UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in International Economics and Commerce

RECOMMENDER SYSTEMS AND BUSINESS APPLICATIONS

SISTEMI DI RACCOMANDAZIONE E APPLICAZIONI PER I BUSINESS

Relatore: Chiar.ma
Prof. Claudia Pigini

Tesi di Laurea di:
Marie Juliette Lizzardi

Anno Accademico 2022 – 2023

Abstract

In un’era definita dalla quantità illimitata e dal sovraccarico di informazioni, prodotti e scelte, gli individui navigano in un paesaggio labirintico in cui la ricerca dell’articolo, del servizio o del contenuto ideale diventa un compito sempre più sfuggente. L’era digitale ha portato ad una notevole proliferazione di opzioni, offrendo ai consumatori una vasta gamma di alternative in ogni aspetto della quotidianità, dalle piattaforme di e-commerce e servizi di streaming alle notizie e ai post sui social media. Con l’espansione esponenziale del regno digitale, la necessità di orientamento e personalizzazione nei consigli sui contenuti e sui prodotti è diventata fondamentale. Questo compito monumentale di smistare il vasto mare di possibilità, adattato alle preferenze e ai bisogni individuali, è il luogo in cui i Recommender Systems emergono come una pietra miliare del nostro moderno ecosistema informativo.

I Recommender Systems, o RS, rappresentano un'innovazione tecnologica fondamentale che affronta il problema del sovraccarico di informazioni. Si tratta di applicazioni software o algoritmi specializzati progettati per analizzare il comportamento e le preferenze degli utenti, consentendo loro di fornire consigli personalizzati, semplificando così il complesso processo di scelta. Questi sistemi, spesso lavorando dietro le quinte, facilitano la scoperta di prodotti, servizi o contenuti in linea con i gusti, gli interessi e le esigenze di un individuo.

Lo scopo di questa tesi è quello di immergersi nel panorama dei Recommender Systems, esplorandone l'evoluzione, i meccanismi, gli algoritmi e, soprattutto, le loro enormi implicazioni per i business. Per affrontare questa indagine in modo efficace, la discussione è organizzata in cinque capitoli, ciascuno dedicato ad un aspetto specifico.

I primi tre capitoli presentano la storia, l'"anatomia" e le varie caratteristiche di questi meccanismi di raccomandazione. Vi è una generale comprensione dello stato dell'arte, degli algoritmi, nonché delle difficoltà (anche su un piano etico) che devono essere affrontate.

Il quarto e quinto capitolo spostano l'attenzione sul mondo del business. Viene fatta luce sulle applicazioni più comuni e improntanti dei RS in ambiti quali l'e-commerce, il sistema sanitario e il mercato del lavoro, assieme ad un'applicazione pratica per evidenziarne il funzionamento; per poi passare ad un'analisi del valore che questi sistemi generano per le aziende. Approfondiremo i vantaggi tangibili e immateriali che vengono conferiti alle imprese, evidenziando il loro ruolo nella generazione di entrate, nella fidelizzazione dei clienti e nel vantaggio competitivo. Scopriamo come i RS migliorano l'esperienza degli utenti, stimolano il coinvolgimento e catalizzano la crescita del business. Questo ultimo capitolo funge da testimonianza del potenziale di trasformazione dei RS nel plasmare il panorama imprenditoriale contemporaneo, soprattutto nel mondo altamente globalizzato e capitalistico in cui viviamo.

Nelle conclusioni sintetizziamo i risultati e gli approfondimenti dei capitoli precedenti, offrendo una prospettiva olistica sull'impatto generale dei Recommender Systems. Inoltre, esploriamo brevemente le strade per la ricerca e lo sviluppo futuri in questo campo dinamico, evidenziando il potenziale non sfruttato e le tendenze emergenti che promettono di rimodellare i RS negli anni a venire.

Index

Introduction **5**

CHAPTER 1: Recommender Systems: history and development **9**

CHAPTER 2: Recommender Systems: definition and structure **16**

2.1. Recommender Classification 20

2.1.1. Collaborative Filtering 21

2.1.2. Content-based systems 23

2.1.3. Hybrid systems 26

2.2. Minor Recommender Systems 28

2.3. Recommender Systems' challenges 35

CHAPTER 3: Recommender Systems' algorithms **47**

3.1. Collaborative Filtering algorithm 48

3.1.1. Neighborhood-based Collaborative Filtering 49

3.1.2 Model-based Collaborative Filtering 52

3.2. Content-based Recommendation algorithm 58

3.3. Hybrid Approaches 66

CHAPTER 4: Recommender Systems' applications for businesses **70**

4.1. Streaming services 73

4.2. Social Network Service 77

4.3. Tourism Service 79

4.4. Healthcare Service 82

4.5. Education Service 86

4.6. E-Commerce Service 91

4.7. Job market 96
4.7.1. Job Recommender System: the application 101

CHAPTER 5: Recommender Systems and their business value..................107

Conclusions and Future Research...118

Bibliography...124

INTRODUCTION

In an era defined by the limitless amount and overload of information, products, and choices, individuals navigate a labyrinthine landscape where the quest for the ideal item, service, or content becomes an increasingly elusive task. The digital age has resulted in a remarkable proliferation of options, presenting consumers with an overwhelming array of alternatives in virtually every facet of life, from e-commerce platforms and streaming services to news articles and social media posts. As the digital realm expands exponentially, the need for guidance and personalization in content and product recommendations has become critical. This monumental task of sorting through the vast sea of possibilities, tailored to individual preferences and needs, is where Recommender Systems (RS) emerge as an indispensable cornerstone of our modern information ecosystem.

Recommender Systems, or RS, represent a pivotal technological innovation that addresses the information overload problematic. They are specialized software applications or algorithms designed to analyze user behavior and preferences, enabling them to provide personalized recommendations, thereby simplifying the complex process of choice. These systems, often working behind the scenes, facilitate the discovery of products, services, or content that align with an individual's tastes, interests, and needs.

The scope of this thesis is to dive into the multifaceted landscape of Recommender Systems, exploring their evolution, mechanics, algorithms, and, most crucially, their huge implications in the realm of business and their value. To navigate this investigation effectively, the discussion is organized into five chapters, each dedicated to a distinct facet of RS.

The first chapter explores the historical backdrop and the foundational principles of Recommender Systems. Here, we unravel the historical origins and turning points that have paved the way for the development of recommendation engines as we know them today and of the state-of-the-art algorithms. Understanding the historical context is essential to grasp the trajectory of RS growth and their important role in modern society.

The second chapter provides an analysis of RSs' "anatomy", offering an in-depth exploration of their definition, categories, and underlying mechanisms. All the major and minor types of recommendation systems are presented, as well as the advantages and disadvantages of each of them. Lastly, this chapter is dedicated to a digression on the ethical problems related to RS. This is a critical matter, since privacy and huge volumes of users' data are now managed by the so-called big tech on a daily basis, but still nowadays there is a lack of regulatory norms to protect consumers, but also to guarantee fair competition within the market.

Then we proceed on a comprehensive exploration of RS algorithms in chapter three. This chapter examines the intricate algorithms powering RS, shedding light on their different and mostly used strategies for filtering data, such as collaborative filtering and content-based systems.

The fourth chapter is a juncture with which we shift our focus to the pragmatic applications of Recommender Systems in the world of business. Here, we delve into the real-world scenarios where RS are deployed, ranging from e-commerce platforms and online streaming services to the application within the healthcare system and the utilization of RS for diagnosticate and treat a disease, for example. In the last part of this chapter, I present the recommender application within the job market; after some theoretical considerations I present a rudimental empirical application with Rstudio with a content-based approach for matching the right candidate to the right job offer.

The fifth chapter is devoted to investigating the business value of RS. We delve into the tangible and intangible benefits that RS confer upon enterprises, highlighting their role in revenue generation, customer retention, and competitive advantage. We unravel how RS augment user experiences, drive engagement, and catalyze business growth. This chapter serves as a testament to the transformative potential of RS in shaping the contemporary business landscape, especially in the highly globalized and capitalistic world we live in.

In the conclusions, we synthesize our findings and insights from the preceding chapters, offering a holistic perspective on the overarching impact of Recommender Systems. Moreover, we explore briefly avenues for future research and development in this dynamic field, highlighting the untapped potential and emerging trends that promise to reshape RS in the years to come.

I. RECOMMENDER SYSTEMS: HISTORY AND DEVELOPMENT

Although the idea of Recommender Systems may be new in the academic world, it has been prevailing in society since ages. It is known that humans have evolved certain characteristics such as complex thought, language use, tool making, etc. over the past 100,000 years. The concept of recommendations can be seen in the case of cavemen, ants and other creatures too. For instance, ants have genetically evolved to leave markers for other ants which further serve as a recommender to other ants, showing them the way to food.

Recommendations in ancient society might have included things like what crop to grow and when to do it, what religion to practice, etc. Later, during the time of colonization (between the 11th and 18th centuries), the advice was transformed into what region to annex in accordance with the fruitful aspects like land fertility, manpower (for slavery) and numerous other resources.

Also, when families used to arrange marriages, usually someone from the family itself used to come up with some match. Again, those were recommendations. Similarly, people asking for opinions from others as to where to purchase the best clothes from or where to go on vacation, the quickest route to get somewhere etc. are all recommendations which have been in the society for a long time. There was no internet back then, but still people used to rely on recommendations from their peers.

With the Industrial Revolution, the worldwide market was revolutionized with the advent of computers. Because there were so many options available to people, it was frequently difficult for them to determine which product could best satisfy their desires and needs. Therefore, the need for a system that could simplify this selection criteria and eliminate the masses' dilemma was recognized, and eventually the Recommender Systems of the modern world were introduced.

Recommender Systems (RS) are therefore software tools that help us narrow down our options while providing us with the best suggestions based on our requirements.

Modern RS has its roots in the early 1990s, when it was primarily used experimentally for information filtering and personal email.

In 1992, the Palo Alto Research Center (PARC) introduced the expression "Collaborative filtering" and demonstrated how users could create custom filters using both explicit annotation and implicit behavioral data that had been gathered into a queryable database. The first recommender system that ever came into existence was Tapestry. The motivation that led to its development was the growing number of incoming emails, most of which were pointless, annoying, and challenging to manage. So users had to create their mailing lists and just like how email accounts operate today, only those on the contact list could send emails to the user or those they might want to hear from, while everyone else was sent to the spam list.

In 1994 GroupLens system showed that the collaborative filtering approach could be both distributed across a network and automated; and they used it to filter Usenet new messages. With GroupLens, one of the first systems was proposed, which employed machine learning to predict whether a user would like particular unseen messages based on explicit ratings given by community users.

The World Wide Web expanded quickly in the 1990s, opening up a growing number of potential application areas for Recommender Systems.

All of these systems made use of similar automation techniques: algorithms that identified other users with similar tastes and combined their ratings together into a personalized, weighted average. This simple “k-nearest neighbor” algorithm proved so effective that it rapidly established itself as the benchmark by which collaborative filtering algorithms were measured.

Recommender Systems' viability and effectiveness led to a great deal of enthusiasm for the field's advancement in both academic study and practical application.

One of the key events took place in 1996: the collaborative filtering workshop at Berkeley. This event brought together individuals working on personalized and non-personalized systems, on various algorithmic techniques (from statistical summaries to k-nearest neighbor to Bayesian clustering), and on various domains, which helped the community coalesce together.

Recommender Systems were developed in an Internet business environment that was quickly expanding, and commercialization began almost immediately. Many businesses were formed. They had to demonstrate their capacity to offer insightful advice and valuable recommendations.

New algorithms were developed to reduce online computation time, such as dimensionality-reduction techniques and item-based correlation algorithms, both of which are still in use today. In fact, the early GroupLens system relied on a comparably simple nearest-neighbor approach. Since then, however, all sorts of machine learning methods were applied or tailored to the problem setting.

The evaluation of recommenders using "top-N" recommendation list metrics piqued researchers' attention.

Even before the decade ended, a number of success stories regarding the use of Recommender Systems in E-commerce were reported for example with Amazon.com being one of the first adopters of recommendation technology at large scale.

In the first years of the 21st century, research in Recommender Systems exploded with a flood of people and approaches from many disciplines. New analyses and methodologies have evolved from various fields, including artificial intelligence, information retrieval, data mining, security and privacy, and business and marketing research.

In 2006 Netflix announced a 1 million dollars prize for an improvement in prediction accuracy of their algorithm. The Netflix Prize's excitement inspired numerous academics to collaborate in determined efforts to increase prediction accuracy. Netflix challenged academics and industry professionals to create Recommender Systems that could surpass the precision of the business's existing one, Cinematch, by releasing a data set comprising roughly 100 million anonymous movie ratings. Despite only making up a tiny portion of the company's rating data, the released dataset quickly rose to prominence in the data mining and machine learning communities due to its size and high quality. The dataset included dates along with ratings on an integer scale from 1 to 5. Title and year of release information were given for each film and users' personal information was not disclosed. More than 2817,131 unknown ratings formed the dataset, and submitted predictions were assessed based on their root mean squared error (RMSE) on this data set. 2000 teams out of the 20,000 registered teams submitted at least one set of answers. On September 21, 2009, a team that outperformed the Cinematch's accuracy by 10% won the top prize of $100,000.

Recommender Systems quickly gained popularity and are now crucial to companies like Amazon.com, Pandora, Netflix, YouTube, Tripadvisor, and others. For a variety of industries, including entertainment, social networking, e-learning, book or article suggestions, e-filtering, matching, e-commerce, and tourism, researchers have developed Recommender Systems.

Personalized recommendations are commonplace today, 30 years later, and research in this very successful AI application area is booming more than ever. Machine learning technology advancements have been a major driving force behind research over the past few decades. A sophisticated general-purpose algorithm alone won't be enough to create an effective Recommender System, though. It necessitates a thorough comprehension of the particulars of the application environment as well as the anticipated consequences the system will have on its users. Making suggestions is ultimately a problem of human-computer interaction, where a computerized system assists users in situations involving information search or decision-making.

The Recommender System is currently one of the most popular web programs, providing daily content recommendations to billions of users including news feeds, videos, e-commerce items, music, movies, books, friends, jobs, etc. These triumphant tales have demonstrated the ability of Recommender Systems to convert large amounts of data into high values.

II. RECOMMENDER SYSTEMS: DEFINITION AND STRUCTURE

There are many definitions of RSs. These can be defined as:

1. "Tools to mine items and/or collect users' opinions to help users in their search process and suggest items related to their preferences". (Hdioud et al., 2013)
2. "A program or software for content filtering that attempts to reduce the information overload problem, where users encounter a flood of data on the Web, by recommending personalized items to users depending on the items' information and/or users' preferences". (D'Addio & Manzato, 2015)
3. "A system to manage information overload problems by collecting information, guiding users in a personalized way, and providing individualized recommendations as output when there are many possible alternatives to choose from". (Chen et al., 2015)

Recommender Systems typically comprises three components: user/customer data (i.e., personal data such as interest, purchase history, ratings, and reviews), item data (i.e., item specifications and features), and filtering techniques that use the two aforementioned data to filter the items and recommend those that are in line with the user interests.

Many common operational and technical goals of Recommender Systems can be identified.

1. **Relevance**: A Recommender System's main operational goal is to make recommendations that are pertinent to the current user. Users are more likely to consume items they find interesting. Even though relevance is the primary operational goal of a Recommender System, it is not sufficient on its own.

2. **Novelty**: Recommender Systems work best when the item being recommended is something that the user has never seen before. In fact, repeated recommendation of popular items can also lead to a reduction in sales diversity.

3. **Serendipity**: Serendipity is a similar idea in which the items suggested are a little unexpected, so there is a small amount of fortuitous discovery, as opposed to obvious suggestions. Serendipity differs from novelty since the users are actually surprised by the recommendations, rather than simply learning about certain products for the first time. It may often be the case that a particular user may only be consuming items of a specific type, although a latent interest in items of other types may exist which the user might themselves find surprising. Serendipitous strategies, in contrast to novelty, concentrate on locating such recommendations. For example, if a new Chinese restaurant opens in a neighborhood, then the recommendation of that restaurant to a user who normally eats Chinese food is novel but not necessarily serendipitous. On the other hand, when the same user is recommended Indian food, and it was unknown to the user that such food might

appeal to him, then the recommendation is serendipitous. Serendipity has the positive side effect of increasing the variety of sales or beginning a new consumer interest trend. Increasing serendipity often has long-term and strategic benefits to the seller because of the possibility of discovering entirely new areas of interest. On the other hand, algorithms that provide serendipitous recommendations often tend to suggest irrelevant items. In many cases, the longer term and strategic benefits of serendipitous methods outweigh these short-term drawbacks.

4. **Increasing recommendation diversity**: Recommender Systems normally suggest a list of top-k items. The likelihood that the user won't like any of the suggested items rises when said items are all very similar. On the other hand, when the recommended list comprises items of several sorts, there is a higher possibility that the user might like at least one of these items. Diversity has the advantage of preventing the user from getting bored from frequent recommendations of the same products.

Aside from these concrete goals, a number of soft goals are also met by the recommendation process both from the user's and merchant's perspective. From the perspective of the user, recommendations can help improve overall user satisfaction with the website. From the merchant point of view, the recommendation process can provide insights into the customer demand and further customize the user's experience. Finally, providing the user an explanation for why a particular item is recommended is often useful. For instance, in the case of Netflix recommendations are provided along with previously watched movies.

Therefore, the main objective of a Recommender System is to provide users with lists of customized "recommended" objects. In order to do this, evaluations can be predicted or, as an alternative, recommendation scores can be assigned to items that a specific user is not yet familiar with. The suggestion list that the target user receives is made up of the items with the highest expected ratings or highest recommendation scores. The generated recommendation list can be assessed using a wide range of performance indicators.
A Recommender System understands human behavior and suggests content based on that. It uses sophisticated machine learning algorithms to give consumers the most accurate recommendations.

We can distinguish four different phases in the making process of a recommendation:

- **Data Collection**. The first step of a Recommender System starts with data collection. It collects implicit and explicit data from its users.

  Explicit data is data provided directly by users (such as ratings, comments, preferred genre, etc.), while implicit data is created by the Recommender

System itself with data on browsing history, clicks, watchlist events, etc. of each user.

- **Data Storage**. After data collection, a Recommender System stores the data. Storage space should always be well managed where all data can be stored securely.
- **Data analysis**: After collecting the explicit and implicit data and storing the same in the database, the recommender system starts analyzing the stored data. Data analysis is an important step towards accurate predictions as similar data is filtered leading to more precise recommendations.
- **Data Filtering**. This is the final and fundamental step, where the necessary data will be filtered from the entire database and only the relevant ones will be shown to the users. A Recommendation System basically uses three different methods to filter data and predict accurate recommendations: content-based filtering, collaborative filtering, and hybrid filtering.

2.1. RECOMMENDER CLASSIFICATION

The usual classifications of Recommender Systems comprehend three main categories, as we just saw: collaborative filtering, content-based systems and hybrid systems. Other less-used categories can be identified too.

2.1.1. Collaborative filtering

We can refer to collaborative filtering as a "people-to people correlation". The basic concept of collaborative filtering is that two or more users or individuals sharing some similar interests in one area tend to gravitate towards similar goods or services from other areas as well. The similarity between these users can be determined by their browsing behavior (click-through rate), browsing pattern and ratings (explicit, implicit).

The concept of collaborative filtering can be easily grasped with the aid of a basic example: consider Facebook. On your homepage you can always see a list of "people you may know". The basic criteria behind making these suggestions are based on the concept of Recommender Systems only. The suggestions are filtered out on multiple parameters such as number of mutual friends you have with that user, number of similar pages you both have liked or common groups or places. The approach used here is the collaborative filtering one: if a user and you have a number of friends in common, then it is likely that you two may also know each other. Therefore, it is called people to people co-relation.

So, collaborative filtering models use the collaborative power of the ratings provided by multiple users to make recommendations. The main challenge in designing collaborative filtering methods is that the underlying ratings matrices are *sparse*. In fact, there is frequently a very small (or no) overlap between two users since there is frequently an excessively large pool of available items. Also, even if we have a high average number of evaluations for each user or item, they are distributed among these users and items unevenly (frequently with a power-law distribution or a Weibull distribution) and so many items or users may have expressed or received only few ratings. Therefore, an effective and accurate recommender algorithm must take into account the concept of data sparsity.

Most collaborative filtering models either use inter-item correlations or inter-user correlations for the prediction process. Some models use both kinds of correlations. Some models also use carefully designed optimization techniques to generate a training model, such as a classifier does with the labeled data. The model is then used to approximate the missing values in the matrix, much like a classifier approximates the missing test labels. In the algorithms presentation and analysis, we will examine in greater detail two categories of methods, model-based methods and memory-based methods, which are frequently employed in collaborative filtering:

a. In the *memory-based collaborative filtering* objects that are recommended are those that users with similar preferences to the "target user" have chosen, or those that are similar to other items the target user has chosen.

b. With *model-based collaborative filtering* the recommended objects are the ones selected on models that are trained to identify patterns in the input data.

2.1.2. Content-based systems

Content-based systems are based on the principle of: “show me more of what I have enjoyed in the past”. Such systems' fundamental premise is to suggest goods or services to a specific user that are comparable to those the person has previously enjoyed. Based on their related features, two or more things' similarities can be estimated. To understand better, continuing with the example discussed before, when you watch any video on Facebook like the ones posted on a sustainability page, after you finish watching it, links of similar videos are shown on your home page. Additionally, when you put a like on some pages you will get suggestions of similar pages. So, the true driving force behind all of this is content-based filtering, or the idea that if you like an item from a certain category, *xyz*, you could also like an item from another category, *abc*, that is related to it in some way.

The descriptive properties of objects are utilized in content-based Recommender Systems to generate recommendations. These descriptions are referred to as "content". In content-based methods, the available content information is combined with user ratings and purchasing trends.

For example, consider that the user Lucy has rated the movie “A beautiful mind” highly, but we do not have access to other users’ ratings. Therefore, collaborative filtering methods cannot be used here and are consequently ruled out. However, the item description of “A beautiful mind” contains similar genre keywords as other dramatic movies, such as “Good Will Hunting” or “The Imitation game”. Here, these movies can be recommended to Lucy. In the content-based methods, the item descriptions, which are labeled with ratings, are used as training data to create a user-specific classification or regression modeling problem.

The training materials for each user match the descriptions of the goods they have purchased or rated. The ratings or purchasing patterns specified correspond to the class (or dependent) variable. A classification or regression model that is unique to the user at hand (or active user) is made using these training materials. This user-specific model is used to forecast an individual's liking of a product for which her rating or purchasing tendencies are unknown.

When there are insufficient rating data for an item, content-based approaches offer some advantages when providing suggestions for new items. This is due to the possibility that the active user has already rated other items with comparable characteristics.

Therefore, even when there is no rating history for that item, the supervised model will be able to use these ratings along with the item attributes to generate recommendations.

There are a number of drawbacks to content-based techniques as well:

1. Because of the usage of keywords or content, content-based approaches frequently offer obvious recommendations. For instance, an item with a specific combination of keywords has no possibility of being recommended if a user has never consumed it. This is due to the built-in user-specificity of the model and the lack of exploitation of the collective wisdom of similar users. This behavior has the unintended effect of making the recommended items less diverse.

2. Even though content-based methods are effective at providing recommendations for new *items*, they are not effective at providing recommendations for new *users*. This is because the target user's training model must take into account her rating history. In fact, in order to produce reliable forecasts without overfitting, it is typically crucial to have a significant number of ratings accessible for the target user.

Therefore, compared to collaborative filtering systems, content-based methods have different trade-offs.

A broader view of these methods is occasionally employed, even if the aforementioned description offers the traditional learning-based view of content-based methods. Users can, for instance, include pertinent keywords in their own profiles. In order to generate suggestions, these profiles can be matched with item descriptions. Such a method is helpful in cold-start conditions because it does not use ratings in the recommendation process.

However, because the similarity metrics are frequently based on domain knowledge, such methods are frequently viewed as a distinct class of Recommender Systems, known as **knowledge-based systems**. It is sometimes questioned if there is a clear distinction between the two groups of methodologies because knowledge-based Recommender Systems are sometimes thought to be closely related to content-based ones.

2.1.3. Hybrid systems

As the name implies, hybrid systems are combinational systems. The objective is to mix the characteristics of two systems (recommended techniques) so that the weaknesses of one are made up for by the other. It also offers the best of both worlds, you could say. Take Netflix as an example; it combines collaborative filtering and content-based filtering, suggesting movies to the user based on both his preferences and his similarities to other users.

So, if a person likes movies such as *The Notebook, PS I Love You*, etc., he can receive content-based recommendations for romantic movies the next time he visits the website (content-based filtering).

Additionally, if user X and user Y have a lot of shared favorite movies, then the subsequent film that either of them enjoys will be recommended to the other (collaborative filtering).

The systems mentioned above use many sources of input, and depending on the situation, they might function better. For instance, content-based approaches rely on textual descriptions and the target user's own ratings, whereas knowledge-based systems depend on interactions with the user in the context of knowledge bases. Collaborative filtering systems also rely on community ratings.

Similarly, as we will see later, demographic systems base their suggestions on the user's demographic profiles. It is important to note that these various systems employ different inputs and have different strengths and weaknesses. In cold-start situations where a sizable amount of data is not available, some Recommender Systems, such as knowledge-based systems, perform better. When there is a ton of data accessible, other recommender systems, such collaborative techniques, are more successful.

When a wider variety of inputs is available, one has the choice to employ a variety of Recommender Systems for the same task in various circumstances.

In these circumstances, there are numerous opportunities for hybridization, in which various elements from various kinds of systems are combined to achieve the best of both worlds. Hybrid Recommender Systems are closely related to the field of ensemble analysis, in which the strength of multiple types of machine learning algorithms is combined to create a more robust and reliable model. By merging numerous models of the same type, ensemble-based Recommender Systems are able to increase not only the power of many data sources but also the efficiency of a particular class of Recommender Systems (such as collaborative systems).

This situation is not all that different from ensemble analysis in the context of data classification.

Then, we can identify some minor and less used Recommender Systems: Knowledge-based systems; demographic systems and community-based systems.

2.2. MINOR RECOMMENDER SYSTEMS

Knowledge-based systems are based on the concept “tell me what fits my needs”. Such systems provide recommendations to the user based on a specific domain knowledge. To do this, the system collects the user’s requirements, compares them with its knowledge base about that particular domain and recommends those items that, in its opinion, while also taking into account the user's choices, are the most relevant and useful. Let’s use an example to help with the understanding: consider online shopping websites. If you wish to purchase something, such as a smartphone, you will be asked for your specifications. Once you do that, you will then be given recommendations for the best products that not only fulfill your specifications, but also those that the system believes would be the most beneficial to you.

Knowledge-based recommender systems are really helpful when dealing with items that are not purchased frequently. Examples include expensive luxury products, real estate, cars, travel demands, and financial services. In these circumstances there wouldn’t probably be enough ratings available for the recommendation process, since the items are bought rarely and also with different types of detailed options.

In addition, when dealing with such products, consumer tastes and preferences may change over time. For instance, the model of a smartphone or a car could change dramatically over a few years and, as a result, preferences might evolve correspondingly. In some other situations, it could be challenging to accurately and fully capture user interest using historical data such as ratings. The many properties of an object may be matched by attributes, and a user may be only interested in items with a certain set of properties. Cars, for instance, may come in a variety of types, models, colors, engine configurations and interior options. Depending on how these options are combined, user preferences may be affected, and a customer could be interested only in items that have a specific combination of these properties or options. Consequently, the item domain in these situations tends to be complex, and it is really hard to associate sufficient ratings with the large number of combinations at hand.

Such situations can be addressed with knowledge-based Recommender Systems, which are based on the similarities between customer requirements and item descriptions, or the use of constraints specifying user requirements. The process is facilitated with the use of *knowledge bases*, which hold information on the rules and similarity functions to employ throughout the retrieval process. In fact, the strategy gets its name from the fact that knowledge bases are crucial to the efficient operation of these systems. User control over the recommendation process is increased as a result of the clear statement of requirements.

Both collaborative filtering and content-based systems, recommend items to users based on the user's past actions or ratings, the action or ratings of their peers, or a combination of the two. Knowledge-based systems are unique in that they allow the users to explicitly specify what they want.

The interactivity in knowledge-based recommender systems can be achieved with one or more of the following methods:

1. *Conversational systems:* here, a feedback loop is used to iteratively determine the user preferences. The primary reason for this is that the item domain is complex and makes it not possible to discern user preferences outside of an iterative conversational system.

2. *Search-based systems:* in search-based systems, user preferences are elicited by answering a predetermined series of questions, such as: “Do you prefer a diesel-fueled or an electric car?” To enable the flexibility to set user constraints, it may occasionally be necessary to build personalized search interfaces.

3. *Navigation-based recommendation:* here, the user specifies a number of change requests to the item being currently recommended. Through a series of incremental modification requests, it is possible to arrive at a desirable item. These Recommender Systems are also referred to as *critiquing recommender systems*.

It is important to highlight that both knowledge-based and content-based systems depend significantly on the attributes and characteristics of the items. In fact, knowledge-based systems are considered to be the closest to content-based systems. The main difference is that content-based Recommender Systems learn from *past user behavior*, whereas knowledge-based systems recommend based on active user *specification of their needs and interests*.

Demographic systems, as the name says clearly, are based on the user's demography or on the location they are from. The idea behind this is that the recommendations made are based on the demographic region of the user: for example, for any given online shopping website, the user will have to insert their region. As a result, the user would only receive recommendations for items that were offered in the selected region, and the price would also be shown in the appropriate currency. Therefore, only pertinent suggestions would be offered.

In demographic recommender systems, the user's demographic data is used to train classifiers that can relate particular demographics to ratings or purchasing tendencies. In many situations, it is possible to direct the recommendation process by combining demographic data with additional context.

More recent techniques have focused on using classifiers for making recommendations. One of the interesting systems in this regard was a method that deduced and extracted features from the users' home pages in order to forecast their likelihood of liking specific restaurants.

Rule-based classifiers are frequently used to interactively link the demographic profile to purchasing behavior. Although demographic Recommender Systems typically do not provide the best results when used alone, when included in hybrid models, they considerably increase the power of the Recommender System. Demographic techniques can be eventually combined with knowledge-based systems to increase their robustness.

Community-based systems follow the: “tell me who your friends are, and I will tell you who you are”. In contrast with collaborative filtering, where recommendations are based on users' similarities (random or friends and peers), here the recommendations are solely based on the similarity with the user's friends. Think of Facebook, for instance. You are much more likely to add someone to your profile who was recommended by a friend than you are to add someone who is listed under "People You May Know". And this is what community-based systems are all about, which clearly is based on the fact that we rely and depend more on the recommendations that take into account our friends than those that consider random people.

Clearly, the various types of RS have advantages and disadvantages, as presented in the image below.

Table 1 - Pros and cons of the different RS discussed - Sharma, Singh, 2016

TYPE	PROS	CONS
CF	Provides better results when the no. of users and the no. of ratings are available in abundance. No knowledge engineering required.	Cold-start problem both in case of new users and new items. Sparsity problems
CBF	It does not rely on other users' preferences. Comparison between items possible.	Users with thousands of purchases or varying tastes, make it a little difficult to make the valid recommendations. Cold start for new users. No surprises.
KBS	Deterministic recommendations Assured quality No cold start	Cost of knowledge acquisition from domain experts, users etc. are too high. Basically static
HS	Avoids some of the shortcomings.	Most datasets do not allow to compare different recommendation paradigmz.
DS	Recommendations valid to only a particular region are made, thus avoiding invalid ones.	Relatively less research in this particular technique.
CBS	It enables simple & comprehensive data acquisition related to users' social relations.	Recommendations are not always accurate.

Here,

CF- Collaborative Filtering
CBF- Content-Based Filtering
KBS- Knowledge Based Systems
HS- Hybrid Systems
DS- Demographic Systems
CBS- Community Based systems

Therefore, each business or anyone, when implementing a Recommender System, will have to select the most appropriate data filtering system for the objectives they want to achieve. Each RS faces obstacles that it will have to overcome, not only on a practical level, but also on an ethical and social level.

2.3. RECOMMENDER SYSTEMS' CHALLENGES

We can detect here the major challenges that the most used Recommender Systems have to face and overcome.

First, we have ***data sparsity.*** We already talked about it when introducing the collaborative filtering method. As we saw, this problem relates to the fact that users usually rate a limited number of items; consequently, data and ratings will not be sufficient for determining neighbors and for elaborating accurate suggestions.

A second problem is ***scalability***. While the data is mostly sparse, for major websites it includes millions of users and items. Therefore, it is crucial to take into account the concerns with computational cost and look for recommender algorithms that are either not too demanding or simple to parallelize (or both). Another option is to use incremental versions of the algorithms, which would allow suggestions to be updated incrementally rather than globally (using the entire data) as the data grows.

With the ***cold start*** problem, we have to take into account that typically, whenever a new user enters the platform or system for the first time, there is not enough data to generate an accurate recommendation for them. The typical approaches to solving this issue rely on hybrid recommender techniques that combine content and collaborative data. Occasionally, these approaches also involve asking users for some basic data (such as their age, location, and favored genres).

Diversity vs. accuracy. When we want to recommend items which are likely to be appreciated by a particular user, it is usually most effective to recommend popular and highly rated items. However, this recommendation has very little benefit for the consumers because popular objects are simple to locate and frequently difficult to avoid, even without a Recommender System. Therefore, a good list of suggested items should also include less obvious suggestions that are unlikely to be made by the users themselves. The usage of hybrid RS and direct diversification of the suggestion list are two solutions to this issue.

Vulnerability to attacks. Recommender Systems are likely targets of malicious attacks seeking to unfairly promote or restrict specific items because of their significance and relevance in e-commerce applications. There are many strategies available to stop this kind of conduct, from advanced resistant RS to preventing malevolent evaluations from entering the system.

This is a difficult undertaking, nevertheless, as attackers' tactics advance along with the development of defense mechanisms.

The value of time. Real users have interests with widely diverse time scales (for example, short-term interests related to a trip or vacation and long-term interests related to the place to live in or political preferences). Most recommendation algorithms neglect and ignore the timestamps of evaluations. The question of whether and how the worth of outdated judgments should deteriorate with time, as well as what are the typical transient patterns in user evaluations and item relevance, is still being researched.

Then there is the problem related to the ***evaluation of the Recommender Systems.*** Although there are many different metrics available, it is still unclear how to select the ones that are most appropriate for the particular work and situation. Comparisons of various recommender algorithms are particularly difficult because they might just be used to handle different problems and tasks.

User interface. It has been demonstrated that suggestions need to be transparent in order for consumers to accept them: consumers value recommendation when it is clear the reason why a particular item has been suggested to them. Another challenge is that because the list of potentially interesting objects may be extremely extensive, the interface must be presented simply and navigated through easily so that users can browse various recommendations, which are frequently obtained via different approaches.

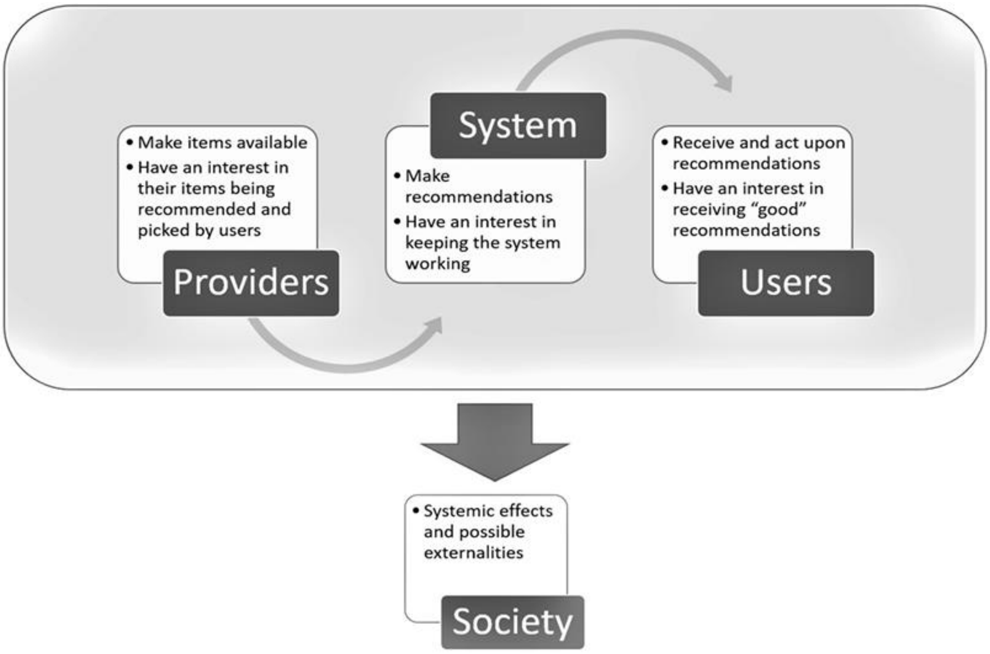
Along with the aforementioned persistent problems, several new ones have recently emerged. Scientists began to think about the implications of network structure on suggestion and how to leverage known structural elements to improve recommendation thanks to the development of methods in adjacent fields of science, particularly the new tools in network analysis. Using consumer-product networks as an example, Huang et al. proposed an enhanced recommendation algorithm that favors edges that increase the local clustering property. New obstacles are also brought forth by advancement and the spread of new techniques. For instance, GPS-enabled mobile phones are already commonplace, and Internet connectivity is available everywhere; as a result, location-based recommendations are now possible and becoming every day more important. Both the great predictability of human movement and quantitative approaches to defining commonalities between places and people are necessary for accurate recommendation. Last but not least, effective Recommender Systems ought to consider the various behavioral patterns of various people. Older users typically have more focused interests than new users, for instance, and users behave very differently when engaging in low-risk (such as bookmarking, downloading music, etc.) and high-risk (such as purchasing a computer, renting a home, etc.) activities.

Among all of this, when analyzing Recommender Systems, we also have to take into consideration the consequences and concerns in terms of ethics and society at large.

Recommender Systems gather, curate, and utilize enormous volumes of personal data in order to function successfully and efficiently. They inevitably have an impact on how each person interacts with others and experiences digital settings.
RS have therefore a wider influence on users and on society more broadly. After all, they shape and influence user preferences and direct decisions, both on an individual and social level. Given that RS can be used in morally charged situations like healthcare, lifestyle, insurance, and the labor market, this impact is important and merits ethical attention.

In order to properly analyze the ethical challenges faced by RS, in the figure are displayed the stakeholders involved in the recommendation process, who usually have different and possibly contrasting interests.

Figure 1 - The different stakeholders involved in RS - Milano et al., 2020



We take into account just actions and consequences while assessing ethical dilemmas. We can pinpoint two ways in which a Recommender System may have an ethical impact by considering actions and their results.

First, its operations a) could have a detrimental influence on any of its stakeholders' utility and/or b) may infringe upon their rights.

Second, these two types of ethical influence may have an immediate effect (for instance, an inaccurate advice may reduce the utility for the user) or they may expose the concerned parties to hazards in the future. This is presented in Table 1.

Table 2 - Possible ethical problems with RS - Milano et al., 2020

	Immediate harm	Exposure to risk
Utility	Inappropriate content	Opacity
		Questionable content
Rights	Unfair recommendations	Privacy
	Encroachment on individual autonomy and identity	Social effects

Given the previous analysis, we can now divide the ethical problems brought on by recommender systems into two categories: whether a particular feature of an RS negatively affects the utility of some of its stakeholders or, instead, constitutes a rights violation, which is not always measured in terms of utility; and whether the negative impact results in immediate harm or exposes the relevant party to risk of harm or rights violation in the future.

The ethical challenges of Recommender Systems

We can identify six main areas of ethical concerns, which often overlap.

1. Inappropriate content. Focusing on e-commerce applications, Paraschakis suggests that there are five ethically problematic areas: user interface design, the practices of user profiling, algorithm design, data publishing, and online experimentation or A/B testing. This last one is the practice of exposing chosen groups of users to modifications of the algorithm, with the aim of gathering feedback on the efficacy of each version from the user responses.

The risks he cites include privacy violations (such as data leaks or data collection without explicit consent), anonymity violations, behavior manipulation and bias in user recommendations, content censorship, exposure to side effects, and unequal treatment in A/B testing with minimal user awareness, all of which could erode user trust. The recommendations rely on a user-centered design methodology, introducing movable tools for users to actively control how RS use their personal data, filter out marketing biases or content censorship, and opt out of online trials and experiments.

2. Privacy

One of the main issues for Recommendation Systems is user privacy. Given that the bulk of the most commercially successful RS rely on hybrid or collaborative filtering approaches and create models of their users in order to produce tailored recommendations, this may be considered as inevitable. Risks to your privacy come in at least four stages. They can first appear when user agreement is not explicitly obtained before data is gathered or shared.

Second, when data sets are stored, there is a chance that they could be exposed to de-anonymization attempts or leak to outside parties. At both stages, privacy violations put users at risk of losing utility (for instance, if specific users are singled out by malicious agents as a result) or violating their rights (for instance, if their private information is used in ways that put their individual autonomy in jeopardy).

Third, privacy concerns also surface at the point where the machine might draw conclusions from the data, which is unrelated to how securely data is collected and stored. Users might not be aware of the nature of these inferences, and if they were, they might object to this use of their personal information. Privacy hazards extend beyond data gathering because, for instance, a third party could deduce sensitive information about a user by looking at the system's recommendations for that user.

Lastly, the system can create a model of the user based on the information it has obtained on how other users interact, and this raises another subtle but significant systemic privacy concern at the collaborative filtering stage. In other words, the system may be able to create a reasonably accurate profile even for those people about whom it has less data, as long as enough users interact and share their data with it. This suggests that it might not be possible to totally protect specific users from the types of inferences the system may be able to make about them. In some fields, like medical research, it might be a benefit, but it might also turn out to be problematic in other domains, like recruitment or finance.

3. Autonomy and personal identity

By making suggestions that push users in a specific direction, by seeking to "addict" them to certain types of information, or by limiting the range of possibilities they are exposed to, Recommender Systems may infringe on the autonomy of specific users. These interventions can range from being problematic (persuasion, nudging), to potentially harmful (manipulative and coercive), to being benign (allowing individual agency and helping better decision-making by screening out irrelevant possibilities).

One notable example is YouTube's recommendation algorithm, which has drawn a lot of attention recently for its propensity to highlight biased content and "fake news" in an effort to keep people on the platform.

The question to ask is not how users may avoid the traps of Recommender Systems, but rather how users can make the traps work for them in light of this captological approach and their effectiveness and pervasiveness.

4. Opacity

Theoretically, giving users access to the justifications for why the system "thinks" that particular alternatives are pertinent to them could assist to reduce the risk of infringing on their autonomy. Additionally, it would help make algorithmic decisions about how to categorize and model users more transparently.

Herlocker pointed out that recommendations produced by collaborative filtering approaches can, at a basic level, be regarded as comparable to "word of mouth" recommendations among users. However, the offline word of mouth basically works on the basis of trust and shared personal experience. In contrast, Recommender Systems do not give users access to the identities of other users or the models the system uses to produce the recommendations. As we previously stated, this is problematic because it reduces the user's autonomy.

For instance, telling a user that a particular item is suggested because it is the most popular among other users may make the item more desirable, leading to a spiral and a self-reinforcing pattern, where the item is recommended more frequently because it is well-liked and popular. This in turn increases its popularity and the result is a winner-takes-all situation that, depending on the intended sphere of application, can have negative consequences on the variety of possibilities, plurality of choices, and the establishment of competition.

5. Fairness

Fairness in algorithmic decision-making is a complex subject that is made more difficult by the existence of various, incompatible fairness definitions.

Yao and Huang (2017) focus on three ideas of fair recommendations, from the perspectives of either the user or consumer (C-fairness), who wants to receive the most pertinent recommendations, the provider (P-fairness), who wants to have their own products or services recommended to users who might be interested, or a combination of the two (CP-fairness).

Also, the idea of fairness is closely related to the social context in which the system collects its data and provides recommendations.

6. Social effects

A much-discussed effect and consequence of some Recommender Systems is their transformative impact on society. Particularly, news Recommender Systems and social media filters run the risk of shielding users from exposure to opposing viewpoints by virtue of their design, leading to "filter bubbles" and self-reinforcing biases that are detrimental to the normal operation of public debate, group deliberation, and democratic institutions more generally. The social usefulness of this aspect of RS may be negatively impacted. The propagation of misinformation against vaccines, which has been connected to a decline in herd immunity, is a relatively recent but alarming example. Protecting these systems from manipulation by (often even small yet) particularly active user groups, whose interactions with the system might produce intensely positive feedback and increase the system's rate of recommendations for particular goods, is a closely connected topic.

III. RECOMMENDER SYSTEMS' ALGORITHMS

Different Recommender Systems utilize different methods of data analysis to create sets of user and item affinities that can be used to find matches. While, as we saw earlier, Content-based Filtering systems are based on profile traits, and Hybrid approaches aim to combine both of these concepts, Collaborative Filtering systems evaluate past interactions alone. An important topic of research, with a dedicated Association for computing machinery (ACM) conference, is the design of Recommender Systems and how real-world problems are evaluated using them. This area crosses various sub-disciplines, including statistics and machine learning.

Figure 2 – Matrix with users and items rating, where we want to predict the missing rating for the active user - Melville, Sindhwani, 2017

Items

Users	*1*	*2*	...	*i*	...	*m*
1	5	3		1	2	
2		2				4
:			5			
u	3	4		2	1	
:					4	
n			3	2		
a	3	5		?	1	

The most typical setting for recommender systems is depicted in the figure. A matrix of *n* users and *m* items is used to reflect known user preferences, with each column representing the user u's rating of item i. As most users do not rate the majority of items, this user ratings matrix is often sparse.

The goal of the recommendation is to predict what rating a specific user would give to a previously unrated item. The most highly rated items are typically displayed as recommendations, and ratings are typically anticipated for all products that have not yet been examined or observed by a user. The term "active user" refers to the person who is currently being taken into account for suggestions. We can focus on the different algorithms utilized by the three main categories of filtering systems: Collaborative Filtering, Content-based and Hybrid systems.

3.1. COLLABORATIVE FILTERING ALGORITHM

Collaborative Filtering systems gather user feedback in the form of ratings for items in a specific domain and use patterns of rating behavior shared by many users to decide how to recommend a product. Using a database of user votes from a sample or population of other users (the user database), the objective of collaborative filtering is to predict the usefulness of items to a specific user (the active user). A common criterion of differentiation is whether a collaborative filtering system uses implicit or explicit votes. We talk about explicit voting when a user expresses his preference openly, typically on a discrete numerical scale. Implicit voting is the process of attributing a vote or preference based on user behavior or choices. Implicit votes may be based on surfing history (such as in Web apps), purchasing behavior (such as in brick-and-mortar or online stores), or other information access patterns. Collaborative filtering algorithms must deal with the issue of missing data, regardless of the kind of vote data that is provided.

Collaborative Filtering methods can be further subdivided into neighborhood-based (also referred to as memory-based approaches) and model-based approaches.

3.1.1. Neighborhood-based Collaborative Filtering

Memory-based algorithms perform predictions on the whole user database. In neighborhood-based approaches, a pool of users is selected based on how closely they resemble the active user, and predictions for the user in question are generated using a weighted combination of their ratings. The algorithm outlined in the following steps can be used to generalize the majority of these techniques:

1. Assign each user a weight based on how similar they are to the active user. The weight *wa,u* used in this phase serves as a measure of how similar the user *u* and the active user *a* are. The Pearson correlation coefficient between the ratings of the two users is the most frequently used metric of similarity and is as follows:

$$w_{a,u} = \frac{\sum_{i\in I}(r_{a,i} - \overline{r_a})(r_{u,i} - \overline{r_u})}{\sqrt{\sum_{i\in I}(r_{a,i} - \overline{r_a})^2 \sum_{i\in I}(r_{u,i} - \overline{r_u})^2}}$$

where I is the collection of elements that both users evaluated, ru,i is the rating that user u gave item i, and ru is the mean rating that user u gave.

2. In the second step we choose the k users, often referred to as the neighborhood, who are the most like the active user.

3. Create a prediction using a weighted average of the ratings of the chosen neighbors. Predictions are typically calculated in this stage as the weighted average of deviations from the neighbor's mean, as in the following formula:

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u\in K}(r_{u,i} - \overline{r_u}) \times w_{a,u}}{\sum_{u\in K} w_{a,u}}$$

where $w_{a,u}$ is the degree of similarity between users a and u, K is the neighborhood of users who are most similar and $p_{a,i}$ is the prediction for the active user a for item i.

Pearson correlation-based similarity evaluates the degree of linear dependency between two variables. As an alternative, one can compute similarity based on the cosine of the angle between two users' ratings as a vector in an m-dimensional space, as shown by:

$$w_{a,u} = \cos(\vec{r}_a, \vec{r}_u) = \frac{\vec{r}_a \cdot \vec{r}_u}{\|\vec{r}_a\|_2 \times \|\vec{r}_u\|_2} = \frac{\sum_{i=1}^{m} r_{a,i}\, r_{u,i}}{\sqrt{\sum_{i=1}^{m} r_{a,i}^2}\sqrt{\sum_{i=1}^{m} r_{u,i}^2}}$$

One cannot have negative ratings when calculating cosine similarity, and unrated items are given a rating of zero. According to empirical research, Pearson correlation typically performs better with respect to the cosine similarity.

Item-based Collaborative Filtering. Due to the computational difficulty of the search for similar users, traditional neighborhood-based CF algorithms do not scale effectively when applied to millions of users and products. Linden et al. (2003) instead proposed item-to-item Collaborative Filtering, which matches a user's rated items to similar items rather than matching similar individuals. In reality, this strategy speeds up online systems and frequently produces better recommendations. This method uses Pearson correlation to calculate similarities between pairs of items i and j off-line as follows:

$$w_{i,j} = \frac{\sum_{u \in U}(r_{u,i} - \overline{r_i})(r_{u,j} - \overline{r_j})}{\sqrt{\sum_{u \in U}(r_{u,i} - \overline{r_i})^2 \sum_{u \in U}(r_{u,j} - \overline{r_j})^2}}$$

where *U* is a set of all users who have given ratings to both items i and j, ru,i is the user u's rating of item i, and ri is the sum of all user ratings for item it. Now, it is possible to forecast the rating for item *i* for user a using a straightforward weighted average, as in:

$$p_{a,i} = \frac{\sum_{j \in K}(r_{a,j} - w_{i,j})}{\sum_{j \in K}\left|w_{i,j}\right|}$$

where K is the neighborhood set of the k items that user *a* has rated and that are most comparable to item *i*. Alternative similarity metrics, such as significance weighting, default voting, inverse user frequency and case amplification, can also be used for item-based collaborative filtering.

3.1.2. Model-based Collaborative Filtering

Machine learning and data mining techniques are applied in the context of predictive models in model-based methodologies. The parameters of this model are learned within the scope of an optimization model, when the model is parameterized. We can talk about, for instance, decision trees, rule-based models, Bayesian techniques and latent factor models. Even for sparse ratings matrices, several of these techniques, such as latent factor models, offer a high level of coverage.

The user database is utilized in model-based collaborative filtering to estimate or learn a specific model, which is subsequently applied to predictions. Through the estimation of statistical model parameters for user ratings, these strategies offer different recommendations.

For instance, one early method to convert CF to a classification problem involved creating a classifier for each active user that represented items as feature vectors over users and the available ratings as labels. Dimensionality reduction techniques may have also been used to address the problem of sparse data in this method. Similar applications of other predictive modeling strategies have also been made.

In this family of approaches, **latent factor** and **matrix factorization** models have lately emerged as state-of-the-art methodologies. Latent Factor models, in contrast to neighborhood-based approaches, presume that the similarity between users and objects is simultaneously driven by some hidden lower-dimensional structure in the data.

For instance, it may be believed that a user's rating of a movie is influenced by a few implicit elements, such as the user's preferences for certain movie genres. Users and things are simultaneously represented as unknown feature vectors (column vectors) $w_u, h_i \in \Re^k$ along k latent dimensions in Latent Factor models, using matrix factorization techniques. These feature vectors are learned in order for the inner products $w_u^T\, h_i$ to be close to some loss measure to the known preference ratings $r_{u,i}$. Standard loss function selections include squared loss, in which case the following objective function is minimized:

$$J(W, H, \{b_u\}_{u=1}^{n}, \{b_u\}_{i=1}^{m}) = \sum_{(u,i)\in L} (r_{u,i} - w_u^T\, h_i)^2$$

with $W = [w_1 \ldots w_n]^T$ as an $n \times k$ matrix, while $H = [h_1 \ldots h_m]$ is a $k \times m$ matrix and L is the set of the known ratings for the user-item pairs.

Although the two may not necessarily be the same, a ratings matrix is occasionally referred to as a utility matrix. In a strict sense, the utility of a user-item combination refers to the profit made by recommending that particular item to that specific user, where the utility relates to the quantity of profit. It is feasible for the application to explicitly convert the ratings to utility values based on domain-specific criteria, even though utility matrices are frequently specified to be the same as the ratings matrices. The utility matrix is then used rather than the ratings matrix to apply all collaborative filtering techniques. In reality, however, collaborative filtering algorithms hardly never take this route and instead operate directly on the ratings matrix.

The matrix is known as a positive preference utility matrix for situations where the ratings are unary since it only permits the expression of positive preferences.

Unary matrices are also known as implicit feedback matrices since they are typically produced by user activities, such as purchasing an item. As there is no way to tell if a user likes or dislikes an item, unary ratings have a big impact on the current Recommendation System. It is frequently advised to carry out the analysis simply in the case of unary matrices by initially interpreting the missing entries as 0. But if the item matches user interests, the algorithm's ultimate estimated value may be significantly higher than 0.

Thus, the entries with the biggest positive prediction error relative to the initial "zero" assumption are the ones on which the recommended items are based on. In reality, severe overfitting is conceivable if the missing items are not replaced with 0s. This particular form of overfitting is a byproduct of the fact that there is frequently insufficient differentiation between the multiple observed ratings values. Ratings correspond to (more discriminated) preferences in explicit feedback matrices, whereas ratings correspond to (less discriminated) confidences in implicit feedback matrices.

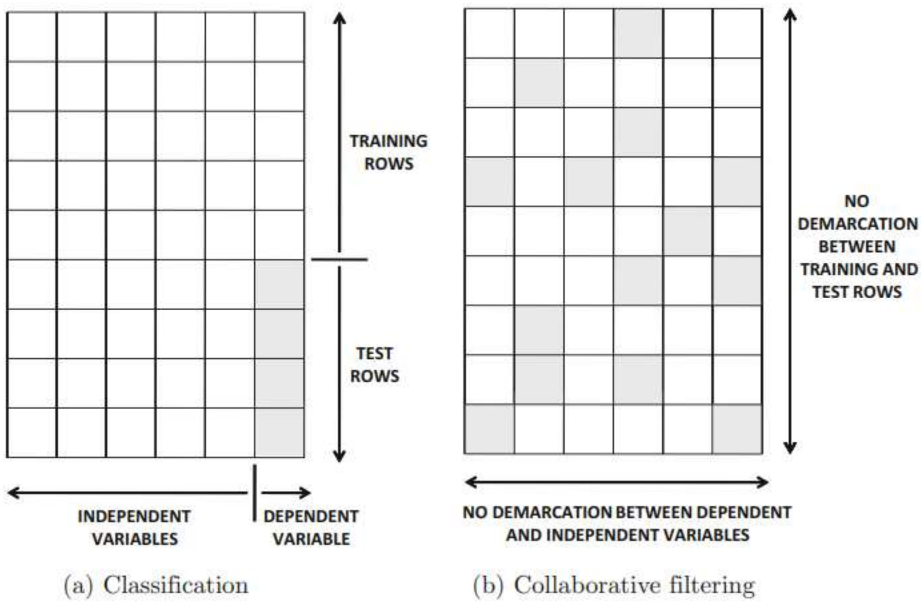
The analysis is always significantly biased when missing items are replaced with any value (such as 0 or the row/column/data mean) in explicit ratings matrices that include both likes and dislikes.

Although it is frequently negligible, in the unary scenario, replacing missing entries with 0s introduces some bias because implicit feedback data, such as buying data, by default assumes that the user won't purchase the majority of the goods. In the unary situation, one is frequently ready to put up with this bias because the substitution significantly reduces overfitting.

Relationship with Missing Value Analysis. Missing value analysis and Collaborative Filtering models are closely connected. The issue of imputation of entries in a data matrix with insufficiently detailed data is examined in the classical literature on missing value analysis. The (difficult) specific example of this problem occurs when the underlying sparse and big data matrix. Generally speaking, collaborative filtering can also be employed with many traditional missing value estimate techniques.

Collaborative Filtering and the Regression and Classification Modeling. This framework can be seen as a generalization of the collaborative filtering problem since any column, as opposed to (just) the class variable, is permitted to have missing data. Because each feature functions as both a dependent and an independent variable in the recommendation issue, it is difficult to distinguish between class variables and feature variables. Only because the missing entries are limited to a certain column does this distinction arise in the categorization problem. Furthermore, because any row could include missing entries, there is no differentiation in collaborative filtering between training and test rows. As a result, it makes more sense to refer to training and test entries rather than training and test rows when discussing collaborative filtering. The generalization of classification/regression modeling of CF performs predictions entry-wise rather than row-wise. It is crucial to bear in mind the connection between classification/regression modeling and CF since Recommender Systems can benefit from many of the generalizations made by classification and regression modeling techniques. The figure shows how the two issues are related to one another.

Figure 3 - Relation between Classification and Collaborative filtering. Grey boxes are the ones that need to be predicted - Aggarwal, 2016



The transductive setting in classification and regression and the matrix completion problem both share a few traits. It is frequently challenging to create predictions for test examples that are not available at the time of training when using the transductive setup because the test instances are also included in the training process (usually using a semi-supervised approach). On the other hand, inductive models are those in which it is simple to make predictions for novel occurrences. A naïve Bayes model for classification, for instance, is intrinsically inductive since it may be used to predict the label of a test instance even when the features were unknown while the Bayes model was being built. The m n ratings matrix R's strong integration of the training and test data makes the matrix completion setting inherently transductive, and many models struggle to predict ratings for out-of-sample users and/or items. For instance, many commercially available approaches won't be able to anticipate Lucy's behavior if she is added to the ratings matrix (with numerous predefined ratings) after the collaborative filtering model has already been built. This is particularly true of methods of model-based collaborative filtering. The ratings for users and/or products outside of the sample can, however, be predicted in some newer inductive matrix completion models.

3.2 CONTENT-BASED RECOMMENDATION ALGORITHM

Recommenders using pure collaborative filtering make use of only the user ratings matrix, either directly or indirectly through inducing a collaborative model. These methods treat each user and object as a single atomic unit, and predictions are made without taking into account the unique characteristics of any one user or item. However, by knowing more about a user, such as demographic details, or about an item, such as the director and genre of a movie, one may offer a better personalized recommendation. Recommender systems that use comparisons between representations of content describing an item and representations of content that the user is interested in are referred to as *content-based recommenders*. These methods are also sometimes known as content-based filtering.

As the name suggests, content-based filtering makes use of data items' content to estimate their relevance in light of the user's profile. Information retrieval and artificial intelligence, in particular, are two areas of computer science that are intersected by research on content-based recommender systems.

A collection of features, also known as attributes or qualities, are used to represent items that can be recommended to the user. Examples of elements used to characterize a movie include cast, genres, year release etc. in a movie recommendation app. Structured data is used to represent items when they are all described by the same set of attributes and when the possible values for those attributes are known.

Numerous ML techniques can be applied in this situation to learn a user profile. The majority of content-based filtering algorithms take textual information from product descriptions, emails, news stories, and web pages to create item descriptions. There are no attributes with clearly defined values, in contrast to structured data. Due to the ambiguity of real language, text features make it difficult to learn a user profile. One of the more creative and intriguing solutions to these difficulties suggested in the literature involves the incorporation of semantic analysis into customization models.

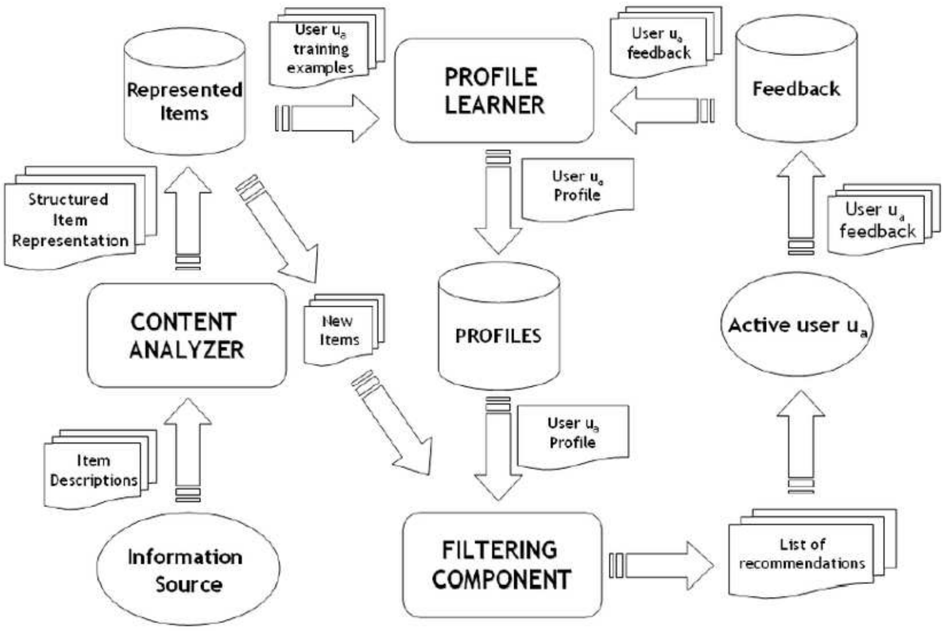
When the web pages themselves or related content like descriptions and user reviews are accessible, much of the study in this field has concentrated on recommending products with connected textual content, such web pages, books, and movies.

Accordingly, a number of ways have approached this issue as an information retrieval (IR) task, where the content related to the user's preferences is regarded as a query and the unrated documents are evaluated with relevance to or similarity to this query. To create a prototype vector of each rating category for a user in NewsWeeder, documents in each rating category are first transformed into tf-idf word vectors and then averaged. A new document is categorized by being compared to each prototype vector and receiving a predicted rating based on how similar it is to each category on a cosine basis.

The treatment of recommendation as a classification job, where each example represents the content of an item and user prior ratings are utilized as labels for these examples, offers an alternative to IR techniques. Mooney et al. trained a multinomial naïve Bayes classifier in the area of book recommendation using text from fields including the title, author, synopses, reviews, and subject phrases. In a probabilistic binary classification context, ratings on a scale of 1 to k can be directly transferred to k classes, or alternatively, the numeric rating can be utilized to weight the training sample. Other classification techniques, such as k-nearest neighbor, decision trees, and neural networks have also been applied to pure content-based recommendation.

The complex architecture of a content-based RS algorithm is presented in the figure.

Figure 4 - Koene et al., 2015



Three steps make up the recommendation process, and each is handled by a different component:

1. **Content Analyzer**. In order to extract structured pertinent information from unstructured material (such as text), some sort of pre-processing step is required. The component's primary duty is to represent information from items (such as papers, Web pages, news articles, product descriptions, etc.) obtained from information sources in a format appropriate for the following processes. In order to change the representation of data items from the original information space to the target one (for example, Web pages represented as keyword vectors), feature extraction techniques are used to examine the data items. The <u>profile learner</u> and the <u>filtering component</u> use this representation as input.
2. **Profile learner.** In order to build the user profile, this module gathers information that is indicative of the user's preferences and makes an effort to generalize it. The generalization strategy is typically implemented using machine learning techniques, which can infer a model of customer interests based on past purchases made or rejected. For instance, a Web page recommender's profile learner can employ a relevant feedback method, where the learning methodology merges vectors of good and bad instances into a prototype vector reflecting the user profile. Web pages

with user-provided positive or negative comments serve as training examples;

3. **Filtering component.** By comparing the user profile representation to that of the recommended items, this section uses the user profile to suggest appropriate items. The outcome is a binary or continuous relevance assessment, the latter of which generates a ranked list of potentially interesting objects (computed using some similarity metrics). The matching is accomplished in the aforementioned example by computing the cosine similarity between the prototype and item vectors.

Indeed, the content analyzer, who often uses methods from information retrieval systems, performs the initial step of the recommendation process. This content analyzer processes item descriptions arriving from Information Source, extracting features (keywords, n-grams, concepts, etc.) to create a structured item representation that is kept in the repository Represented Items. The responses to items are somehow gathered and recorded in the repository Feedback in order to construct and update the profile of the active user. These responses, also known as annotations or feedback, are used in conjunction with the descriptions of related objects to build a model that may be used to forecast the true relevance of newly presented items.

As part of their initial profile, users can also expressly list their areas of interest without offering any feedback. Positive information (inferring features the user likes) and negative information (i.e., inferring elements the user is not engaged in) are typically two categories of relevance feedback that can be distinguished.

As we saw earlier, there are two methods that can be used to record user feedback. The term "explicit feedback" refers to a technique where a system asks the user to explicitly rate items. The term "implicit feedback" refers to a technique where feedback is derived from observing and analyzing user behavior without the need for active user involvement. An item's explicit assessments reveal how intriguing or relevant it is to the user.

The following are the three basic methods for receiving explicit relevance feedback:

1) Like/dislike: using a straightforward binary rating scale, topics are categorized as "relevant" or "not relevant";

2) Ratings: to evaluate items, a discrete numeric scale is typically used. Alternatively, a numeric scale is used to map symbolic ratings;

3) Text comments: to assist users in making decisions, remarks regarding a specific item are gathered and given to them. Customers' reviews on websites like Amazon.com, for instance, may aid customers in determining whether a product

has received favorable feedback from the general public. Textual comments are useful, but the active user may become overburdened by them since he must read and analyze each one to determine whether it is positive or bad and to what extent. The literature suggests cutting-edge methods from the emotional computing field to enable content-based recommenders to carry out this type of analysis automatically.

Although the adoption of numeric/symbolic scales increases the user's cognitive burden and may not be sufficient for capturing the user's feelings about items, explicit feedback has the advantage of simplicity. The foundation of implicit feedback techniques is the assignment of a relevance score to particular user actions on a piece of content, including saving, dismissing, printing, bookmarking, etc. Even if biasing is likely to happen, such as phone calls being answered while reading, their main advantage is that they do not require direct user engagement.

The training set for this active user needs to be defined in order to create the profile for that user. The profile learner uses supervised learning techniques to create a user profile from a set of item representations that have been labeled with ratings. User profiles are typically saved in a profile repository for subsequent use by the filtering component. By comparing features in the item representation to those in the representation of user preferences (stored in the user profile), the filtering component can determine if a new item is likely to be of interest to the active user. The filtering component typically uses a few techniques to rank potentially interesting items according to their relevance to the user profile.

A list of suggestions is given to the active user and includes the top-ranked items. Since user preferences frequently vary over time, current information must be kept up to date and supplied to the profile learner in order for the user profile to be automatically updated. By allowing consumers to express their satisfaction or dissatisfaction with products, more input on the generated recommendations is acquired. Following the gathering of such feedback, the learning procedure is repeated on a fresh training set, and the resulting profile is modified to take into account the most recent changes in user interests. The system can accommodate for the user preferences' dynamic nature thanks to the feedback-learning cycle's iterative process over time.

3.3. HYBRID APPROACHES

Hybrid recommender algorithms integrate two or more recommendation strategies to improve performance while minimizing the shortcomings of each one separately. Collaborative filtering is frequently used in conjunction with another method in an effort to get around the ramp-up issue. There have been various hybrid techniques developed that mix the two in order to take advantage of the "pros" of content-based and collaborative recommenders.

There's not a reason why many distinct strategies of the same kind couldn't be combined; for instance, two distinct content-based recommendations could cooperate, and indeed a number of initiatives have looked into this kind of hybrid RS. However, since they are the most frequently used and show the greatest promise for overcoming the cold-start issue, we are particularly focused on recommenders that incorporate data from many sources. Seven different types of hybrids were found in the earlier RS survey:

• **Weighted**: A numerical total is created by adding the scores of the various suggestion components.

• **Switching**: The system selects one of the recommendation components and uses it.

• **Mixed**: A combination of recommendations from several recommenders.

• **Feature Combination**: A single recommendation algorithm is given a set of features that were derived from many sources of knowledge.

• **Feature augmentation**: A feature or group of features is computed using one recommendation technique and utilized as input for another strategy.

• **Cascade**: Recommenders receive rigorous priority, and the lower priority ones are used to break ties between the higher priority ones in the scoring.

• **Meta-level**: After applying a recommendation approach, a model is created that is used as an input by a subsequent technique.

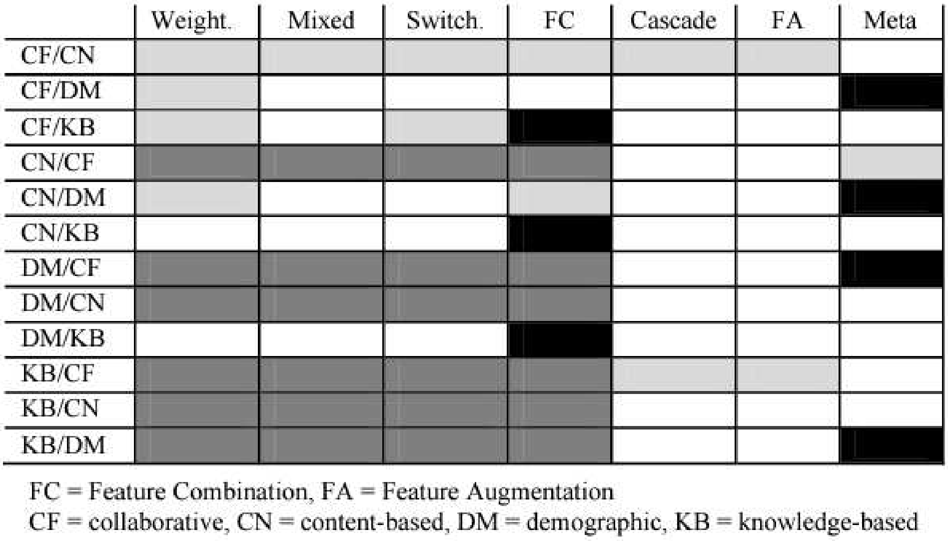
One straightforward strategy is to let the outcomes of content-based and collaborative filtering approaches produce separate ranked lists of suggestions before combining them to create a final list. The two forecasts were for example combined by Claypool et al. using an adaptive weighted average, where the weight of the collaborative component rises as the quantity of users accessing a given item does.

A general framework for content-boosted CF was put forth by Melville et al., in which recommendations are made using a CF approach after content-based predictions are utilized to change from a sparse into a full user ratings matrix. They specifically substitute the unrated items with predictions using a Naive Bayes classifier that was trained on documents that described the rated items of each user. They use the generated pseudo ratings matrix to identify users who are similar to the active user and then use Pearson correlation to make predictions that are correctly weighted to take into account both the active user's content predictions and the overlap of actually rated items. It has been demonstrated that this strategy outperforms linear combinations of the two, pure content-based systems, and pure CF.

Other hybrid strategies preserve a user's content-based profile while also being based on classic collaborative filtering. To find similar individuals, these content-based profiles are employed rather than co-rated goods.

The recommendation process is treated by a number of hybrid techniques as a classification task that includes collaborative aspects.

Figure 5 - Possibilities for hybrid RS - Burke, 2007



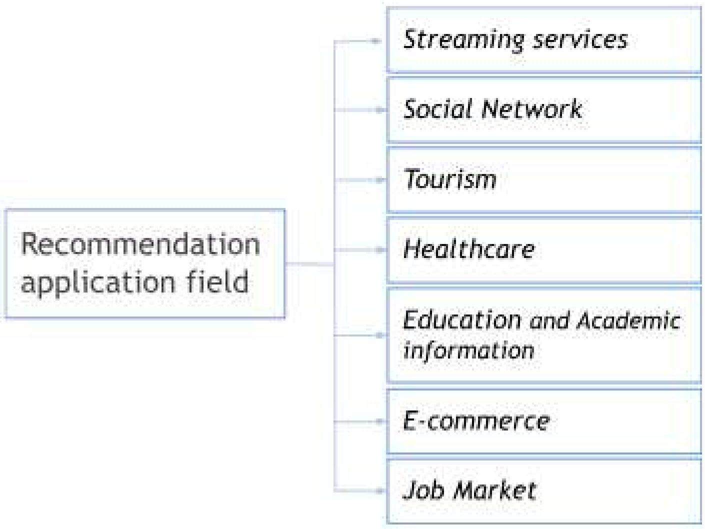
According to earlier research, the five suggestion ways and seven hybridization techniques together can produce 53 different two-part hybrids, as indicated in the table. Due to the order-sensitivity of several of the procedures, this number exceeds the expected 5x7=35. Because some hybrids cannot be logically distinguished from one another and because some combinations are impossible, the taxonomy is more complicated than it needs to be.

IV. RECOMMENDER SYSTEMS' APPLICATIONS FOR BUSINESSES

The study of Recommender Systems has grown over the years and is now applied in many service industries. Numerous web and application services have increased in popularity as a result of the growth and adoption of the Internet, smart devices, and Social Network Services (SNS). The expansion of these services has made it necessary to create a number of Recommendation Systems that can assist users in quickly receiving item information and making selections amidst the fast-growing amount of item information. As a result, RS for a variety of application sectors that make use of clickstream and real-time data collected by wearable technology frequently result in superior outcomes, in terms of accuracy and precision. The recommendation algorithm is generally separated into two technological components: a data mining area that analyzes data gathered about items and users, and a recommendation filter engine component. Every model and technology have undergone research and development to become more specialized to the service industry by utilizing the Recommendation System.

Studies on techniques including Text Mining, KNN, Clustering, Matrix Factorization, and Neural Network have been examined in relation to the filtering model of the Recommendation System. The technologies of text mining and clustering, which analyze user or location data of a comparable group for recommendations, have been thoroughly investigated over a long period of time. However, interest in the high likelihood of using neural networks in Recommendation Systems has arisen only recently, and modeling research to further secure or supplement data are also on the rise. As a result, research to enhance the effectiveness of Recommendation Systems is being broadened and actively carried out.

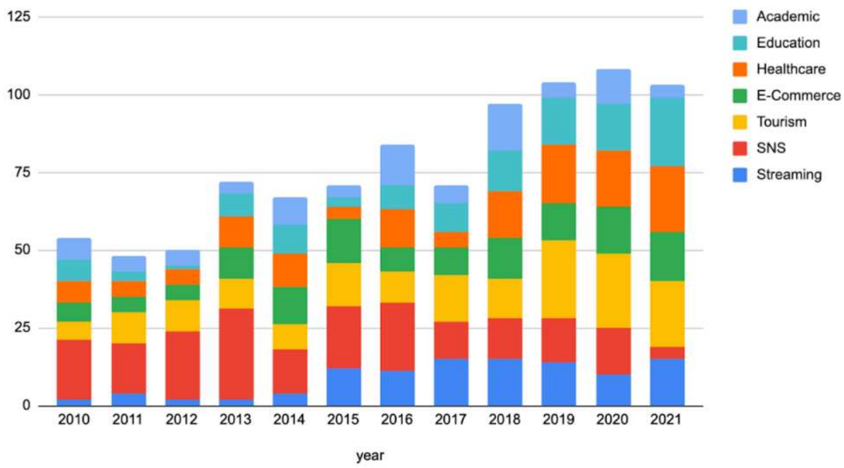
As we can see in the figure, we can distinguish seven major categories in which RS are applied successfully nowadays: streaming services such as Netflix or Spotify, SNS, services for tourism, for e-commerce services, but also for healthcare, education and in the job market (for job seekers but also for



employers).

This list represents the businesses where the utilization of Recommender Systems has seen a growing user base and business value, even though with different impacts and growth for each sector. We can see this distinction in the following graph.

Figure 6 - Development from 2010 to 2021 of different fields of research in RS - Ko et al., 2022



The active research on these application fields for RS has grown a lot in the last decade, in most of these categories, and continues to grow daily.

Now we are able to analyze more specifically some of these fields, in order to present the advancements and utilization of these algorithms in different businesses and practical applications. These will also be important examples for the introduction of the business value generated by RS, which will be discussed in chapter 5.

4.1. STREAMING SERVICES

In the past, users primarily watched TV or went to the movies. However, in recent years, major amounts of video content have been watched via streaming websites and platforms like Netflix and YouTube. Streaming services are one of the most common and used applications of Recommender Systems nowadays.

In fact, the COVID-19 pandemic has increased the number of "virtual cinemas" and online film festivals, as well as the competitive dynamics of the Streaming Wars, enabling services such as Netflix and Disney+ to accrue subscriber numbers that are record-breaking.

The way that audio information is consumed is also evolving, moving away from downloading and playing files to a user's local device and toward streaming services like Spotify. In order to deliver content that is personalized for each user and to ease consumers' concerns about selecting from a massive amount of content, streaming services for media content have been developed alongside Recommendation Systems. User preference data is typically gathered in the streaming service industry with a focus on the user's media content service usage history data. After mapping user preference to all of the content owned by the streaming service, recommendations are generated in the order of the content most similar to the user's preference.

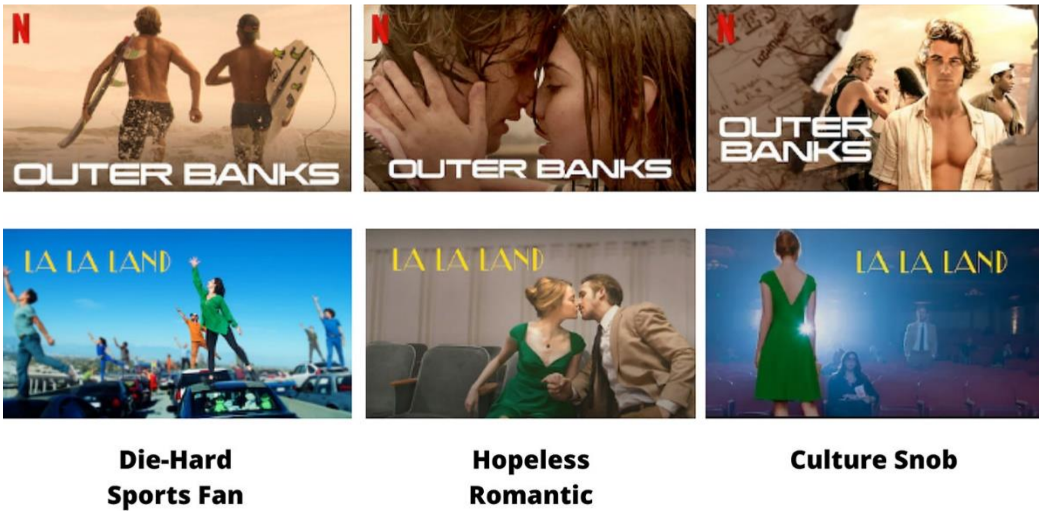
Up to the early 2000s, Recommendation Systems in the streaming service industry frequently used the Content-Based Filtering paradigm. Ontology was utilized in text mining during the Web 2.0 era for more than just finding movie products that consumers wanted, it was also used to make recommendations taking into account users' situational information.

The number of people using streaming services has grown significantly recently. For streaming services, research has been done to enhance several recommendation methods and filtering algorithms. In order to alleviate significant issues like scalability and data sparsity, Barragáns-Martinez et al. adopted the Singular Value Decomposition (SVD) algorithm to reduce computational overload when determining user preferences. This led to the proposal of a hybrid recommendation algorithm that suggests TV shows with higher user preference.

Based on the user's favorite movie genres and movie ratings, Walek et al.'s study (2023) indicated a hybrid recommendation model that suggests movies that match user tastes using the SVD algorithm and Fuzzy Logic. Consequently, promising results were obtained for Precision (81%), Recall (83%), and F-measure (82%).

The Netflix case is even more interesting. A lot of different studies analyze how this company invested so much in their Recommendation System that the majority of the details shown about a particular title are personalized, including the match score, artwork, trailer, synopsis, and metadata (such as awards, cast, etc.). According to Netflix, their "deep personalization" has allowed them to have hundreds of millions of products instead of just one Netflix product: one for each member profile. We can see, for example, how the artwork shown for a specific movie or TV series changes based on the "personas" each user is associated with. Three distinct Netflix taste personalities can be distinguished, each with distinctive identities and particular likes for movies and television shows. These three are the "Culture Snob" , the "Hopeless Romantic", and the "Die-Hard Sports Fan". We can highlight how for each of them a different movie cover is shown, based on their interests and on what attracts them the most.

Figure 7 - Different movie covers for different personas, based on their preferences - Pajkovic, 2021



The distribution of movies and television, and hence our consumption habits, will be more and more in the hands of semi-autonomous algorithmic tools as the Streaming Wars develop. Recommendation systems will become crucial competitive elements for every major OTT streamer. This makes the present, just before these systems become the predominant way that people watch movies and television, a crucial time for asking questions about how they operate, what functions they actually serve, and what impact they have had and will continue to have on movie and television culture.

Since the industry of music streaming services has a greater impact on user preference than other content does, high predictive performance is required for user preferences. Wang et al. (2015) used text mining to examine the musical item qualities, extract audio signal elements including rhythm and melody, and suggest items. This study suggested a hybrid recommendation algorithm that uses a neural network to suggest music that consumers like. In addition, using data samples chosen by similar user groups and obtained through collaborative filtering, McFee et al. (2021) proposed a study to examine audio similarity through content analysis. A hybrid recommendation model that permits the recommendation of several less-known musical genres was proposed using the similarity analysis results for each audio signal of the optimized audio content. In other words, by addressing the limitation of Content-Based Filtering - namely, the limited range of user preference items - this study demonstrated the expandability of the Recommendation System for music streaming services. The table presents different RS models and techniques for different Streaming Services.

Table 3 - Ko et al., 2022

Streaming Service	RS Model	RS Techniques
Video	CF	Clustering
		Matrix Factorization
	Hybrid System	Text Mining
		Matrix Factorization
Music	CB	Text Mining
		Neural Network
	CF	KNN
	Hybrid System	Matrix Factorization
		Neural Network

4.2. SOCIAL NETWORD SERVICE

Social networking websites are becoming far too common. These websites not only help people to stay in touch, but they also let users know what their friends, family, and coworkers are up to. The best examples include Facebook, Twitter, LinkedIn, and others. Such websites display recommendations in the form of "People you may know," "Pages you may like," "Suggestions for you," and other options. Here, recommendations are produced based on user tags, click-through rates, and browsing habits.

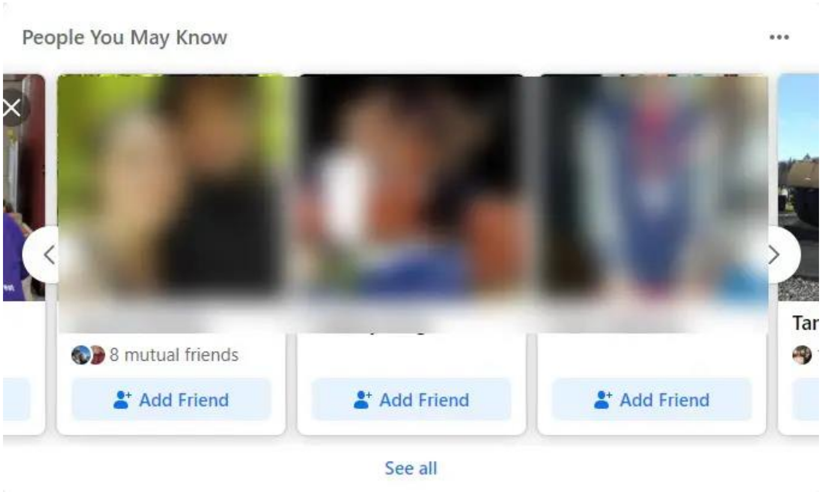


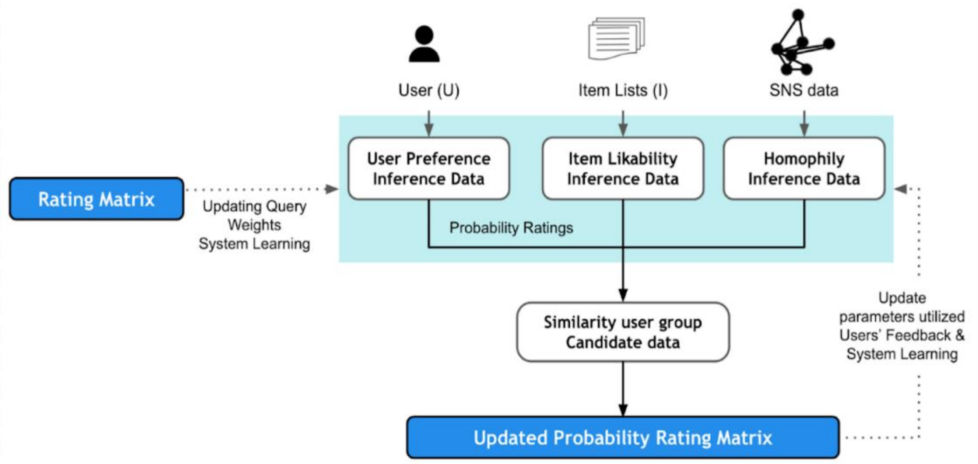
Figure 8 - https://jamesmcallisteronline.com/disappear-people-you-may-know/

Some researchers put out the idea of a neighborhood-based recommender system that provides Twitter users with recommendations in the form of URLs. As a proxy for URL relevance in this case, hashtags are employed, and the similarity between two hashtags is determined using the Euclidean, Cosine, Jaccard, and Dice coefficients.

Online social network services (SNS) like Facebook, Instagram, Twitter, and LinkedIn are enormous digital social networks that allow users to connect with one another in addition to lifelogging their everyday activities, interests, and so forth. A significant increase in user-related data coincided with a significant growth in the use of SNS. It is feasible to gather content information from postings made by users on social networking sites. Additionally, user assessment information can be gathered, along with rating information, this includes several kinds of feedback information, likes and comments. The data collected are available to use in Recommendation Systems for other businesses in addition to SNS, as diverse data obtained through SNS are directly tied to the development of Recommendation Systems.

According to the online user's friend list close to his friend's list activity of other users, collaborative filtering is primarily used for suggesting friends or products. Collaborative filtering's weaknesses, such as cold start, sparsity, and gray sheep, must be addressed, and Model-Based Collaborative Filtering for SNS Recommendation System research has been done for this reason. The figure represents an overview of an Hybrid Recommender engine of overall social network service.

Table 4 - Ko et al., 2022



4.3. TOURISM SERVICE

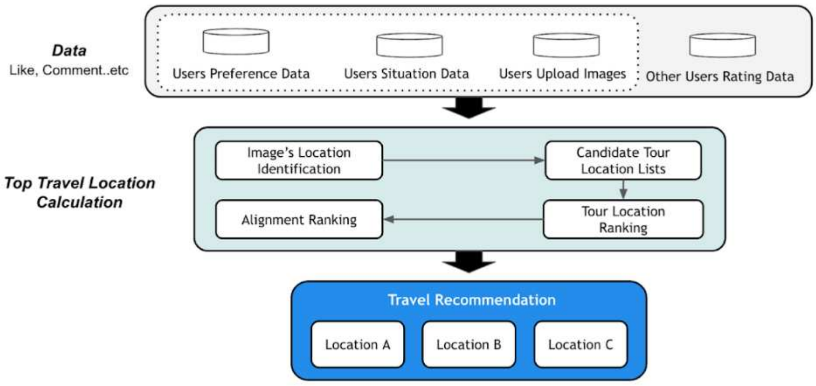
Then we have an entire category of recommendations for service-based systems such as travel services, recommendation of houses to buy or rent and matchmaking services. As the need for travel has grown, recommendation systems are now being utilized in the tourism industry to suggest travel routes, tourist attractions, and means of transportation and a lot of different suggestions. Research on Recommendation Systems utilizing SNS has increased in the field of tourist services because the travel-related recommendation system incorporates situational data, including review and location data, user location, time, and weather, gathered through SNS. Tourism services can use SNS as a dataset for recommending tourist destinations and itineraries because it stores the user's check-in information and the location of the post that was uploaded by the user. The trip information is tailored to the user's preferences after the analysis of these

SNS data by the travel RS, enhancing user happiness and tourism loyalty and enticing users to return.

Sun et al. (2019) used spatial clustering to identify significant tourist destinations using geo-tagged pictures posted on Flickr, a photo-sharing SNS. The rankings were subsequently arranged in ascending fashion, and routes were suggested to users by taking into account both the popularity of places to visit and route length data through the use of machine learning. By assessing empirically the effectiveness of the model suggested, it was demonstrated that by taking into account the photograph captured on the road, the point of interest (POI), and other factors, users can be advised the best tourist itineraries that matched their preferences.

On the other hand, some studies incorporate the current situation data of users in the recommendation system model for the tourism service industry in addition to the assessment data of tourists.

The figure represents an example of the hypothetical structure of a RS for e-tourism, with the aim of suggesting famous landmarks and the best routes. We can cite some other practical



applications of RS in the e-tourism field, such as the Talking Museum Project, which makes use of the Internet of Things to let museum artifacts employ multimedia tools to discuss their histories. Based on the visitor's interests, the museum items' stories are communicated to his/her mobile device while they are there. A subsequent visit is then suggested and organized using the information from the user profile that was gathered.

Another mobile-based recommender system called iTravel was created to give travelers recommendations for on-tour attractions. The principles of CF and mobile peer-to-peer communication were merged in this system. Three data sharing mechanisms for users to trade their ratings of attractions they have visited were presented in order to take advantage of the information of other travelers with similar interests in mobile tourism.

In conclusion, several recommendation approaches are used in e-tourism recommender systems depending on the complexity and specifications of the things they suggest. CB and CF approaches are typically used to promote relatively straightforward goods, such restaurants. KB and hybrid recommendation approaches with domain knowledge are used to recommend more complicated objects, such as transport routes and timetables. Context awareness-based techniques are used to

Table 5 - RS models and Techniques - Ko et al., 2022

Tourism Service	RS Model	RS Techniques
Tourist Attractions or Tourist Information Recommendation	CB	Clustering
	CF	Text Mining
		Clustering
		Matrix Factorization
	Hybrid System	Text Mining
		Clustering
Tourist Route or Transportation Recommendation	CB	Clustering
	CF	Text Mining

suggest products with urgent needs, such as gas stations. The table presents different RS models and techniques for different tourism services.

4.4 HEALTHCARE SERVICE

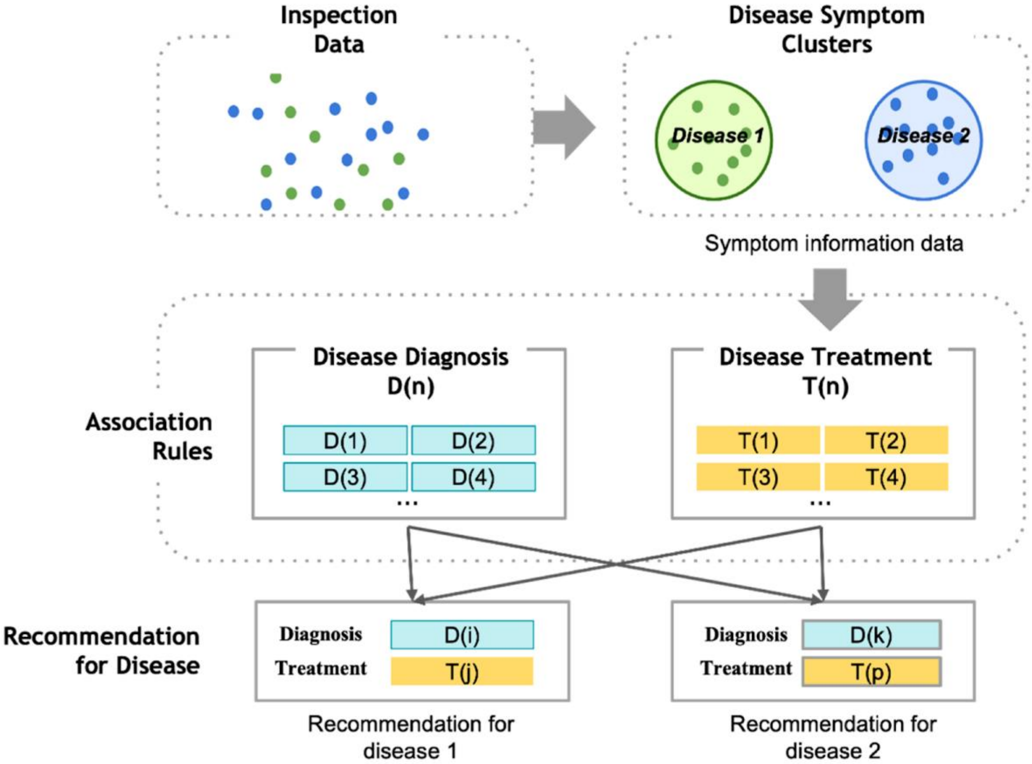
Since the technology has become compatible with smartphones and has become more convenient to use, as interest in health has grown, the number of users of smart wearable devices has begun to rise. These wearable technologies effectively track the user's bio-state. A typical wearable, the Smart Watch, measures the user's body data on a regular basis, enabling self-diagnosis, and aids users who do not have specialist medical expertise in disease prevention. These wearable devices gather a sizable amount of user biometric information to aid in research on diseases or accurate diagnosis through particular scenarios. In addition, it has proved advantageous for research that suggests different kinds of treatments. Health-related RS studies examined the connection between disease patterns and patient symptom patterns to give users information into more effective treatment alternatives.

The major objective of health Recommendation Systems, which aid users in receiving professional care, is to provide appropriate treatment approaches in accordance with the symptoms of many types of diseases and the stages of each condition. In order to achieve this, the health Recommendation System examines the patient's data and the features of the condition, provides the patient with an accurate diagnosis of the disease, and suggests a suitable course of treatment in light of the diagnosed ailment. The analysis of the patient's information and the features of the patient's sickness is required in this regard, leading to the widespread use of the Content-Based Filtering model. The research by Duan et al. (2011) is typical of studies on Recommendation Systems that employ the Content-Based Filtering approach. The goal of Duan's study was to suggest a patient care management strategy that would combine item ranking with clinical decision support, nurse education, and impression quality control. A "disease diagnosis and treatment recommendation system" (DDTRS) that suggests an appropriate disease diagnosis and treatment plan was proposed in a study by Chen et al. (2022), that classified a patient's disease into stages. Large historical disease diagnostic and treatment data sets were collected from patient examination reports and grouped using the DDTRS. Additionally, by using the association analysis method, the similarity of the disease's content information and diagnosis and treatment data related to the user were searched, and a precise disease diagnosis and efficient treatment strategy were advised.

Effectiveness, chronological effect, economy, non-harmful side effects, and patient satisfaction were used as indicators for the assessment of the standard of treatment recommendations made by DDTRS. Evaluation of the indicators was done using physician comments. Comparing the four indexes, the efficacy index had the highest average value of 4.33. The testing findings showed that the suggested system offers top-notch therapy recommendations.

The figure represents the hypothetical structure of a RS for the diagnosis and treatment of a disease.

Figure 10 - Recommendation process in HRS for diagnosis and treatment of a disease - Ko et al., 2022



Additionally, utilizing techniques like Collaborative Filtering, the area of health recommendations might advise other patients with comparable symptoms to the patient to pursue similar treatment options. After determining the degree of similarity between patients using the fuzzy clustering technique, a medical diagnosis is made using the relational information obtained in this manner.

A medical diagnosis suggestion system that divides patients into groups based on the traits of their diseases was suggested on the basis of the relational data obtained in this manner. In this study, the Collaborative Filtering suggestion model was utilized to diagnose the patient by examining the medical diagnosis record of the matching patient group after identifying a patient group with disease characteristics similar to the diagnosed patient.

Studies have recently been conducted on the topic of recommending an appropriate diet in response to patient information such as disease, age, gender, and weight, individual nutritional expectations, and food preference in the field of specialized treatment, in addition to diagnosing a patient's disease and providing treatment based on professional medical knowledge.

The research conducted by Iwendi et al. (2020) suggested solutions based on deep learning that make use of different Neural Network approaches, like Logistic Regression and Recurrent Neural Network, to share information among patients with the same illness and advise which type of diet they should follow. The model performed exceptionally well all around.

The use of user-specific health data in the context of e-health services, on the other hand, enables the user to autonomously regulate their health data. In the area of e-health services, the focus is on enabling users to access both personalized health information and general health information at anytime and anywhere, providing a variety of material delivering the health data needed by users without having to go through a specialist.

Table 6 - Ko et al., 2022

Healthcare Service	RS Model	RS Techniques
Medical Treatment or Diet Recommendation	CB	Text Mining
		Clustering
		Neural Network
	CF	Clustering
	Hybrid System	Neural Network
Health Information Recommendation Using E-Health	CB	Text Mining
		Neural Network

The table presents the most important RS models and techniques for different healthcare service fields.

4.5 EDUCATION SERVICE

Since more than a decade ago, e-learning has grown in popularity in educational institutions. These systems are designed to give users learning materials depending on their learning activities and preferences. The user can find a wide variety of knowledge sources and information in e-libraries, also known as digital libraries, which are the sources of e-learning. Indeed, a new educational trend known as *Smart Learning* has emerged from the traditional style of education in the classrooms or lecture halls to e-learning, in which learning is conducted through an online environment. Due to the COVID pandemic as well as the growing use of wireless networks and various smart gadgets in education, smart education has begun to be implemented gradually. Smart education may smoothly provide individualized learning adapted to the requirements, goals, talents, and interests of learners without being constrained by time or geography by accessing huge digital resources. Throughout the whole e-learning life cycle, learners' needs must be taken into account while creating tailored learning e-environments. It is also stated that personalized support through recommendations in Web-based educational systems can be designed and evaluated using the e-learning life cycle.

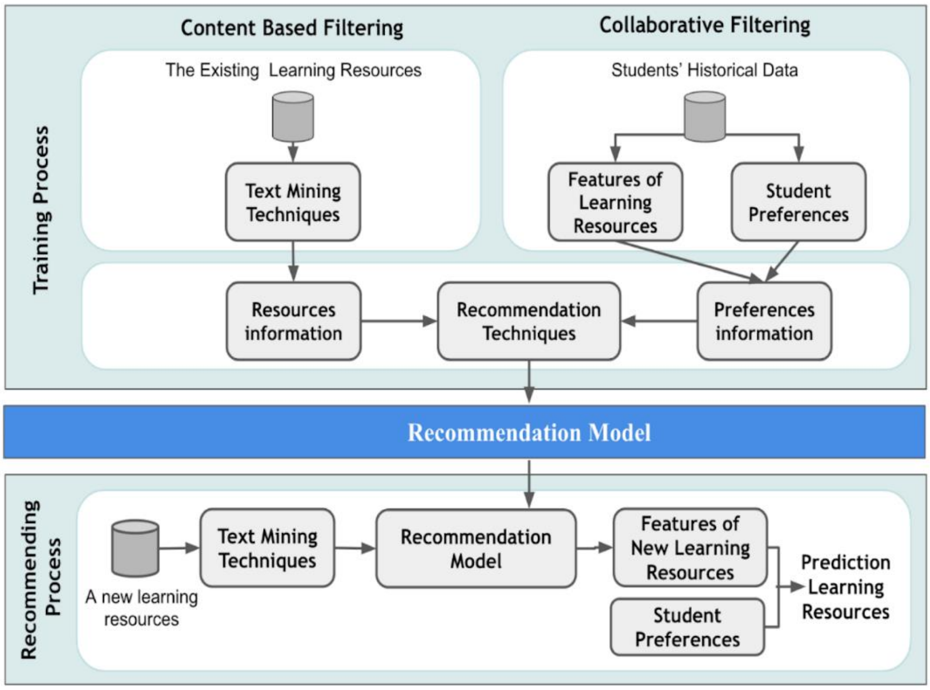
An e-learning recommender system's architecture typically consists of three components: 1) collecting student profiles and identifying their individualized demands utilizing web analytic tools; 2) gathering the metadata of educational goals to identify the features; and 3) getting relevant pedagogical information to assess the degree of alignment between learners and learning objectives.

In order to give learners an effective and efficient learning experience, the area of education services uses Recommendation Systems to deliver learning resources that take into account the students' learning styles and levels of expertise. In other words, learners can receive content that is specifically tailored to them.

The Content-Based Filtering recommendation model was primarily utilized in the study of proposing learning content suited for learners with an emphasis on the similarity between learners and learning objects after assessing the learner's profile information and learning object information. A study by Shu et al. (2018) made use of a Content-Based Filtering recommendation algorithm, which combines learner preferences with text data from learning resources to deliver learning materials to students at the proper level.

The figure represents the hypothetical structure of a RS for recommending educational content.

Figure 11 - Ko et al., 2022



By taking into account the learner's sequential learning patterns, Tarus et al. proposed an online learning resource RS based on a Knowledge-Based

Collaborative Filtering model. They used the ontology technique and the Sequential Pattern Mining (SPM) algorithm to determine similarities between users based on the learner's knowledge and learning resources.

The smart education system in the e-learning environment has been the subject of several studies in the field of education services. These studies covered not only a system that recommends learning content appropriate for learners but also a system that recommended college preparation topics and course choices so that students could receive a more thorough education.

Esteban et al. laid out a hybrid multi-criteria lecture recommendation algorithm that combines a Collaborative Filtering model that takes into account student information, such as student ratings and grades for university courses, and a Content-Based Filtering model that takes into account lecture information, such as professors and lecture contents, in order to be able to choose a suitable lecture. It has been demonstrated that recommendations utilizing numerous criteria are more successful than those using just one.

The table presents the major RS models and techniques for different education services.

et al., 2022

Education Service	RS Model	RS Techniques
E-Learning and Customized Learning Recommendation	CB	Text Mining
		Neural Network
	CF	Text Mining
	Hybrid System	Text Mining
		Clustering
Education Course Recommendation	Hybrid System	Text Mining

Academic Information Service. In addition, academic scholars have to dedicate a lot of time and energy in finding academic material in their area of study considering that the volume of academic information is growing continuously. Research on Recommendation Systems has been done in the area of academic information services to provide data and tools that can aid scholars in their research.

The Digital Library, an information collecting system that enables users to quickly and conveniently search for and utilize diverse digital assets around the world, is a typical service to which the Recommendation System in the academic information sector is applied. In particular, the Recommendation System is regularly used by the University Digital Libraries (UDL), a service that promotes academic learning, research, and education. The fundamental goal of RS research in the area of academic information is to support the research itself while also recommending and providing academic information appropriate for a range of users, including scientific communities, research institutions, and development practitioners.

A Content-Based Filtering recommendation model through text similarity was proposed as a UDL integrated Recommendation System. In interdisciplinary resource-related tasks contained in UDL, the system gives outstanding recommendation performance and promotes the transmission of user and resource information.

In addition, a study was carried out to offer a list of publications or conferences that would suit research. As an example, Wang et al. (2015) suggested a system that uses the thesis' abstract to suggest a list of journals or conferences with relevant material. The proposed method was tested, and the results showed that it can recommend the best journal or conference for study in an average of roughly 5 seconds with an accuracy of 61.37%.

4.6. E-COMMERCE SERVICE

In the past, consumers primarily purchased goods like clothing, food, and books from offline retailers. The way that people consume goods has altered recently, though, because of the emergence of digital platforms like the web and applications such as Amazon, eBay, and Alibaba. Consumers can find a wide range of goods and options through e-commerce, which also gives retailers a simple way to sell their products. Customers in particular were unable to patronize offline stores because of the COVID-19 lockout procedures that prevented them from going outside. As a result, consumption on digital channels rose dramatically. Additionally, the types of goods sold on digital platforms started to expand.

There are indeed numerous Recommender Systems available for e-business applications.

Business-to-consumer (B2C) systems, in general, concentrate on making suggestions to specific customers, whereas business-to-business (B2B) systems seek to provide recommendations regarding products and services to business users. Recommender systems for B2C applications are referred to as e-commerce/e-shopping systems. These B2C e-commerce companies offer recommendations for products based on the best-selling items overall, customer demographics, or an examination of prior consumer purchasing patterns to forecast future behavior.

Recommender Systems are being made available to mobile users as mobile-based recommender systems, in addition to being available to Web users due to the growing use of mobile phones and the advancements in wireless networks.

The e-commerce service actively uses data in a Recommendation System and actively gathers information about diverse users in order to grow its business. By examining auxiliary user data like gender and age group, the service can forecast user preference. It is employed for item recommendations based on user preferences. Additionally, research is being done on how to harness the Virtual Community (VC) offered by each business to gather user-shared thoughts on reviews or products that reflect their subjective ideas. The recommendation System currently employs tracking data that is produced by tracing the user's mouse and keyboard movements while utilizing the service.

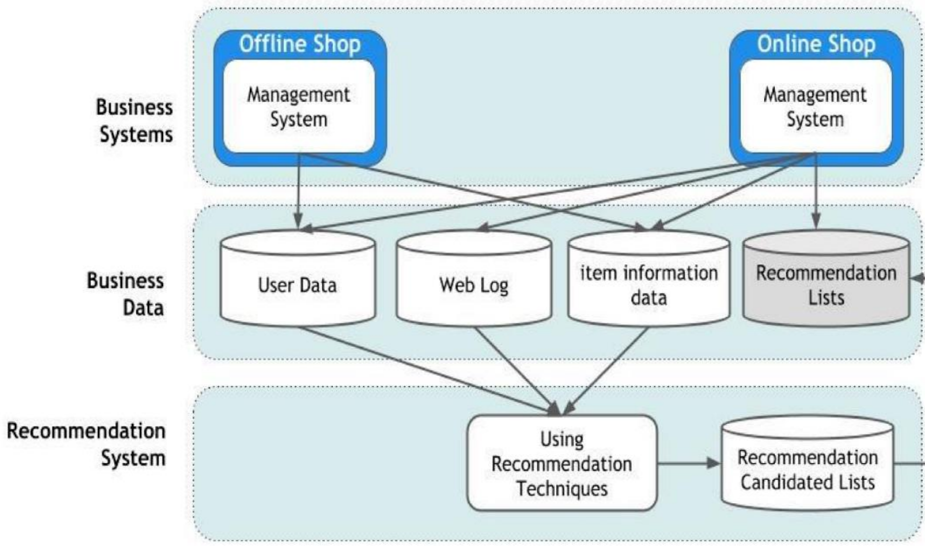
By observing how users interact with browsers and applications to ascertain their purchase intentions, mouse and keyboard data can be used to assess user preferences.

Therefore, it recommends the item that captures the user's preference utilizing the information gathered from the VC and the Recommendation System. Additionally, it is feasible to suggest a product to a user based on the tastes of a group of other users who share the same likes as the user.

The most important aspect of the online shopping experience is that customers typically display a pattern of spending that complements the goods they have previously chosen or purchased. As a result, locating things that are comparable to items the user has already bought can assist in recommending items that are right for the user. Because of this, the Collaborative Filtering and Hybrid recommendation models are mostly utilized in the service.

The figure represents the structure of a RS utilized for recommending products with data provided by online and offline stores.

Figure 12 - Ko et al., 2022



Specifically, a variety of hybrid recommendation models that advocated using all of the user preference analysis data gathered from VC's reviews, comments, and reviews were used. In a study, Zhang et al. (2018) suggested a 'Hybrid Probabilistic Matrix Factorization' model-based recommendation system that models user preferences using user auxiliary information and items through a Neural Network and separates keywords from text information about items. This method forecasts consumers' ratings of things while also taking into account their individual preferences and emotional characteristics.

By combining neural networks and information from several sources to calculate the weight of recommended items and examine the needs of mobile e-commerce customers, researchers created a system to recommend products acceptable for consumers. Every year, research into RS for use in e-commerce is growing quickly. Recent research has also been done on cutting-edge suggestions for the digital platform's user interface. The user's degree of interest rapidly declines as more products are advertised online. In order to improve user engagement, research is being done on how to set up an effective interface. In the study by Sulikowski et al., the "Evaluation of a Recommending Interface" (PERI) framework was used, and for a predetermined amount of time, equivalent performances were demonstrated regardless of the vertical and horizontal directions of the interface. However, the vertical layout's recommended performance more than doubled in terms of purchase commitments.

Offering a vertical layout and a slow-blinking visual appearance also served to further heighten the user's interest in the object.

In conclusion, e-shopping Recommender Systems - both Web and mobile-based - are typically used for online purchases of both physical and digital commodities.

Table 8 - Ko et al., 2022

E-Commerce Service	RS Model	RS Techniques
Web	CF	Text Mining
		Clustering
	Hybrid System	Neural Network
Mobile	Hybrid System	Neural Network

In order to put their innovative algorithms to use, academics have created a number of successful online shopping platforms.

These systems offer developers instructions on how to effectively design recommender systems for online buying.

The table represents some e-commerce service field Recommendation Systems.

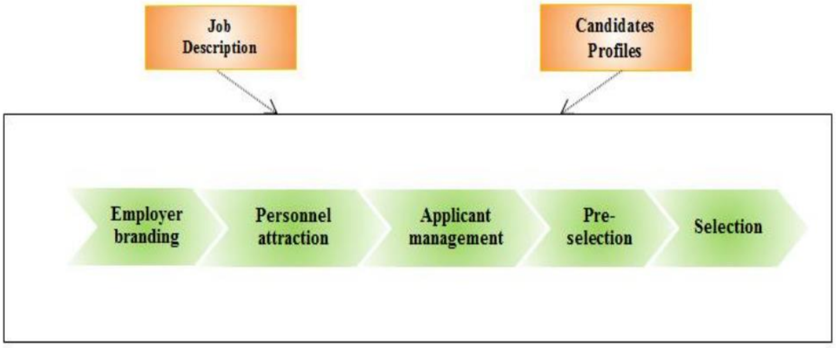
4.7. JOB MARKET

A key component of human resource management is the hiring procedure, which treats labor as one of the key contributors to production and a source of value for the company. The main goal of the hiring process is to bring on applicants who will be beneficial to the business.

There are two distinct points of view: those of job seekers and recruiters. The prerequisites and restrictions on abilities, expertise levels, and degrees are determined by the recruiters, who then create the job description. On the other side, a job seeker creates a CV by outlining their educational background, prior employment history, and talents. From attracting and finding talent to selecting and keeping applicants, IT support is provided for recruiting efforts. The complexity of using e-recruitment solutions is represented by the degree of process integration.

The attraction phase and the selection phase are the two primary stages of the hiring process, as shown in the figure below.

Figure 13 - Phases of the recruiting process - Al-Otaibi, 2018



The objective of the attraction phase is to create job position descriptions. Pre-screening of resumes and other submitted materials ushers in the selection phase. The remaining group of candidates who were not eliminated during the screening process is then compared to make the final selection of candidates. The communication with applicants, the maintenance of applicant data, and related procedures like forwarding applications to organization members who are involved in the selection decision comprise the applicant management, which serves as a secondary function.

The subject of how employee recruiting could use the Internet to improve job seeker and vacancy matching was brought up from the beginning of the commercialization of the internet in the late 1980s.

One of the most influential e-business innovations, the online recruiting platform, also known as e-recruitment platform, altered how businesses hire applicants. These platforms have grown in popularity recently since it can be difficult for businesses to find the right applicants in some skill sets, and this problem has long been recognized as a major barrier to business success. Plus, from the economic point of view one of the major causes of unemployment is given by the timing problems in matching the right candidate to the right vacant job offer.

This usage of online platforms for the hiring process has been fueled by online channels like social networking applications, online job portals, and company career websites. These portals are used by employers to post job openings, and by job seekers to post their resumes. Companies receive thousands of resumes for each job that is posted. As a result, there are now a ton of job descriptions and resumes for candidates online. Due to the limited search functionality in recruiting applications, this enormous amount of data presents a huge opportunity for improving the matching quality. As a result, a lot of candidates pass up the chance to get hired. This search kind is insufficient for creating a good fit between candidate aptitudes and job requirements, according to actual practices and theoretical considerations.

Applying Recommender System technologies that can assist recruiters in effectively managing this information is becoming more and more necessary. Numerous studies have been done to examine various problems with recruiting as well as the use of RS technologies. The field of job recommendations is still complex, and research in this area is expanding.

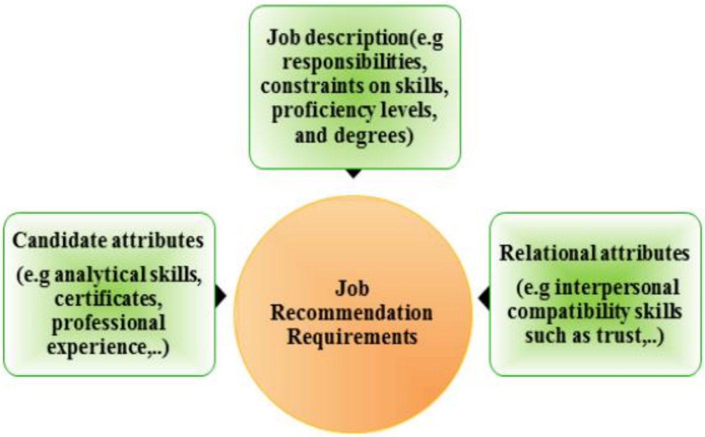
When recommending individuals for a certain job, there are key criteria that should be taken into consideration that are described in the literature.

1. The skills and talents that people should possess are used to match people to employment.

2. People recommendations are a two-way process that must take into account the preferences of both the candidate and the recruiter.

3. Recommendations should be based on the candidate's qualifications as well as the interpersonal factors that affect how well the candidate will get along with the other team members.

Figure 14 - Requirements for making a job recommendation - Al-Otaibi, 2018



4. People are thought to be unique; we cannot pick the same individual repeatedly, as in a play or a novel. The issue with employment recommendations is that they must be given both ways between employers and candidates.

A candidate's profile typically consists of three parts.

1. The employee's name, last name, and home address are examples of personal information.

2. Information on the professional jobs the candidate has had in the past and now. Companies, job titles, job descriptions, start and end dates for jobs, and company names could all be found in this area.

Additional firm details, such as the number of employees, may be included in the company description section.

3. Details about educational experiences, such as start and end dates, university names, degrees, and educational subjects. Candidates may also be requested to assess the job profiles on a scale of 1 to 5 for collaborative purposes.

Numerous attributes can be derived from these meta-data to train and test recommendations. Conversely, the job profile should be created to outline the qualifications and a list of all pertinent abilities that a candidate for this position should possess. Additionally, statistical accuracy measurements like the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or Correlation computations can be used to evaluate the effectiveness of the Recommendation System.

4.7.1. Job Recommender System: the application

Here I present the results and outputs of a basic empirical application in RStudio of a content-based Recommender System for Job seekers.

The datasets considered are two: one containing 2527 job offers in the data science field. Each job posting presents data such as: job title, job description, minimum and maximum salary, location, information about the company (name, rating, industry etc.) and requirements such as specific degrees.

The second dataset presents candidates' resumes, comprehensive of hard and soft skills, previous experiences etc.

The first steps consist of uploading the data and subsetting the dataset in order to consider a smaller-sized set (of only 500 job postings, to avoid computational problems).

After cleaning the text, removing unnecessary punctuations and symbols, putting words into lower case and removing excessive blank spaces and all the coherent steps needed, I proceeded with the tokenization.

The matrix contains the term frequency for each job description in the sample, and the terms are weighted using the *term frequency-inverse document frequency (tfidf)* formula. The tfidf offsets the number of documents in the corpus that contain a word from the number of times it appears in a document. Since a term can merely appear more often if a document contains more words, doing this prevents terms that simply occur more frequently than others from being mistakenly judged to be significant.

The base of the recommendation generated lies in the **cosine similarity.**

I basically calculated the cosine similarity between the attribute “job description” for each job offer and the candidates’ CV. In fact, the cosine similarity calculates the cosine of the angle formed between the two vectors considered: so, the closer the cosine of this angle is to 1, the closer these two objects will be (the more similar they will be). The figure presents only a few of the cosine similarities calculated between the vectors. The matrix presented below has 501 rows and 501 columns: the first 500 elements are drawn from the job descriptions, while the last one contains the candidate’s resume. So, what interest us for the job recommendation is actually the last row, where the cosine similarity between job description and candidate’s CV is computed.

Table 9 - Personal elaboration on RStudio, part of cosine similarity calculated. The last row presents the cosine similarity between the jobs descriptions and the candidate's resume.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V489	0.024307888	0.016543999	0.03456521	0.01922848	0.010256543	0.01738715	0.044339738	0.015543170	0.034216732	0.023456936	0
V490	0.027082840	0.039436807	0.06192381	0.01545638	0.021180950	0.03320719	0.017196148	0.028918774	0.063241986	0.013655486	0
V491	0.043565672	0.044845710	0.04199906	0.03757339	0.047934517	0.05277645	0.052967982	0.060317224	0.048820631	0.051528671	0
V492	0.039720907	0.043336902	0.04264100	0.04781707	0.023195206	0.13843481	0.053854678	0.079369705	0.053212756	0.049454441	0
V493	0.038237502	0.033503771	0.03809454	0.03621828	0.053377820	0.03553287	0.030652847	0.035672259	0.053002341	0.030577190	0
V494	0.048268281	0.047035222	0.03933038	0.04191337	0.039586194	0.05317839	0.031677505	0.053312745	0.049532747	0.023654511	0
V495	0.076637080	0.063286579	0.04816499	0.04972589	0.042547576	0.06316750	0.031605757	0.063359044	0.073714736	0.032345693	0
V496	0.003567568	0.001311204	0.00000000	0.00000000	0.000000000	0.03289225	0.002084791	0.012488299	0.003162832	0.001547241	0
V497	0.035618801	0.028478082	0.03293475	0.01613158	0.024048416	0.02976409	0.016120021	0.042006656	0.018983431	0.024510123	0
V498	0.038177512	0.021125985	0.03750789	0.02509666	0.041937096	0.03268305	0.043742661	0.025003307	0.024620294	0.026307590	0
V499	0.085182318	0.063386296	0.06134130	0.06565892	0.091369042	0.07838714	0.066461068	0.079912690	0.107813773	0.023633662	0
V500	0.003567568	0.001311204	0.00000000	0.00000000	0.000000000	0.03289225	0.002084791	0.012488299	0.003162832	0.001547241	0
V501	0.164370958	0.082954646	0.15423666	0.13087848	0.143665855	0.11946278	0.105970983	0.078826650	0.115812666	0.095338980	0

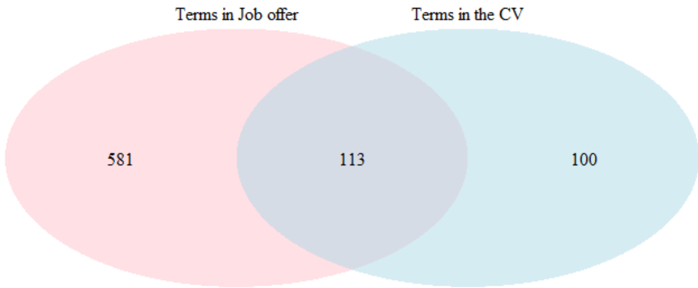
From this point on, the elaboration of the recommendation is smooth. After rearranging the job description in a descendent order (based on the cosine similarity) a function is displayed that results in three outputs: the corresponding recommendation, a Word Cloud and a Venn Diagram.

Since we have a descendent order, by running "rec(1)", we will receive the output predicted for the first recommendation associated with that specific resume. So, we can easily detect the job that would be the best fit for that specific candidate considered.

In the console the job recommended appears as follows:

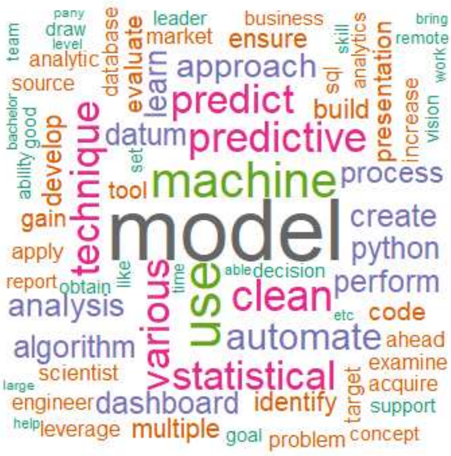
```
> rec(1)
# A tibble: 1 × 8
  job_title        min_salary max_salary city  state company_name company_industry
  <chr>                 <dbl>      <dbl> <chr> <chr> <chr>        <chr>
1 Machine Learn…       122998     151838 San … CA    price.com    NA
# ℹ 1 more variable: company_rating <dbl>
```

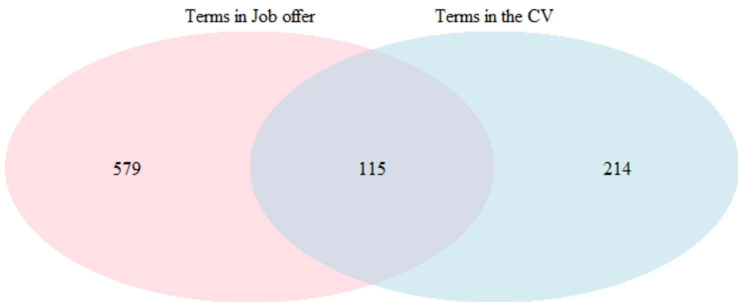
Additionally, we obtain a WordCloud and a Venn Diagram for the first recommendation. Both of these present the common and frequent terms between the job offers and the terms in the CV.



I did this also for the second job recommendations:

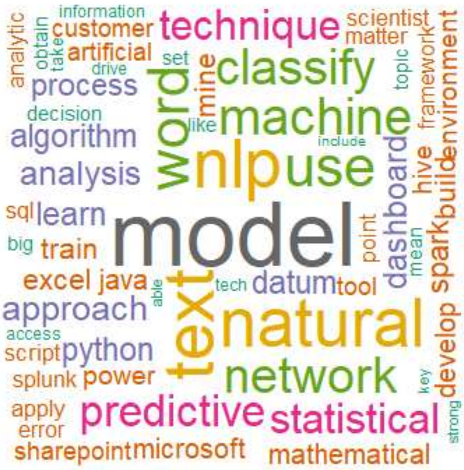
```
> rec(2)
# A tibble: 1 × 8
  job_title       min_salary max_salary city  state company_name company_industry
  <chr>                <dbl>      <dbl> <chr> <chr> <chr>        <chr>
1 Data Scientis…       95925     152004 Oakl… CA    Center for … Oil, Gas, Energ…
# ℹ 1 more variable: company_rating <dbl>
```

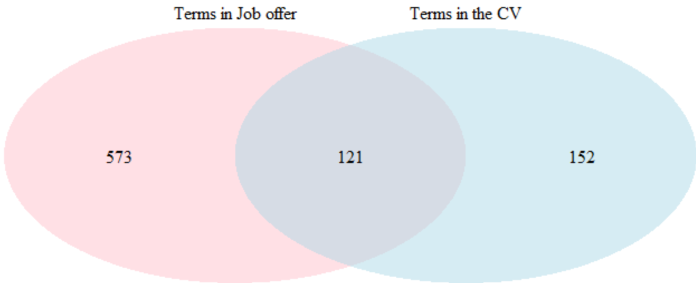




And also, for the third:

```
> rec(3)
# A tibble: 1 x 8
  job_title        min_salary max_salary city  state company_name company_industry
  <chr>                 <dbl>      <dbl> <chr> <chr> <chr>        <chr>
1 Data Scientist           NA         NA Palo… CA    Landing AI   Information Tec…
# i 1 more variable: company_rating <dbl>
```





This is a very basic application of a Job Recommender System, especially since in the last decade there has been an impressive increase in the application of neural networks and deep learning in the RS field. The algorithms actually utilized by businesses are much more complex, yet this application demonstrates practically how cosine similarity can be utilized to determine the closeness between two attributes taken into consideration for the recommendation.

This application draws the attention to the future possibilities for improvements in the effectiveness of the "matchmaking" process of the Job Market, which is nowadays very critical and the main source of unemployment. The hope is to further automatize in the future part of the hiring process (at least for a first screening of the candidates / companies), to reduce the search time and costs of finding the right job offer or the right candidate.

V. RECOMMENDER SYSTEMS AND THEIR BUSINESS VALUE

Joe Pine makes the case in his book *Mass Customization* that businesses must transition from the traditional model of mass production, where standard goods, homogeneous marketplaces, and long product life and development cycles were the norm, to the brand-new world, where differentiation and personalization take the place of standard items. Pine contends that producing just one product is no longer sufficient. Companies must be able to create a baseline of different items that respond to various consumer demands. The shift toward online shopping has made it possible for businesses to provide customers additional choices. However, by going to this additional degree of personalization, companies add more information for customers to comprehend before they can decide which products best suit their needs.

The use of RS is one remedy for this information overload issue. E-commerce websites employ RS to make product recommendations to their users. The products can be suggested based on the top sellers on a website, the customer's demographics, or an analysis of past purchasing patterns that predicts future purchasing patterns. In general, these methods are a part of website personalization because they enable the website to adjust to every user. By automating web personalization, RS makes it possible to cater to each customer specifically.

Today's websites and online services frequently include automated recommendations, which are frequently a key component of the entire user experience, such as on e-commerce and media streaming websites. As a result, what we see online is frequently highly customized. The recommendation service provider, which customizes its own material for users, typically plays a significant role in determining the personalization strategy that is used.

How to keep consumers satisfied with personalized items and services, while achieving economically motivated goals like conversion rates, improved revenue, or client retention, is the problem. In cases of information overload, one frequently expressed objective of a Recommender System is to assist users in finding items of interest.

This is in accordance with literature that is frequently judged on its capacity to precisely forecast the relevance of certain items for particular users.

There are at least two different types of stakeholders for any Recommender System: the organization that offers the recommendations as part of their service (i) and the consumers or users (ii) who receive them. It's crucial to distinguish between these stakeholders. Even though a Recommender System should, in the best case scenario, generate value concurrently for all relevant stakeholders, there may be conflicting objectives at play. For example, the advice that helps users discover new items may be the best for the consumer, but it may not be the best for the supplier in terms of short-term revenue generation. So, it is fundamental to take into consideration all those actors that influence and are influenced by the architecture of the RS. Indeed we need to consider not only consumers and suppliers, but also RS providers and society at large.

Talking about users or customers, we can say that thanks to the increasing development of these recommender engines, there is an opportunity for "subtle personalization," in which the website offers the user a completely organic individualized experience. This is made possible by recommender system algorithms that employ a variety of sorts of data. The customer uses the website exactly how he would have done without personalisation; he does not have to express his likes or desires on the website explicitly. Without his knowledge, the website slowly alters the interface in almost undetectable ways to give him a more customized experience.

Rather than being employed as marketing tools, Recommender Systems are now more often deployed as **virtual salesmen**. The majority of RS in use today are "buy-side" systems. In other words, they are made to act on the customer's behalf while picking which things to buy.

Modern marketing, at the same time, seeks to provide the greatest value for both the company and the client at the same time. The E-commerce site might offer each product at the price that optimizes the lifetime value of the consumer to the site by using the Recommender System's ability to create an indication of the client's price sensitivity for a specific product. For instance, one consumer might be willing to buy the goods at a cost that would net the store a profit of 20 cents, while another might buy it at a cost that nets them a profit of $1. Using technologies like these to analyze consumer data to determine ways to increase revenue from customers raises difficult ethical considerations.

In a related idea, "sell-side" Recommender Systems might let companies choose which customers to target with special offers.

Three approaches exist where Recommender Systems could increase e-commerce sales:

“From browsers to buyers”: Many website visitors browse the site without ever making a purchase. Customers can find the items they want to buy with the use of recommender systems.

Cross-selling: By urging customers to buy additional products, Recommender Systems enhance cross-selling. The average order size ought to rise if the recommendations are sound. For instance, a website may suggest further items during the checkout process based on the items you already have in your shopping cart.

Lastly, **gaining customer loyalty** is a crucial part of any business strategy in a world where rival websites can be reached with only a few clicks. By fostering a relationship between the website and the user that adds value, Recommender Systems increase client loyalty.

Websites make an investment in user research, operationalize that research using RS, and deliver tailored user interfaces that satisfy user requirements. Customers reward these websites by visiting the ones that best meet their demands again and again. Customers become more devoted to a website the more they use the suggestion system and teach it what they want. Even if a rival built exactly the same capabilities, the customer would still have to spend a lot of time and effort teaching the rival what the company already knows. Additionally, building relationships among clients can boost loyalty.

Customers will come back to the website that suggests users they might want to get in touch with.

According to some researchers, RS increase revenue by, as we just witnessed, turning browsers into purchasers, boosting cross-sell chances, and cultivating client loyalty. It is more difficult for customers to select the product that best suits their likes and needs because of the amount of items and product-related information available online, which raises the cost of fit. Online Recommender systems can assist customers in sifting through the vast selection of products accessible to find those that are relevant to their present interests, hence cutting the cost of processing product-related data. This viewpoint predicts that the strength of recommendations will have a beneficial impact on the sales of the recommended items.

The impact of the strength of suggestions on sales, however, is also significantly influenced by the legitimacy of Recommender Systems.

The ability of Recommender Systems to gather enough data to generate useful recommendations for new users is one of its drawbacks. Sharing user data across websites is one method for speeding up the shift. Users benefit from shared information because they receive more accurate recommendations faster, while individual websites suffer since users are less devoted to them. Websites have no motivation to exchange information with rivals because they own the data they gather.

Groups of non-competing websites could nevertheless establish with the intention of sharing data to boost the value to businesses inside the consortia. Customers of these consortiums will require guarantees that their privacy will be carefully safeguarded, even though their data will be shared outside of the confines of a single site.

Additionally, recommendations only have an impact on consumers' choices when they are viewed as credible and objective. Customers are likely to doubt the validity of online Recommender Systems because of the possibility of retailer manipulation (i.e., recommendations that differ from the results produced by the collaborative filtering algorithms), since retailers have complete control over what recommendations to make and how to present them.
Anecdotal evidence of retailers faking the results of Recommender Systems furthers this notion.

But how can this mass personalization be done with RS? Joe Pine enumerates the essential techniques for obtaining mass personalization. Using Recommender Systems, the first four of these objectives can each be accomplished:

1. "Customize services around standardized products and services" (Pine, 1994): RS offers a customized service that enables E-commerce companies to more effectively sell their primarily commodity products.

2. "Create customizable products and services" (idem): The e-commerce site's RS are a customisable product.
3. "Provide point of delivery customization" (idem): For the Ecommerce site, the RS directly customizes the point of delivery.
4. "Provide quick response throughout the value chain" (idem) They anticipate that RS will be utilized in the future to forecast product demand, allowing for early supply chain communication.

A vital component of automating bulk customization for e-commerce sites is Recommender Systems. Future firms will place more emphasis on the long-term value of customers to the company, making them more significant. The worth of a customer to an e-commerce site will be maximized by offering the pricing and services that they believe will result in the best possible connection with the consumer. This relationship will frequently be advantageous to both the consumer and the site, but not always. For the sites, customer retention will be particularly crucial. Finding a balance between the value of recommendations to the site and to the customer is providing and will provide significant ethical problems.

There are numerous methods for putting Recommender Systems into practice, and each method can be applied almost independently of how the Recommender Systems is meant to boost a website's revenue. E-commerce companies can first decide how to boost sales, then decide how persistent and automated they want to be, and then decide on a Recommender Systems approach that suits that description.

Information technologies are being used more frequently by online shops to offer customers value-added services. These services, which aim to reduce consumer search costs and uncertainty related to the purchase of untested items, include online RS and customer feedback methods. Online merchants can now offer new services to improve customer experience and boost sales thanks to internet-based information technologies. Customers have had access to review and rating systems on store websites for a long time, allowing them to contribute and share their feedback about products.

It has been intuitively assumed that making suggestions would boost sales by giving buyers access to excellent, practical information.

The retailer's pricing strategy, which reflects both the quality of the product and the amount of customer service received, mediates the indirect effect of the strength of recommendations on sales.

To represent the value-creating interactions between networks of business actors in virtual markets, different business models have been created.

Efficiency, complementarities, lock-in, and innovation are essentially grouped together as the four value-creating components of e-commerce business models.

Efficiency is typically defined as the decrease of information asymmetries on virtual marketplaces, speed gains, scaling effects, or reductions in transaction costs or time. Efficiencies can indeed create value when combined with Recommender Systems in e-commerce businesses. It can lead to information being delivered in an unbiased way, it can help users by recommending relevant items, matching both their short and long-term preferences. Also, they can assist the decision making process, reducing choice difficulties and they can establish real-time transparency.

When several sales channels or product catalogs are combined, **complementary** effects are produced that create extra opportunities, such as cross-selling of products or follow-up sales offers to customers. The value generated by RS in terms of complementarities can be identified as the help in discovering new items not known by the user, but also as the presentation of the different available options. RS can create additional demand, show accessories products and generate a logical sequence based on previous interactions between the user and the product in question.

Because switching costs or favorable network externalities are avoided, **lock-in** effects add value by keeping customers. This can be displayed as the increase in the user engagement, by providing satisfying entertainment and valuable "add-on" services. As an illustration of such a beneficial network externality, consider how collaborative filtering systems improve their accuracy as they employ more data and input from larger groups of users.

Last but not least, **innovation** refers to fresh architectural arrangements and opportunities that were not possible on non-virtual markets. Novelty is represented by the stimulation of a desirable behavior of the users, but also by the promotion of revenue favorable items.

Different measures for determining the effects of RS in terms of value can be defined: we can talk about Click-through rate (CTR), the adoption and conversion measures; sales and revenue; user engagement; effects on consumption; distribution and many other variables. On the one hand these measures allow us to understand the intrinsic value of Recommender Systems for a business; on the other hand they are sometimes difficult to isolate from other sources of value and to comprehend in their entirety.
There is no doubt that with the one-to-one market of these years, Recommender engines are an essential asset for many businesses, still, there is a lot of room for improvement.

CONCLUSIONS AND FUTURE RESEARCH

The digital economy is rapidly being driven by Recommender Systems, a specific class of AI applications. These, as we saw, are elaborately crafted systems that use sophisticated algorithms to filter through enormous amounts of data, methodically analyzing user preferences, actions, and a wide range of other variables. As a result, they produce suggestions for goods, services, or information that are specifically catered to the requirements and tastes of the users. They serve important roles in many different industries, from suggesting best-choices in the entertainment platform to helping with medical diagnosis, assisting with financial investment decisions, and a whole lot more. Future advances will boost the value of Recommender Systems, which can be a very effective tool in a business's toolbox.

A potential application is the ability to predict seasonal sales based on recommendations, identify significant purchases, and provide customers with better recommendations that can improve retention and brand loyalty. These systems' growing importance is due to their capacity to increase client interaction, enhance decision-making, and increase service efficiency as a whole.

Recommender systems are anticipated to keep developing in the future, offering precise and effective decision-making. Hybrid models that incorporate several recommendation techniques and perhaps human judgment could become widespread. The demand for systems that are able to effectively handle enormous data volumes while protecting privacy will only increase as the volume of data rises and users demand increasingly personalized experiences.

The way that choice can be tailored, wrapped, laid out, and appreciated has undergone a revolution due to Recommendation Systems. But it's important to comprehend this revolution alongside these alternative architectures. The future is framed by it.

Business recommenders such as Amazon or Zara and many others provide helpful guidance on potential purchases. The genuine recommender revolution, however, is more concerned with **who you would like to be** than with what you would like to purchase. An introspective cycle of greater self-awareness is generally encouraged by properly effective, computationally calibrated, and full of information recommenders. And this is granted not only by what people like, but also by what people decide to ignore and not to buy. Better decisions encourage better results, both in principle and in practice.

Aristotele clearly told us that our future is shaped by our choices and not our chances or luck. Is this still valid today? Or do people follow recommendations religiously, as robots? At the end of the day people decide whether or not to follow the suggestions given to them.

The ultimate power of the Recommender Systems goes beyond awareness of ourselves and leads us into self-exploration. Effective recommenders are seen by people as openings to deeper self-awareness as well as personalized counselors. Their counsel serves as a virtual window that encourages and facilitates contemplation and research, which is powerful.

One self-awareness value of RS is the fact that they simply deliver substantially **better choices**. In other words, users immediately realize that they would not have found these better options on their own. Additionally, these recommendation tools also benefit the businesses that hire them when their suggestions pique interest and prompt more research; the urge to find out more about that movie or gain more knowledge about those employers.

Then there are the **network externalities**. These Recommender Systems become of greater value as more people use them, and as the more valuable they become, the more the people will use them. This beneficial loop is accelerated by machine learning skills so that recommendations and guidance are always compelling and current.

These possibilities and chances most accurately reflect what people desire to accomplish next.

Sharper self-awareness results from better decisions and higher empowerment. Just as their recommenders learn more about people, people can also learn more about themselves. What do their music choices, television viewing habits, and book annotations reveal about who they are and what they aspire to be? Self-reflection engines can develop from recommendation systems.

What revolutionary discoveries might surface if one could scan, combine, and evaluate all the recommenders, suggestions, and digital advice one encountered? The machines that learn alongside us, from us, and for us will hold the key to knowing who we truly are and who we truly want to become.

Previous chapters have demonstrated that there are various factors that influence how Recommender Systems add value for their users but in general for different stakeholders. Considering this variety, academic research seems to have a limited scope. First, a lot of attention is paid to computational recommendations that are mainly assessed using past data. Although critical, this accuracy focus only touches on a small number of value-creating aspects, leaving unanswered questions like customer experience or the long-term impacts of sales patterns. Without a doubt, industries have consistently adopted ideas from academia and have effectively integrated and enhanced unique machine learning models to support specific organizational objectives. A notable example is the application of matrix factorization techniques, which were first investigated in the late 1990s, rose to prominence in the context of the Netflix Prize, and were afterwards widely applied in industry.

However, there are a number of alternative research tools that are accessible, and there are a number of ways that the community's research efforts could be refocused to be more effective.

Future research on Recommender Systems will become without any doubt increasingly important, so some observations must be made. First, we can see that a lot of research papers were found for example in movie recommendations, but just a small number of recommender systems for health, tourism, and education were found. This is because public domain movie datasets are readily available. Consequently, it is essential to create datasets in other fields as well.

Additionally, the most popular programming languages used to create Recommender Systems are Python and Java. This is because there are many standard Java and Python libraries available, which facilitate the programming workflow.

Lastly, the use of neural networks and deep learning-based techniques for creating Recommender Systems is the subject of extensive research. It is discovered that systems created utilizing these techniques attain high-performance accuracy. Still, there is a lot of room for improvement and future research.

I am really confident in the prospects of development of this field of research, also for socially valuable practices. Society could profoundly benefit from a proper usage and application of these systems, and I believe that almost every business should invest in research and development for the improvement of Recommendation engines.

BIBLIOGRAPHY

[1] Abdollahpouri, H. Burke, R, Mobasher, B, *Managing Popularity Bias in Recommender Systems with Personalized Re-ranking*, arXiv preprint, Cornell University, 12 Aug 2019

[2] Barragáns-Martínez, B., Costa-Montenegro, E., & Juncal-Martínez, J., *Developing a recommender system in a consumer electronic device*. *Expert Systems with Applications, 42(9),* 2015

[3] Beel, J., Gipp, B., Langer, S., & Breitinger, C, *Paper recommender systems: a literature survey*. International Journal on Digital Libraries, 17, 305-338. 2016

[4] Bin, C., Gu, T., Sun, Y., & Chang, L. *A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining.* Multimedia Tools and Applications, *78*, 2019.

[5] Burke, R. *Hybrid Recommender Systems: Survey and Experiments Department* of Information Systems and Decision Sciences, California State University, Fullerton, CA 92834, USA, 2001

[6] Burke, R. *Hybrid web recommender systems. The adaptive web: methods and strategies of web personalization* (2007): 377-408.

[7] Charu C. Aggarwal, *Recommender Systems: the textbook*, Springer International Publishing, 2016

[8] Chen, L., Chen, G. & Wang, F. *Recommender Systems Based on User Reviews: The State of the Art.* User Modeling and User-Adapted Interaction 25: 99-154, 2015.

[9] D'addio, R., Conrado, M., Resende, S. & Manzato, M.. *Generating Recommendations Based on Robust Term Extraction from Users' Reviews.* Proceedings of the 20th Brazilian Symposium on Multimedia and the Web, 55-58, 2014

[10] De Ruijt, Corné, and Bhulai, *Job recommender systems: A review.* ArXiv preprint, 2021.

[11] Dong, Zhenhua, et al., A brief history of recommender systems, arXiv preprint, 2022.

[12] Duan, L., Street, W. N., & Xu, E. *Healthcare information systems: data mining methods in the creation of a clinical recommender system.* Enterprise Information Systems, *5*(2), 169-181, 2011.

[13] Garfinkel, R., et al. *Empirical analysis of the business value of recommender systems*. Available at SSRN 958770 (2006)

[14] Gomez-Uribe, C.A., Hunt, N., *The Netflix Recommender System: Algorithms, Business Value, and Innovation* , Netflix, Inc., 2015

[15] Hdioud, F., Frikh, B. & Ouhbi, B., *Multi-Criteria Recommender Systems Based on Multi-Attribute Decision Making*. Proceedings of international conference on information integration and web-based applications & services, 2013.

[16] Heap, Bradford, et al. *Combining career progression and profile matching in a job recommender system*. PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014.

[17] Iwendi, C., Khan, S., Anajemba, J. H., Bashir, A. K., & Noor, F. *Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model*. IEEE access, 8, 2020.

[18] Jannach, D., P. Pu, F. Ricci, and M. Zanker. *"Recommender systems: Past, present, future."* AI Magazine, 2021

[19] Jannach D, Zanker M., *Value and Impact of Recommender Systems*, Recommender Systems Handbook. (519-546), 2022

[20] Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[21] Ko, H., Lee, S.; Park, Y.; Choi, A. *A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields*. Electronics, 2022

[22] Koene, Ansgar, et al., *Ethics of personalized information filtering, Internet Science: Second International Conference*, INSCI 2015, Brussels, Belgium, May 27-29, 2015.

[23] Lops, P. , de Gemmis, M. & Semeraro G., *Content-based Recommender Systems: State of the Art and Trends*, 2011

[24] Lu, J., Wu, D., et al., *Recommender system application developments: A survey*, Decision Support Systems, Volume 74, 2015, Pages 12-32

[25] Melville, P., Sindhwani, V., *Recommender Systems*, Encyclopedia of machine learning, Watson Research Center, Yorktown Heights, 2010.

[26] Milano, S., Taddeo, M. and Floridi, L., *Recommender systems and their ethical challenges*. *Ai & Society* 35: 957-967, 2020.

[27] Nanehkaran, Y. A., Licai, Z., Chen, J., Zhongpan, Q., Xiaofeng, Y., Navaei, Y. D., & Einy, S. *Diagnosis of chronic diseases based on patients' health records in IoT healthcare using the recommender system*. Wireless Communications and Mobile Computing, 1-14, 2022.

[28] Pajkovic, N. *Algorithms and taste-making: Exposing the Netflix Recommender System's operational logics*. Convergence, 28(1), 214–235. 2022

[29] Resnick, P., and Varian, H.R., *Recommender systems*, Communications of the ACM 40.3 (1997): 56-58.

[30] Roy, D., Dutta, M. *A systematic review and research perspective on recommender systems*. J Big Data 9, 59 (2022).

[31] Schafer, J. Ben, et al. *"Collaborative filtering recommender systems." The adaptive web: methods and strategies of web personalization.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. 291-324.

[32] Schedl, M., Knees, P., McFee, B., & Bogdanov, D., *Music recommendation systems: Techniques, use cases, and challenges.* In Recommender Systems Handbook (pp. 927-971). New York, NY: Springer US, 2021

[33] Seaver, N, *Captivating algorithms: Recommender systems as traps*, Journal of Material Culture 2019, Vol. 24(4) 421 –436

[34] Sharma, R. Singh, R.K. *Evolution of Recommender Systems from Ancient Times to Modern Era: A Survey.* Article in Indian Journal of Science and Technology, May 2016.

[35] Shu, J., Shen, X., Liu, H., Yi, B., & Zhang, Z. *A content-based recommendation algorithm for learning resources*. Multimedia Systems, 24(2), 163-173, 2018

[36] Smith, B., & Linden, G, *Two decades of recommender systems at Amazon. com*. Internet computing, 21(3), 12-18. 2017

[37] Su, Xiaoyuan, & Taghi M. Khoshgoftaar, *A survey of collaborative filtering techniques*, Advances in artificial intelligence, 2009.

[38] Walek, B., & Fajmon, P., A *hybrid recommender system for an online store using a fuzzy expert system. Expert Systems with Applications, 212,* 2023