People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

**IBN KHALDOUN UNIVERSITY OF TIARET**

# Dissertation

Presented to:

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE

in order to obtain the degree of :

**MASTER**

Specialty: Software Engineering

Presented by:

**BARIK Mohamed Noureddine**

**LASFER Yaaqoub**

On the theme:

# Developing a Drug Recommendation System: Design, Implementation and Evaluation

Defended publicly on 16/06/2025 in Tiaret in front the jury composed of:

| | | | |
|---|---|---|---|
| Mr MEGHAZI Hadj Madani | MCB | Tiaret University | Chairman |
| Mr BOUDAA Boudjemaa | MCA | Tiaret University | Supervisor |
| Mrs ZENATI Anissa | MCB | Tiaret University | Examiner |

2024-2025

# Abstract

The choice of the right drug remains a complex decision for physicians, despite floods of patient data and continuously growing drug lists. We see recommendation systems helping in many online spaces. But when it comes to health, especially drug choices, these systems need to be much alert. Many current drug recommendation systems face problems and challenges. This turns their suggestions to be not precise or clinically helpful as they could be.

This research addresses precisely these issues, by building a new kind of drug recommendation system, rooted in deep learning. By adopting the highly regarded Neural Collaborative Filtering (NCF) model and significantly improved it. The system does not just look at demographics; it digs into what patients say in their reviews using sentiment analysis and groups similar patient experiences together through a custom clustering. The concept is simple: combine clinical data with patient feedback to get a more complete picture, providing more personalized and accurate drug recommendations.

The system is tested with a large, real-world collection of drug reviews. This approach clearly did better than a range of existing methods in important measures such as accuracy, precision, and F1 score. This work shows a promising path toward making drug recommendations genuinely more helpful and precise.

**Keywords:** Recommender System, Healthcare, Drug Recommendation System, Neural Collaborative Filtering, Sentiment Analysis, U-KMeans Clustering.

# Dedication

.

To my beloved parents ***Assid Fatma*** and ***Abdalkader***, your unconditional love, constant support, and belief in me have inspired and paved the way for all my achievements. This work is a testament to your enduring presence in my life. *Thank you for encouraging me.*

To my supportive siblings, ***Sadek, Sliman, Hadjer, Akila***, for your inspiration, companionship, understanding, and for always believing in me.

To ***Khaled, Nasreddine, Abdelbasset, Hocine, Moncef, Amine***, and all my ***friends*** and ***colleagues*** for their invaluable companionship, for celebrating my successes, and for standing by me during challenging times. Your friendship and support have been a constant source of strength.

***This accomplishment is a tribute to all of you who have shaped who I am today. May Allah grant healing to the sick and mercy on the departed. With heartfelt gratitude and sincere thanks***

***Barik Mohamed Noureddine***

To my dear parents ***Lounis Samira*** and ***Abdenour***, thank you for all you have done for me and for your dedication and sacrifices for me. Thank you for being there and supporting me. Thank you ***Mam***, thank you ***Dad***, You are the pillar of who I am today, and even if I sometimes fail to express it, I am lost without you. All my thanks are not enough for everything you have given me.

To my dear brothers ***Muhammad*** and ***Mahdi***, and my sister ***Maryem***, I thank you for your support and encouragement all these years, these few words are not enough to describe you but I take advantage of this passage to thank you and express to you all my gratitude for what you have done for me.

To my family, the ***Lasfer*** and ***Lounis*** family, thank you for being there and for supporting, encouraging and helping me when I needed it.

To my grandfather ***Rachid*** and my grandmother ***Assila*** thank you for all the love, continued support, and encouragement. I ask Allah grant you all health ,wellness and longevity.

To ***Marzouk, Yacine***, and all my ***friends*** for your invaluable companionship, for celebrating my successes, and for standing by me during challenging times. Your friendship and support have been a constant source of strength.

***I would also like to extend my sincere thanks to all the people who have helped and supported me from near and far. May Allah grant you all health and wellness***

**Lasfer Yaaqoub**

# Acknowledgments

We express our profound gratitude to ***Allah***, the Most Gracious and the Most Merciful, for bestowing on us the strength, guidance, and perseverance to complete this research journey successfully.

Our sincere and heartfelt gratitude extends to our esteemed supervisor,
Dr. ***Boudaa Boudjemaa***. His guidance, support, patience, and profound expertise were essential in shaping this work. We consider it a great privilege and honor to have worked under his guidance.

We would also like to thank the distinguished members of the jury, Dr. ***Meghazi Hadj Madani*** and Dr. ***Zenati Anissa***, for their time, insightful questions, and valuable feedback during the defense of this thesis.

To our beloved ***families***, we are forever indebted. Your unconditional love, encouragement, constant belief in our abilities, and countless sacrifices have been the foundation of our academic pursuits and personal growth.

Our deepest thanks go to our ***friends*** and ***colleagues*** who have stood by us, offering encouragement and a shoulder to lean on during both challenging and joyful moments. Your presence has made this journey more meaningful and enjoyable.

We are immensely thankful to the PhD Student ***Djenane Mouloud Amine***, for his collaboration, shared insights, and mutual support throughout this demanding endeavor.

We also extend our appreciation to the staff of the ***Faculty of Mathematics and Computer Science***, at ***Ibn Khaldoun University of Tiaret***

# Contents

**Abstract** i

**Dedication** iii

**Acknowledgments** iv

**List of Tables** vi

**List of Figures** viii

**Acronyms** ix

**Introduction** 1

**1 Recommendation Systems in Healthcare** 3
- 1.1 Introduction . . . 3
- 1.2 Recommendation Systems . . . 3
  - 1.2.1 Types . . . 4
  - 1.2.2 Applications . . . 8
  - 1.2.3 Evaluation . . . 14
- 1.3 Health Recommender System . . . 16
  - 1.3.1 Scenarios . . . 16
  - 1.3.2 Evaluation . . . 20
- 1.4 Conclusion . . . 21

**2 Drug Recommendation Systems: A Literature Review** 22
- 2.1 Introduction . . . 22
- 2.2 Overview of Drug Recommendation Systems . . . 22
  - 2.2.1 Taxonomy . . . 23
  - 2.2.2 Data Sources . . . 27
  - 2.2.3 Evaluation Strategies . . . 29
- 2.3 Challenges and Opportunities . . . 33
  - 2.3.1 Data-related . . . 33
  - 2.3.2 Model-related . . . 34
  - 2.3.3 Ethical, Interpretability, and Implementation . . . 35

2.4 Conclusion . . . . . 36

**3 The Proposed Drug Recommendation System: Design, Implementation, and Evaluation 37**

3.1 Introduction . . . . . 37
3.2 The Proposed Drug Recommendation System: Architecture and Design . . . . . 37
3.2.1 Conceptual Framework . . . . . 37
3.2.2 Main Architectural Components . . . . . 38
3.3 Experiments and implementations . . . . . 42
3.3.1 Dataset and Pre-processing . . . . . 43
3.3.2 Implementation Details . . . . . 46
3.3.3 Evaluation Metrics . . . . . 47
3.3.4 Baseline Models . . . . . 49
3.4 Results and Discussion . . . . . 50
3.4.1 Results . . . . . 50
3.4.2 Comparison with Baselines . . . . . 51
3.4.3 Limitations of the Proposed Model . . . . . 52
3.5 Conclusion . . . . . 53

**General Conclusion 56**

# List of Tables


3.1 Key features in the original dataset. . . . . 44
3.2 General confusion matrix for recommendation results. . . . . 48
3.3 Confusion matrix for The proposed model. . . . . 50
3.4 Comparison of classification performance with baseline models. . . . . 51
3.5 Comparison of error metrics (Root Mean Square Error (RMSE) and Relative Root Mean Squared Error (RRMSE)) with baseline models. . . . . 51

# List of Figures


1.1 The workflow of a modern recommendation system . . . . . 4
1.2 Types of Recommendation Systems (RSs) . . . . . 4
1.3 Illustration for Content-based Recommendation System (CB) Recommendation Systems . . . . . 5
1.4 Illustration for Collaborative Filtering (CF) Recommendation Systems . . . . . 6
1.5 Illustration for CF Recommendation techniques . . . . . 7
1.6 Illustration for Knowledge-based Recommendation System (KB) Recommendation Systems . . . . . 8
1.7 Illustration of Hybrid Recommendation System . . . . . 8
1.8 The workflow of a Electronic Commerce (E-commerce) Recommendation System (RS) . . . . . 9
1.9 The impact of data sparsity . . . . . 12
1.10 Illustration of confusion matrix . . . . . 15
1.11 Healthy recipes recommendation app . . . . . 17
1.12 Drug recommendation system workflow . . . . . 18
1.13 Healthcare professional recommendation system . . . . . 19
1.14 Workout recommendation system . . . . . 20

2.1 Core objectives of Drug Recommendation Systems (DRSs) . . . . . 23
2.2 Methodologies used in DRSs . . . . . 24
2.3 Taxonomy of data types used in DRSs . . . . . 25
2.4 Target users of DRSs . . . . . 26
2.5 Area Under the Receiver Operating Characteristic Curve (AUROC) . . . . . 30
2.6 Example of coverage illustration in DRS . . . . . 32
2.7 Illustration of Data Sparsity . . . . . 33
2.8 Illustration of Personalization vs. Generalization . . . . . 34

3.1 The proposed model Conceptual Framework . . . . . 38
3.2 Conceptual flow of Valence Aware Dictionary and sEntiment Reasoner (VADER) model in sentiment analysis . . . . . 39
3.3 Clustering concept . . . . . 40
3.4 Architecture Diagram of the Enhanced NCF (Multi-inputs) . . . . . 43

3.5 Learning curves for the proposed model: (Left) Training and validation MSE over epochs. (Right) Training and validation MAE over epochs. . . . . 50

# Acronyms

**Adam** Adaptive Moment Estimation

**ADE** Adverse Drug Event

**ADR** Adverse Drug Reaction

**AI** Artificial Intelligence

**AP** Average Precision

**AUROC** Area Under the Receiver Operating Characteristic Curve

**BCE** Binary Cross-Entropy

**CB** Content-based Recommendation System

**CF** Collaborative Filtering

**ChEMBL** Chemical Exploration of Molecule Bioactivity Large-scale

**ConvMF** Conventional Matrix Factorization

**CPC** Constrained Pearson Correlation

**CUR** Consolidated User Rating

**DCG** Discounted Cumulative Gain

**DDI** Drug-Drug Interaction

**DL** Deep Learning

**DOE** Degree of Effectiveness

**DOS** Degree of Side Effects

**DRS** Drug Recommendation System

**E-commerce** Electronic Commerce

**EHR** Electronic Health Record

**EMA** European Medicines Agency

**FAERS** FDA Adverse Event Reporting System

**FDA** Food and Drug Administration

**FN** False Negative

**FP** False Positive

**GDPR** General Data Protection Regulation

**GMF** Generalized Matrix Factorization

**HB** Hybrid-based Recommendation System

**HIPAA** Health Insurance Portability and Accountability Act

**HR** Hit Rate

**HRS** Health Recommender System

**IBI** Item Based Indicator

**ICU** Intensive Care Unit

**IDCG** Ideal Discounted Cumulative Gain

**IoT** Internet of Things

**IT** Information Technology

**KB** Knowledge-based Recommendation System

**KEGG** Kyoto Encyclopedia of Genes and Genomes

**KG** Knowledge Graph

**KMeans-CF** K-Means Collaborative Filtering

**LIME** Local Interpretable Model agnostic Explanations

**MAE** Mean Absolute Error

**MAP** Mean Average Precision

**MF** Matrix Factorization

**MIMIC** Medical Information Mart for Intensive Care

**E-governance** Electronic Governance

**ML** Machine Learning

**MLP** Multi-Layer Perceptron

**NCF** Neural Collaborative Filtering

**NCHS** National Center for Health Statistics

**NDCG** Normalized Discounted Cumulative Gain

**NeuMF** Neural Matrix Factorization

**NHANES** National Health and Nutrition Examination Survey

**NLTK** Natural Language Toolkit

**NLP** Natural Language Processing

**NN** Neural Network

**NSGA-III** Non-dominated Sorting Genetic Algorithm III

**PUC** Polarity of User Comment

**RBI** Rating Based Indicator

**RMSE** Root Mean Square Error

**RRMSE** Relative Root Mean Squared Error

**ROC** Receiver Operating Characteristic

**RS** Recommender System

**RS** Recommendation System

**SAR** Structure-Activity Relationship

**SHAP** SHapley Additive exPlanations

**SIDER** Side Effect Resource

**SVM** Support Vector Machine

**TN** True Negative

**TP** True Positive

**VADER** Valence Aware Dictionary and sEntiment Reasoner

**XAI** Explainable AI

# Introduction

## Context and Motivation

In recent years, the integration of technology into healthcare has led to the development of systems that help diagnoses disease and plan treatment. Recommendation systems, which have been widely used in E-commerce, are now being adapted to healthcare to provide personalized medication suggestions. These systems analyze patient data, including medical history and current symptoms, to recommend appropriate drugs, thus supporting healthcare professionals and improving patient care. For example, a study by Granda Morales et al. (2022) developed a Drug Recommendation System (DRS) for diabetes patients using collaborative filtering and clustering techniques, demonstrating the potential of such systems to improve treatment outcomes.

## Problem Statement

Despite advances in healthcare technology, selecting the most appropriate medication for individual patients remains a complex challenge. Factors such as patient specific characteristics, potential drug interactions, and the vast array of available medications complicate the decision making process. Existing systems may not adequately account for the personalized needs of patients or may lack the integration of comprehensive data sources. This gap highlights the need for a robust drug recommendation system that can provide tailored suggestions to healthcare providers, therefore improving patient outcomes. This dissertation is therefore guided by a central research question:

> *Baseline drug recommendation models suffer from data sparsity, inability to address subjective patient experiences, and inaccurate personalization. How can a deep learning architecture, combined with insights from patient reviews and user ratings, overcome these limitations to improve drug recommendations accuracy?*

## Objectives

The primary objectives of this dissertation are:

- To design a drug recommendation system that integrates patient data to provide personalized medication suggestions.

- To implement the system using advanced machine learning techniques, ensuring accuracy and reliability in recommendations.

- To evaluate the system's performance through rigorous testing and validation against existing benchmarks.

### Dissertation Organization

After this introduction, the rest of the dissertation is organized as follows:

**Chapter 1: Recommendation Systems in Healthcare**
This chapter introduces the role of recommendation systems in healthcare, covering their types, common applications, and the challenges facing their implementation. We also explain how these systems are evaluated. Then we focus more closely on Health Recommender Systems (HRSs) and their importance, illustrating their relevance in clinical contexts, and discussing specific evaluation strategies used for them.

**Chapter 2: Drug Recommendation Systems: A Literature Review**
This chapter provides a comprehensive review of existing DRSs, highlighting the latest advancements in the field. It critically analyzes previous works methodologies from traditional approaches to advanced Deep Learning (DL) models, the data sources and the evaluation strategies that are commonly employed, identifying their strengths and challenges. Finally, it discusses existing gaps and potential research opportunities to improve DRSs.

**Chapter 3: The Proposed Drug Recommendation System: Design, Implementation, and Evaluation**
This chapter presents our major contribution: a novel DL-based drug recommendation system. We describe the system architecture, which extends the Neural Collaborative Filtering NCF framework by incorporating sentiment analysis from user reviews and a custom KMeans clustering algorithm for user segmentation. We outline the data preprocessing steps, feature engineering techniques, model implementation, and experimental setup. Furthermore, we conduct a comprehensive evaluation of the model performance against established baselines, highlighting how our approach addresses the specific limitations identified in Chapter 2.

**General Conclusion**

We conclude the dissertation by summarizing the key findings, contributions, and insights derived from this research. It revisits the initial objectives and discusses the extent to which they were met. Furthermore, this chapter outlines directions for future research and development in the field of DRSs.

# Chapter 1

# Recommendation Systems in Healthcare

## 1.1 Introduction

Today, a large collection of clinical data spread across the Internet makes it difficult for users to find useful information to improve their well being. In addition, the overload of medical information (for example, on drugs, medical tests and treatment suggestions) has brought many difficulties to medical professionals in making patient oriented decisions. These issues raise the need to apply Recommendation Systems (RSs) in the healthcare domain to help end users and medical professionals make more efficient and accurate health related decisions.

In this chapter, we will explore different types of RSs and their underlying principles, advantages and limitations; then we will discuss their applications in different domains and highlight some of the problems and challenges in the field. In addition, we will provide an overview of the evaluation metrics and methods for RSs and introduce the domain of Health Recommender Systems (HRSs) and their scenarios.

## 1.2 Recommendation Systems

The use of RSs is a crucial aspect in addressing the problem of online information overload and improving customer relationship management. These systems are designed to provide personalized recommendations to users of online products and services, enhancing the user's online experience. The applications of RSs can be seen in various online platforms, such as product recommendations for customers and content recommendations for readers, among others. Figure 1.1 illustrates a modern RS workflow. The objective of these systems is to identify new and relevant items that align with the user's preferences.

The fundamental principle of RSs is to suggest relevant items to users by using feature engineering techniques on user preferences, item features and their interactions (such as purchases or clicks).

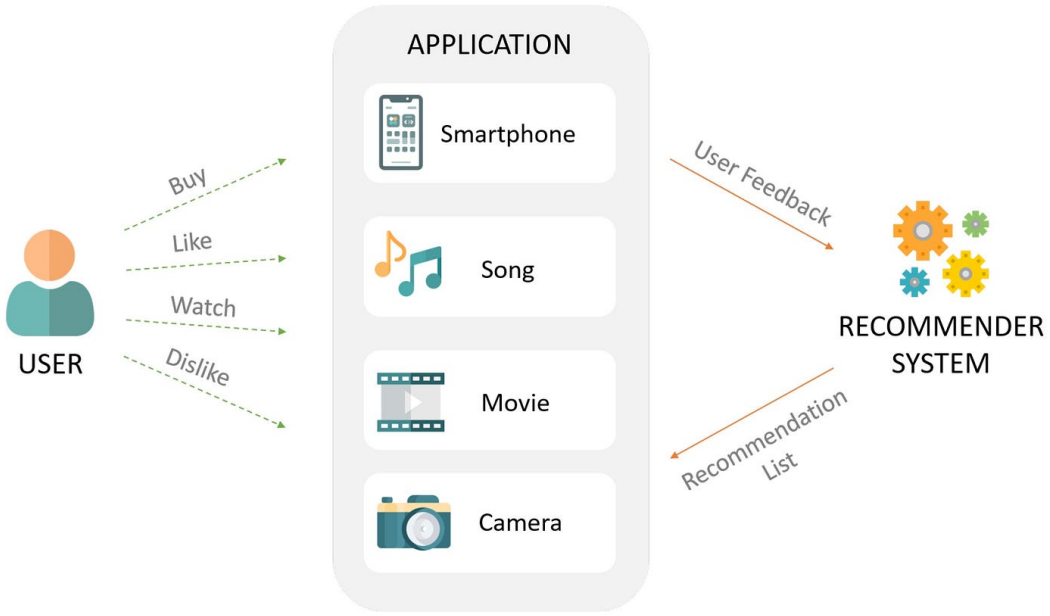


Figure 1.1: The workflow of a modern recommendation system

### 1.2.1 Types

RSs play a crucial role in filtering and suggesting relevant items to users based on their preferences, behaviors, or interactions. These systems are widely used in various domains, such as E-commerce, healthcare and entertainment. Depending on the approach used to generate recommendations, RSs can be categorized into different types, each with its strengths and limitations.

Figure 1.2 provides a visual representation of the main types of RSs, illustrating their classification based on different methodologies. In this section, we will explore these types in detail and discuss their underlying techniques.

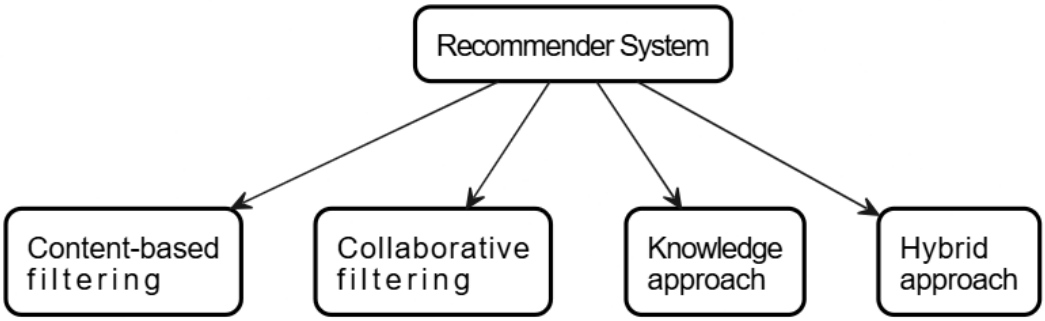


Figure 1.2: Types of Recommendation Systems (RSs)

#### 1.2.1.1 Content-based Recommendation System

The main idea of Content-based Recommendation System (CB) is to recommend items based on the similarity between different users or items (Lops et al. (2011b)). This algorithm determines

and differentiates the main common attributes of a particular user's favorite items by analyzing the descriptions of those items. Then, these preferences are stored in the user profile. The algorithm then recommends items with a higher degree of similarity to the user profile.

Figure 1.3 provides an illustration of the CB recommendation approach, demonstrating how items are recommended based on their similarity to user preferences. In addition, CB can capture the specific interests of the user and can recommend rare items that are of little interest to other users. However, since the feature representations of items are designed manually to some extent, this method requires a lot of domain knowledge. In addition, CB can only recommend based on user existing interests, so the ability to expand user existing interests is limited.

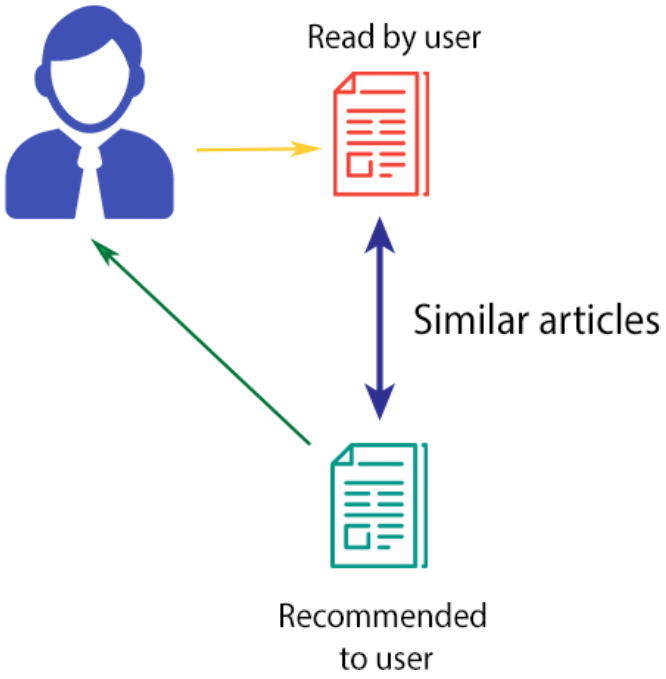


Figure 1.3: Illustration for CB Recommendation Systems

#### 1.2.1.2 Collaborative Filtering-based Recommendation System

CF-based methods are mainly used in big data processing platforms due to their parallelization characteristics (Elahi et al. (2016)). The basic principle of CF-based RSs is illustrated in Figure 1.4. CF RSs use the behavior of a group of users to recommend items to other users (Hassanieh et al. (2018)). Figure 1.5 illustrates the two mainly types of CF techniques, user-based and item-based.

- **User-based CF:** In user-based CF RSs, users receive recommendations for products that similar users have liked (Burke et al. (2006)). Many similarity metrics can be used to calculate the similarity between users or items, such as the Constrained Pearson Correlation (CPC), cosine similarity and adjusted cosine similarity (Manning et al. (2008)).

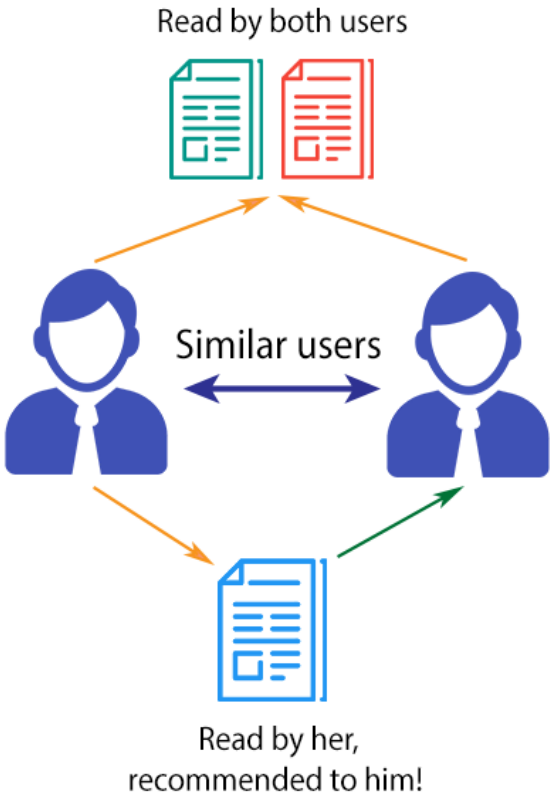


Figure 1.4: Illustration for CF Recommendation Systems

One of the most commonly used similarity measures is cosine similarity, which determines the angle between two vectors. It is mathematically defined as:

$$\cos(\theta) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{1.1}$$

where $x_i$ and $y_i$ are the components of vectors $x$ and $y$ respectively, $n$ is the dimension of the vectors and $\theta$ is the angle between the two vectors.

Similarly, the Pearson correlation coefficient measures the strength of the linear relationship between two variables and is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1.2}$$

where $x_i$ and $y_i$ are the individual sample points indexed with $i$, $n$ is the sample size, $\bar{x}$ is the mean of the $x$ values, $\bar{y}$ is the mean of the $y$ values and $r_{xy}$ is the Pearson correlation coefficient between variables $x$ and $y$.

- **Item-based CF:** In contrast, item-based CF algorithms predict user ratings for items based on the similarity between items rather than between users. Generally, item-based CF tends to yield better results than user-based CF because user-based CF suffers from

data sparsity and scalability issues. However, both techniques may face challenges such as the cold-start problem (Zhang et al. (2016)).

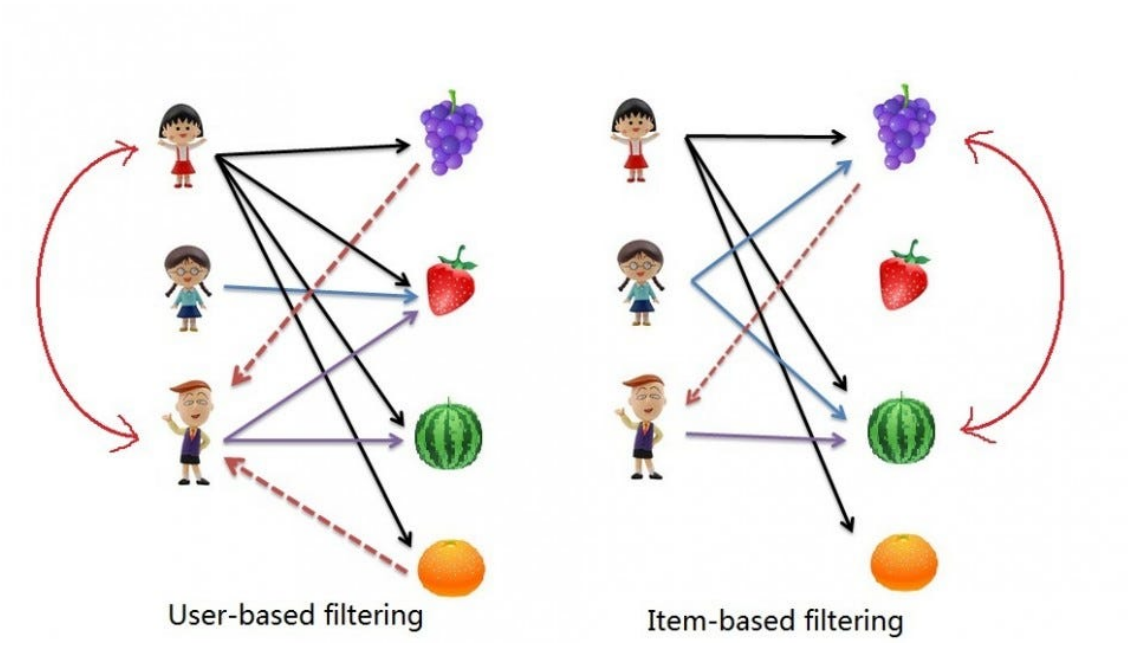


Figure 1.5: Illustration for CF Recommendation techniques

#### 1.2.1.3 Knowledge-based Recommendation Systems

The main idea of KBs as illustrated in Figure 1.6 is to recommend items to users based on basic knowledge of users, items and relationships between items (Aggarwal (2016a)). Since KBs do not require user ratings or purchase history, there is no cold start problem for this type of recommendation (Cabezas et al. (2017)). KBs are commonly used in complex domains where items are not typically purchased, such as cars and apartments (Tarus et al. (2018)). In such scenarios, users may face constraints like budget limits when making a purchase. KBs rely on domain knowledge, which includes user insights, item details, or the interactions between users and items (Chen et al. (2019)). But the acquisition of required domain knowledge can become a bottleneck for this recommendation technique (Dong et al. (2020)).

#### 1.2.1.4 Hybrid Approach Recommendation System

Hybrid-based Recommendation System (HB) combine the advantages of multiple recommendation techniques and aim to overcome the potential weaknesses of traditional RSs (Ribeiro et al. (2012)). There are seven basic hybrid recommendation techniques (Ibrahim et al. (2021)): weighted, mixed, switching, combination of characteristics, augmentation of characteristics, cascade and meta-level methods (Zhang et al. (2016)).

Among these methods, the most commonly used approach is the combination of CF with other recommendation methods, such as content-based or knowledge-based approaches. This

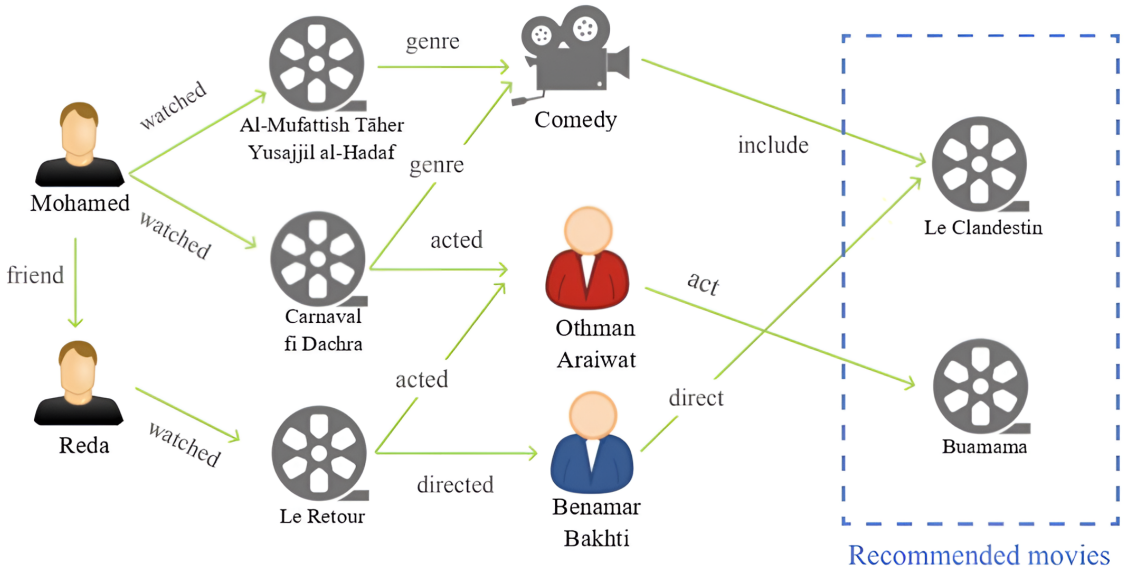


Figure 1.6: Illustration for KB Recommendation Systems

combination helps mitigate issues such as data sparsity, scalability and cold start problems (Zagranovskaia and Mitura (2021)). Figure 1.7 illustrates the hybrid recommendation approach, demonstrating how different recommendation techniques are integrated to enhance accuracy and overcome the limitations of individual methods.

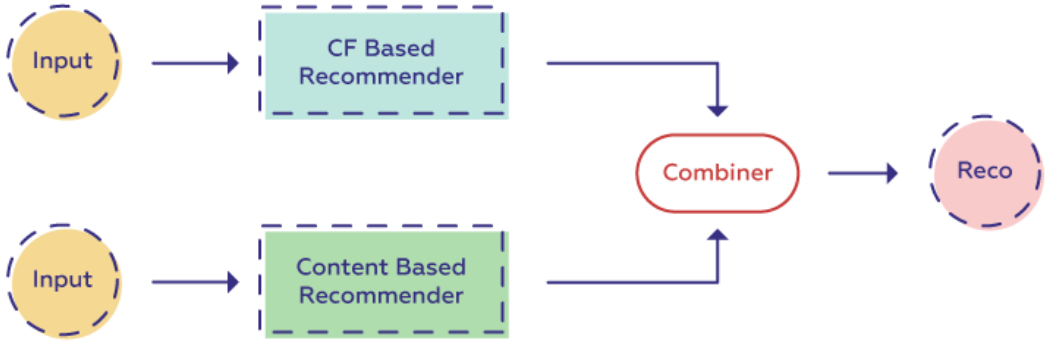


Figure 1.7: Illustration of Hybrid Recommendation System

### 1.2.2 Applications

RSs have become everywhere and in various sectors, such as search engines, digital media platforms and E-commerce sites on the Internet (Wang et al. (2020a)). The progression of Information Technology (IT) has led to their significant evolution, embracing increasingly complex models.

#### 1.2.2.1 Recommendation System in Electronic Commerce (E-commerce)

RSs have evolved from specialized tools utilized by a select few E-commerce platforms to vital commercial assets that significantly transform the E-commerce landscape (Ali et al. (2017)). Figure 1.8 illustrates The workflow of a E-commerce Recommendation System (RS). Major online platforms and applications, such as `Amazon` and `TikTok`, now harness the power of big data to refine their recommendation algorithms for users Rismanto et al. (2020) previously highlighted the need for advancements in data collection and analytics to expand the operational advantages in the marketing sector, prompted by the advent of new applications for information agents (Li et al. (2021)). With the advancement of big data, the application of information agents has shifted towards providing accurate and tailored recommendations to customers on online markets and platforms through innovative models like topic modeling and sentiment analysis (Numnonda (2018)). Therefore, in an era dominated by big data, it's evident that RSs have become widely integrated into various aspects of E-commerce operations. This extensive adoption confirms the pivotal role these systems play in improving user experience, personalizing customer interactions and boosting the overall efficiency of E-commerce RS platforms.


Figure 1.8: The workflow of a E-commerce Recommendation System (RS)

#### 1.2.2.2 Recommendation System in Electronic Governance (E-governance)

E-governance stands as a fundamental challenge in the realm of smart city initiatives, integrating IT and big data in the public sector to elevate the delivery of services and information. This approach not only aims to enhance government transparency, accountability and trustworthiness but also to engage citizens in the governance process (Xiao et al. (2018)). In the digital era, particularly highlighted by the rise of epidemics, the demand for E-governance in society is on the rise. This underscores the necessity for governments to establish adept online information

systems to meet the goals of effective E-governance (Huang et al. (2019)). Businesses, including pharmaceutical companies, are also recognizing the need for digital governance. For instance, they can implement RSs, built on blockchain and Machine Learning (ML) technologies, to streamline drug shipment monitoring (Zhang et al. (2018)). Additionally, these systems have applications in demand-side management, like energy management, where they utilize big data analytics to identify residential users' preferences for energy-efficient appliances (Benouaret and Amer-Yahia (2020)). Thus, the deployment of RSs, bolstered by big data technologies, plays a crucial role in improving the digital governance framework, optimizing business processes and facilitating efficient energy management in the digital age.

#### 1.2.2.3 Recommendation System in Sustainable Lifestyle

The digital transformation created by the Internet revolution has significantly facilitated the transition to a lifestyle centered on digital interactions, concurrently with a greater focus on sustainability in the environment, societal and other sectors (Xia et al. (2023)).

These systems are instrumental in refining the practices of daily activities, such as online shopping, dietary habits and transportation, aiming to promote sustainable living by prioritizing options. thus, facilitating a shift towards sustainability among both providers and consumers. Addressing the progression of technological innovation, consumer behavior and environmental responsibility, recent studies advocate the integration of green marketing strategies with online retail platforms through the deployment of sophisticated RSs (Felfernig et al. (2023)). The work of Zhang et al. (2022) highlights the critical capacity of RSs to advocate for environmentally sustainable choices, green building emerges as a significant aspect that profoundly influences our connection to sustainable living (Xu (2021)). With the advancement of RSs, big data and the Internet of Things (IoT), several challenges associated with green building can be addressed through the integration of RSs and ML technologies. These technologies have the potential to enhance various aspects of green building.

#### 1.2.2.4 Recommendation System in Healthcare

The healthcare industry is experiencing a paradigm shift with the incorporation of advanced technologies and RSs have emerged as a key component in this transformation. Leveraging ML algorithms and data analytics, RSs in healthcare offer a wide range of applications, from improving clinical decision-making to enhancing patient engagement. This subsection explores the diverse facets of RSs in healthcare and their impact on various stakeholders within the ecosystem.

1. **Clinical Decision Support:** One of the primary applications of RSs in healthcare is in clinical decision support. These systems analyze Electronic Health Records (EHRs), medical literature and patient data to assist healthcare professionals in making informed decisions about diagnostics, treatment plans and interventions. By providing relevant and evidence-based information, RSs contribute to more accurate and personalized patient

care, potentially reducing diagnostic errors and improving overall healthcare outcomes (Shojania et al. (2009)).

2. **Patient Centered Care:** In the era of patient centered care, RSs play a crucial role in tailoring healthcare services to individual patient needs. These systems analyze patient preferences, demographics and health histories to generate personalized recommendations for treatment options, preventive measures and lifestyle modifications. By facilitating patient engagement and empowerment, RSs contribute to a more collaborative and effective healthcare relationship between providers and patients (Jr et al. (2009)).

3. **Resource Optimization:** RSs help optimize healthcare resources by optimizing workflows and improving operational efficiency. For instance, in hospital management, these systems can suggest optimal bed allocation, appointment scheduling- resource utilization based on historical data and real-time information. By minimizing bottlenecks and enhancing resource allocation, RSs contribute to cost-effectiveness and improved service delivery Li and Wang (2017).

4. **Remote Monitoring:** With the rise of remote monitoring, RSs support healthcare providers in delivering virtual care. These systems analyze patient generated health data from wearable devices, monitoring tools and telehealth platforms to provide timely recommendations for interventions, medication adjustments, or lifestyle modifications. This realtime support contributes to proactive healthcare management, particularly for patients with chronic conditions (Fatehi and Wootton (2018).)

5. **Healthcare Collaboration Networks:** RSs facilitate collaboration and knowledge sharing among healthcare professionals through the creation of collaborative networks. By analyzing expertise, research interests and clinical experiences, these systems connect healthcare professionals for consultations, research collaborations and second opinions. This fosters a culture of continuous learning and knowledge sharing within the healthcare community (Al-Shammary et al. (2019). )

As RSs continue to evolve and offer a significant help and guidance with their diverse applications and scenarios 1.3, their development and deployment do face some difficulties. Section 1.2.2.5 looks at the key problems and challenges that must be overcome to fully exploit the potential of these systems.

#### 1.2.2.5 Common Problems and Challenges in Recommendation Systems (RSs)

To build competent RSs, developers must address several inherent problems and challenges that can limit their effectiveness. The following highlights the key issues crucial to improving performance and advancing research in RSs.

- **Data Sparsity:** Many commercial RSs utilize large datasets where the user-item interaction matrix (e.g., purchases, views, ratings) is often very sparse, meaning most users have

interacted with only a small fraction of items. Figure 1.9 demonstrates how data sparsity can severely compromise the performance of CF techniques, which rely on user interactions that overlap to identify similarity and make predictions (Isinkaye et al. (2015)). Content-based recommendations are generally less affected by this particular issue as they rely on item/user features.

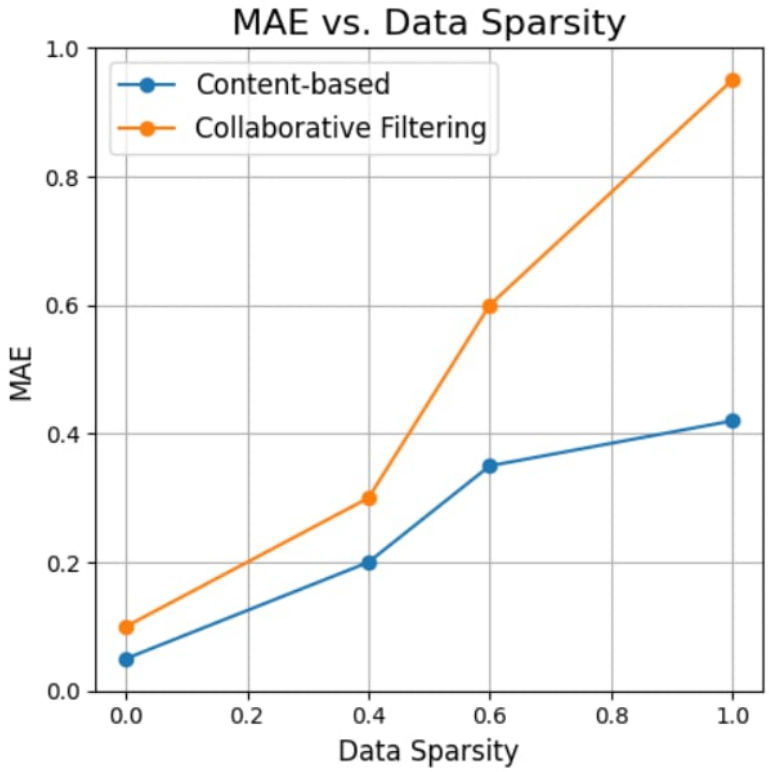


Figure 1.9: The impact of data sparsity

- **Cold Start Problem:** This well-known issue occurs when the system has insufficient data to make reliable recommendations (Bobadilla et al. (2013)). It manifests in two primary forms:
  - *User Cold-Start:* Occurs when a new user enters the system and there is little to no information about their preferences.
  - *Item Cold-Start:* Occurs when a new item is added to the system and there is little to no interaction data for it.

  A common initial strategy for new systems is to employ content-based filtering, gradually transitioning to or incorporating CF as more user data becomes available (Lops et al. (2011a)).
- **Lack of User Activity:** Closely related to data sparsity and cold-start, insufficient overall user activity (e.g., few ratings, limited browsing) makes it challenging for the system to learn user preferences accurately and generate meaningful recommendations.
- **Privacy Concerns:** RSs often collect extensive user data, including ratings, preferences and browsing history, to improve recommendation accuracy. This raises significant concerns

regarding data storage, processing, security and potential sharing with third parties. Building user trust through transparent data handling practices and robust privacy-preserving techniques is crucial (Hossain et al. (2023)).

- **Synonymy:** Synonymy arises when a single item is represented by multiple names or entries that have equivalent meanings (e.g., "laptop" vs. "notebook computer"). The RS might fail to recognize these as the same entity, leading to fragmented interaction data and potentially inefficient or inaccurate recommendations (Su (2009)).

- **Shilling (Profile Injection) Attacks:** These attacks involve malicious users or competitors manipulating the system by providing false ratings or interaction data. The goal is to artificially inflate or deflate the popularity of specific items, thereby undermining the system's trustworthiness and degrading recommendation quality (Mobasher et al. (2007)). CF techniques are particularly vulnerable.

- **Gray Sheep Problem:** Users classified as "gray sheep" possess diverse or idiosyncratic preferences that do not align well with any significant user group. Consequently, CF methods often struggle to provide accurate or relevant recommendations for these users, as their tastes are not easily predictable based on community behavior (Su and Khoshgoftaar (2009)).

- **Long Tail Problem (Item Popularity Bias):** Recommendation algorithms, especially those based on popularity, tend to recommend popular items (the "head" of the distribution) more frequently, often neglecting a vast number of less popular or niche items (the "long tail"). This leads to an imbalance where many available items remain underrepresented or undiscovered by users, limiting diversity and serendipity Park and Tuzhilin (2008).

- **Over-Specialization:** This occurs when the system predominantly recommends items that are very similar to a user's past interactions or explicitly stated preferences, thereby limiting exposure to new, diverse, or serendipitous options. While aiming for relevance, extreme over-specialization can lead to filter bubbles and reduce user satisfaction by failing to broaden horizons (Pariser (2011)). Content-based systems can be particularly prone to this if user profiles are narrow.

- **Achieving Novelty and Serendipity:** Beyond mere accuracy, effective RSs should aim to provide:

  - *Novelty:* Recommending items that are new and unfamiliar to the user.
  - *Serendipity:* Recommending items that are not only novel but also surprisingly relevant and useful, items the user might not have discovered on their own or thought to search for (Ge et al. (2010)).

  Balancing relevance with novelty and serendipity is a significant design challenge.

- **Scalability:** As the number of users and items grows (often into millions or billions), traditional algorithms can face significant scalability issues. RSs must efficiently manage

large-scale data and computations to provide timely responses to users (Linden et al. (2003a)). This often requires distributed computing architectures and optimized algorithms.

- **Diversity in Recommendations:** Related to avoiding over-specialization and addressing the long-tail problem, actively promoting diversity in recommendations helps users discover a broader range of items. This can improve user satisfaction and exploration but needs to be balanced with relevance (Ziegler et al. (2005)).

- **Evaluating RSs:** Comprehensive evaluation of RSs is itself a challenge. It requires considering multiple dimensions beyond simple prediction accuracy, including ranking quality, diversity, novelty, serendipity, user satisfaction and system scalability. The choice of appropriate metrics (as discussed in Section 1.2.3.1) and evaluation methodologies (e.g., offline vs. online A/B testing) is critical (Shani and Gunawardana (2011)).

The field of RSs is shaped by an interplay of these problems and challenges, each demanding targeted solutions while acknowledging their interconnected nature. For instance, issues like data sparsity, cold starts and shilling attacks directly degrade recommendation quality, while challenges such as scalability and achieving diversity reflect systemic considerations in designing robust, effective and user-centric systems.

### 1.2.3 Evaluation

RSs aim to provide personalized suggestions to users based on their preferences and needs. Evaluating RSs involves measuring how well they achieve this goal using different metrics and methods.

#### 1.2.3.1 Evaluation Metrics

In terms of evaluation metrics, RSs can be classified into Rating Based Indicators (RBIs) and Item Based Indicators (IBIs). RBI evaluates the recommendations based on a predicted rating score, while IBI evaluates the recommendations based on a set or list of predicted items. This taxonomy allows for the classification of most existing evaluation indicators. Additionally, other key evaluation metrics, such as coverage and novelty, provide further insights into recommendation performance (Kaminskas and Bridge (2016)). Further metrics evaluate the ranking quality, such as Mean Average Precision (MAP), Normalized Discounted Cumulative Gain (NDCG) and Hit Rate (HR), often considering the top-k items. The AUROC is also a valuable metric for assessing ranking or classification performance.

1. **Indicators:** Which is divided into two categories:

   - **Rating Based Indicator (RBI):**
     The rating-based indicator is to evaluate the quality of the prediction rating score. The direct way is to calculate the gap between implicit/explicit labels. One of the

most popular measurements is RMSE when the rating score is an explicit value. The equation format for the RMSE can be expressed in Equation 3.10.

Similarly, Mean Absolute Error (MAE) is another popular measurement that can be expressed in Equation 1.3.

$$MAE = \frac{1}{|U||I|} \sum_{u \in U, i \in I} |\hat{r}_{ui} - r_{ui}| \tag{1.3}$$

Mean Squared Error (MSE) is also a fundamental measurement, which squares the differences before averaging. It can be expressed as:

$$MSE = \frac{1}{|U||I|} \sum_{u \in U, i \in I} (\hat{r}_{ui} - r_{ui})^2 \tag{1.4}$$

where U is the set of the users, I is the set of items, $\hat{r}_{ui}$ denotes the predicted rating and $r_{ui}$ denotes the true rating.

RMSE, MAE and MSE are non-negative; a lower value for each is better than a higher one. While each error term ($(\hat{r}_{ui} - r_{ui})^2$ in RMSE, $|\hat{r}_{ui} - r_{ui}|$ in MAE) contributes to the final error, RMSE penalizes larger errors more severely due to the squaring term, making both sensitive to outliers.

- **Item Based Indicator (IBI):**
  When the RS output is a set or list of items and if there is no inherent ranking information considered for these initial metrics, a confusion matrix as illustrated in (Figure 1.10) AI (n.d.), and listed in Table 3.2, can be adopted in the evaluation.

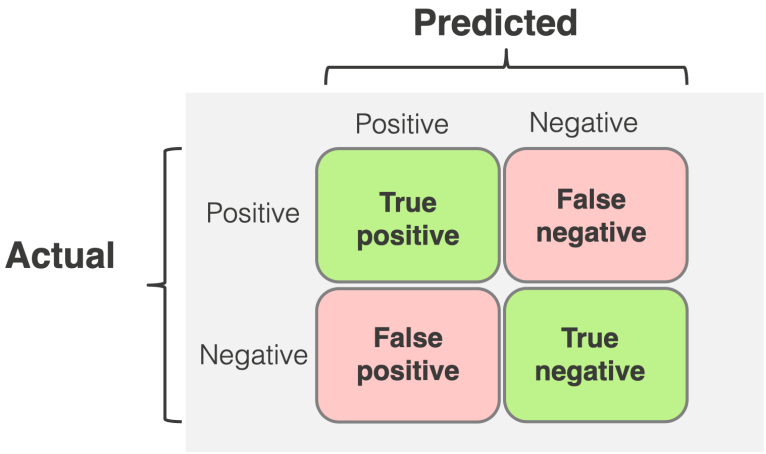


Figure 1.10: Illustration of confusion matrix

Generally, more comprehensive compositions of these four values are adopted, such as Accuracy, Precision, Recall, F1-Score and Specificity. A detailed discussion of these metrics can be found in Chapter 3.3.3.

2. **Coverage:** is how many items or users can be recommended by the system. This can be measured by calculating the percentage of items or users that receive at least one recommendation. Coverage reflects how broad or diverse the system is in providing recommendations (Ge et al. (2010)).

3. **Novelty:** is how surprising or unexpected the recommendations are for the users. This can be measured by calculating the average popularity or familiarity of the recommended items among the users. Novelty reflects how creative or innovative the system is in providing recommendations (Hossain et al. (2023)).

#### 1.2.3.2 Evaluation Methods

Some common methods for evaluating RSs Shani and Gunawardana (2011) are:

- **Offline Testing:** Using historical data to simulate user feedback allows to compare different algorithms or parameters efficiently and cost-effectively. However, this approach may not accurately represent real user behavior or account for dynamic shifts in their preferences.

- **Online Evaluation:** Deploying different versions of the system to real users and analyzing their responses. This approach provides more realistic and reliable insights but can be costly, risky and raise ethical concerns.

- **User Studies:** Gathering user opinions and feedback through surveys or interviews with a sample group. This approach offers valuable qualitative insights but may be limited by low response rates, biases and scalability challenges.

## 1.3 Health Recommender System

Today, a large amount of clinical data scattered across different sites on the Internet hinders users from finding useful information for their well-being improvement. In addition, the overload of medical information (for example, on drugs, medical tests and treatment suggestions) has brought many difficulties to medical professionals in making patient-oriented decisions. These issues raise the need to apply RSs in the healthcare domain to help end-users and medical professionals make more efficient and accurate health-related decisions. This section explores the various facets of RSs in healthcare, such as food recommendation, drug recommendation, health status prediction, healthcare service recommendation and healthcare professional recommendation. Finally, we discuss challenges concerning the future development of healthcare RSs.

### 1.3.1 Scenarios

Health Recommender System (HRS) can be applied in a variety of real-world healthcare settings, each highlighting specific needs within the healthcare domain, offering personalized recommendations. In this section, we will present the key areas where HRS being developed and applied.

1. **Food recommendation:** Due to the extensive growth of food options and increasingly busy lifestyles, people have been struggling and facing the issue of making healthy food choices, which are essential for to reduce the risk of chronic diseases (Ge et al. (2015)). In this context, food RSs can motivate users to change their eating behaviors or suggest healthier food choices (Figure 1.11) (Yang et al. (2017)).

2. **Drug recommendation:** Drug RSs assist caregivers in improving medication selection, accuracy and safety by considering patient conditions and characteristics, Figure 1.12 (AI-generated image created using SORA by OpenAI) illustrate the workflow of a DRS. The following points discuss their use in recommending treatments for diseases and predicting potential side effects.

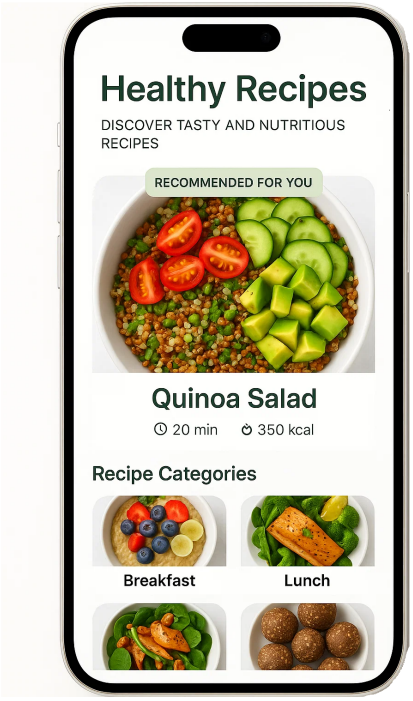


Figure 1.11: Healthy recipes recommendation app

(a) **Medication recommendations for treating diseases:** Prescribing errors considered to be one of the most serious medical errors that could endanger patients' lives (Gorgich et al. (2015)). More than 42% of these errors are caused by doctors who have limited experiences/knowledge about drugs and diseases (Bao and Jiang (2016)). Another reason lies in the increasing number of available drug information, which has brought obstacles concerning the discovery of relevant drugs and drug-disease interactions (Doulaverakis et al. (2012)). In this context, drug RSs have been developed to assist end-users and healthcare professionals in identifying accurate medications for a specific disease.

(b) **Predict drug side effects:** Adverse Drug Reactions (ADRs) as known as drug side effects are a leading cause of misery and mortality, with 100,000 deaths annually in

the USA (Galeano and Paccanaro (2018)). Early prediction methods used Structure-Activity Relationships (SARs), such as effect spectra Fliri et al. (2006) and gene pathway analysis Fukuzaki et al. (2009). ML approaches like *in silico* methods leverage drug structures and biological features for prediction (Lafta et al. (2015)). However, these methods face challenges like data availability, high computational demands and false positives (Deshpande and Butte (2011)). Since clinical trials may miss side effects, improved prediction models are needed (Zhang et al. (2016)). A RS approach has been proposed, using neighborhood-based methods to predict side effects from similar drugs.

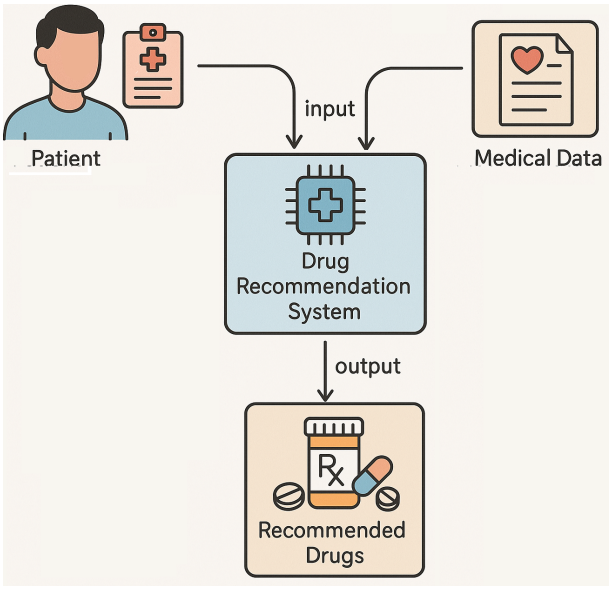


Figure 1.12: Drug recommendation system workflow

3. **Healthcare professional recommendations:** In recent years, there has been a significant increase in the amount of available medical information, which results in some difficulties for patients when searching for suitable doctors. What concerns patients greatly is how to find medical professionals with the best expertise for resolving their health issues (Narducci et al. (2015)). Most existing healthcare providers do not provide patients with full infrastructure or service design implementations that assist them in fulfilling this task. This gap raises an open topic on patient-doctor matchmaking, in which patients can find the right doctors to build a trust relationship Figure 1.13. Han et al. (2018) proposed a hybrid RS, in which family-doctor recommendations are made based on the level of available information about users.

4. **Workout recommendation:** Besides treatment recommendations, HRSs now focus on physical-activity suggestions to prevent fragility and health complications (Valdez et al. (2016)). These recommendations help users meet calorie-burn goals through personalized plans Figure 1.14.

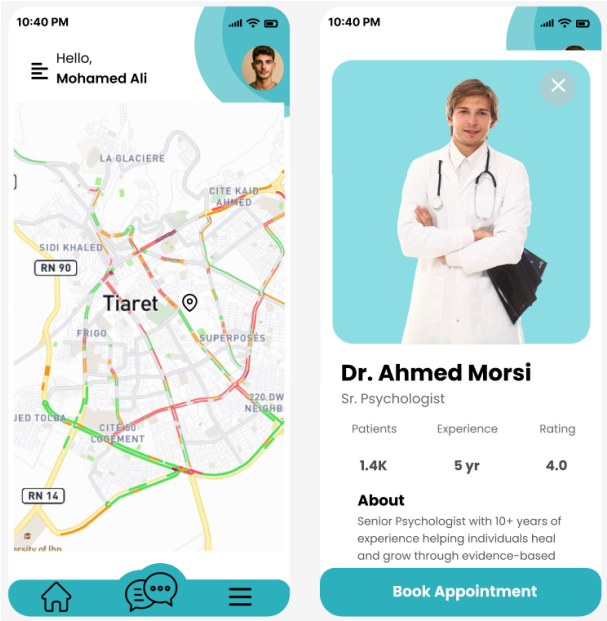


Figure 1.13: Healthcare professional recommendation system

Systems like RUNNER Donciu et al. (2011) and SHADE Faiz et al. (2014) offer food and exercise recommendations, considering that meal timing is crucial for exercise effectiveness. Recommendations are tailored using user data from various sources, including health status, goals and preferences. Ontologies and semantic technologies help manage data heterogeneity (Orgun and Vu (2006)). The process involves selecting exercises based on health status and goals, then refining them using usage history and feedback before delivering them to users.

5. **Health status prediction:** Over the past few decades, researchers have spent a lot of time studying how to predict the risks of certain diseases. In particular, studies on chronic diseases have increased a lot because these illnesses are spreading quickly worldwide (Hussein et al. (2012)). Chronic diseases can make it hard for people to stay active and can also be expensive and time-consuming to treat (Nasiri et al. (2016)).

   To help people avoid these diseases, HRSs have been created. These systems can detect symptoms early and assist doctors in planning the best treatments for patients. Davis et al. (2009) developed RSs that predict possible risks, such as complications or other diseases, that a person with a chronic illness might face in the future. These systems use CF, which is based on the idea that "patients with similar diseases and health conditions may have similar risks." The system makes predictions by comparing a patient's information with data from similar patients. Traditional CF techniques have been adjusted to work better in the healthcare field.

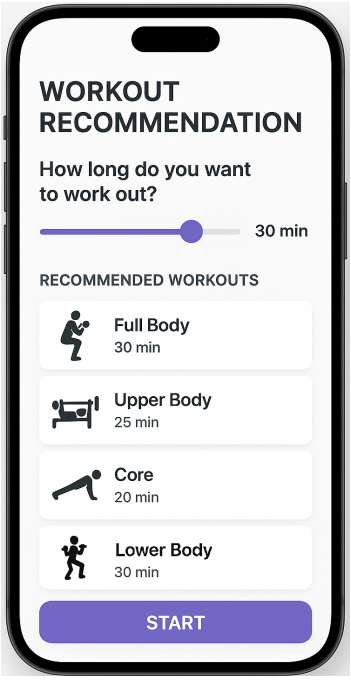


Figure 1.14: Workout recommendation system

### 1.3.2 Evaluation

Evaluating HRSs is crucial to ensure their effectiveness, trustworthiness and impact on users. Below, we discuss key evaluation criteria based on recent research.

#### 1.3.2.1 Evaluation Metrics and Real World Implementation

HRSs are often evaluated based on traditional accuracy metrics. However, a recent scoping review highlights the lack of real-world assessments measuring actual health outcomes. Future studies should incorporate longitudinal evaluations and user feedback to assess practical effectiveness (Chen et al. (2024)).

#### 1.3.2.2 Beyond Accuracy: Trust, Ethics and Privacy

HRSs must be evaluated not just on accuracy but also on trust, privacy and ethical considerations. Researchers suggest that causability, robustness and interpretability should be key components of evaluation (Hmida et al. (2020)).

#### 1.3.2.3 User Engagement and Reproducibility

A systematic review of 73 HRSs found that many focus on lifestyle and nutrition but lack standardized evaluation frameworks. The study proposes five guidelines for improving reporting clarity and reproducibility (Smith et al. (2021)).

## 1.4 Conclusion

In conclusion, this chapter provides a comprehensive overview of recent developments in RSs. The field has seen significant progress in the past few years due to the attention of researchers and academicians. We have identified and explain different topics in RSs domain like different application fields, techniques used, performance metrics, challenges of different RSs and the research gaps and challenges were put forward to explore the future research perspective on RSs.

Furthermore, we aimed more attention in our research on the HRS domain, especially the DRS which will be discussed in depth in the next chapter.

# Chapter 2

# Drug Recommendation Systems: A Literature Review

## 2.1 Introduction

Hospitals have a large amount amount of patient data and a constantly growing number of treatment options, making it difficult for doctors to choose the best course of action. A recommendation engine solves this by analyzing the history, symptoms, and tests of a patient to find similar cases in its database. By identifying what treatments were most successful for those similar patients, DRS helps medical professionals make more precise and effective decisions.

In this chapter, we explore the field of drug RS, highlighting traditional approaches as well as modern techniques. We will explore the various models that have shaped this field, from the early rule-based systems to the latest DL based methods. Furthermore, we will discuss the data sources used to train these models, and the evaluation strategies that ensure their effectiveness. In addion, we well acknowledge the challenges and gaps in drug recommendation in current research and potential opportunities for improving these systems will also be highlighted.

## 2.2 Overview of Drug Recommendation Systems

DRSs represent a pivotal application of IT and Artificial Intelligence (AI), specifically ML, within the modern healthcare ecosystem. At their core, DRSs are sophisticated computational tools designed to assist healthcare professionals by suggesting appropriate medications, dosages, or comprehensive treatment regimens tailored to individual patients (Waleed et al. (2021)). These systems operate by processing and analyzing vast quantities of patient-specific data, which can include EHRs, demographic details, genomic information, past medical history, current symptomatology, and laboratory test results. This patient data is often contextualized with extensive medical knowledge bases, such as pharmacological databases, disease ontologies, and established clinical guidelines.

The imperative for such systems stems from the confluence of rapidly expanding medical knowledge and the increasing complexity of patient care. Healthcare facilities are inundated with an ever-growing volume of patient data, while the pharmaceutical landscape sees a continuous influx of new drugs, and treatment protocols evolve with ongoing research (Shameer et al. (2018)). This `information overload` makes it exceptionally challenging for clinicians to assimilate all relevant variables for every patient, potentially impacting the quality and timeliness of care.

As noted in the introduction, "it becomes increasingly difficult for health professions to decide which treatment to provide to a patient based on his symptoms, test results or previous medical history".

It is within this intricate environment that the importance of DRSs becomes particularly salient. DRSs offer significant potential to:

- Enhance Personalized Medicine
- Improve Clinical Decision Support
- Mitigate Information Overload
- Reduce Medication Errors and Adverse Events
- Increase Efficiency in Clinical Workflows

### 2.2.1 Taxonomy

To better understand the diverse landscape of DRSs, it is useful to classify them along several key dimensions. This taxonomy helps in categorizing existing systems, identifying research trends, and pinpointing areas for future development. The primary dimensions for classifying DRSs include their recommendation objective, the underlying approach or methodology, the types of data utilized, and the target users (Zheng et al. (2021a)).

- **Recommendation Objective:** This dimension defines the primary goal or task the DRS is designed to accomplish. Figure 2.1 summarizes the main objectives of DRSs. Common objectives include:

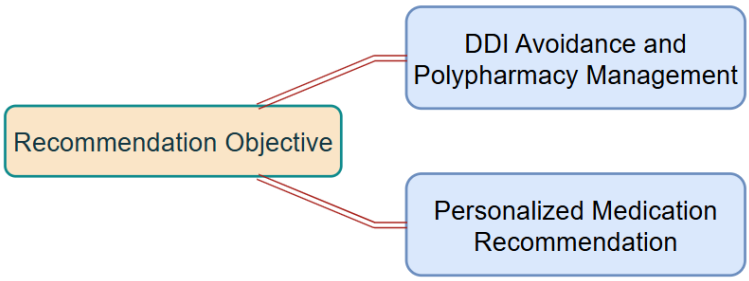


Figure 2.1: Core objectives of DRSs

  - *Drug-Drug Interaction (DDI) Avoidance and Polypharmacy Management:* Systems designed to identify and flag potentially harmful interactions between multiple drugs a patient might be taking. This is crucial for patients with multiple comorbidities requiring polypharmacy, aiming to enhance safety by preventing adverse events.
  - *Personalized Medication Recommendation:* This involves tailoring drug prescriptions and dosages based on an individual patient's comprehensive profile. Such profiles may include demographics, genetic markers, lifestyle, comorbidities, and treatment history, moving towards precision medicine.

- **Recommendation Approach / Methodology:** This describes the core algorithmic or technical strategy employed by the DRS to generate recommendations. As shown in Figure 2.2, DRSs can be classified by the methodology they use to generate recommendations. Broad categories include:

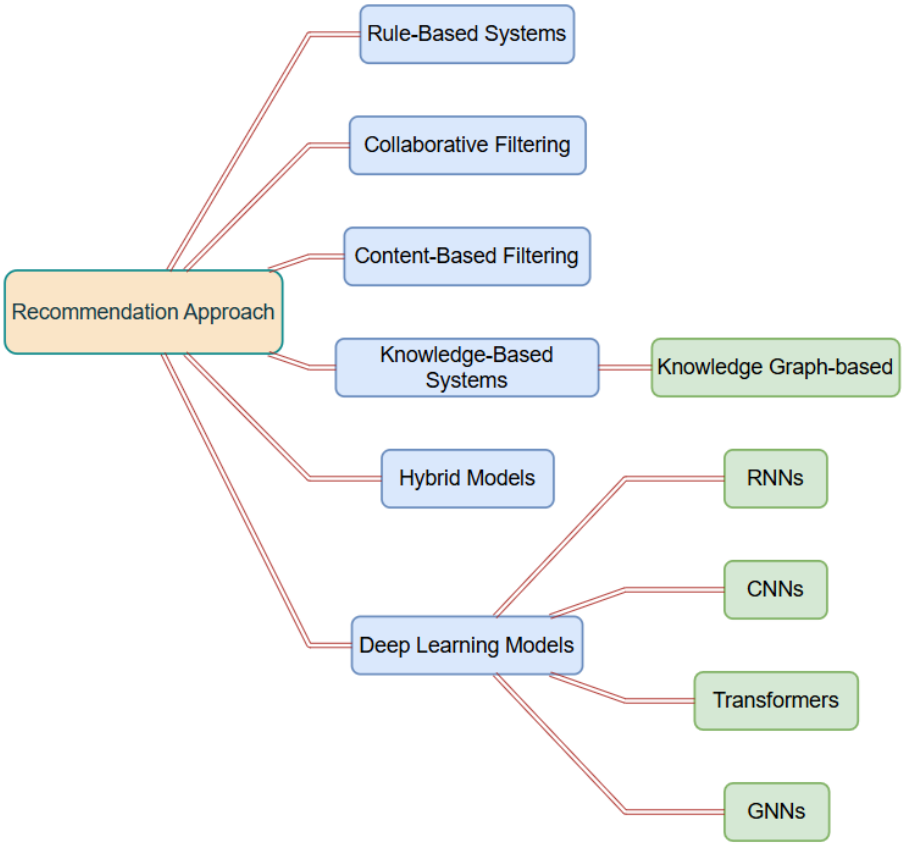


Figure 2.2: Methodologies used in DRSs

  - *Rule-Based Systems:* Utilize predefined rules, often derived from clinical guidelines, expert knowledge, or pharmacological databases (e.g., for DDIs from sources like `DrugBank` (Wishart et al. (2018))). Related work often involves encoding clinical pathways or safety alerts directly into the logic of system (Wright and Sittig (2009)).
  - *Collaborative Filtering (CF):* Predicts drug suitability by identifying patterns from a large group of patients, recommending drugs that were effective for similar patients or that similar patients rated highly (Adomavicius and Tuzhilin (2005)). In DRSs, this often involves analyzing EHRs to find patient clusters with similar treatment responses (Zheng et al. (2021b)).
  - *Content-Based Filtering:* Recommends drugs based on the characteristics (content) of the drugs themselves (e.g., chemical structure, mechanism of action from databases like `Chemical Exploration of Molecule Bioactivity Large-scale (ChEMBL)` (Mendez et al. (2019))) and/or the features of the patient (e.g., specific symptoms, biomarkers). Related work includes predicting drug efficacy based on molecular fingerprints or patient

genomic data (Gottlieb et al. (2011)).

- *Knowledge-Based Systems (including Knowledge Graphs (KGs)-based):* Leverage structured medical knowledge, often represented in ontologies (e.g., `SNOMED CT`, `MeSH`) or KGs, to infer relationships between diseases, drugs, genes, and symptoms for recommendation (Hogan et al. (2021)). Work in this area often focuses on drug repurposing, polypharmacy side-effect prediction, or explaining recommendations through paths in the KG (Xiong et al. (2021)).
- *Hybrid Models:* Combine two or more of the above approaches (e.g., CF with content-based features, or rule-based systems augmented with ML) to leverage their respective strengths and mitigate individual weaknesses (Burke (2002)). In DRSs, a hybrid model might use a KG to enrich patient/drug features for a CF algorithm or use rules to post-filter ML-generated recommendations for safety (Sun et al. (2022)).
- *Deep Learning (DL) Models:* Employ various neural network architectures (e.g., RNNs for sequential patient data, CNNs for medical imaging or molecular structures, Transformers for clinical text, GNNs for KGs or interaction networks) to learn complex patterns from large-scale and often heterogeneous health data (Gao et al. (2022)). Applications in DRSs include predicting drug-target interactions, drug adverse events, or personalized treatment effects from multi-modal data (Wang et al. (2020b)).

- **Data Type Utilized:** The nature of the input data significantly influences the design and capabilities of a DRS. As illustrated in Figure 2.3, DRSs often rely on various types of data, which can be broadly categorized into structured, unstructured, and multi-modal formats.

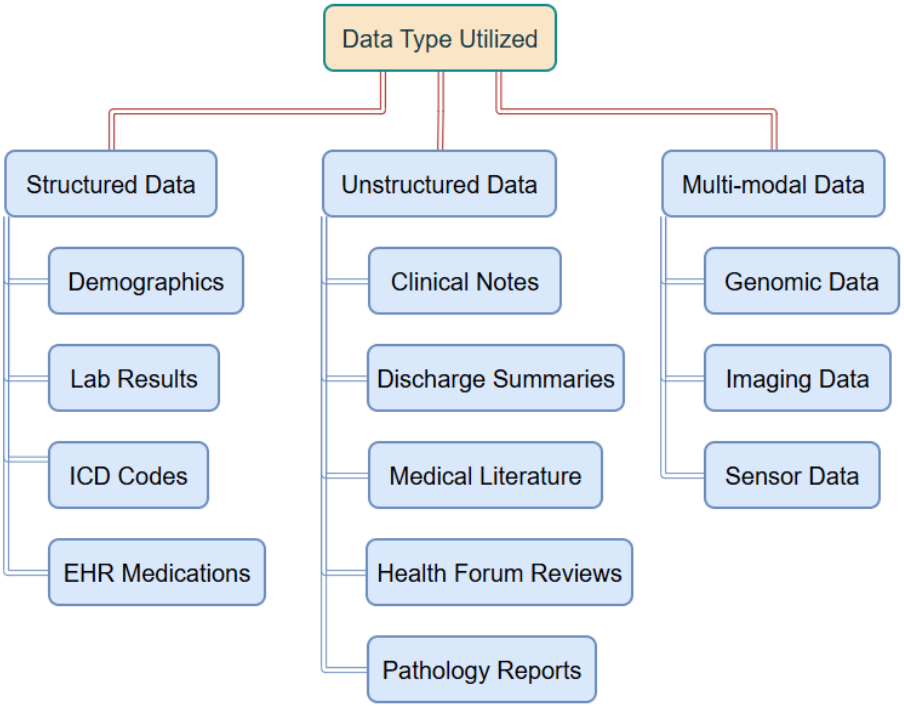


Figure 2.3: Taxonomy of data types used in DRSs

Data sources can be broadly categorized as:

- *Structured Data:* Includes organized and easily queryable data such as patient demographics, lab results, codified diagnoses (e.g., `ICD codes`), medication administration records from EHRs, and data from structured biomedical databases (e.g., `DrugBank`, `SIDER`).
- *Unstructured Data:* Comprises free-text data that requires Natural Language Processing (NLP) techniques for information extraction. Examples include clinical notes, discharge summaries, medical literature, patient-generated reviews from health forums, and pathology reports.
- *Multi-modal Data:* Some advanced systems aim to integrate diverse data types, such as genomic data (e.g., SNPs), imaging data (e.g., MRI, CT scans), and sensor data (e.g., from wearables), alongside clinical data.

- **Target User:** DRSs can be designed with different end-users in mind, which affects their interface, the type of recommendations provided, and the level of detail in explanations. Figure 2.4 presents the types of target users for DRSs. Depending on whether the system is designed for clinicians, patients, or researchers, its functionalities, interface, and level of explanation will vary accordingly.

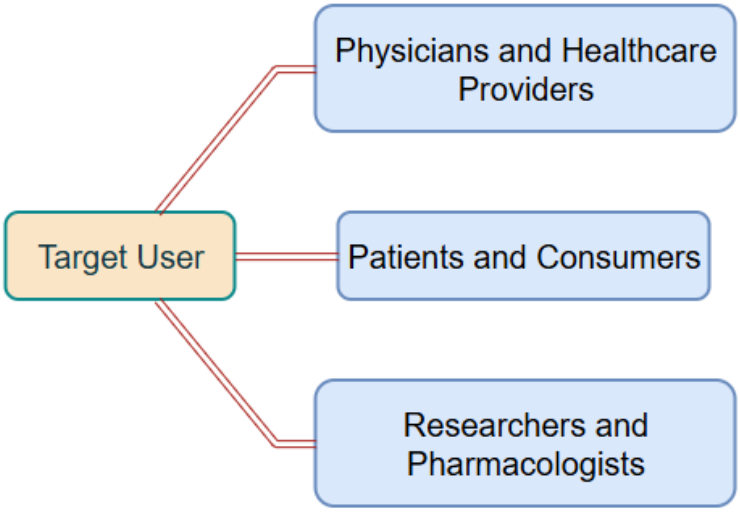


Figure 2.4: Target users of DRSs

- *Physicians and Healthcare Providers:* Systems tailored for clinicians typically focus on providing decision support at the point of care, integrating with EHRs, and offering evidence-backed recommendations that can be critically evaluated by the expert.
- *Patients and Consumers:* DRSs aimed at patients might focus on medication adherence, understanding potential side effects, providing information on over-the-counter drugs, or supporting self-management of chronic conditions through health apps. These systems usually prioritize user-friendliness and easily understandable information.

– *Researchers and Pharmacologists:* Some systems may be designed for drug discovery, repositioning, or pharmacovigilance research, providing insights into drug efficacy, safety profiles, or novel therapeutic targets.

Understanding these classifications is essential for navigating the literature on DRSs and for designing new systems that effectively address specific clinical needs or research questions. Each combination of objective, approach, data, and target user presents unique challenges and opportunities.

### 2.2.2 Data Sources

The effectiveness of DRSs largely depends on the quality, diversity, and appropriate use of the datasets employed for their training and evaluation. The nature of these data sources—whether structured clinical records, unstructured patient narratives, or curated knowledge bases—profoundly influences the choice of recommendation methodologies and the types of insights that can be derived. This section highlights prominent data sources commonly utilized in DRS research.

- **Medical Information Mart for Intensive Care (MIMIC):** The MIMIC datasets are among the most extensively used public data sources in healthcare-related ML and DRS research. These datasets contain comprehensive health records of patients admitted to Intensive Care Units (ICUs), encompassing demographics, vital signs, laboratory measurements, medication administration records, diagnoses, procedures, and lengthy clinical notes (Johnson et al. (2016))[1]. For drug recommendation tasks, tables such as `PRESCRIPTIONS` (detailing drug orders), `INPUTEVENTS` (covering drug administrations, particularly for IV medications), and `NOTEEVENTS` (containing free-text clinical narratives) are particularly relevant and widely exploited.

- **Drug Review Datasets (e.g., `Drugs.com`, `WebMD`):** Datasets comprising user-submitted reviews of pharmaceutical drugs, often scraped from consumer health websites like `Drugs.com`[2] or `WebMD`[3], offer a rich source of real-world patient experiences. Each entry typically includes the drug name, the condition treated, a patient-assigned rating (e.g., for efficacy or satisfaction), reported side effects, and free-text reviews detailing personal experiences (Garg (2021)). Such datasets are invaluable for studies leveraging sentiment analysis and NLP to incorporate patient perspectives, satisfaction levels, and side effect into recommendation models.

- **FDA Adverse Event Reporting System (FAERS):** Maintained by the USA Food and Drug Administration (FDA), the FAERS[4] database is a crucial repository for post-marketing surveillance of drug safety. It contains voluntary reports of Adverse Drug Events (ADEs) and medication errors submitted by healthcare professionals, consumers, and manufacturers (fae

[1] MIMIC-III Dataset on PhysioNet
[2] `https://www.drugs.com/`
[3] `https://www.webmd.com/`
[4] `https://open.fda.gov/data/faers/`

(2021)). FAERS data is frequently employed in pharmacovigilance research and for developing safety-aware RSs. These systems aim to identify and potentially avoid recommending drugs that have a historical pattern of causing adverse events in patients with similar profiles or co-medications.

- **Side Effect Resource (SIDER):** The `SIDER`[5] database consolidates information on marketed medicines and their documented ADRs, extracted from public documents and package inserts (Kuhn et al. (2016)). It systematically links drugs to their known side effects, often categorized by frequency. `SIDER` is a valuable resource for developing DRSs that prioritize safety, particularly in scenarios involving polypharmacy where the risk of cumulative side effects or interactions is high. It can also be used to inform content-based models by providing drug feature information related to safety profiles.

- **UK Biobank:** The UK Biobank[6] is a large-scale, prospective biomedical database containing in-depth genetic and health information from over 500,000 UK participants (Sudlow et al. (2015)). Data includes prescription records, diagnoses (linked to primary and secondary care data), extensive genotyping and exome sequencing data, lifestyle questionnaires, and various physiological measurements. Its rich, multi-modal nature makes it exceptionally valuable for building highly personalized DRS models, particularly those aiming to integrate pharmacogenomic insights to predict drug response or ADR susceptibility.

- **National Health and Nutrition Examination Survey (NHANES):** NHANES[7] is a program of studies designed to assess the health and nutritional status of adults and children in the USA, conducted by the National Center for Health Statistics (NCHS). It uniquely combines interviews (including prescription drug use, dietary information, and health conditions) with physical examinations and laboratory tests (nha (2019)). The prescription drug component of NHANES provides population-level data on medication usage patterns, which can be used to build or validate DRSs, understand medication trends, and explore associations between drug use and health outcomes in a representative sample.

In addition, several other data sources (e.g., `Kyoto Encyclopedia of Genes and Genomes (KEGG)`, `DrugBank`, `ChEMBL`, `PubChem`) are essential for many DRSs. Databases like `KEGG`[8] Kanehisa and Goto (2000) provide information on metabolic pathways and drug targets. `DrugBank` Wishart et al. (2018)[9] offers comprehensive details on drugs, including chemical data, pharmacology, interactions, and targets. `ChEMBL`[10] and `PubChem`[11] are rich sources of chemical structures and bioactivity data. This "side information" is crucial for Constructing feature, build and enrich medical KGs.

[5] http://sideeffects.embl.de/
[6] https://www.ukbiobank.ac.uk/
[7] https://www.cdc.gov/nchs/nhanes/index.html
[8] https://www.genome.jp/kegg/
[9] https://go.drugbank.com/
[10] https://www.ebi.ac.uk/chembl/
[11] https://pubchem.ncbi.nlm.nih.gov/

The wise selection, preprocessing, and integration of these diverse data sources are paramount for developing robust, effective, and clinically relevant DRSs. The subsequent challenge lies in appropriately evaluating the performance and utility of systems built upon such data.

### 2.2.3 Evaluation Strategies

The rigorous evaluation of DRSs is important to ensure their efficacy, safety, and clinical utility before any consideration for real-world deployment. Given the critical nature of healthcare decisions, evaluation cannot just rely on computational metrics; it must also include clinical relevance and user acceptance. This section outlines common quantitative metrics, many of which are adaptations of general metrics mentionned in Section 1.2.3.1, and essential qualitative and clinical evaluation approaches for DRSs.

#### 2.2.3.1 Quantitative Metrics

Quantitative metrics provide objective measures of a DRSs performance, typically by comparing its recommendations against a ground truth (e.g., drugs actually prescribed, drugs known to be effective from literature or trials). These are often adapted from the fields of information retrieval and ML (Manning et al. (2008)), with foundational concepts introduced in Section 1.2.3.1 and further detailed in Section 3.3.3.

- **Precision@k, Recall@k, and F1-score@k:** These are fundamental metrics for evaluating set-based recommendation tasks (Powers (2011)), building upon the definitions of Precision (Equation 3.13), Recall (Equation 3.14), and F1-score (Equation 3.15) that will be see Chapter 3. In DRSs, it is common to evaluate the top-$k$ recommendations, as clinicians typically review a limited list.
  - *Precision@k:* Measures the proportion of recommended drugs within the top-$k$ list that are relevant. In DRSs, this indicates how many of the $k$ suggested drugs are appropriate for the patient/condition. It is calculated as:

$$\text{Precision@k} = \frac{|\{\text{Relevant Recommended Drugs}\} \cap \{\text{Top-k Recommended Drugs}\}|}{k} \tag{2.1}$$

    where $k$ is the number of top recommendations considered.
  - *Recall@k:* Measures the proportion of all relevant drugs (that should have been recommended) that are actually present in the top-$k$ recommendations. In DRSs, this shows how many of the truly appropriate drugs the system managed to find within the top-$k$. Its formula is:

$$\text{Recall@k} = \frac{|\{\text{Relevant Recommended Drugs}\} \cap \{\text{Top-k Recommended Drugs}\}|}{|\{\text{All Relevant Drugs}\}|} \tag{2.2}$$

– *F1-score@k:* The harmonic mean of Precision@k and Recall@k, providing a single measure that balances both for the top-$k$ items. It is useful when both false positives and false negatives within the top-$k$ are important, and is defined as:

$$F1\text{-score@k} = 2 \cdot \frac{\text{Precision@k} \cdot \text{Recall@k}}{\text{Precision@k} + \text{Recall@k}} \tag{2.3}$$

where $k$ is the number of top recommendations considered.

The choice of $k$ (the number of recommendations considered) is crucial and should reflect realistic clinical scenarios.

- **Area Under the Receiver Operating Characteristic Curve (AUROC):** The Receiver Operating Characteristic (ROC) curve, as represented in Figure 2.5 (Martinez-Ríos et al. (2021)), plots the True Positive (TP) Rate (Recall) against the False Positive (FP) Rate at various threshold settings. The AUROC represents the probability that the model ranks a randomly chosen positive instance higher than a randomly chosen negative instance (Fawcett (2006)). In DRSs, this can be used if the task is framed as binary classification (e.g., will this drug be effective/safe for this patient: yes/no) or for evaluating ranked lists where a threshold determines recommendation. A higher AUROC (closer to 1) indicates better discrimination.

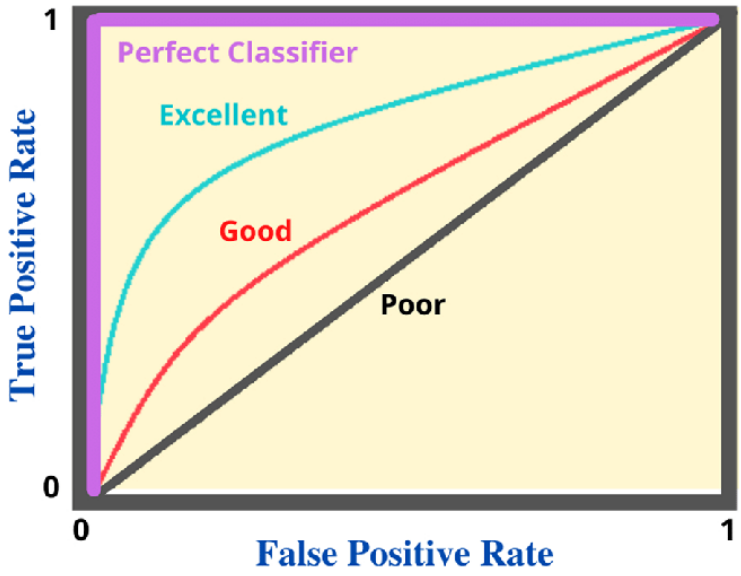


Figure 2.5: AUROC

- **Mean Average Precision (MAP):** A popular metric for evaluating ranked retrieval results. It considers the order of recommendations and is the mean of Average Precision (AP) scores across all queries (or patients). AP for a single patient rewards systems that rank relevant drugs higher in the recommendation list. This is particularly important in DRSs as clinicians are more likely to consider drugs listed at the top. For a set of queries

$Q$, MAP is defined as:

$$\text{MAP} = \frac{1}{|Q|} \sum_{q \in Q} \text{AP}(q) \tag{2.4}$$

where AP for a query $q$ is:

$$\text{AP}(q) = \frac{\sum_{k=1}^{N_q} (P(k) \times rel(k))}{\text{Number of relevant drugs for } q} \tag{2.5}$$

Here $N_q$ is the number of recommended drugs for query $q$ (or total drugs in the list considered for ranking), $P(k)$ is the precision at cut-off $k$ in the list of recommendations for query $q$, $rel(k)$ is an indicator function equaling 1 if the item at rank $k$ is relevant for query $q$, and 0 otherwise. "Number of relevant drugs for $q$" is the total count of truly relevant drugs for patient $q$.

- **Normalized Discounted Cumulative Gain (NDCG)@k:** Evaluates ranked lists but allows for multiple levels of relevance (e.g., highly relevant, moderately relevant, irrelevant drug) (Järvelin and Kekäläinen (2002)). Discounts the value of relevant drugs found lower in the list. The "normalized" aspect compares the system's Discounted Cumulative Gain (DCG) to the Ideal Discounted Cumulative Gain (IDCG) (if all relevant drugs were ranked perfectly), resulting in a score between 0 and 1. This is highly suitable for DRSs where not all correct drugs are equally optimal.

$$\text{DCG@k} = \sum_{i=1}^{k} \frac{rel_i}{\log_2(i+1)} \tag{2.6}$$

$$\text{NDCG@k} = \frac{\text{DCG@k}}{\text{IDCG@k}} \tag{2.7}$$

where $k$ is the number of recommendations considered, $rel_i$ is the graded relevance of the drug at position $i$ in the recommended list, and IDCG@k is the Ideal DCG at $k$.

- **Hit Rate (HR)@k:** This simpler metric measures the proportion of cases (e.g., patients) for which at least one of the ground-truth relevant drugs appears in the top-$k$ recommended drugs. It indicates whether the system can successfully suggest at least one correct option within a limited list size.

$$\text{HR@k} = \frac{1}{|U|} \sum_{u \in U} \text{hit}(u, k) \tag{2.8}$$

where $|U|$ is the total number of users or test cases (e.g., patients), $u$ is an individual user or test case, and $\texttt{hit}(u, k)$ is an indicator function, which is 1 if at least one relevant drug for user $u$ is found within the top-$k$ recommendations, and 0 otherwise. $k$ is the number of top recommendations considered.

- **Coverage:** As introduced in Section 1.2.3.1 and illustrated in Figure 2.6, coverage measures the percentage of the total drug formulary (or relevant drug set) that the DRS

is capable of recommending. Low coverage might indicate the system is overly specialized or biased towards popular drugs, which is a critical concern in healthcare.

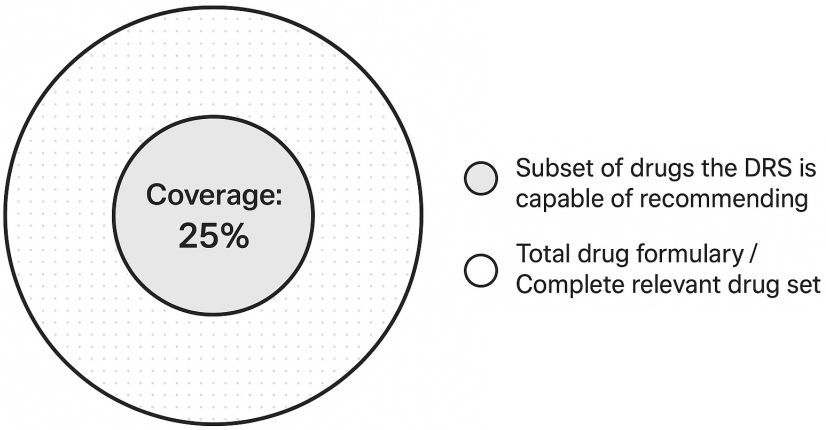


Figure 2.6: Example of coverage illustration in DRS

It is important to note that the definition of `relevance` or `ground truth` in DRS evaluation can be complex, depending on whether it is based on historical prescriptions (which may not always be optimal), clinical trial outcomes, or expert-defined appropriateness.

#### 2.2.3.2 Qualitative and Clinical Evaluation

While quantitative metrics are essential for assessing algorithmic performance, they do not capture the full picture of a DRSs value in a clinical setting. Qualitative and clinical evaluations are crucial for assessing safety, real-world utility, and trustworthiness. This involves having domain experts (e.g., physicians, clinical pharmacists, specialists) review the recommendations generated by the DRS.

- *Methodologies:* This can be done through expert panel reviews, where a group of clinicians assesses recommendations for given patient scenarios (real or simulated cases). They evaluate aspects like appropriateness for the diagnosis, safety (considering DDIs, ADRs, contraindications), alignment with clinical guidelines, and potential efficacy. Scoring rubrics or Likert scales are often used[12].
- *Focus:* Experts can identify nuances that quantitative metrics might miss, such as whether a recommended drug, while technically correct for a disease, is inappropriate due to a patient's specific comorbidity, age, or other contextual factors not fully captured by the model.

A comprehensive evaluation strategy for DRSs should thus combine offline quantitative assessments with rigorous online or expert-driven qualitative and clinical validation to ensure the system is not only accurate but also safe, useful, and trustworthy in practice.

[12]Scoring rubrics are tools that assess performance based on defined criteria, while Likert scales measure attitudes using agreement ratings (e.g., 1 = strongly disagree to 5 = strongly agree).

## 2.3 Challenges and Opportunities

Despite significant advancements, the field of DRSs faces numerous challenges that hinder their widespread adoption and optimal performance. Addressing these challenges concurrently presents exciting opportunities for future research and development, ultimately aiming to create more effective, safer, and personalized clinical decision support tools. Many of these challenges are specific manifestations or intensifications of general issues discussed in Section 1.2.2.5.

### 2.3.1 Data-related

The foundation of any DRS is data, and issues related to data quality, availability, and characteristics are paramount.

- **Data Sparsity and Missing Values:** EHRs and patient profiles are often incomplete, with many patients having data for only a few drugs or conditions as illustrated in Figure 2.7. This high data sparsity hinders CF and other models from identifying reliable patterns (Lu et al. (2015)) (see Section 1.2.2.5).

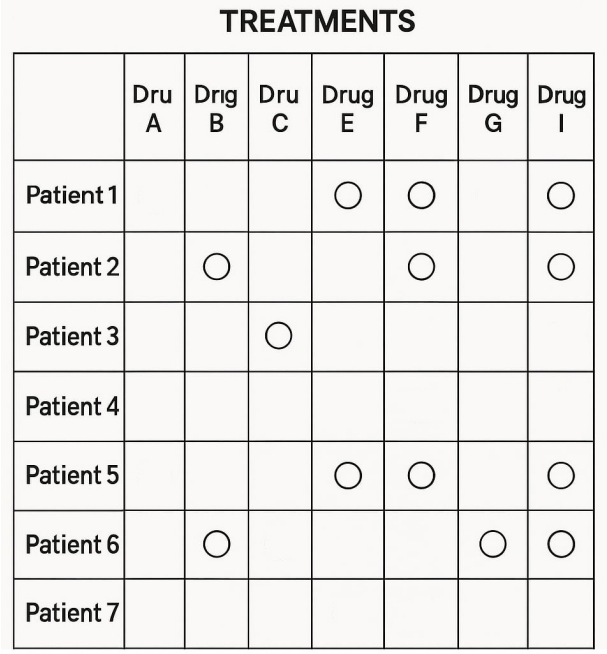


Figure 2.7: Illustration of Data Sparsity

- **Data Imbalance:** Healthcare datasets are typically imbalanced, overrepresenting common diseases and drugs while underrepresenting rare but potentially more appropriate treatments. This bias, related to the long-tail problem (Section 1.2.2.5), can reduce recommendation effectiveness for less common cases.

- **Label Noise and Quality:** Training data such as diagnoses, prescriptions, and outcomes may be inaccurate, incomplete, or subjective. Issues like miscodings and non-adherence introduce noise and affect model reliability.

- **Privacy and Security:** Patient data is highly sensitive and regulated (e.g., Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR)). Ensuring privacy and security through techniques like de-identification, federated learning[13] and differential privacy[14] is essential for compliance and trust (El Emam and Arbuckle (2013)), building upon general RS privacy concerns (Section 1.2.2.5).

- **Lack of Standardized Data Formats and Interoperability:** Inconsistent data formats and terminologies (e.g., `SNOMED CT`, `ICD`, `RxNorm`) across healthcare systems limit data integration and model generalizability, amplifying synonymy-related issues (Section 1.2.2.5) (Benson (2012)).

### 2.3.2 Model-related

The algorithms and models themselves present several intrinsic challenges when applied to the DRS domain.

- **Personalization vs. Generalization:** Achieving a balance between personalized recommendations (based on genomics, comorbidities, lifestyle, etc.) and models that generalize well to unseen patients is critical (Figure 2.8). Over-personalized systems may fail to generalize, reflecting a subtle facet of the cold-start problem (Section 1.2.2.5).

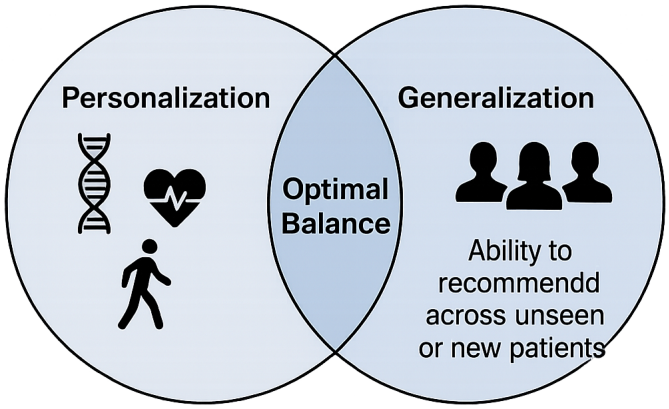


Figure 2.8: Illustration of Personalization vs. Generalization

- **Cold-Start Problem in DRSs:** The cold-start issue is especially prominent in DRSs. Making drug recommendations for new patients (with limited history) or for newly introduced drugs (with insufficient real-world data) requires specialized techniques beyond conventional CF (Section 1.2.2.5).

[13] Federated learning is a machine learning approach where multiple decentralized devices or servers collaboratively train a shared model without exchanging their raw data. Only model updates are communicated, preserving data privacy.

[14] Differential privacy is a technique that adds small amounts of noise to data or computations to protect individuals' private information while still allowing useful insights to be learned from the data.

- **Scalability for Clinical Utility:** DRS models need to be scalable to support large-scale patient populations, comprehensive drug databases, and fast-growing clinical data while still delivering timely recommendations suitable for clinical settings (Section 1.2.2.5).
- **Computational Complexity:** Sophisticated models like deep neural networks, which handle high-dimensional and heterogeneous data, demand significant computational power for both training and inference. This can limit their feasibility in resource-constrained healthcare environments.
- **Temporal Dynamics:** Patients health conditions, treatment responses, and drug effectiveness often change over time. DRS models must capture these temporal trends to offer relevant recommendations, surpassing the static approaches used in traditional RSs.

### 2.3.3 Ethical, Interpretability, and Implementation

Beyond data and models, broader considerations impact the adoption and responsible use of DRSs.

- **Interpretability and Explainable AI (XAI):** As discussed in evaluation contexts, the `black box` nature of many advanced models is a significant barrier to clinical trust and adoption. Clinicians need to understand the rationale behind a drug recommendation to critically assess it, ensure it aligns with their clinical judgment, and take responsibility for the treatment decision (Holzinger et al. (2019)).
- **Bias and Fairness:** AI models can accidentally learn and perpetuate biases present in historical medical data related to race, gender, socioeconomic status, or geographic location, potentially leading to health disparities (Obermeyer et al. (2019)). Ensuring fairness and equity in DRSs is a critical ethical concern that goes beyond general discussions of item popularity bias (Section 1.2.2.5).
- **Accountability and Responsibility:** Determining who is accountable if a DRS provides an incorrect or harmful recommendation (the algorithm developer, the deploying institution, the clinician who accepts or rejects it) is a complex medico-legal and ethical issue that needs clear frameworks.
- **Clinical Workflow Integration:** Seamlessly integrating DRSs into existing clinical workflows and EHR systems without causing alert fatigue, increasing clinician charge, or undue disruption is crucial for practical adoption. The user interface and interaction design are key to success.
- **Regulatory Hurdles and Validation Standards:** Establishing clear regulatory pathways (e.g., via FDA, European Medicines Agency (EMA)) and standardized validation protocols for AI-driven medical devices, including DRSs, is essential for ensuring their safety, efficacy, and reliability before widespread clinical use.

Addressing these multifaceted challenges will be crucial for realizing the full potential of DRSs to transform patient care and improve health outcomes.

## 2.4 Conclusion

This chapter has systematically reviewed the field of DRSs, emphasizing their critical role in advancing personalized medicine and improving clinical decision-making. We explored the taxonomy of DRSs, the evolution of methodologies from traditional to advanced Deep Learning (DL) and hybrid models, and the essential data sources that fuel them. The necessity of rigorous, multifaceted evaluation, confirming both quantitative metrics and clinical interpretability analysis, was also highlighted. Despite significant advancements, major challenges persist in relation to data quality, model robustness (especially in DDI/ADR prediction), and ethical considerations like bias and accountability. However, these challenges also define inspiring direction for future research, focusing on XAI, multi-modal data integration, and fairness-aware systems.
In the third chapter, by recognizing the previews gaps and opportunities, we will propose a novel approach developed to address several of these key limitations.

# Chapter 3

# The Proposed Drug Recommendation System: Design, Implementation, and Evaluation

## 3.1 Introduction

Personalized drug RSs are increasingly critical in modern healthcare, aiming to support clinical decision making by suggesting the most appropriate treatment for individual patients. These systems leverage MLs models to predict the effectiveness of specific drugs based on patient features such as condition, sex and prior experiences with medications. However, many current DRS approaches struggle to effectively associate the richness and subjectivity of patient experiences and the small difference characteristics that define patient groups. This often limits their ability to achieve truly personalized and accurate drug suggestions.

This chapter presents the design, implementation, and evaluation of our proposed model, describing our algorithm and its underlying principles. Finally, we will discuss the results obtained from our experiments, providing insights into the strengths and limitations of our approach.

## 3.2 The Proposed Drug Recommendation System: Architecture and Design

This section outlines the foundational principles and methodologies that rise our proposed drug RS. We cover the three main areas: Sentiment Analysis, User/Item Clustering in Recommendation and Enhanced NCF. Each principle plays a crucial role in capturing accurate user preferences and interaction patterns, ultimately contributing to a more accurate and context aware drug recommendation model. By integrating these concepts, our system aims to move beyond traditional recommendation approaches to provide more personalized and effective drug suggestions.

### 3.2.1 Conceptual Framework

Building upon the concepts and techniques of sentiment analysis, clustering and enhanced NCF that will be see in Section 3.2.2, we introduce the framework of our proposed drug recommendation model 3.1. That aims to capture both implicit interaction patterns and explicit feature characteristics to predict the suitability of a drug for a given user context.

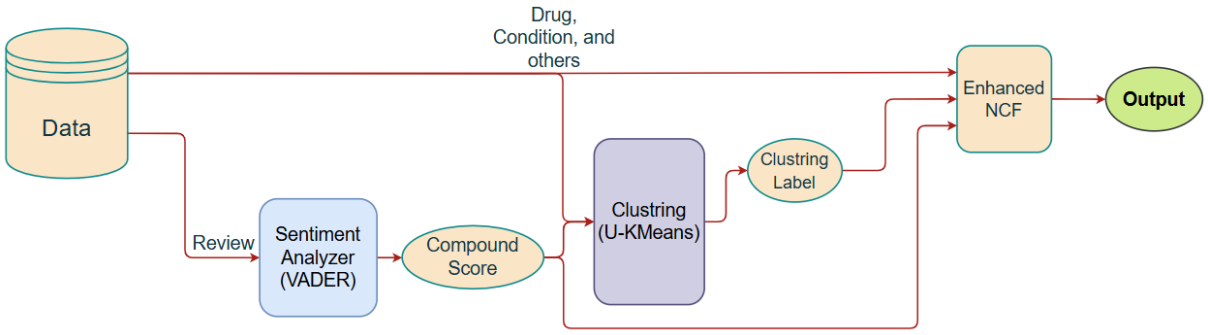


Figure 3.1: The proposed model Conceptual Framework

The model first leverages sentiment analysis to transform each user review into a quantitative compound score that captures overall sentiment polarity on a continuous scale from –1 (most negative) to +1 (most positive). This compound score,provides a robust representation of textual opinions.

Next, these sentiment scores are combined with various data such as drug and medical condition and other additional engineered features to feed a modified clustering algorithm (U-KMeans), which partitions the data into homogeneous groups based on both semantic and contextual similarity. The result of clustering is then introduced as input, alongside drug, condition and other featuers into an enhanced NCF (multi-input NCF).

In sum, this three-stage pipeline lexicon based sentiment extraction, feature clustering, and enhanced deep NCF,constitutes a cohesive architecture optimized for personalized drug recommendation.

### 3.2.2 Main Architectural Components

This section details the main elements that constitute our proposed methodology, each playing a crucial role in enhancing the recommendation process. We begin by exploring Sentiment Analysis in User Reviews (Section 3.2.2.1). Next, we discuss User/Item Clustering in Recommendation (Section 3.2.2.2). Finally, we introduce an Enhanced NCF model (Section 3.2.2.3). Together, these components form an integrated framework for more accurate and effective recommendations.

#### 3.2.2.1 Sentiment Analysis in User Reviews

Sentiment analysis, or opinion mining, computationally studies opinions and emotions in text Liu (2012). In RSs using textual feedback, it helps understand the qualitative aspects of user experiences beyond numerical ratings Pang and Lee (2008).

Integrating sentiment analysis into RSs offers several advantages:

- **Enhanced User Preference Modeling:** Augments ratings by providing deeper insights into why a user liked or disliked an item, helping to disambiguate experiences (e.g., a high rating with negative sentiment might indicate sneer or a mixed experience).

- **Improved Recommendation Accuracy:** By combining sentiment as a feature, models can learn more exact user preferences, leading to more accurate predictions and relevant recommendations Zhang et al. (2014).

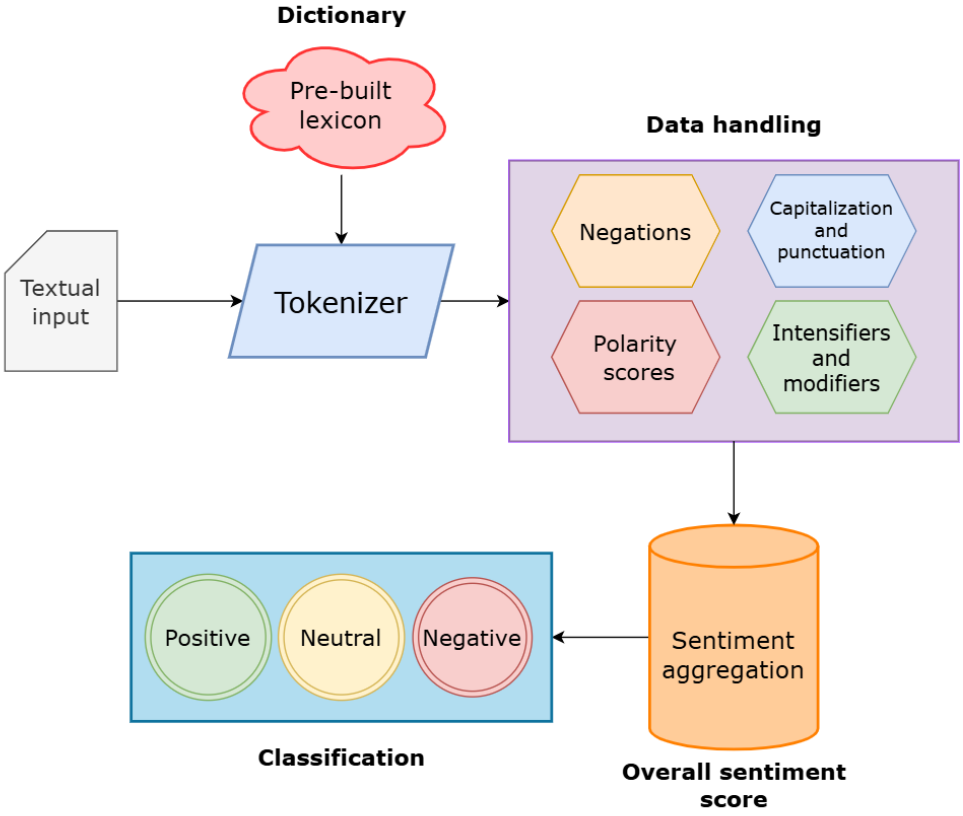


Figure 3.2: Conceptual flow of VADER model in sentiment analysis

- **Feature Engineering:** Sentiment polarity or intensity scores can be directly used as features in recommendation models.
- **Addressing Review Sparsity:** Sentiment from reviews can act as an implicit preference signal, even without explicit ratings.
- **Explainability:** Sentiment can contribute to more explainable recommendations by highlighting positive aspects from reviews.

The model first leverages VADER, a lexicon and rule based sentiment analyzer Hutto and Gilbert (2014), to transform each user review into a quantitative compound score that captures overall sentiment polarity. This compound score, computed by aggregating and normalizing valence scores from a sentiment lexicon, provides a robust single feature representation of textual opinions without extensive preprocessing. The choice of VADER is motivated by its effectiveness in analyzing sentiments expressed in social media and short informal texts, which often share characteristics with user reviews.

#### 3.2.2.2 User/Item Clustering in Recommendation

User/item clustering is a key technique in modern RSs, addressing challenges like data sparsity, scalability, and the cold start problem Linden et al. (2003b). It involves grouping similar users or items to generate recommendations based on collective cluster preferences rather than sparse individual data.

Clustering offers multiple benefits:

- **Scalability:** Reduces computational load by operating on clusters instead of individual entities in large datasets Aggarwal (2016b).
- **Alleviating Data Sparsity:** Creates denser neighborhoods for recommendations by pooling preferences within clusters.
- **Cold Start Problem Mitigation:** Assigns new users or items to existing clusters based on available features to enable initial recommendations.
- **Improved Recommendation Quality and Diversity:** Can lead to more diverse and unexpected recommendations by understanding group behaviors.

Common clustering techniques employed in RSs include K-Means (and its variants), DBSCAN, and hierarchical clustering Jain (2010). The choice depends on data characteristics and recommendation goals. For example, K-Means is efficient for large datasets but needs a predefined number of clusters, while hierarchical clustering offers a nested cluster structure.

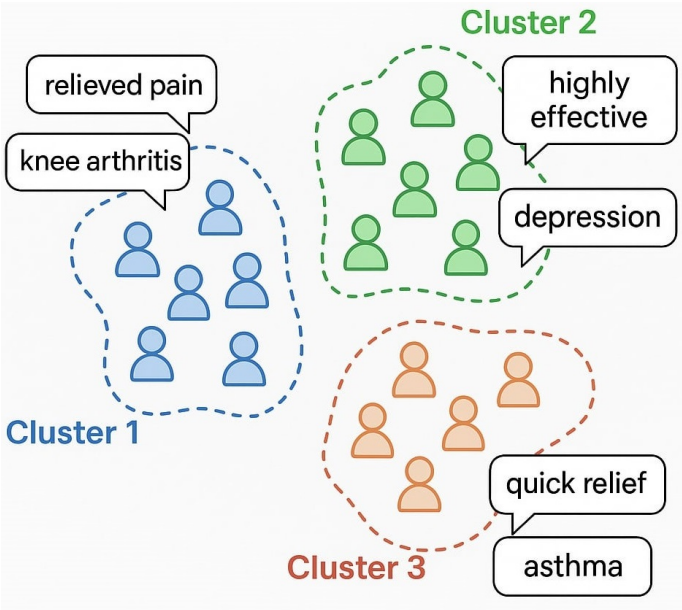


Figure 3.3: Clustering concept

This research use a modified K-Means approach, termed U-KMeans, specifically developed for robust user segmentation. The methodology involves integrating diverse user features, including demographic data, sentiment scores derived from reviews, drug information, medical conditions,

and other engineered features, as input for the U-KMeans algorithm. This algorithm then partitions users into homogeneous groups by considering both semantic (e.g., sentiment) and contextual (e.g., medical condition) similarities as illustrated in Figure 3.3. A key methodological distinction from standard K-Means is that U-KMeans assigns specific weights to the sentiment dimension, ensuring that the resulting clusters effectively capture the small difference of patient experiences alongside other defining characteristics.

#### 3.2.2.3 Enhanced Neural Collaborative Filtering

The NCF model was proposed by He et al. (2017) to leverages DL to model user/item interactions, moving beyond the simple dot product. Its core idea is to replace the dot product with a learnable Multi-Layer Perceptron (MLP) for more expressive interaction modeling.

The NCF framework typically includes:

- **Input Layer:** Takes user and item IDs (e.g., one hot encoded or integer indices).
- **Embedding Layer:** Maps sparse input IDs to dense, lower dimensional embeddings ($\mathbf{p}_u$ for user $u$, $\mathbf{q}_i$ for item $i$) that capture latent features. These embeddings can be initialized randomly and learned during training.
- **Neural CF Layers (Interaction Layers):** The core, where user and item embeddings are fed into neural network layers to learn the interaction function He et al. (2017). Variants include:
  - **Generalized Matrix Factorization (GMF):** Learns user/item interactions via element wise multiplication of their embeddings, generalizing the dot product in traditional Matrix Factorization (MF). The GMF interaction vector is:

$$\mathbf{h}_{\text{GMF}} = \mathbf{p}_u \odot \mathbf{q}_i \tag{3.1}$$

    where $\odot$ denotes the element wise product.
  - **Multi-Layer Perceptron (MLP):** Concatenates user/item embeddings and passes them through MLP layers with non linear activation functions to learn complex interactions. The input to the MLP path is:

$$\mathbf{z}_0 = [\mathbf{p}_u; \mathbf{q}_i] \tag{3.2}$$

    This concatenated vector is then processed through multiple dense layers:

$$\mathbf{z}_1 = \phi_1(\mathbf{W}_1^T \mathbf{z}_0 + \mathbf{b}_1) \tag{3.3}$$
$$\mathbf{z}_2 = \phi_2(\mathbf{W}_2^T \mathbf{z}_1 + \mathbf{b}_2) \tag{3.4}$$
$$\vdots$$
$$\mathbf{h}_{\text{MLP}} = \phi_L(\mathbf{W}_L^T \mathbf{z}_{L-1} + \mathbf{b}_L) \tag{3.5}$$

where $\mathbf{W}_l$, $\mathbf{b}_l$, and $\phi_l$ are the weight matrix, bias vector, and activation function (e.g., ReLU) for the $l$-th layer, respectively. $\mathbf{h}_{\text{MLP}}$ is the output of the MLP path.

– **Neural Matrix Factorization (NeuMF):** A hybrid model combining GMF and MLP. It typically learns separate embeddings for the GMF and MLP paths (e.g., $\mathbf{p}_{u,\text{GMF}}, \mathbf{q}_{i,\text{GMF}}$ and $\mathbf{p}_{u,\text{MLP}}, \mathbf{q}_{i,\text{MLP}}$ respectively). The outputs of the GMF path ($\mathbf{h}_{\text{GMF}}$, derived from its specific embeddings) and the MLP path ($\mathbf{h}_{\text{MLP}}$, derived from its specific embeddings) are then fused, usually by concatenation:

$$\mathbf{v}_{\text{fusion}} = [\mathbf{h}_{\text{GMF}}; \mathbf{h}_{\text{MLP}}] \tag{3.6}$$

– **Output Layer:** Produces the predicted interaction score, often with a sigmoid activation for binary classification (e.g, predicting implicit feedback like click or purchase) or linear activation for solving the regression problem.

The main advantage of NCF are its flexibility, enhanced modeling capacity, and ability to automatically learn complex feature interactions. It often outperforms traditional MF-based methods, especially with sparse data He et al. (2017); Zhang et al. (2019).

Since the DRS is very complex due to the non-linear relationships between drugs, disease conditions, and demographics, to capture all these relationships, we need to improve NCF to capture all interactions from multiple inputs rather than just user/item inputs (vanilla[1] NCF). Therefore, we proposed an improved NCF (multi-input NCF) as shown in Figure 3.4, which takes categorical inputs. In this stage, embedding layers map each categorical feature (drug, condition, ...) into dense latent vectors. These embeddings are then processed through parallel GMF and MLP branches. GMF applies element-wise multiplication to capture latent interactions, while MLP captures higher-order feature interactions via dense layers. The outputs are fused and passed through a final dense layer with linear activation.

## 3.3 Experiments and implementations

This section outlines the specific details of the experimental environment, including the dataset processing, the software stack, development tools, and hardware utilized. It also defines the

[1]The term `vanilla` refers to the simplest or most basic version of a model or algorithm, implemented without any modifications, enhancements, or additional features.

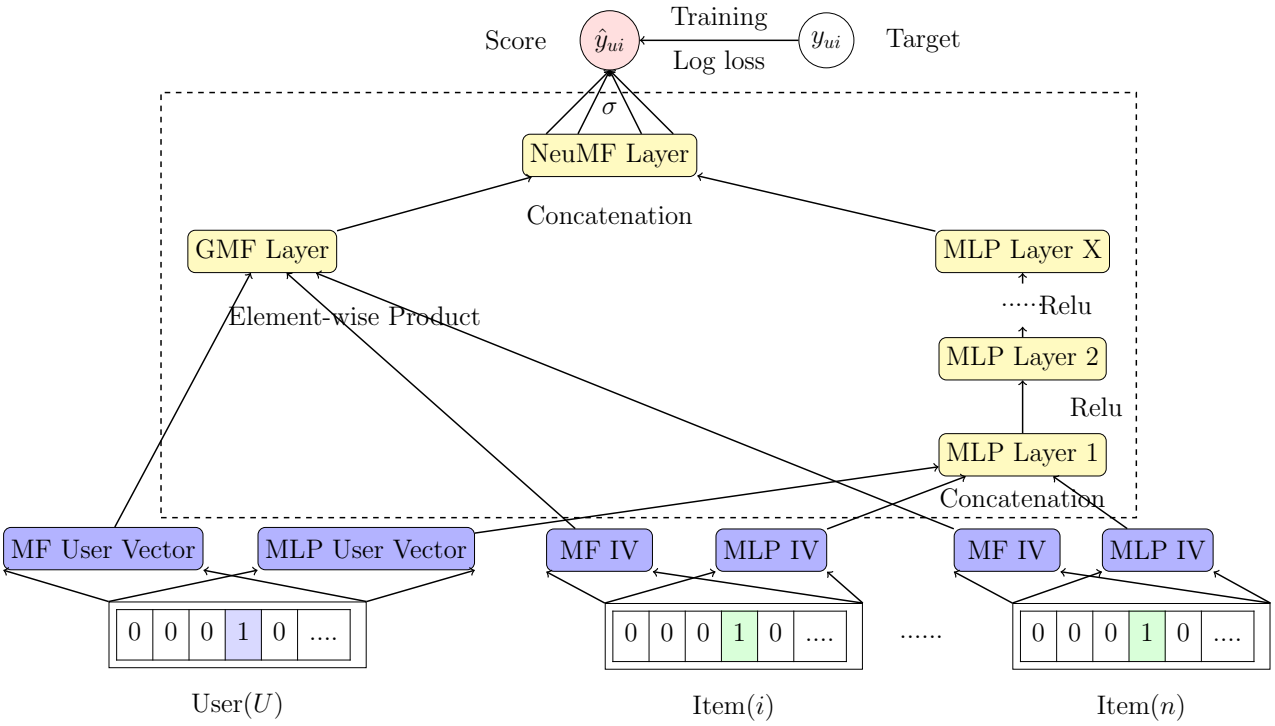


Figure 3.4: Architecture Diagram of the Enhanced NCF (Multi-inputs)

evaluation metrics used to assess the performance of the proposed model and discusses baseline models used for comparison.

### 3.3.1 Dataset and Pre-processing

This section details the dataset processing that utilized for developing and evaluating the proposed drug recommendation model. The primary dataset used in this study is sourced from Github, titled `Recomed Dataset` and is available as `Drug Rating.csv`. This dataset contains patient-reported drug reviews, ratings, and related information. Initially, the dataset comprises 3292 instances (rows) and 15 features (columns).

#### 3.3.1.1 Key Features

The raw dataset contains several features that are pertinent to the task of drug recommendation. The key features utilized in this work are summarized in Table 3.1.

#### 3.3.1.2 Data Cleaning

Prior to any feature engineering, several data cleaning steps were performed to ensure data quality and consistency. The dataset was first processed to remove any rows containing missing values across any of the columns. This was achieved using the pandas `dropna()` method on the entire DataFrame.

| Feature Name | Data Type (Original) | Description |
|---|---|---|
| `DrugName` | Text | The name of the drug reviewed. |
| `Condition Reason` | Text | The medical condition. |
| `Age` | Text/Numeric | The age of the patient. |
| `Genus` | Text | The gender of the patient. |
| `Effectiveness` | Categorical Text | Patient's perceived effectiveness of the drug. |
| `SideEffect` | Categorical Text | Severity of side effects experienced by the patient. |
| `Comment Review` | Text | The textual review or comment by the patient. |
| `OverallRating` | Numeric | An overall numerical rating given by the patient. |
| `Category` | Text | The broader category to which the drug belongs. |

Table 3.1: Key features in the original dataset.

#### 3.3.1.3 Feature Engineering

Following data cleaning, several new features will be engineered, and existing features were transformed to better capture the underlying information relevant for drug recommendation.

- **Text Preprocessing for Review Comments:** The raw textual data in `Comment Review` requires significant preprocessing to be suitable for sentiment analysis. The following steps were applied:
  - **Lowercasing:** Converting all text to lowercase for consistency.
  - **Special Character Removal:** Eliminating punctuation and special characters, retaining only alphanumeric characters and whitespace to simplify text to word tokens.
  - **Stopword Removal:** Removing common English stopwords (e.g., "the", "is") using the Natural Language Toolkit (NLTK) library, as they typically lack significant sentiment meaning.
  - **Lemmatization:** Reducing words to their base form (lemma[2]) using `NLTKs` WordNet lemmatizer to consolidate different word forms.
- **Sentiment Analysis (Polarity of User Comment Score):** To quantify sentiment in user reviews, the VADER sentiment intensity analyzer Hutto and Gilbert (2014) was employed, selected for its effectiveness in handling short texts such as reviews, which are similar in style to social media content.
  - **Methodology:** An instance of `NLTKs Sentiment Intensity Analyzer` (with the `vader_lexicon`) was initialized. For each preprocessed review, VADER `polarity_scores` method yielded positive, negative, neutral, and a composite `compound` score.

[2]A lemma is the base or dictionary form of a word. For example, the lemmas of "running" and "better" are "run" and "good," respectively.

– **Sentiment Indicator:** The `compound` score, a normalized weighted composite ranging from -1 (most negative) to +1 (most positive), was extracted as the primary sentiment measure.

This compound sentiment score was stored in a new column named `PUC`.

- **Mapping Categorical Features (Effectiveness and Side Effects):** The categorical text features `Effectiveness` and `SideEffect` were converted to numerical scales to capture their ordinal nature.

  – **`Effectiveness` to `DOE` (Degree of Effectiveness):** The `Effectiveness` categories (from `Ineffective` to `Highly Effective`) were mapped to numerical values from 0 to 4, respectively. This raw score was then normalized by dividing by 4, scaling it to a 0-1 range to produce the `DOE`.

  – **`SideEffect` to `DOS` (Degree of Side Effects):** Similarly, the `SideEffect` categories (from `No Side Effects` to `Extremely Severe Side Effects`) were mapped to numerical values from 0 to 4. This score was also normalized by dividing by 4, yielding a 0-1 scaled value referred to as `DOS`.

- **Consolidated User Rating Calculation:** A novel Consolidated User Rating (CUR), as proposed in the RECOMED framework Zomorodi et al. (2024),was developed to comprehensively represent user experience, integrating overall rating, effectiveness, side effects, and Polarity of User Comment (PUC).

  The `CUR` was then calculated using the following formula:

$$\texttt{CUR} = \frac{\left(\frac{\text{OverallRating}_{\text{norm}} + \texttt{DOE}}{2}\right) - \texttt{DOS} + \texttt{PUC}}{2} \tag{3.7}$$

  where:

  – $\text{OverallRating}_{\text{norm}}$ is the normalized overall rating (scaled to [0, 1]).

  – `DOE` is the Degree of Effectiveness (DOE) (scaled to [0, 1]).

  – `DOS` is the Degree of Side Effects (DOS) (scaled to [0, 1]).

  – `PUC` is the PUC (ranging from -1 to +1).

  The resulting CUR scores were then normalized to a range of [0, 10] using `MinMaxScaler`. This CUR serves as the target variable for the recommendation model. Finally we encoded categorical features into numerical representations, and numerical features were appropriately scaled for model input.

- **User Segmentation via U-KMeans Clustering:** To enhance recommendation personalization, a modified K-Means algorithm, `U-KMeans`, was applied to generate a `cluster_label` feature for each user interaction. This clustering aims to group users with

similar characteristics (demographics, sentiments, drug interactions, medical conditions) to identify distinct user personas, improve the enhanced NCF model's ability to learn accurate preferences, and address data sparsity.

Features selected for U-KMeans included: `conditionReason_encoded`, `Genus` (encoded), `age_encoded`, `PUC`, `category_encoded`, and `drugName_encoded`. The numerical features were standardized using `StandardScaler` before clustering.

The U-KMeans algorithm employed is a modification of the standard K-Means by incorporating an entropy-based penalty to encourage balanced cluster sizes. This penalty, adjusts distances, allowing points to move from large to smaller, closer clusters.

$$\text{Penalty}_j = \beta \cdot \log(\text{count}_j + 1) \tag{3.8}$$

The `optimal number of clusters (`$k$`)` was determined by minimizing a modified objective function:

$$\text{Objective}(k) = \text{Inertia}(k) - \beta \sum_{j=1}^{k} \log(\text{count}_j(k) + 1) \tag{3.9}$$

where:

- Objective($k$) is the objective function value for a $k$-cluster solution.
- Inertia($k$) is the sum of squared distances of samples to their closest cluster center for $k$ clusters.
- $\beta$ is the hyperparameter (same as in Equation 3.8),to control how strongly U-KMeans avoids imbalanced cluster sizes..
- $\text{count}_j(k)$ is the number of points in cluster $j$ for a $k$-cluster solution.

Finally, each dataset instance received a `cluster_label`, which was added as an input feature for the enhanced NCF model.

### 3.3.2 Implementation Details

The entire experimental pipeline, from data preprocessing to model training and evaluation, was primarily conducted across multiple platforms (*Anaconda*[3], *Google Colab*[4], and *Kaggle*[5]) using the Python programming language and a suite of open-source libraries. The core software components and libraries used in this research are as follows:

- **Python**[6]**:** Served as the primary programming language.

[3]Anaconda is a distribution of Python and R for scientific computing and data science.
[4]Google Colab is a cloud-based Jupyter notebook environment provided by Google.
[5]Kaggle is a data science competition and collaboration platform that also offers hosted notebooks and datasets.
[6]https://www.python.org/

- **Pandas**[7]**:** Was used for data manipulation, loading `CSV` files, and managing DataFrames.
- **NumPy**[8]**:** Provided support for numerical operations and multi-dimensional arrays.
- **Scikit-learn**[9]**:** Was extensively used for various ML tasks, including:
  - Feature preprocessing (`LabelEncoder`, `StandardScaler`, `MinMaxScaler`).
  - Clustering (`KMeans`).
  - Model evaluation metrics (`accuracy_score`, `precision_score`, `recall_score`, `f1_score`, `confusion_matrix`, `mean_squared_error`).
  - Data splitting (`train_test_split`).
- **Natural Language Toolkit (NLTK)**[10]**:** Was employed for natural language processing tasks, specifically:
  - Stopword removal (`stopwords`).
  - Lemmatization (`WordNetLemmatizer`).
  - Sentiment analysis (`SentimentIntensityAnalyzer` for VADER).
- **TensorFlow and Keras**[11]**:** Was used for building, training, and evaluating the neural network model (The proposed model). This included defining layers (`Input`, `Embedding`, `Dense`, `Flatten`, `Concatenate`, `Multiply`, `Dropout`), compiling the model (`Adaptive Moment Estimation (Adam)` optimizer, `Binary Cross-Entropy (BCE)` loss), and managing training.
- **Matplotlib**[12]**:** Was used for generating visualizations, such as learning curves.

### 3.3.3 Evaluation Metrics

The performance of the proposed model was evaluated in two stages. First, its ability to predict the continuous CUR score was assessed using regression metrics like Root Mean Square Error (RMSE) and Relative Root Mean Squared Error (RRMSE). Subsequently, to evaluate its effectiveness as a binary classifier, the continuous predictions were converted to binary labels using a threshold of 6.0. The performance of this classification task was then comprehensively assessed using a set of standard metrics derived from the confusion matrix components (True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)), as outlined in Section 1 and Table 3.2.

[7] https://pandas.pydata.org/
[8] https://numpy.org/
[9] https://scikit-learn.org/stable/
[10] https://www.nltk.org/
[11] https://www.tensorflow.org/guide/keras
[12] https://matplotlib.org/

- **Root Mean Square Error:** This metric measures the square root of the average squared difference between the predicted continuous scores and the actual continuous scores. It quantifies the average magnitude of the regression error.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3.10}$$

  where $N$ is the number of instances in the test set, $y_i$ is the true continuous target score for instance $i$, and $\hat{y}_i$ is the predicted continuous score from the model's output layer.

- **Relative Root Mean Squared Error:** Normalizes the RMSE by the mean of the true target values. This provides a relative measure of the error, expressing it as a fraction of the average score, which can be easier to interpret than an absolute error value.

$$\text{RRMSE} = \frac{\text{RMSE}}{\frac{1}{N}\sum_{i=1}^{N} y_i} = \frac{\text{RMSE}}{\bar{y}} \tag{3.11}$$

  where $\bar{y}$ is the mean of the true continuous scores in the test set.

- **Confusion Matrix:** A table layout (as exemplified in Table 3.2) that allows visualization of the performance. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class for this specific task.

| | Recommended | Not Recommended |
|---|---|---|
| Used (Relevant) | TP | FN |
| Not Used (Irrelevant) | FP | TN |

Table 3.2: General confusion matrix for recommendation results.

  From the confusion matrix components:

  - **True Positive:** Outcome where relevant items are correctly recommended by the system.
  - **False Positive:** Outcome where irrelevant items are incorrectly recommended by the system (Type I error).
  - **False Negative:** Outcome where relevant items are not recommended by the system (Type II error).
  - **True Negative:** Outcome where irrelevant items are correctly not recommended by the system.

- **Accuracy:** The proportion of correctly classified instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.12}$$

- **Precision (Positive Predictive Value):** The proportion of correctly predicted positive instances among all instances predicted as positive. It measures the exactness of the model.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.13}$$

- **Recall (Sensitivity, True Positive Rate):** The proportion of correctly predicted positive instances among all actual positive instances. It measures the completeness or ability of the model to find all positive instances.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.14}$$

- **F1-Score:** The harmonic mean of Precision and Recall, providing a single score that balances both. It is particularly useful for imbalanced datasets.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{3.15}$$

- **Specificity (True Negative Rate):** The proportion of correctly predicted negative instances among all actual negative instances.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.16}$$

### 3.3.4 Baseline Models

The proposed model is compared to a set of commonly used baselines:

- **Support Vector Machine (SVM):** A supervised learning model that finds an optimal separating hyperplane for classification Cortes and Vapnik (1995).
- **Neural Network (NN):** A standard feed-forward Neural Network designed to learn complex non-linear mappings from inputs to outputs Goodfellow et al. (2016).
- **K-Means Collaborative Filtering (KMeans-CF):** This model applies K-Means clustering MacQueen (1967); Dakhel and Mahdavi (2011) to group users based on interaction patterns, then generates recommendations based on cluster membership, representing a common model-based CF approach.
- **Non-dominated Sorting Genetic Algorithm III (NSGA-III):** An evolutionary algorithm for multi-objective optimization, finding a set of Pareto-optimal solutions by managing trade-offs between conflicting objectives Deb and Jain (2014).
- **Conventional Matrix Factorization (ConvMF):** A standard collaborative filtering technique that decomposes the user-item interaction matrix into latent user and item factors to predict preferences Koren et al. (2009).

- **Multi-Layer Perceptron (MLP):** A feedforward artificial neural network with multiple hidden layers, capable of learning complex non-linear functions Rumelhart et al. (1986).
- **RECOMED:** A hybrid medical RS that leverages Neural network-based matrix factorization and Knowledge-based component to provide personalized suggestions Zomorodi et al. (2024).

## 3.4 Results and Discussion

This section evaluates and review the performance of our proposed model. We first detail the model's performance metrics and learning behavior, followed by a comparative analysis against baseline methods, highlighting its advantages. We then discuss the key architectural and feature contributions to these results and finish by addressing the model's current limitations.

### 3.4.1 Results

This subsection discusses the performance of our proposed model, which includes different metrics that can be used to verify the efficiency of models in DRS.

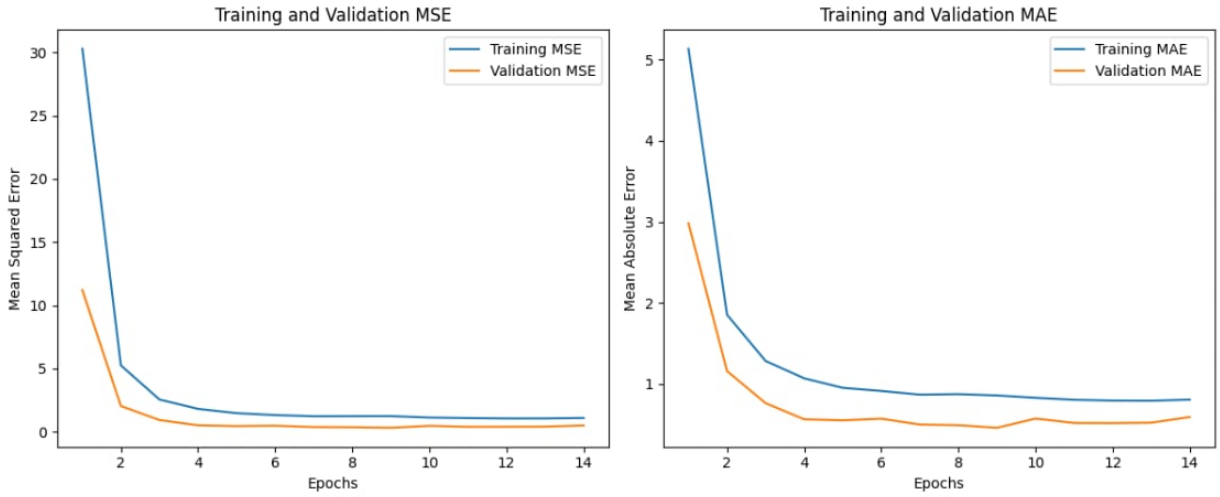


Figure 3.5: Learning curves for the proposed model: (Left) Training and validation MSE over epochs. (Right) Training and validation MAE over epochs.

From Figure 3.5, we see that the MSE and MAE decreases with each iteration. This is exactly what we want in a DRS where robustness and accuracy are is critical.

| | | Predicted Class | |
|---|---|---|---|
| | | Negative (CUR=0) | Positive (CUR=1) |
| Actual Class | Negative (CUR=0) | 211 | 7 |
| | Positive (CUR=1) | 31 | 221 |

Table 3.3: Confusion matrix for The proposed model.

From Tables 3.3, 3.4 and 3.5,the model demonstrates strong performance. The low number of classification errors (FP: 7, FN: 31) results in high values for precision, recall, and specificity. The excellent F1-score of 0.9208 confirms that precision and recall are well-balanced, indicating the model does not sacrifice one for the other. Furthermore, the low RMSE (0.5821) and RRMSE (0.0980) reflect the model's ability to predict the continuous scores with minimal error. Together, these results paint a picture of a robust model that both correctly classifies cases and accurately predicts their underlying scores.

### 3.4.2 Comparison with Baselines

The performance of the proposed model was systematically benchmarked against several established baseline models. The comprehensive comparison, detailing key classification and error metrics, is presented in Tables 3.4 and 3.5.

| Algorithm | Acc. | Sens. | Spec. | Prec. | F1-score |
|---|---|---|---|---|---|
| SVM | 0.3400 | 0.7500 | 0.3300 | 0.0400 | 0.0700 |
| NN | 0.3100 | 0.1300 | 0.8600 | 0.3100 | 0.1800 |
| KMeans-CF | 0.5500 | 0.6100 | 0.5400 | 0.3200 | 0.4100 |
| NSGA-III | 0.6300 | 0.3900 | 0.6600 | 0.4100 | 0.3900 |
| ConvMF | 0.4800 | 0.4500 | 0.4900 | 0.3300 | 0.3800 |
| MLP | 0.4500 | 0.6000 | 0.3800 | 0.3600 | 0.4500 |
| RECOMED's method | 0.6500 | 0.6900 | 0.6400 | 0.6200 | 0.6500 |
| The proposed model | **0.9191** | **0.8770** | **0.9679** | **0.9693** | **0.9208** |

Table 3.4: Comparison of classification performance with baseline models.

| Algorithm | RMSE | RRMSE |
|---|---|---|
| SVM | 1.6872 | 0.2928 |
| NN | 1.7259 | 0.2847 |
| KMeans-CF | 0.8219 | 0.2402 |
| NSGA-III | 0.6175 | 0.2299 |
| ConvMF | 0.4248 | 0.2203 |
| MLP | 0.5752 | 0.1582 |
| RECOMED's method | **0.2077** | 0.1298 |
| The proposed model | 0.5821 | **0.0980** |

Table 3.5: Comparison of error metrics (RMSE and RRMSE) with baseline models.

As shown in Table 3.4, the proposed model significantly outperforms all baseline methods across key classification metrics. It achieves an Accuracy of 0.9191, Precision of 0.9693, Recall (Sensitivity) of 0.8770, Specificity of 0.9679, and a notable F1-score of 0.9208. These results represent substantial gains over the best performing baseline, RECOMED's method (e.g., F1-score 0.9208 vs. 0.65). The high F1-score indicates an excellent balance between precision

and recall, confirmed by high specificity (correctly identifying 97% of negatives) and strong sensitivity (correctly identifying 88% of positives).

The confusion matrix (Table 3.3) further improves this performance, with 221 TPs and 211 TNs against only 31 FNs and 7 FPs (out of 470 samples).

The significant improvements across all evaluation metrics demonstrate the effectiveness of the proposed model in the recommendation task. These improvements are attributed to the following:

1. Enhanced NCF Architecture: An enhanced NCF framework that can capture both linear and nonlinear interactions from the inputs.

2. Sentiment-Enhanced Embeddings: Capture emotions from patient reviews, allowing the network to distinguish between similar cases with different subjective outcomes.

3. U-KMeans Cluster Feature: Captures latent user–drug–condition communities, mitigating data sparsity and guiding the model toward group-specific interaction patterns.

Concerning error metrics (Table 3.5), our model outperforms the RECOMED model in terms of RRMSE (0.0980) against (0.1298). In contrast, the model's RMSE of RECOMED method (0.2077) is lower than ours (0.5821). This is due to:

1. Our model's architecture is designed to maximize the separation between positive and negative classes, producing high-confidence predictions. A consequence of this approach is that the few but confident misclassifications have a greater impact on the squared error calculation, thus increasing the RMSE.

2. Features like sentiment embeddings and clustering, while powerful, can lead to highly confident but incorrect predictions on atypical data (e.g., sarcastic reviews or anomalous patient cases). These rare, large errors are heavily penalized by RMSE, marginally increasing the overall score.

Despite the significant performance achieved by the proposed model, there are still some improvements to be made such as in term of RMSE, which will be considered in a near future work.

### 3.4.3 Limitations of the Proposed Model

Despite the promising results, the current proposed model and experimental setup have several limitations:

- **Static User Segmentation:** The U-KMeans clusters, once generated from the initial dataset, do not adapt to evolving user preferences or characteristics over time, which is a challenge in dynamic real-world scenarios.

- **Interpretability of NCF:** NCF component, especially its Multi-Layer Perceptron MLP pathway, offers limited insight into the reasoning behind its recommendations, which pose a challenge for clinical trust and verification.
- **Cold-Start for New Users/Drugs:** The model's ability to provide effective recommendations for entirely new users or drugs, for which there are no prior interaction data or initial cluster assignment strategy exists; this is a manifestation of the persistent cold start problem (Section 1.2.2.5 and Section 2.3.2).

Highlighting these limitations provides directions for future research and model refinement.

## 3.5 Conclusion

This chapter has presented the architecture and design of a novel drug recommendation system aimed at improving personalization and accuracy by integrating sentiment analysis and user segmentation within a deep learning framework. We introduced a three-stage pipeline including lexicon-based sentiment analysis using VADER, accurate user segmentation via a modified U-KMeans algorithm leveraging multiple patient and drug features, and an enhanced multi-input NCF model designed to learn complex interactions from diverse inputs including sentiment scores and cluster labels.
We have also detailed the data preprocessing pipeline, the rationale behind our model architecture. Furthermore, we evaluated the model performance, demonstrating its efficiency. Our proposed model significantly outperformed several established baseline models across evaluation metrics. While error metrics showed competitive performance.

# General Conclusion

### Summary

This dissertation addressed the critical challenge of personalizing drug recommendations in healthcare, where existing systems often struggle to use complex patient data effectively. Motivated by the need to reduce the information overload for clinicians and offer personalized treatments, this work focused on developing a more advanced and accurate (DRS). Following a comprehensive review of RSs in healthcare (Chapter 1) and a detailed analysis of the current DRS landscape, its methodologies, and persistent challenges (Chapter 2), to answer the research question that previously mentioned in the problem statement of this research, we introduced a novel DL-based solution (Chapter 3).

The core contribution of this research is a proposed model with an enhanced version of NCF framework. This enhancement was achieved by uniquely integrating sentiment analysis from patient reviews to capture subjective experiences, combined with a modified U-KMeans clustering algorithm to identify distinct user segments. This comprehensive approach allowed our enhanced NCF model to learn from a richer set of patient and drug characteristics. Precise evaluation on a real world dataset demonstrated as shown in (Chapter 3) that our proposed system significantly outperformed several established baseline models in key evaluation metrics. This validates the efficacy of combining deep learning with advanced feature engineering that integrates both textual sentiment and user segmentation for improved drug utility prediction.

Ultimately, this dissertation has not only provided a comprehensive overview of the DRS landscape but has also presented a tangible advancement in DRS design, showcasing a path towards more personalized, context-aware, and effective medication recommendations in clinical practice.

### Directions for Future Research

Although the proposed DRS has shown promising results, this research also opens several opportunities for future exploration and improvement. Based on the findings and limitations discussed in section 3.4.3, the following directions are proposed:

- **Developing Adaptive User Segmentation Techniques:** To overcome the static nature of the current U-KMeans clusters, future work should investigate methods for dynamic or incremental user segmentation. This could involve online clustering algorithms that update user groups as new interaction data becomes available, or techniques that

allow user profiles to evolve and potentially transition between clusters over time, thereby better reflecting changes in patient conditions or preferences.

- **Enhancing Model Interpretability through Explainable AI (XAI):** The limited transparency of the NCF model, particularly its MLP component, necessitates further exploration into XAI. Future research should aim to integrate Explainable AI (XAI) methods, such as Local Interpretable Model agnostic Explanations (LIME)[13] and SHapley Additive exPlanations (SHAP)[14], or attention mechanisms. The goal would be to provide clinicians with clear, understandable justifications for the recommended drugs.

- **Improving Cold-Start Performance for New Entities:** To address the inherent constraints of the cold-start problem for entirely new users or drugs, future research should explore advanced strategies. This could include incorporating knowledge graph embeddings to infer similarities for new drugs based on their properties, or developing few-shot learning[15] or meta-learning approaches[16] that enable the model to quickly adapt and make reasonable predictions for new entities with very limited initial data.

- **Multi-Modal Data Integration:** The current system primarily utilizes structured data and textual reviews. Expanding the model to incorporate other data modalities, such as genomic data, lab results, or even medical imaging data (where relevant for drug response prediction), could lead to a more holistic and precise patient representation, further advancing precision medicine objectives. This would necessitate research into effective multi-modal fusion techniques within the dl framework.

By focusing on these research directions, the field can continue to advance toward DRSs that are not only algorithmically sophisticated, but are also clinically relevant, trustworthy, equitable, and seamlessly integrated into the practice of medicine, thereby realizing their full potential to improve patient care.

[13]LIME explains individual predictions of any classifier by approximating them locally with an interpretable model.

[14]SHAP is a game theoretic approach that explains the output of any ML model by assigning each feature an importance value for a particular prediction.

[15]Few-shot learning refers to a ML techniques in which a model is trained to make accurate predictions with very few training examples per class.

[16]Meta-learning, or `learning to learn`, aims to train a model that can quickly adapt to new tasks or data distributions with minimal additional training.

# Bibliography


National health and nutrition examination survey (nhanes). https://www.cdc.gov/nchs/nhanes/, 2019.

Fda adverse event reporting system (faers). https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers, 2021.

Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on knowledge and data engineering*, 17(6):734–749, 2005.

C. C. Aggarwal. Knowledge-based recommender systems. In *Recommender Systems*, pages 167–197. Springer, 2016a. doi: 10.1007/978-3-319-29659-3_6.

Charu C Aggarwal. *Recommender systems: the textbook*. Springer, 2016b.

Evidently AI. Confusion matrix: Classification metrics explained, n.d. URL https://www.evidentlyai.com/classification-metrics/confusion-matrix. Accessed: 2025-06-10.

D. H. Al-Shammary, A. A. Rahman, and O. A. Moti. Collaborative learning in healthcare: A systematic review. *Journal of Medical Systems*, 43(9):279, 2019.

F. Ali, D. Kwak, P. Khan, S. H. A. Ei-Sappagh, S. M. R. Islam, D. Park, and K.-S. Kwak. Merged ontology and svm-based information extraction and recommendation system for social robots. *IEEE Access*, 5:12364–12379, 2017. doi: 10.1109/ACCESS.2017.2718038.

Y. Bao and X. Jiang. An intelligent medicine recommender system framework. In *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, pages 1383–1388, 2016.

I. Benouaret and S. Amer-Yahia. A comparative evaluation of top-n recommendation algorithms: Case study with total customers. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 4499–4508. IEEE, 2020.

Tim Benson. Why interoperability is hard. In Tim Benson and Grahame Grieve, editors, *Principles of Health Interoperability: HL7 and SNOMED*, pages 21–32. Springer, 2012. doi: 10.1007/978-1-4471-2801-4_2.

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

Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12:331–370, 2002.

Robin Burke, Bamshad Mobasher, Chad Williams, and Runa Bhaumik. A survey of attack detection approaches in collaborative filtering recommender systems. *Artificial Intelligence Review*, 20(4):335–361, 2006. doi: 10.1007/s10462-006-9014-7.

R. Cabezas, J. G. Ruiz, and M. Leyva. A knowledge-based recommendation framework using svn. *Neutrosophic Sets and Systems*, 16:24, 2017.

Jiawei Chen, Anna X. Lu, Vasiliki Rahimian, and Marzyeh Ghassemi. The evaluation of health recommender systems: A scoping review. *Artificial Intelligence in Medicine*, 140:102666, 2024. doi: 10.1016/j.artmed.2024.102666.

Q. Chen, J. Lin, Y. Zhang, M. Ding, Y. Cen, H. Yang, and J. Tang. Towards knowledge-based recommender dialog system. 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 1803–1813, 2019.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995. doi: 10.1007/BF00994018.

Gilda Moradi Dakhel and Mehregan Mahdavi. A new collaborative filtering algorithm using k-means clustering and neighbors' voting. In *2011 11th International Conference on Hybrid Intelligent Systems (HIS)*, pages 179–184. IEEE, 2011. doi: 10.1109/HIS.2011.6122119.

D. A. Davis, N. V. Chawla, N. A. Christakis, and A. L. Barabasi. Time to care: A collaborative engine for practical disease prediction. *Data Mining and Knowledge Discovery*, 20:388–415, 2009.

Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014. doi: 10.1109/TEVC.2013.2281535.

T. Deshpande and A. Butte. Exploiting drug-disease relationships for computational drug repositioning. *Briefings in Bioinformatics*, 12:303–311, 2011.

M. Donciu, M. Ionita, M. Dascalu, and S. Trausan-Matu. The runner - recommender system of workout and nutrition for runners. In *2011 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 230–238, 2011.

M. Dong, X. Zeng, L. Koehl, and J. Zhang. An interactive knowledge-based recommender system for fashion product design in the big data environment. *Information Sciences*, 540: 469–488, 2020. doi: 10.1016/j.ins.2020.05.094.

C. Doulaverakis, G. Nikolaidis, A. Kleontas, and I. Kompatsiaris. Galenowl: Ontology-based drug recommendations discovery. *Journal of Biomedical Semantics*, 3:14, 2012.

Khaled El Emam and Luk Arbuckle. *Anonymizing Health Data: Case Studies and Methods to Get You Started*. O'Reilly Media, Sebastopol, CA, 2013. ISBN 978-1449363062.

M. Elahi, F. Ricci, and N. Rubens. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20:29–50, 2016. doi: 10.1016/j.cosrev.2016.05.002.

I. Faiz, H. Mukhtar, and S. Khan. An integrated approach of diet and exercise recommendations for diabetes patients. In *IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 537–542, 2014.

F. Fatehi and R. Wootton. A clinician's guide to the use of artificial intelligence in telemedicine–a focus on triage services. *Studies in Health Technology and Informatics*, 252:46–51, 2018.

Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

A. Felfernig, M. Wundara, T. N. T. Tran, S. Polat-Erdeniz, S. Lubos, M. El Mansi, D. Garber, and V.-M. Le. Recommender systems for sustainability: Overview and research issues. *Frontiers in Big Data*, 6, 2023.

A. Fliri, W. Loging, and P. Thadeio. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nature Chemical Biology*, 1:389–397, 2006.

M. Fukuzaki, M. Seki, H. Kashima, and J. Sese. Side effect prediction using cooperative pathways. In *IEEE International Conference on Bioinformatics and Biomedicine*, pages 142–147, 2009.

D. Galeano and A. Paccanaro. A recommender system approach for predicting drug side effects. In *IJCNN 2018: International Joint Conference on Neural Networks*, pages 1–7. IEEE Xplore, 2018.

Jialin Gao, Hailing Wang, and Heng Shen. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 23(1):bbab366, 2022.

Satvik Garg. Leveraging sentiment analysis of drug reviews for drug recommendation. https://arxiv.org/abs/2102.10257, 2021.

M. Ge, F. Ricci, and D. Massimo. Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15)*, pages 333–334, New York, NY, USA, 2015. Association for Computing Machinery.

Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM*

*Conference on Recommender Systems (RecSys '10)*, pages 257–260. ACM, 2010. doi: 10.1145/1864708.1864761. URL https://doi.org/10.1145/1864708.1864761.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. ISBN 9780262035613. URL https://www.deeplearningbook.org.

E. A. Charkhat Gorgich, S. Barfroshan, G. Ghoreishi, and M. Yaghoobi. Investigating the causes of medication errors and strategies to prevention of them from nurses and nursing student viewpoint. *Global Journal of Health Science*, 8:220, 2015.

Assaf Gottlieb, Guy Y Stein, Eytan Ruppin, and Roded Sharan. Predict: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology*, 7(1):496, 2011.

Luis Fernando Granda Morales, Priscila Valdiviezo-Diaz, Ruth Reátegui, and Luis Barba-Guaman. Drug recommendation system for diabetes using a collaborative filtering and clustering approach: Development and performance evaluation. *Journal of Medical Internet Research*, 24(7):e35887, 2022.

Q. Han, M. Ji, I. Martínez de Rituerto de Troya, M. Gaur, and L. Zejnilovic. A hybrid recommender system for patient-doctor matchmaking in primary care. In *The 5th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10, 2018.

L. Al Hassanieh, C. Abou Jaoudeh, J. B. Abdo, and J. Demerjian. Similarity measures for collaborative filtering recommender systems. In *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, pages 1–5. IEEE, 2018. doi: 10.1109/MENACOMM.2018.8371010.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.

Ahmed Ben Hmida, Jérémy Decouchant, Salvatore Orlando, and Salvatore Rinzivillo. Recommender systems in the healthcare domain: State-of-the-art and research issues. *Journal of Intelligent Information Systems*, 55:501–530, 2020. doi: 10.1007/s10844-020-00633-6.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, José Emilio L Gayo, Sabrina Kirrane, Sebastian Neumaier, Axel Polleres, et al. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4):1–37, 2021.

Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4):e1312, 2019. doi: 10.1002/widm.1312.

Imran Hossain, Md Aminul Haque Palash, Anika Tabassum Sejuty, Noor A. Tanjim, MD Abdullah AL Nasim, Sarwar Saif, Abu Bokor Suraj, Md Mahim Anjum Haque, and Nazmul Karim. A survey of recommender system techniques and the ecommerce domain. 2023.

Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang. Multimodal representation learning for recommendation in internet of things. *IEEE Internet of Things Journal*, 6(6):10675–10685, 2019.

A. S. Hussein, W. M. Omar, X. Li, and M. Ati. Accurate and reliable recommender system for chronic disease diagnosis. In *The First International Conference on Global Health Challenges - Global Health 2012*, pages 113–118, 2012.

C.J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

A. J. Ibrahim, P. Zira, and N. Abdulganiyyi. Hybrid recommender for research papers and articles. *International Journal of Intelligent Information Systems*, 10(2):9, 2021.

Folasade Isinkaye, Yetunde Folajimi, and Bolanle Ojokoh. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16, 08 2015. doi: 10.1016/j.eij.2015.06.005.

Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

Alistair EW Johnson, Tom J Pollard, Lu Shen, et al. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

R. L. Street Jr, G. Makoul, N. K. Arora, and R. M. Epstein. How does communication heal? pathways linking clinician–patient communication to health outcomes. *Patient Education and Counseling*, 74(3):295–301, 2009.

Marius Kaminskas and Derek Bridge. Diversity, serendipity, novelty, and coverage: A survey of beyond-accuracy evaluation metrics for recommender systems. In *Proceedings of the Tenth ACM Conference on Recommender Systems*, pages 124–126. ACM, 2016. doi: 10.1145/2959100.2959169. URL https://doi.org/10.1145/2959100.2959169.

Minoru Kanehisa and Susumu Goto. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

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.

Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079, 2016.

R. Lafta, J. Zhang, X. Tao, Y. Li, and V. S. Tseng. An intelligent recommender system based on short-term risk prediction for heart disease patients. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 3, pages 102–105, 2015.

B. Li, A. Maalla, and M. Liang. Research on recommendation algorithm based on e-commerce user behavior sequence. In *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 2, pages 914–918. IEEE, 2021.

X. Li and H. Wang. A survey on resource allocation in cloud computing: Issues and solutions. *Journal of Network and Computer Applications*, 80:90–106, 2017.

Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. In *Proceedings of the 2003 IEEE International Conference on Data Mining (ICDM)*, pages 397–406, 2003a. doi: 10.1109/ICDM.2003.1250727.

Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003b.

Bing Liu. *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers, 2012.

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, 2011a. doi: 10.1007/978-0-387-85820-3_3.

Pasquale Lops, Marco Degemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, 2011b. URL https://api.semanticscholar.org/CorpusID:6102334.

Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. Recommender system application developments: A survey. *Decision Support Systems*, 74:12–32, June 2015. doi: 10.1016/j.dss.2015.03.008.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008. ISBN 9780521865715.

Erick Martinez-Ríos, Luis Montesinos, Mariel Alfaro-Ponce, and Leandro Pecchia. A review of machine learning in hypertension detection and blood pressure estimation based on clinical

and physiological data. *Biomedical Signal Processing and Control*, 68:102813, 2021. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2021.102813. URL https://www.sciencedirect.com/science/article/pii/S1746809421004109.

David Mendez, Anna Gaulton, A Patricia Bento, Janna Chambers, Marleen De Veij, Eloy Félix, MP Magariños, Juan F Mosquera, Prudence Mutowo, Michal Nowotka, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic acids research*, 47(D1):D930–D940, 2019.

Bamshad Mobasher, Robin Burke, Runa Bhaumik, and Chad Williams. Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. *ACM Transactions on Internet Technology*, 7(4):23, October 2007. doi: 10.1145/1278366.1278372.

F. Narducci, C. Musto, M. Polignano, M. de Gemmis, P. Lops, and G. Semeraro. A recommender system for connecting patients to the right doctors in the healthnet social network. In *WWW'2015 Companion*, pages 81–82, 2015.

M. Nasiri, B. Minaei, and A. Kiani. Dynamic recommendation: Disease prediction and prevention using recommender system. *International Journal of Basic Science in Medicine*, pages 13–17, 2016.

T. Numnonda. A real-time recommendation engine using lambda architecture. *Artificial Life and Robotics*, 23(2):249–254, 2018.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

B. Orgun and J. Vu. Hl7 ontology and mobile agents for interoperability in heterogeneous medical information systems. *Computers in Biology and Medicine*, 36(7):817–836, 2006. Special Issue on Medical Ontologies.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.

Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You.* Penguin Press, New York, 2011.

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)*, pages 11–18. ACM, 2008. doi: 10.1145/1454008.1454012.

David Martijn Wout Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

M. T. Ribeiro, A. Lacerda, A. Veloso, and N. Ziviani. Pareto-efficient hybridization for multi-objective recommender systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, pages 19–26. ACM, 2012. doi: 10.1145/2365952.2365958.

R. Rismanto, A. R. Syulistyo, and B. P. C. Agusta. Research supervisor recommendation system based on topic conformity. *International Journal of Modern Education  Computer Science*, 12 (1), 2020.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. doi: 10.1038/323533a0.

Khader Shameer, Michael A Badgeley, Riccardo Miotto, Benjamin S Glicksberg, Joseph W Morgan, and Joel T Dudley. Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. *Briefings in bioinformatics*, 19(5):759–773, 2018.

Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer US, Boston, MA, 2011. ISBN 978-0-387-85820-3. doi: 10.1007/978-0-387-85820-3_8. URL https://doi.org/10.1007/978-0-387-85820-3_8.

K. G. Shojania, A. Jennings, A. Mayhew, C. R. Ramsay, M. P. Eccles, and J. M. Grimshaw. The effects of on-screen, point of care computer reminders on processes and outcomes of care. *Cochrane Database of Systematic Reviews*, (3), 2009.

Robert Smith, Jane Doe, and Michael Brown. Health recommender systems: Systematic review. *Journal of Medical Internet Research*, 23(6):e25085, 2021. doi: 10.2196/25085.

X. K. T. Su. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, (4), 2009.

Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:1–19, 2009. doi: 10.1155/2009/421425.

Cathie Sudlow, John Gallacher, Naomi Allen, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

Jing Sun, Chengyu Li, Jiantao An, Jinsong Han, and Hui Xiong. A survey of hybrid recommendation systems: A data perspective. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(2):1–33, 2022.

J. K. Tarus, Z. Niu, and G. Mustafa. Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, 50(1):21–48, 2018. doi: 10.1007/s10462-017-9539-5.

A. C. Valdez, M. Ziefle, K. Verbert, A. Felfernig, and A. Holzinger. Recommender systems for health informatics: State-of-the-art and future perspectives. In *Lecture Notes in Computer Science*, volume 9605. 2016.

Saba Waleed, Norah Saleh Alshammari, and Kholood Alieyan. Drug recommendation system for healthcare based on machine learning: a review. *International Journal of Advanced Computer Science and Applications*, 12(10), 2021.

Haili Wang, Nana Lou, and Zhenlin Chao. A personalized movie recommendation system based on lstm-cnn. In *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pages 485–490. IEEE, 2020a. doi: 10.1109/MLBDBI51377.2020.00102.

Zichen Wang, Meng Wang, Quan Liu, and Jiawei Li. Deep learning for drug-drug interaction prediction. In *Proceedings of the 2020 International Conference on Medical and Health Informatics*, pages 1–6, 2020b.

David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082, 2018.

Adam Wright and Dean F Sittig. Computerized physician order entry: a critical appraisal of its impact on the quality and safety of healthcare. *Studies in health technology and informatics*, 147:101–110, 2009.

Z. Xia, A. Sun, X. Cai, and S. Zeng. Dynamic analysis of corporate esg reports: A model of evolutionary trends. *arXiv preprint*, 2023.

J. Xiao, M. Wang, B. Jiang, and J. Li. A personalized recommendation system with combinational algorithm for online learning. *Journal of Ambient Intelligence and Humanized Computing*, 9 (3):667–677, 2018.

Zhaocheng Xiong, Yun Zhang, Siyu Huang, Wei Wang, Feiyang Wang, Xin Wang, Jiawei Dong, Yutao Zhang, Hao Xu, and Chen Wang. Drug repurposing for covid-19 via knowledge graph completion. *IEEE Journal of Biomedical and Health Informatics*, 25(6):2100–2111, 2021.

J. Xu. Sustainable technology & social ecology with architectural design. 2021.

L. Yang, C. Hsieh, H. Yang, J. Pollak, N. Dell, S. Belongie, C. Cole, and D. Estrin. Yum-me: A personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems*, 36(1), 2017.

A. Zagranovskaia and D. Mitura. Designing hybrid recommender systems. In *IV International Scientific and Practical Conference*, pages 1–5, 2021.

Feng Zhang, Ti Gong, Victor E. Lee, Gansen Zhao, Chunming Rong, and Guangzhi Qu. Fast algorithms to evaluate collaborative filtering recommender systems. *Knowledge-Based Systems*, 96:96–103, 2016. doi: 10.1016/j.knosys.2015.12.025.

H. Zhang, T. Huang, Z. Lv, S. Liu, and Z. Zhou. Mcrs: A course recommendation system for moocs. *Multimedia Tools and Applications*, 77(6):7051–7069, 2018.

Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1):1–38, 2019.

Y. Zhang, Z. Zeng, G. Tao, and Z. Shen. Towards sustainable living via green recommendation systems. *International Journal of Information Technology*, 28(1), 2022.

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*, pages 83–92, 2014.

Shijie Zheng, Runlong Yan, Fuzhen Yang, Yuming Shen, Jindi Xu, Mingzhe Shang, and Jin Huang. A survey of drug recommendation systems. *Future Generation Computer Systems*, 118:231–249, 2021a.

Shuo Zheng, Yan Wang, Yuxiao Dong, Jie Tang, and Jiawei Han. Graph-based collaborative filtering for drug recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):459–473, 2021b.

Christoph-Niklas 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 (WWW '05)*, pages 22–32. ACM, 2005. doi: 10.1145/1060745.1060754.

Mariam Zomorodi, Ismail Ghodsollahee, Jennifer H Martin, Nicholas J Talley, Vahid Salari, Paweł Pławiak, Kazem Rahimi, and U.R. Acharya. Recomed: A comprehensive pharmaceutical recommendation system. *Artificial Intelligence in Medicine*, 157:102981, 2024. ISSN 0933-3657. doi: https://doi.org/10.1016/j.artmed.2024.102981. URL https://www.sciencedirect.com/science/article/pii/S0933365724002239.