electronic Commerce Research and Applications*, 48: 101071, July 2021. ISSN 1567-4223. doi: 10.1016/j.elerap.2021.101071. URL https://www.sciencedirect.com/science/article/pii/S1567422321000430.

[227] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S. Yu. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 31(2):357–370, February 2019. ISSN 1558-2191. doi: 10.1109/TKDE.2018.2833443. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[228] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 297–305, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-5901-6. doi: 10.1145/3240323.3240361. URL https://dl.acm.org/doi/10.1145/3240323.3240361.

[229] Xiaoli Tang, Tengyun Wang, Haizhi Yang, and Hengjie Song. AKUPM: Attention-Enhanced Knowledge-Aware User Preference Model for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 1891–1899, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330705. URL https://dl.acm.org/doi/10.1145/3292500.3330705.

[230] Qingyao Ai, Vahid Azizi, Xu Chen, and Yongfeng Zhang. Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation. *Algorithms*, 11(9): 137, September 2018. ISSN 1999-4893. doi: 10.3390/a11090137. URL http://www.mdpi.com/1999-4893/11/9/137.

[231] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preferences. In *The World Wide Web Conference*, WWW '19, pages 151–161, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313705. URL https://dl.acm.org/doi/10.1145/3308558.3313705.

[232] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y. Chang. Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pages 505–514, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-5657-2. doi: 10.1145/3209978.3210017. URL https://dl.acm.org/doi/10.1145/3209978.3210017.

[233] Zuoxi Yang and Shoubin Dong. HAGERec: Hierarchical Attention Graph Convolutional Network Incorporating Knowledge Graph for Explainable Recommendation. *Knowledge-Based Systems*, 204:106194, September 2020. ISSN 0950-7051. doi: 10.1016/j.knosys.2020.106194. URL https://www.sciencedirect.com/science/article/pii/S0950705120304196.

[234] Weizhi Ma, Min Zhang, Yue Cao, Woojeong Jin, Chenyang Wang, Yiqun Liu, Shaoping Ma, and Xiang Ren. Jointly Learning Explainable Rules for Recommendation with Knowledge Graph. In *The World Wide Web Conference*, pages 1210–1221, San Francisco CA USA, May 2019. ACM. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313607. URL https://dl.acm.org/doi/10.1145/3308558.3313607.

[235] Xiting Wang, Kunpeng Liu, Dongjie Wang, Le Wu, Yanjie Fu, and Xing Xie. Multi-level Recommendation Reasoning over Knowledge Graphs with Reinforcement Learning. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pages 2098–2108, New York, NY, USA, April 2022. Association for Computing Machinery. ISBN 978-1-4503-9096-5. doi: 10.1145/3485447.3512083. URL https://dl.acm.org/doi/10.1145/3485447.3512083.

[236] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable Reasoning over Knowledge Graphs for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):5329–5336, July 2019. ISSN 2374-3468. doi: 10.1609/aaai.v33i01.33015329. URL https://ojs.aaai.org/index.php/AAAI/article/view/4470. Number: 01.

[237] Xiaowen Huang, Quan Fang, Shengsheng Qian, Jitao Sang, Yan Li, and Changsheng Xu. Explainable Interaction-driven User Modeling over Knowledge Graph for Sequential Recommendation. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, pages 548–556, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6889-6. doi: 10.1145/3343031.3350893. URL https://doi.org/10.1145/3343031.3350893.

[238] Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. Reinforcement Knowledge Graph Reasoning for Explainable Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, pages 285–294, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6172-9. doi: 10.1145/3331184.3331203. URL https://doi.org/10.1145/3331184.3331203.

[239] Yikun Xian, Zuohui Fu, Handong Zhao, Yingqiang Ge, Xu Chen, Qiaoying Huang, Shijie Geng, Zhou Qin, Gerard de Melo, S. Muthukrishnan, and Yongfeng Zhang. CAFE: Coarse-to-Fine Neural Symbolic Reasoning for Explainable Recommendation. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1645–1654, October 2020. doi: 10.1145/3340531.3412038. URL http://arxiv.org/abs/2010.15620. 28 citations (Semantic Scholar/arXiv) [2022-02-14] 28 citations (Semantic Scholar/DOI) [2022-02-14] arXiv: 2010.15620.

[240] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 950–958, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330989. URL https://dl.acm.org/doi/10.1145/3292500.3330989.

[241] Hongxu Chen, Yicong Li, Xiangguo Sun, Guandong Xu, and Hongzhi Yin. Temporal Meta-path Guided Explainable Recommendation. In *Proceedings of the 14th ACM In-*

ternational Conference on Web Search and Data Mining, WSDM '21, pages 1056–1064, New York, NY, USA, March 2021. Association for Computing Machinery. ISBN 978-1-4503-8297-7. doi: 10.1145/3437963.3441762. URL https://dl.acm.org/doi/10.1145/3437963.3441762.

[242] Weiping Song, Zhijian Duan, Ziqing Yang, Hao Zhu, Ming Zhang, and Jian Tang. Ekar: An Explainable Method for Knowledge Aware Recommendation. *arXiv:1906.09506 [cs]*, January 2022. URL http://arxiv.org/abs/1906.09506. arXiv: 1906.09506.

[243] Xin Xin, Xiangnan He, Yongfeng Zhang, Yongdong Zhang, and Joemon Jose. Relational Collaborative Filtering: Modeling Multiple Item Relations for Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 125–134, Paris France, July 2019. ACM. ISBN 978-1-4503-6172-9. doi: 10.1145/3331184.3331188. URL https://dl.acm.org/doi/10.1145/3331184.3331188.

[244] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S. Yu. Leveraging Meta-path based Context for Top- N Recommendation with A Neural Co-Attention Model. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 1531–1540, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 978-1-4503-5552-0. doi: 10.1145/3219819.3219965. URL https://dl.acm.org/doi/10.1145/3219819.3219965.

[245] Qinqin Wang, Elias Tragos, Neil Hurley, Barry Smyth, Aonghus Lawlor, and Ruihai Dong. Entity-Enhanced Graph Convolutional Network for Accurate and Explainable Recommendation. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '22, pages 79–88, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 978-1-4503-9207-5. doi: 10.1145/3503252.3531316. URL https://dl.acm.org/doi/10.1145/3503252.3531316.

[246] Ting Ma, Longtao Huang, Qianqian Lu, and Songlin Hu. KR-GCN: Knowledge-Aware Reasoning with Graph Convolution Network for Explainable Recommendation. *ACM Transactions on Information Systems*, 41(1):4:1–4:27, 2023. ISSN 1046-8188. doi: 10.1145/3511019. URL https://dl.acm.org/doi/10.1145/3511019.

[247] Ziyu Lyu, Yue Wu, Junjie Lai, Min Yang, Chengming Li, and Wei Zhou. Knowledge Enhanced Graph Neural Networks for Explainable Recommendation. *IEEE Transactions*

on Knowledge and Data Engineering, 35(5):4954–4968, May 2023. ISSN 1558-2191. doi: 10.1109/TKDE.2022.3142260. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[248] Chang-You Tai, Liang-Ying Huang, Chien-Kun Huang, and Lun-Wei Ku. User-Centric Path Reasoning towards Explainable Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, pages 879–889, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 978-1-4503-8037-9. doi: 10.1145/3404835.3462847. URL https://dl.acm.org/doi/10.1145/3404835.3462847.

[249] Yuyue Zhao, Xiang Wang, Jiawei Chen, Yashen Wang, Wei Tang, Xiangnan He, and Haiyong Xie. Time-aware Path Reasoning on Knowledge Graph for Recommendation. *ACM Transactions on Information Systems*, 41(2):26:1–26:26, 2022. ISSN 1046-8188. doi: 10.1145/3531267. URL https://dl.acm.org/doi/10.1145/3531267.

[250] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. Do Graph Neural Networks Build Fair User Models? Assessing Disparate Impact and Mistreatment in Behavioural User Profiling. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pages 4399–4403, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 978-1-4503-9236-5. doi: 10.1145/3511808.3557584. URL https://doi.org/10.1145/3511808.3557584.

[251] Christian Scheel, Angel Castellanos, Thebin Lee, and Ernesto William De Luca. The Reason Why: A Survey of Explanations for Recommender Systems. In Andreas Nürnberger, Sebastian Stober, Birger Larsen, and Marcin Detyniecki, editors, *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation*, Lecture Notes in Computer Science, pages 67–84, Cham, 2014. Springer International Publishing. ISBN 978-3-319-12093-5. doi: 10.1007/978-3-319-12093-5_3.

[252] Yunfei Chu, Jiangchao Yao, Chang Zhou, and Hongxia Yang. Graph Neural Networks in Modern Recommender Systems. In Lingfei Wu, Peng Cui, Jian Pei, and Liang Zhao, editors, *Graph Neural Networks: Foundations, Frontiers, and Applications*, pages 423–445. Springer Nature, Singapore, 2022. ISBN 9789811660542. doi: 10.1007/978-981-16-6054-2_19. URL https://doi.org/10.1007/978-981-16-6054-2_19.

[253] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph Neural Networks in Recommender Systems: A Survey. *ACM Computing Surveys*, 55(5):97:1–97:37, 2022. ISSN 0360-0300. doi: 10.1145/3535101. URL https://dl.acm.org/doi/10.1145/3535101.

[254] 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):3:1–3:51, March 2023. doi: 10.1145/3568022. URL https://dl.acm.org/doi/10.1145/3568022.

[255] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in Graph Neural Networks: A Taxonomic Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5782–5799, May 2023. ISSN 1939-3539. doi: 10.1109/TPAMI.2022.3204236. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[256] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3549–3568, August 2022. ISSN 1558-2191. doi: 10.1109/TKDE.2020.3028705. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[257] Giacomo Balloccu, Ludovico Boratto, Christian Cancedda, Gianni Fenu, and Mirko Marras. Knowledge is Power, Understanding is Impact: Utility and Beyond Goals, Explanation Quality, and Fairness in Path Reasoning Recommendation. In Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo, editors, *Advances in Information Retrieval*, Lecture Notes in Computer Science, pages 3–19, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-28241-6. doi: 10.1007/978-3-031-28241-6_1.

[258] Sandeep K. Raghuwanshi and R. K. Pateriya. Recommendation Systems: Techniques, Challenges, Application, and Evaluation. In Jagdish Chand Bansal, Kedar Nath Das, Atulya Nagar, Kusum Deep, and Akshay Kumar Ojha, editors, *Soft Computing for Problem Solving*, Advances in Intelligent Systems and Computing, pages 151–164, Singapore, 2019. Springer. ISBN 9789811315954. doi: 10.1007/978-981-13-1595-4_12.

[259] Shuo Chang, F. Maxwell Harper, and Loren Gilbert Terveen. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, pages 175–182, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959153. URL https://doi.org/10.1145/2959100.2959153. 77 citations (Semantic Scholar/DOI) [2022-10-07] 44 citations (Crossref) [2022-10-07].

[260] Felipe Costa, Sixun Ouyang, Peter Dolog, and Aonghus Lawlor. Automatic Generation of Natural Language Explanations. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, pages 1–2, New York, NY, USA, March 2018. Association for Computing Machinery. ISBN 978-1-4503-5571-1. doi: 10.1145/3180308.3180366. URL https://doi.org/10.1145/3180308.3180366.

[261] Marta Caro-Martínez, Guillermo Jiménez-Díaz, and Juan A. Recio-García. Conceptual Modeling of Explainable Recommender Systems: An Ontological Formalization to Guide Their Design and Development. *Journal of Artificial Intelligence Research*, 71:557–589, July 2021. ISSN 1076-9757. doi: 10.1613/jair.1.12789. URL https://jair.org/index.php/jair/article/view/12789.

[262] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, and Gerard de Melo. Fairness-Aware Explainable Recommendation over Knowledge Graphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, pages 69–78, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-8016-4. doi: 10.1145/3397271.3401051. URL https://doi.org/10.1145/3397271.3401051.