Personalized news recommendation without private investigations

Investigating how relevant news article recommendations by recommender systems with limited user information available are

Name student: Lukas Borgers
Student ID number: 475591

Abstract

Recommender systems have become a necessity for news distribution to help users find the news they are looking for and enhancing the user experience. For recommender systems to make relevant recommendations, user information is required needs to be collected from users. Since the implementation of the GDPR, collecting user information has become more difficult making it harder to provide relevant recommendations. This thesis investigates how relevant news recommendations are from recommender systems which have limited user information available. The Microsoft News Dataset was used to create collaborative filtering based and content based recommendation models. For each model, two variant were made where one had all user information available and the other only a part. All models are compared on both recommendation accuracy and quality. The results show that content-based recommenders are most accurate and collaborative filtering based recommenders provide higher quality recommendations when limited user information is available. This implies that approach selection is made on whether recommendation accuracy or quality is valued more.

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Master Thesis Data Science and Marketing Analytics
Supervisor: Oktay Karabağ
Second assessor: Bas Donkers
Date final version: 10-7-2023

Table of Contents

1 Introduction **2**
1.1 Literature Review . . . 3
1.1.1 Empirical Context . . . 3
1.1.2 Societal Context . . . 5

2 Conceptual Framework **5**
2.1 Recommender Systems . . . 5
2.1.1 What are Recommender Systems? . . . 6
2.1.2 Cold-Start problem . . . 6
2.1.3 Collaborative Filtering (CF) . . . 7
2.1.4 Content-Based Filtering . . . 8
2.2 User Information . . . 9
2.2.1 Composition of User Information . . . 9
2.2.2 How is User Information Collected? . . . 10
2.2.3 Need for Informational Privacy . . . 11
2.2.4 General Data Privacy Regulations Outlines . . . 11
2.3 Online News Setting . . . 12
2.4 This research . . . 13
2.4.1 Recommending Methods . . . 13
2.4.2 Quality Control . . . 14
2.4.3 Data Sparsity . . . 15

3 Experimental Setting **15**
3.1 Experimental Design . . . 16
3.2 Data . . . 16
3.2.1 Impressions . . . 17
3.2.2 News . . . 18
3.2.3 Filtering and Subsetting . . . 20
3.3 Methods . . . 21
3.3.1 Collaborative Filtering Recommendations . . . 21
3.3.2 Content-Based Recommendation . . . 24
3.4 Performance Evaluation . . . 25
3.4.1 Accuracy Evaluation . . . 25
3.4.2 Quality Evaluation . . . 25

4 Results **26**
4.1 Collaborative Filtering . . . 26
4.1.1 Tuning . . . 27
4.1.2 ALS with Implicit Data . . . 28
4.1.3 Funk SVD with Explicit Data . . . 29
4.2 Content-Based recommendation . . . 32

5 Discussion of the Results **34**
5.1 Accuracy . . . 34
5.2 Quality . . . 35
5.3 Data Sparsity . . . 36
5.4 Concluding . . . 36

6 Conclusion **37**
6.1 Summary and Key Findings . . . 37
6.2 Limitations and Future Work . . . 38

Literature List **40**

1 Introduction

News is essential for all people to stay informed about affairs happening outside of their own environment. The distribution of news has always adapted to the latest technology, making it more accessible with each development. The shift from print and broadcast media to digital distribution forms no exception. Forman-Katz and Matsa (2022) reports that in 2022 only 8% of the U.S. population never consumes news from digital devices and over half of the population (53%) prefers digital platforms over other news sources, confirming how digital platforms superseded traditional platforms. Simultaneously with being more accessible, there is also a greater supply of news creating an abundance of choice for news seekers. This is supported by the more than doubled number of newsroom employees for digital native platforms between 2010 and 2020 (Stocking and Khuzam 2023).

The shift towards online platforms changed the constraint for choice from the size of a newspaper or number of news programs to the amount of time someone is willing to search for news. This abundance of options can cause stress in readers, making it challenging to find what they are looking for in a convenient manner (Karimi, Jannach, and Jugovac 2018). In these situations, recommender systems can help people by providing personalized recommendations. They manage the stream of information and make the choice process faster which contributes to a better user experience (Burke 2007). This in turns helps in binding users to the platforms allowing for, for example, more advertisements sales.

The recommendations from a recommender system are based on estimated user preferences. This means that a recommender system with better estimations works faster and therefore gives a better user experience (Adomavicius and Tuzhilin 2005). However, recommender systems depend on user feedback for the estimation of user preferences (Thorat, Goudar, and Barve 2015). In May 2018, the European Union (EU) implemented the General Data Protection Regulation (GDPR) which requires websites to get consent from their users to collect and process their data (European Commission 2016). This affects one of the most powerful tools for collecting user data, namely third-party cookies. These cookies are set by external parties on websites and can track users from website to website, giving a panoramic view of the online user journey. X. Hu and Sastry (2019) report an average drop of 18% in the number of third-party cookies on the Alexa top100 websites. The restrictive effect of the GDPR has created the need for another approach for recommender systems in which they can still make relevant suggestions, but with limited access to user data.

In light of these GDPR effects, this research focuses on recommender systems for news articles in situations when limited data is available to estimate user preferences. The goal is to find out what approaches are most resilient to these circumstances, and which can best be applied for an optimal user experience. This leads to the following research question:

How relevant are suggestions made by news recommender systems for English online news sites in a user information deprived environment?

To answer this research question, recommender system performance is compared between models with full access to user information and models with limited access to user information. The research question is divided into several sub-questions to shed light on different parts of this problem. This subdivision is based on the conceptual background given in later in this research.

This thesis is structured as follows. First, a review of literature is given to form the scope of this research. This is also used to frame the emperical and societal implications of this research. This is followed by a conceptual framework which gives an overview of all concepts related to news recommender systems and privacy. The a more specified scope of this research and several sub-questions with expectations in the form of hypotheses follow from this conceptual framework. The sub-questions are answered by rejecting or not rejecting the corresponding hypothesis. Next, an overview of the experimental setting used to test the hypotheses is given. The methods and data used in this research are also explained in this section. This is followed by the results section and a discussion that includes the explanation and implications of these results. Lastly, a summary of the main conclusions from this research is given, along with the limitations of this research and recommendations for future research.

1.1 Literature Review

The goal of this research is to investigate how relevant recommendations are for news articles when limited user information is available. The topics covered in this research therefore include recommender systems, online user information and news distribution. This section gives an overview of the empirical and societal context in which these topics take place. This frames the scope of this research and identifies the empirical and societal contributions of this research.

1.1.1 Empirical Context

While news was one of the first fields in which recommendation was implemented, it is no longer the most popular subfield within recommender system development. One explanation is the prevalence of high-quality datasets in other subfields on the one hand, such as the Netflix Prize competition and MovieLens dataset for movies, and on the other hand the lack of news equivalents. F. Wu et al. (2020) tries to fill this void and boost research into effective news recommendation methods by introducing the Microsoft News Dataset (MIND). Raza and Ding (2022) confirm a positive impact of the publication of this dataset by reporting a significant increase in research into news recommendation systems since the publication. MIND is larger than alternative datasets for news recommendation and includes more contextual information. Next to publishing an improved dataset, F. Wu et al. (2020) also contributes to news recommendation system development by testing multiple different recommending techniques. The aim of this thesis is to extend the research of F. Wu et al. (2020) among others by exploring how to optimize news recommendation using data from the MIND.

The work of Raza and Ding (2022) covers an overview of the recent research in the field of news recommendations. They describe that the traditional approaches of recommendation are user-oriented collaborative filtering (CF) and item-oriented content-based filtering (CB). The developments in recommendation consist of creating new methods for either approach that is more prolific in successfully linking items to users. An example of this are the matrix factorization methods used for CF developed by Funk (2006) and Koren, Bell, and Volinsky (2009) during the Netflix prize competition. These methods were an improvement over older methods by by being more flexible and resilient to data sparsity which led to more accurate recommendations. These methods were later adapted to other fields than movies of recommendation. Recent examples of developments in news recommendation are C. Wu et al. (2019), Zheng et al. (2018), and Wang et al. (2017) who investigate new techniques to create more personalized news recommendations. In the work of C. Wu et al. (2019), this is done by focusing on the difference in user taste by separately investigating user and item representation. The work of Zheng et al. (2018) explores how different forms of user feedback can improve news recommendation. This research also intends to investigate how different forms of user feedback can affect recommendation performance, similarly to Zheng et al. (2018). The work of Wang et al. (2017) focuses on modeling non-explicit selection criteria of related to publishing and news articles.

In research like that of C. Wu et al. (2019) and Wang et al. (2017), the golden standard in performance assessment is measuring prediction accuracy. This entails how prolific a recommender system is in recommending an article that the targeted person also deems relevant. Zhang et al. (2019) show how this standard is especially prevalent for new fields of recommendation such as audio or visual feature recommendation in music and video. However, contrary to these new fields, items such as news articles include more contextual information and require more quality control for the recommendations to make sense. Raza and Ding (2022) identify only using accuracy as performance measure as an additional current problem in news recommendation next to the lack of high-quality datasets for news recommendation development. As a solution, they propose several quality measures which also can be used to assess the performance of news recommender systems. Similar, beyond-accuracy measures have also been employed in other fields of recommendation with specific item particularities such as music. Schedl et al. (2018) describe how including these measures in the evaluation of music recommender systems improves user experience. By including several of their quality measures, this research builds further on work of Raza and Ding (2022) and Schedl et al. (2018) and differentiates itself from earlier research into news recommendations which do not include such measures like C. Wu et al. (2019) and Wang et al. (2017). By using such quality measures, it also builds further on the work of Ziarani and Ravanmehr (2021) and Gemmis et al. (2015) who also investigate beyond-accuracy measures in recommender systems.

Although research into different forms of feedback and data for recommender systems is available, discussions on the comparison between recommender systems with and without user information included are very limited in general and especially for news recommending at the moment of writing. A lack of user information is identified in the work of Raza and Ding (2022) as a future problem for news recommendation and is amplified by the restrictive measures for collecting user information in the GDPR. As user data is likely to become scarcer in the future, investigating this phenomenon is relevant for future research into improving recommender systems.

1.1.2 Societal Context

Recommender systems have become a necessity for online news platforms to provide a good user experience. Since this research investigates the performance of news recommender systems, the implications of this research are therefore directed at them. Firstly, to explore what methods are most successful helps in generating better recommendations and enhancing user experience. Secondly, this research explores the effectiveness and usefulness of including novel, qualitative measures beside to the most commonly used accuracy measures. By including such beyond-accuracy measures, additional insight into the recommendation process could be obtained to provide better recommendations and improve the user experience. Lastly, by evaluating if and how recommender systems can still make relevant recommendations without user information, this research sets guidelines for online news publishers on how proceed providing relevant recommendations and improving user experience in the future.

To a lesser extent, these principles also apply to other platforms outside of news that use recommender systems, who also depend on user information or who want to be more privacy conscious. In addition, a recommender system that uses less data is strictly more efficient than a system which uses more data at equal performances. The implications of this research into the performance of user information deprived recommender systems can therefore extend beyond news recommendation to improve recommender system performance and user experience in general.

2 Conceptual Framework

The purpose of this conceptual framework is to give background information on the subjects discussed in this research. This provides a rationale for the sub-questions and corresponding expectations. The subjects covered in this section are recommender systems, user information and news.

2.1 Recommender Systems

This section focuses on how recommender systems work. This contributes to understanding how they work and how to implement them for recommending news articles. This section starts by describing what recommender systems are. Also, different forms of recommender system are explained.

2.1.1 What are Recommender Systems?

Recommender systems are a form of information retrieval that has its origins in the 1990s and early 2000s. Resnick and Varian (1997) define recommender systems as tools that take recommendations of people as input, aggregate them, and then redirects them to appropriate recipients. They are used as information filtering systems, purposed to provide users with new items that they might find useful based on past behavior (Lü et al. 2012). Schwartz and Schwartz (2004) describe that humans get quickly overwhelmed when presented with many choices, leading them to make bad choices or none at all. Recommender systems can be helpful by decreasing the number of choice options and only presenting the ones that are most likely to be relevant. For the rest of this research, the definition of a recommender system of Ricci, Rokach, and Shapira (2010) is adapted as theirs is the most generally applicable and is developed based on extensive empirical research. They state that recommender systems are composed of users and items where the users are information profiles built from the dataset and items the recommended objects.

2.1.2 Cold-Start problem

For recommender systems to function, they require input data in order to make predictions and new suggestions. This input data consists of previous interactions that users have had with items. These interactions are used to derive characteristics of the users and items and build personalized profiles (Jannach et al. 2010). Based on these characteristics, preferences are estimated to determine which items should be recommended and which should not. In the case of news recommendation, interactions consist of users reading news articles. Based on this, user preferences for certain topics and categories can be estimated and used for recommending new articles. Articles can be profiled based on what users do and do not read the article.

The cold-start problem occurs for recommender systems when limited input data is available to base recommendations on. The name “cold-start” comes from initializing recommender systems in situations where limited input data has been generated and thus start off cold. All learning-forms of recommender systems suffer from a form of the cold-start problem. In the item variant of the cold-start problem, there is insufficient item information to determine the characteristics of the item. This makes it harder to find similar items to recommend. In the user variant of the cold-start problem, there is insufficient user information available for the recommender system to estimate the preferences of the user and thus provide recommendations aligned with the interests of the user (Schein et al. 2002). This variant describes the problems created by the new restrictive user information collection regulations of the GDPR.

With the rise of the internet, the availability of data and information has seen an exponential increase which in turn leads to a dramatic increase in the number of choice options in almost every aspect of life. The combination of better data availability and the increase of choice options is why recommender systems are nowadays commonly implemented to help users make better choices more quickly. Ricci, Rokach, and Shapira (2010) describe how a faster choice process helps in increasing user satisfaction and leads to more customer retention. Outside of the CF and CB approaches for recommender systems, other approaches like demographic-based and knowledge-based recommendation exist. However, these heavily rely on contextual user information like location and age (Ricci, Rokach, and Shapira 2010). This makes them less applicable for news recommendation compared to CF and CB recommendation and thus not relevant for this research (Raza and Ding 2022). Therefore, only CF and CB approaches are evaluated for this research.

2.1.3 Collaborative Filtering (CF)

CF is one of the first forms of recommender systems. It was first introduced by Goldberg et al. (1992) who created a user-oriented recommender system called Tapestry. CF recommender systems use similar user-item interactions to estimate user preferences and recommend new items to users. The left panel of Figure 1 shows how a group of users consisting of users X, Y and Z have liked both item A and B. User V has so far only liked item A. Similarity in this case is based on what items are liked among users. Since multiple users that have liked item A have also liked item B, the CF recommender system will recommend item B to user V. Instead of a user-oriented approach for linking new items to users, an item-oriented approach can also be applied. This concept was first introduced by Balabanović and Shoham (1997). The right panel of Figure 1 shows how user X has liked items A, B, C and D and user Y items A, B and C. Here similarity is based on that the large overlap in preferences of the two users. Because users X and Y are similar and user X has liked item D, the CF recommender system will recommend D to user Y.

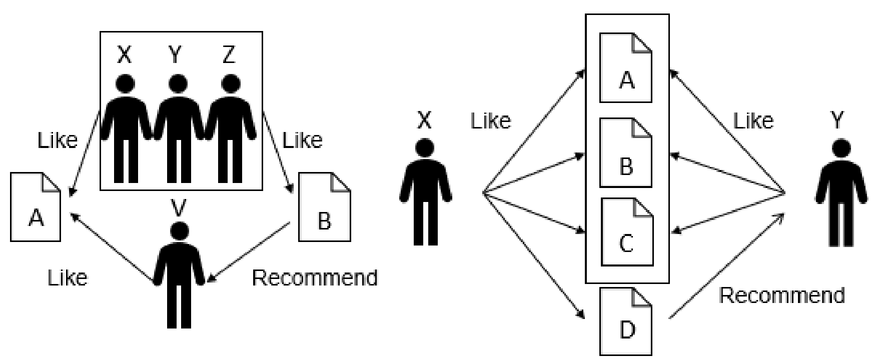

Figure 1: Illustration of item-oriented (left) and user-oriented (right) collaborative filtering.

A real-world example of CF implementation is the item recommendation used by Amazon. It is applied when visiting the site to promote buying additional items by recommending new items based on previous purchases, items in the shopping cart and previously visited item pages (Smith and Linden 2017). These recommendations are based on what items are frequently bought together and therefore likely to be complementary to each other. The result of this is more sales and profit. For news publishers, CF can be used similarly by recommending articles that are frequently read together. This would for example result in the recommendation of background articles.

By focusing on patterns in user behavior, CF provides user-centric recommendations that are personalized at the user level, making it also effective for users with broad or non-mainstream taste. Focusing on the user behavior also allows CF to make recommendations matching with user's taste without the need for domain-specific knowledge and recommend items that users would not have come across on their own. In addition, CF mitigates the item cold-start problem and offers more diversity than content-based recommenders. However, CF becomes less effective in environments with relatively many users and items. This is because an increase in items or users causes an exponential increase in the number of possible interactions. This makes the process more computationally expensive as every interaction needs to be evaluated separately. In addition, CF suffers from popularity bias where popular items tend to be overrepresented in recommendations compared to unpopular items. In extension of this, CF is also susceptible to amplifying the popularity bias and other biases present in user behavior in the recommendations. Lastly, when not enough user information is available, this approach suffers from the user cold-start problem where it is unable to make accurate predictions (Thorat, Goudar, and Barve 2015).

2.1.4 Content-Based Filtering

Whereas CF recommendations are based on user similarity, CB recommendations are based on item similarity. CB relies on the assumption that items with similar characteristics are likely to also be liked by users who have liked similar items in the past (Zhang et al. 2019). Item similarity is determined on the features and characteristics of an item. This method is created by Mooney and Roy (2000) among others and was derived from item-oriented collaborative filtering introduced by Balabanović and Shoham (1997). Figure 2 below gives a visual representation of CB. It shows items A1 and A2 of which are determined that they possess similar features. User X liked item A1. Because items A1 and A2 have similar features and user X has liked A1, the CB recommender system deems it likely that user X will also like item A2 and will therefore recommend it to user X.

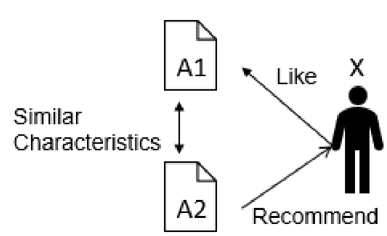

Figure 2: Illustration of content-based recommendation.

An example of how CB recommendation is implemented is the related page Netflix has for all movies. This page contains movies with similar categories of movies, cast members or directors. The related page is used to find movies that are similar to the ones liked by a user (Maddodi et al. 2019). CB recommendation can be used by news publishers to recommend articles that are similar to the articles that the user has just read. These articles could provide new insights and perspective by giving a different view of the matters discussed in the original article. Alternatively, articles with a similar premise can be recommended to meet the specific interests of the user, like for example, a match report of a different game in the same competition.

Not using user-to-user interactions, CB suffers less from the user cold-start problem as it does not need to build up a user profile. Additionally, recommendations are likely to be relevant even if the user has specific taste or when little information on user preferences is available since the recommendation is tailored to one user. However, the features of every added item need to be tagged, thus making this approach computationally expensive in environments with many items with varying features. Furthermore, CB suffers from a lack of novelty as the recommendations are limited to the known interests of the user (Liu, Dolan, and Pedersen 2010).

In summary, the two most popular and appropriate forms of recommending methods for news are CF and CB. CF performs relatively well in providing personalized and serendipitous recommendations without the need for domain-specific knowledge. While it is resilient to the item cold-start problem, it does suffer from the user cold start problem and is also susceptible to biases present in the data. CB is resilient to the user cold-start problem and performs relatively well for users with specific or abnormal preferences. However, it suffers from the user cold-start problem and is less prolific in recommending novel items to users as it is limited to acting upon the known interests of the user.

2.2 User Information

This section explains what user information is in the context of recommender systems and how it is used. In addition, it describes how user information is collected and how and why the collection is restricted by the GDPR.

2.2.1 Composition of User Information

According to Ricci, Rokach, and Shapira (2010), a recommender system requires three types of information: user, item, and transaction information. An overview of all three is given below.

Users are the recipients of the recommendations. Recommender systems build user profiles which encode the preferences of the user by using user information. Gathered user information can vary from sociodemographic features, such as age and gender, to numeric values regarding given ratings of the user. What information is used differs per method and technique. The user profile has a central role in personalized recommendation. The performance of a recommender system therefore heavily depends on how capable it is of making accurate user profiles (Fischer 2001).

Items are what is actually recommended. An item can both be characterized by the combination of features and by the value or utility a user derives from them. The item value can vary per user and dictates whether an item is deemed relevant or irrelevant. A simple economic cost-benefit calculation can be used to determine relevancy since there is always a cost and benefit incurred when the item is acquired. Both item features and values can also be used in a richer form of item representation instead of a more minimalist approach of using an ID code (Ricci, Rokach, and Shapira 2010).

Transactions are logs of the interactions between users and items that signify important information for recommendation creation. The transaction describes an action and may contain the context of where the action occurred. The most common form of transactional information is ratings. These can either be implicit or explicit (Ricci, Rokach, and Shapira 2010). Explicit ratings are actively given rating by a user to an item. This can be any nominal number. Implicit ratings are ratings derived from an action of the user related to the item like a clicking on the item. Implicit ratings are therefore binary with 1 for a positive action and 0 for a negative or no action (Y. Hu, Koren, and Volinsky 2008). Explicit rating data provide more user feedback compared to implicit systems which helps understanding the user rationale. However, it suffers more from user biases and active user participation is required for collection. Implicit rating data can be collected passively, making it the more convenient option for collecting large amounts of data. Unlike movies or music, explicit rating systems do not apply to news and are seldom implemented. News recommendation datasets therefore almost exclusively use an implicit rating system. This phenomenon is described by Feng et al. (2020) who also find improved accuracy for news recommendation when using implicit feedback instead of explicit.

2.2.2 How is User Information Collected?

Sanchez-Rola et al. (2019) describes that digital tracking mostly occurs using either fingerprinting or HTML cookies. Fingerprinting is a form of user identification using unique browser characteristics like screen size or differences in internal clock. However, HTML cookies, more commonly known as cookies, are the more prevalent form of digital tracking. These are small pieces of data stored in web browsers. Each cookie contains a name, value, associated URL and expiration date where the value can vary its purpose. Web applications use cookies to keep track of users and enable complex workflows. Originally this was to create a more seamless navigation, but nowadays it can also be used to collect data for advertising and analytics purposes. These purposes are further aided by the possibility to include third-party cookies on a website. These third-party cookies exist on multiple different websites and can help track user activity between websites and map their entire online journey.

Aguirre et al. (2015) describes how the method of data collection can be categorized as either overt or covert. Overt information collection states that the user is aware that that their information is being collected because an effort was made to inform them (Sundar and Marathe 2010). Covert information collection happens when user information is collected without the knowledge of the user (Montgomery and Smith 2009). Aguirre et al. (2015) argues that overt data is more transparent and therefore more attentive to privacy concerns, but also more obtrusive towards users and therefore detrimental for the customer experience. In addition, Verhoef, Reinartz, and Krafft (2010) argue that covert data leads to less biased data and therefore a better user understanding.

2.2.3 Need for Informational Privacy

Malhotra, Kim, and Agarwal (2004) describe personal information in a digital format as a double-edged sword: it makes it possible to construct more thorough descriptions of individuals. These can either be a serious privacy threat or a beneficial tool for providing personalized service such as recommendation. This as such is also observed in human behavior. Madden (2014) finds that 80% of the United States population has significant concerns about third parties using data they share online for advertising purposes in 2014. At the same time, Carrascal et al. (2013) finds that people value their online personal information significantly less than their offline personal information. This contradiction in behavior is known as the privacy-personalization paradox (Aguirre et al. 2016). It describes that user engagement can either increase or decrease depending on whether it triggers privacy concerns or has an appealing effect.

2.2.4 General Data Privacy Regulations Outlines

The GDPR is a regulation implemented in May 2018 by the European Union (EU) to provide guidelines on digital privacy. It replaced the Data Protection Directive (DPD) issued in 1995 which became outdated due to new technological innovations (Zerlang 2017). The GDPR changed several things on an interterritorial scale compared to the DPD. This research focuses on aspects related to personal data as the data used in recommender systems can reasonably be considered as such (Hildebrandt 2022). A more global overview of all changes is given in Tankard (2016).

GDPR defines personal data as any information related to an identified or identifiable natural person. This definition is intentionally kept broad to include both explicit and implicit forms of personal data. The GDPR states that processing of personal data is exempt unless one of two conditions is met: it is prohibited by law, or the user gives explicit consent for his data to be processed. In order for the consent to be valid, the GDPR states that the contract must be freely given, specific, informed, and unambiguous (European Commission 2016). This holds, in this regulation's context, that the user must give voluntary and informed consent for specific actions. Additionally, the user must be informed of the possible risks related to the data processing and must have the option to withdraw consent at any possible moment.

One of the largest consequences of the GDPR for recommender systems is therefore that personal information needs to be collected overtly which, in the definition of Aguirre et al. (2015), means that data collection is more transparent. On the downside it is also more intrusive and creates a poorer user experience. Where transparency is a clear advantageous and intended effect of this legislation, the decrease in user experience is less beneficial.

2.3 Online News Setting

Mitchelstein and Boczkowski (2009) describe how traditionally the news market was highly competitive and filled with established players. Although journalism and news distribution have always evolved with technology, the shift towards online news was more motivated by concern for competition from online newcomers than by an own desire for innovation (Allan 2006). The transition to online news became more definitive when traditional revenue models, such as subscription sales, were abandoned for more innovative ones like targeted advertisement (Mitchelstein and Boczkowski 2009).

Before the shift towards online news and the developments that came along with it, news was produced for the masses and consumed by largely heterogeneous groups. What content was made available to the public, and therefore what would be consumed, depended on the subjective criteria of journalists; a phenomenon called gatekeeping (Shoemaker, Vos, and Reese 2009). The development of news towards online platforms has increased the level of control someone has on what news he consumes and indirectly what selection of news is published.

Bucy (2003) describes the key difference that distinguishes online news from offline news as the possibility to interact with the news. This possibility provides newsrooms with evaluation metrics to assess what content most closely approaches the wants of their audience. This continuous evaluation is now an important driver in the content selection process (Blanchett Neheli 2018). Harcup and O'neill (2017) describe how the criteria for something to qualify as news has shifted significantly due to the change in news distribution.

An additional consequence of the transition from offline to online is the collapse of the twice-a-day news cycle. Where in the past news could only be consumed in the morning and evening via traditional media outlets, online distribution made it possible to consume news instantaneously and at any desired time (Lawson-Borders 2006). Mitchelstein and Boczkowski (2009) describe how these developments increased the speed of communication in the journalistic work field. This caused additional importance of recency of news relative to other online distributed items.

Raza and Ding (2022) describe the recent developments and challenges in online news recommendation in their work. Next to recency, they describe user modeling and quality control as the three biggest challenges at the moment of writing. The challenge in user modelling is that currently most methods are not successful in capturing the diversified interests of news readers. One of the reasons for this is that most news site visitors are anonymous and it therefore impossible to build elaborate and accurate user profiles. Another reason for this is that most models for estimating user preferences are only checked for accuracy. This is also where quality control comes into play. By not accounting for quality, news recommenders are prone to clickbait, biased news presentation and duplication. In their work, Raza and Ding (2022) describe additional, beyond-accuracy measures relevant for online news distribution and news recommendation in specific. These measures are coverage, novelty, and serendipity.

Coverage is used in most cases of recommendation and depicts the ratio of unique items among the recommended items compared to all available items. In news settings, coverage can be determined for the number of topics (topical coverage), entities or items (Jungkyu Han and Yamana 2017). High coverage is desirable in news recommendations to ensure that a wide range of topics, items or entities are represented in the recommendations. This protects news recommenders from filter bubbles and from becoming echo chambers (Flaxman, Goel, and Rao 2016).

Novelty depicts how familiar the user is with the recommended items. In news settings, novelty extends from recency as old news is likely to be outdated and therefore less likely to be novel. That is also why a high degree of novelty is a desirable attribute for news recommendations.

Serendipity means that something is unexpected, yet desirable. Serendipity can be distinguished from novelty as a list of recommendations can be very novel, yet irrelevant at the same time. High serendipity is thus desirable in recommending items as this indicates that aiming for novelty does not come at the cost of relevancy (Ziarani and Ravanmehr 2021).

2.4 This research

The overview above describes all aspects of news recommendations and what challenges they are accompanied by. Together they form the conceptual framework for this research. This section splits the research question into several sub-questions based on this conceptual framework. For each of these sub-questions, an expected answer is formulated in the form of a hypothesis.

2.4.1 Recommending Methods

The two most popular and appropriate forms of recommending methods for news are CF and CB. Thorat, Goudar, and Barve (2015) describe how CF performs relatively well in providing personalized and serendipitous recommendations without the need for domain-specific knowledge. Where it is also resilient to the item cold-start problem, it does suffer from the user cold start problem and is also susceptible to biases present in the data. CB is resilient to the user cold-start problem and performs relatively well for users with specific or abnormal preferences. However, it suffers from the user cold-start problem and is less prolific in recommending novel items to users as it is limited to acting upon the known interests of the user. Raza and Ding (2022) signify methods used for news recommendation as one of the current challenges in their work. This leads to the first sub-question:

What methods are most accurate in recommending news articles?

Based on the strengths and weaknesses of CF and CB, it can be inferred that what approach performs best in making news articles recommendations differs per situation. In the news recommendation setting, item information is relatively easier to obtain compared to user information. It is expected that CF methods therefore struggle to make relevant recommendations since little user information is available to estimate preferences. This evidence suggests that CB models perform better compared to CF models. However, F. Wu et al. (2020) find in their work that CF outperform CB methods for news recommendation. Feng et al. (2020) find in their work that implicit feedback provides relatively good prediction for news articles. Nevertheless, Zheng et al. (2018) find evidence of how other data than implicit user feedback can improve recommendation accuracy. In addition to this, Ricci, Rokach, and Shapira (2010) describe how CF can leverage more information from explicit data and create a more complex user profile compared to implicit data. More complex user profiles lead to better estimation of user preferences and therefore more accurate recommendations. This leads to this first hypothesis:

H0: Recommender systems using a collaborative filtering approach with explicit data provide a better recommendation accuracy for news articles compared to content-based approaches or collaborative filtering approaches using implicit data.

2.4.2 Quality Control

The data used in recommender systems can be split into three groups: user, item, and transaction data. User data is used to build user profiles describing the preferences of users. Item data is used to characterize items and is used to determine a value linked to consuming the item. Transaction data depicts the relationship between users and items. Together they are used in recommender systems to predict what items a user is interested in. Research into recommender systems, like F. Wu et al. (2020) and Zheng et al. (2018), test different methods of news recommendation based on how accurate the predictions of recommender systems are. Raza and Ding (2022) describe in their overview of progress in news recommendation that next to prediction accuracy, quality is also an important, yet neglected performance measure. They suggest including coverage, novelty, and serendipity as additional, beyond-accuracy measures in the performance evaluation of recommendation models. The second sub-question is aimed at using this suggestion and including quality measures in performance evaluation. This leads to the following sub-question:

What methods produce the highest quality of news articles recommendations?

Both CF and CB recommenders provide personalized lists of recommendations to users. Thorat, Goudar, and Barve (2015) describes that by not requiring domain knowledge, CF based recommenders can make serendipitous recommendations by recommending novel items that align with the interests of the user. CB recommenders are limited to the known interests of users and can therefore provide fewer novel items compared to CF based recommenders. Because of this, CB recommenders are also less able to provide users with serendipitous recommendations. However, Liu, Dolan, and Pedersen (2010) describe how CB recommenders are less prone to biases, with the popularity bias in specific, compared to CF recommenders and can therefore provide wider coverage. By providing more novel and serendipitous articles, CF recommenders are more likely to provide higher quality recommendations than CB recommenders. This leads to the following hypotheses:

H0: Recommender systems using a collaborative filtering approach produce higher quality new article recommendations than recommender systems using a content-based approach.

2.4.3 Data Sparsity

Sanchez-Rola et al. (2019) describe how most online user information is collected through cookies. Malhotra, Kim, and Agarwal (2004) describe how digital user information can either be a beneficial tool in providing personalized service, or as a serious privacy threat. As a response to the latter, the European commission created new guidelines on how to handle personal user information (European Commission 2016). This transformed the collection of user information from mostly covert to overt which is more intrusive and therefore more harmful for the user experience (Aguirre et al. 2015). In the case of news recommendations, it also makes it harder to collect user information and therefore create useful profiles which are used to make relevant recommendations. The difficulty in collecting user information for online news platforms was already identified by Raza and Ding (2022) as a problem before the introduction of the GDPR. This leads to the last sub-question:

What methods are most resilient to data sparsity in user information?

CF approaches are reliant on user information to make recommendations and therefore suffer from the user cold-start problem (Burke 2007). CB approaches mostly rely on item information to make recommendations and mitigate this variant of the cold-start problem. By considering the curtailment of user information as a user cold-start problem, one can postulate that CB approaches provide better recommendations when little data is available than CF approaches. This leads to the following hypothesis:

H0: Recommender systems with a content-based approach are more resilient to data sparsity in user information than recommender systems using a collaborative filtering approach.

3 Experimental Setting

The following section discusses the experimental setting of this study. The setting describes the methods and data used to test the hypotheses and answer the research question formulated in the previous section. First, a description of experimental design is described which elaborates on how the hypotheses are tested. Next, A detailed overview of the data used in this study is given followed by a description of the data filtering process. The methods used for the data analysis are discussed thereafter. Lastly, the evaluation metrics used for model performance assessment are discussed.

3.1 Experimental Design

To establish what method is most accurate and provides the highest quality for news recommendation, a recommender system model is made applying a variety of different methods. Each model produces a list of 20 recommendations based on user data which are evaluated for accuracy and quality using several different measures. The average accuracy and quality are computed for each measure and used to compare the model and determine which method performs best for recommending news. This design is in line with C. Wu et al. (2019) and Zheng et al. (2018) among others. The methods applied in this research are explained in the methods section.

To test what methods can best cope with data deprivation, all different models are made and tested with both user history included and excluded. The effect of the data deprivation can be measured by comparing accuracy and quality measures of the varying models. What model is most resilient to this can also be determined by calculating the lowest drop-off. First, the tuning for the CF models is analyzed to find the best performing model. The best possible model is estimated, and the output is compared for performance to test what the effect including user information is and to find out what the best approach is for CF recommendation. Secondly, results for the CB model are estimated and interpreted. Lastly, the results of the CF and CB models are compared to each other.

3.2 Data

The data used in this research originates from MIND. It was presented by F. Wu et al. (2020) as benchmark dataset for news recommender system research. It distinguishes itself from other news recommendation datasets by its size, the inclusion of contextual data and by using English articles. The extensiveness and amount of contextual detail makes this the best suited dataset for this research. An overview of the dataset is given in Table 1. It shows that MIND consists of over 2 million impression logs mentioning almost 97 thousand unique news articles generated by 711 thousand users. The dataset consists of a behavior part with user impressions and a news part with contextual information of the news articles.

Table 1: Overview of MIND

	Number
Articles	97928
Categories	18
Subcategories	284
Users	711222
Impressions	2232748

3.2.1 Impressions

The user behavior part of the dataset consists of impression logs. Each impression has a unique ID and timestamp and consists of a list of articles presented to a user. Both the articles as the user are represented by an ID in the impressions data. Clicking behavior for each article in the list is indicated with a boolean where 1 is a click and 0 a non-click. Next to clicking behavior of the current impression, articles that the user has clicked on in previous sessions are also included under the name user history. Table 2 shows summary statistics of the impression dataset.

Table 2: Summary Statistics of Articles per Impression

	Mean	Standard Deviation	Max.	Min
Total number of Articles	74.98	47.30	300	4
Read Articles	4.11	1.85	51	3
Unread Articles	70.95	46.78	297	1
Articles in User History	53.54	58.38	801	1

The average impression consists of 75 articles of which approximately 4 are clicked on and read and 71 are ignored. During the dataset collection, the average user has read 54 articles on the Microsoft news site before revisiting.

The basis of every recommender system is user-interaction information. This is most commonly represented in a matrix for analysis which typically exists of rows with users and columns with items where the observations represent the interaction between the user and item. The user-item interaction matrix using implicit data is constructed from the impression dataset where each impression is considered a separate user. This results in a matrix with all impressions as separate rows and all mentioned articles as columns. Articles read in an impression are given a 1 and all other observations are denoted with 0. To test the effect of limited user data, a separate matrix is created which also includes user history. The matrix with user history included is used to proxy data saturation and the matrix without user history to proxy data deprivation (Ilievski and Roy 2013).

Absent observations are handled the same as ignored items in the analysis of the impression dataset. Because of the relatively high number of absent observations an additional user-item interaction matrix is constructed to add more feedback information. The explicit variant is created by denoting all read articles in an impression by 10, all seen but unread articles by 1 and all others by 0. This method gives more contextual information by regarding unread articles as a dislike, read articles as a like and ignoring unseen ones. This is done for both the data deprived and the data saturated model.

3.2.2 News

The news dataset contains all news articles presented to users where each observation is a separate article. Each observation includes an article ID, title, abstract and a category and sub-category which are manually marked by the author of the respective article. The article characteristics are used to add context to the lists of recommendations and to assess their quality. There are 97,928 unique articles in total in the dataset divided into 18 categories. Table 3 shows how many articles and subcategories there are per category.

Table 3: Number of Articles and Subcategories per Category

Category	Number of Articles	Number of Subcategories
Sports	30743	34
News	29720	39
Finance	5666	32
Travel	4799	17
Video	4460	13
Lifestyle	4379	53
Food and Drink	4252	17
Weather	3989	4
Autos	2955	25
Health	2825	23
Music	1252	11
TV	1239	10
Entertainment	789	15
Movies	768	7
Kids	75	6
Middle East	2	1
Games	1	1
North America	1	1
Total	97915	309

Table 3 shows that sports and news are categories with the most published articles. This makes sense as they can also be seen as the broadest categories due to their high number of subcategories. Lifestyle can also be considered as a broad category but has significantly less published articles. How many times an article from a certain category is read is depicted in Figure 3 below.

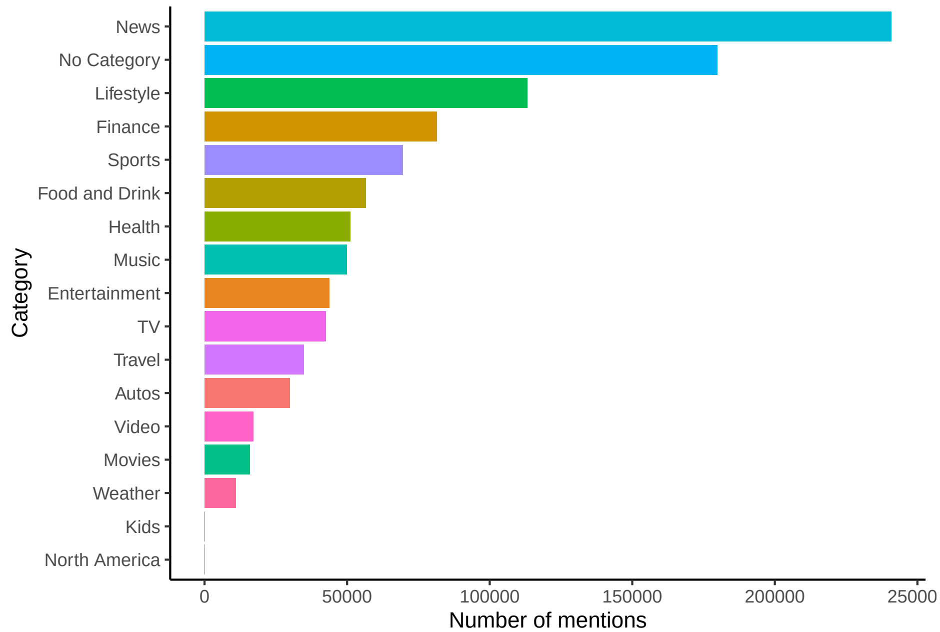

Figure 3: Number of Mentions per Article Category

Figure 3 shows that news is the most read news category followed by lifestyle, finance, and sports. That news is the most read category can be reasonably expected due to its size. Remarkably, sports articles are not read as much despite the large number of articles and subcategories. On the contrary, lifestyle articles are read relatively often even though there are not as many articles published in this category.

In addition to the characteristics already in the dataset, Natural Language Processing (NLP) can be used on the article title and abstract to derive additional characteristics. Table 4 shows an overview of the number of words in the article titles and abstracts.

Table 4: Number of Words in Article Title and Abstract

	Mean	Standard Deviation	Max.	Min
Words in Title	10.68	3.27	95	1
Words in Abstract	133.14	717.14	19898	0

As could be expected, the average length of the article abstract is multitudes longer than that of the title. As the use of abstract text in NLP would lead a much more computational expensive model, only titles are processed using NLP to derive further article characteristics.

3.2.3 Filtering and Subsetting

The datasets are filtered to reduce dimensionality. While higher dimensionality increases the complexity of predicting models, it also increases the computational power requirements and the risk of overfitting. Eliminating dimensions with little explanatory power increases efficiency and improves the fit of the models.

Impressions from the user-item interaction matrix which only contain one read article are filtered out of the dataset as they contain relatively little information about user-item interactions. The chance of irrelevant recommendations is also higher due to the limited amount of information to base these on, negatively skewing the results. This reduces the size of the impressions dataset to 252.644 observations which each contain 3 or more read articles.

The article titles from the news dataset are transformed into bag-of-words for the NLP. Lemmatization is applied to reduce the dimensionality of the bag-of-words by reducing the size of the vocabulary. Lemmatization is the process of reducing words to their base form, called a lemma. This causes different inflections of a word to be treated as the same. In contrast to stemming, which cuts words off to a stemmed form, lemmatization takes context into account and can also revert different forms of the same word to a single form. A dictionary containing different word forms and corresponding lemma is required for this process (Plisson et al. 2004). The dictionary used for lemmatization is based on Mechura's 2016 lemmatization list.

The computational resources for this research are limited. Computing predictive models on the dataset is a resource and intensive process. By using a subset of the data instead, the computational burden and time consumption can be reduced significantly. It also makes it possible to iterate more quickly and investigate more methods with different settings. Therefore, this research only uses a subset of the MIND. The subset is made by sampling impressions from the filtered impression dataset. The news dataset is subsequently filtered by removing all articles that are no longer mentioned in the subsetted impressions dataset. This reduces the size of the news dataset to 2068 articles. The subset has comparable characteristics to that of the entire datasets. This makes the models created on the subset to also be representative for the entirety of the dataset. The characteristics of the subsetted dataset are presented in Figure 4.

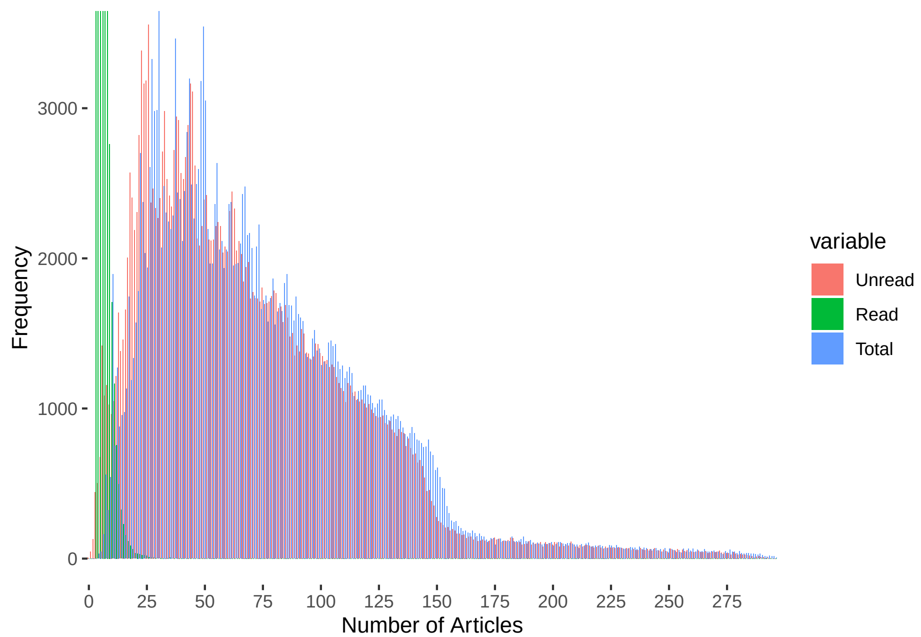

Figure 4: Distribution of the Number of Articles per Impression

Figure 4 shows that most impressions consist of between 3 and 10 read articles and between 15 and 150 articles unread articles. The total number of articles in most impressions is between 20 and 155. Extrapolating from this, most impressions have a ratio of read article to unread articles between 6% and 33%. This distribution is similar to what is found in Table 2 which indicates that the subset is indeed a good representation of the entire dataset.

3.3 Methods

To find out what technique leads to the best performance in news recommendation, multiple different techniques of both CF and CB are applied to produce item recommendations. They are applied on variations of the same dataset to proxy different scenarios and then evaluated on performance to determine how well they handle that scenario. The techniques used in this research consist of model-based CF and CB recommendations. The user-item matrix derived from the impression dataset is used to derive user and interaction information and the news dataset is used to derive item information.

3.3.1 Collaborative Filtering Recommendations

Establishing similarity between items or users can also be done in two separate ways, namely with a memory- and a model-based approach. Memory-based approaches use stored user-to-items interactions and use them to recommend items to similar users. This approach can be called a nearest neighbors approach as it directly uses the values of these interactions and selects the item or user with the highest similarity to make new recommendations (Nguyen et al. 2021). Model-based approaches use the stored interactions in generative models to predict interactions on new items and can therefore make predictions about items that are not interacted with by the user. The most common method for model-based CF is by applying matrix factorization (Nguyen et al. 2021). Matrix factorization in CF is a technique where the user-item interaction matrix is decomposed into two lower rank item and user matrices. In these matrices, the rows represent the users and items and the columns the latent factors of the users and items, respectively. The latent factors represent the underlying characteristics of users and items. The number of latent factors, k, can be tuned where more latent factors lead to larger lower rank sub-matrices and therefore a more elaborate representation of the underlying characteristics. The model also becomes more complex and harder to compute with an increase of latent factors. The matrix decomposition can be represented as

$$M(m \times n) = X(m \times k)Y(k \times n)^T \tag{1}$$

for user-item matrix M (m x n) with user matrix X (m x k) and item matrix Y (n x k). The dot product of the two lower rank matrices represents a predicted score based on the underlying characteristics. The goal of matrix factorization algorithms is to minimize the Sum of Squared Errors (SSE) between the predicted and actual ratings. The SSE is calculated as the difference between dot product of the user and item matrix, and the actual ratings. The calculation of predicted ratings can be denoted as

$$Rating_{ij} = X_i Y_j \tag{2}$$

for user i of user matrix X and column j of item matrix Y. The calculation of SSE can be formulated as

$$SSE = (r_{ui} - x_u^T y_i)^2 \tag{3}$$

with rating r for user u and item i. Matrix factorization algorithms start by randomly initializing the values of the user and item matrices. The user and item matrix are then optimized which can be formulated as

$$\min_{X,Y} \sum_{r_{ui}} \left(r_{ui} - x_u^T y_i\right)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right) \tag{4}$$

where X is the item matrix with factors x, Y the user matrix with factors y and r_{ui} rating of user u for item i. The optimization of the item and user matrices causes the latent factors to best represent the user-item interactions. Next to the first term which calculates the SSE, the formula also includes an L2 regularization term. This regularization term describes how the magnitudes of learned parameters, which are used to compute the predicted rating, are penalized. This term is controlled by λ (lambda), where a lower lambda leads to a more complex model with a closer fit to the training data, but also to a higher risk of overfitting (Koren, Bell, and Volinsky 2009).

Multiple approaches exist for this optimization problem. For this study's purpose, Funk Singular Value Decomposition (SVD) and Alternating Least Squares (ALS) are applied to investigate the effectiveness of model-based CF. Both approaches are commonly used in collaborative filtering since the Netflix Prize competition and are therefore an accurate representation of how recommender systems would perform outside an experimental setting. In addition, with each approach having different strengths and weaknesses, the combination of the two can be used to investigate the wide variety of scenarios of this research.

Funk SVD was first introduced by Simon Funk during the Netflix Prize competition. It optimizes the matrices by using Stochastic Gradient Descent (SGD) making it similar to a least squares problem optimization. The algorithm loops over all training cases and calculates the prediction error for all instances. Afterwards, it updates the prediction parameter by the magnitude of the learning rate. This process is repeated until the algorithm converges or reaches the maximum number of allowed steps. This methodology makes it relatively easy to implement and grants it a relatively fast running time (Funk 2006).

Contrary to Funk SVD, ALS does not optimize both matrices at the same time but starts by fixing the user matrix and optimizing the items matrix. This splits the optimization formula into two separate formulas which are alternatingly optimized. This can be denoted as

$$x_u = \left(\sum_{r_{ui} \in r_{u*}} y_i y_i^{\mathrm{T}} + \lambda I_k \right)^{-1} \sum_{r_{ui} \in r_{u*}} r_{ui} y_i \tag{5}$$

for the items matrix and

$$y_i = \left(\sum_{r_{ui} \in r_{*i}} x_u x_u^{\mathrm{T}} + \lambda I_k \right)^{-1} \sum_{r_{ui} \in r_{*i}} r_{ui} x_u \tag{6}$$

for the user matrix. When the items matrix is optimized, it is updated, and the algorithm repeats the process with the item matrix fixed and the user matrix optimized (Hastie et al. 2015). By splitting the process of optimization, ALS can be faster than Funk SVD in certain situations. In addition, ALS only considers positive data as a signal and regards all other data as 0. By doing so, it considers the absence of a signal as relevant, making it well suited for datasets with implicit data or many missing data points. At the same time, Funk SVD is better suited for data rich datasets with real ratings as it considers missing data as a low rating (Y. Hu, Koren, and Volinsky 2008).

Besides matrix factorization, CF, both user- and item-based, can also be done using cosine similarity. In this method, the cosine of the angle between two vectors in the user-items matrix are calculated to determine their similarity. The cosine similarity can be calculated as

$$Similarity(i, j) = \frac{i \cdot j}{||i|\,|j||'} \tag{7}$$

where i and j represent user and item vectors from the user-item matrix. The numerator represents the dot product between the two vectors and the denominator the product of the vector lengths calculated using the Euclidean Distance. This distance can be calculated as

$$Euclidian\ Norm = \sqrt{x_1^2 + x_2^2 + ... + x_p^2} \tag{8}$$

for vector X with length p. Because of this, the cosine similarity values range from -1 to 1 where -1 indicates opposites, 0 no correlation and 1 perfect correlation between the two vectors. Recommendations using this method are made by finding the most similar users based on the similarity score and aggregating their item interactions. The recommendations are composed by computing the weighted average of the preferences of the aggregated item interactions. This approach gives more weight to the preferences of the most similar users (Jiawei Han et al. 2012).

3.3.2 Content-Based Recommendation

The first step in CB recommendation is representing all items by relevant features. Term Frequency-Inverse Document Frequency (TF-IDF) can be used in the case of text documents to extract item features.

TF-IDF is an NLP technique for text vectorization, the process of text quantification. This enables document comparison and information retrieval in machine learning. The term frequency is how often a term appears in a document and denotes the significance of a term in a document. The inverse document frequency is the logarithmic function of the total number of documents divided by how many times a term is present in all documents, or corpus. The TF-IDF score is obtained by multiplying the term frequency by the inverse document frequency. This score reflects how important a word is in the context of the entire corpus. This causes frequently mentioned words like stop words to be weighted lightly and infrequently mentioned words to be weighted heavily, granting it more explanatory power.

The weight of term i in document j can be calculated as

$$a_{ij} = f_{ij} \cdot \log_2 \frac{N}{d_i} \tag{9}$$

where N is the number of documents in a set and d the document frequency of term i. Applying this to a bag of words, such as article titles, results in vectors with TF-IDF scores. The simplicity of TF-IDF comes at the cost of unawareness of context making it unable to semantics of words (Trstenjak, Mikac, and Donko 2014).

Next, similarity between text vectors is calculated using cosine similarity to determine similar items. Cosine similarity measures the cosine of the angle between two word vectors in an inner product space like that of text vectors created by TF-IDF. A value of 0 means a 90-degree angle between two word vectors and therefore perfect uncorrelation. Therefore, a cosine value closer to 1 means a smaller angle between two word vectors and therefore a higher correlation.

3.4 Performance Evaluation

A variety of model accuracy and quality metrics are assessed to compare performance between models. The accuracy metrics focus on the predicting capabilities of a model and are based on how many relevant recommendations are made the model. Predicting capabilities are evaluated to conform with F. Wu et al. (2020) by using precision and recall for the top 20 recommendations.

3.4.1 Accuracy Evaluation

Determining predicting accuracy starts by counting the number of true positive (TP) and negatives (TN) and the false positives (FP) and negatives (FN). A confusion matrix is typically used to display the prediction performance of a model by showing the number of occurrences for each possibility. For this study, a true positive, where the model accurately predicts a positive value, is considered a relevant recommendation. This is because a true positive reflects how the model recommends an item that the user has read and is therefore in line with the user's preferences. To emphasize this, precision and recall of all models are calculated to determine how effective a model is in making relevant recommendations. Precision is the rate of correctly predicted articles compared to all predicted articles and is calculated as

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{10}$$

and recall the ratio of correctly predicted relevant articles to the total amount of relevant articles in the dataset and is calculated as

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}. \tag{11}$$

For both measures, a higher ratio is preferable as this means that the recommender system recommends relatively more relevant items.

3.4.2 Quality Evaluation

Next to recommendation accuracy, recommendation quality is also evaluated. This is a relatively underexposed and neglected part in news recommendations literature which mostly focusses on accuracy. However, quality is important for user satisfaction and therefore relevant for determining if a recommender system produces desirable results (Raza and Ding 2022). The quality of articles is assessed in accordance with (Raza and Ding 2022) using three: coverage, novelty, and serendipity.

Coverage assesses the extent to which the recommendation list spans the complete list of category possibilities. It measures the proportion of unique item categories in the recommendation list compared to the total number of item categories in the dataset. Higher coverage values indicate a broader coverage of categories, and thus more diversity, which helps circumventing popularity. The calculation of coverage can be formulated as

$$Coverage = \frac{c_L}{C} \tag{12}$$

with the number of categories in list L and C the total number of categories.

Novelty is an indication of how many articles in the list of recommendations are unknown to the user. Higher novelty is desirable in recommender systems as this avoids over-exposure of popular articles, preventing tiring in users. Additionally, lists of recommendations with high coverage and novelty allow for exploration and help recommender systems to adapt to changing interests of users. Novelty is measured as the ratio of expected to unexpected items in a list of recommendations L. Unexpectedness is based on popularity and an item is deemed to be expected if the item belongs to the half of all articles with the highest popularity score. The popularity score is calculated as

$$Novelty = \frac{\sum_{i \in L} U(i)}{N} \tag{13}$$

where U(i) returns 1 for article i if the article is unexpected. The assumption here is that a user is more likely to know an item seen by many, and therefore be less novel, than an item seen by few (Gemmis et al. 2015).

Serendipity is an extension of novelty and signifies unexpected, yet relevant recommendations. Serendipity is a relevant factor as a list of recommendations can be diverse but not helpful if most are irrelevant. Recommendations with higher serendipity are therefore desirable in a recommender system. proposes a valuation method which calculates the number of both unexpected and relevant items. This results in the formula

$$Serendipity = \frac{\sum_{i \in L} S(i)}{N} \tag{14}$$

where S(i) returns 1 for article i if an item is both unexpected and relevant and thus serendipitous.

4 Results

The outputs of different models are analysed and compared in the result section. First, the tuning for the CF models is analyzed to find the best performing model. The best possible model is estimated, and the output is compared for performance to test what the effect including user information is and to find out what the best approach is for CF recommendation. Secondly, results for the CB model are estimated and interpreted. Lastly, the results of the CF and CB models are compared to each other.

4.1 Collaborative Filtering

To assess to accuracy and quality, models applying ALS with implicit data and Funk SVD with explicit data are estimated as described in the methods section and compared to each other on both accuracy and quality measures. The hyperparameters of both models are first tuned to find the model with the best possible prediction accuracy. The model tuning, estimation, and comparison is done by using the recommenderlab package created by Hahsler (2022).

4.1.1 Tuning

Tuning is the process of finding the best possible combination of hyperparameters for a learning model like Funk SVD and ALS. A hyperparameter is a parameter which needs to be defined before the start of the training process in contrast to a model parameter which is learned by the model during the training process. Hyperparameters are used to set model specifications like complexity and learning rate and therefore have a significant effect on model performance (Y. Hu, Koren, and Volinsky 2008). The hyperparameters are tuned using a grid search process in which multiple models with different combinations of hyperparameters are made and assessed for performance. The levels of the hyperparameters are chosen to cover a wide variety of possible models and are in line with the work of Hahsler (2022). The models are made and tested using cross-validation with an 80% training ratio and 5 folds. The hyperparameters used for the matrix factorization algorithms are given in Table 5 below.

Table 5: Tuning Parameter with Explanation and Values

Parameter	Use	Values
Lambda/Gamma	Model regularization; how well the model should fit the training data. Higher lambda/gamma allows for a better fit to the training with more complexity at the risk of overfitting.	Lambda: 0.1, 0.5, 1, 5 Gamma: 0.25, 0.5, 0.75, 1
Alpha	Learning rate; step size of training process. Bigger steps can result in faster convergence, but can also lead to more instability.	0.001 (default
Nstepts/min. epochs and max. epochs	Number of iterations during the training process. It controls model convergence and trade-off between model performance and training time.	Nsteps: 10 (default) Min/max epochs: 50/200 (default)
k	Number of latent factors in the model. More latent factors create larger item and user matrices with more complexity, but is also more computationally expensive.	5, 10, 25, 50

4.1.2 ALS with Implicit Data

The models using ALS and implicit data are tuned first and the results are shown in Figure 5 and 6. The results for tuning the number of latent factors and the degree are shown in Figures 3 and 4, respectively. When applying the grid, it shows for both variants of the model that the number of factors approaches a positive exponential decaying relation with both prediction precision and recall. This holds that extra latent factors result in higher precision and recall and that the first additional latent factors contribute relatively more than later additional latent factors.

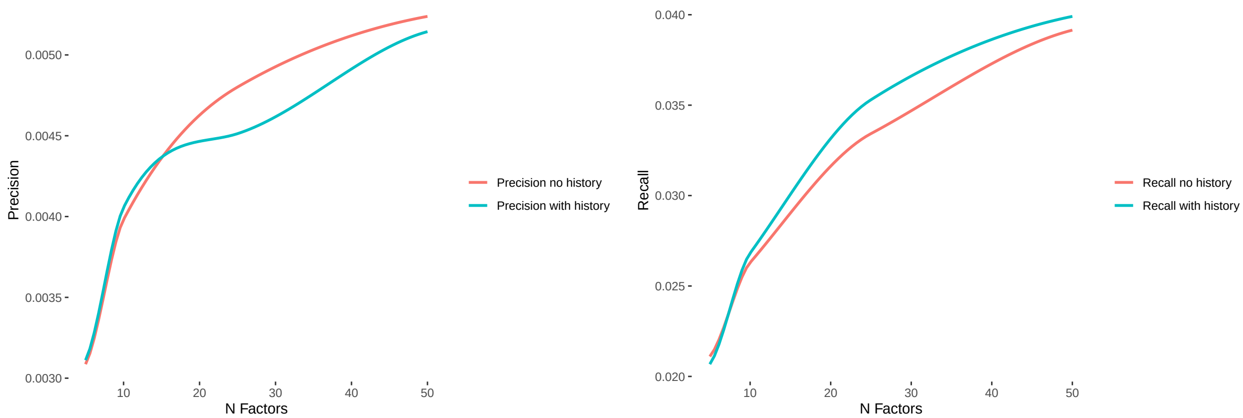

Figure 5: Tuning of the number of Latent Factors

The results for tuning the regularization term show a negative trend for a higher level of regularization. This means that the models relatively quickly suffer from overfitting which causes lesser prediction performances. The highest precision and recall are, on average, achieved at lower levels of the regularization term.

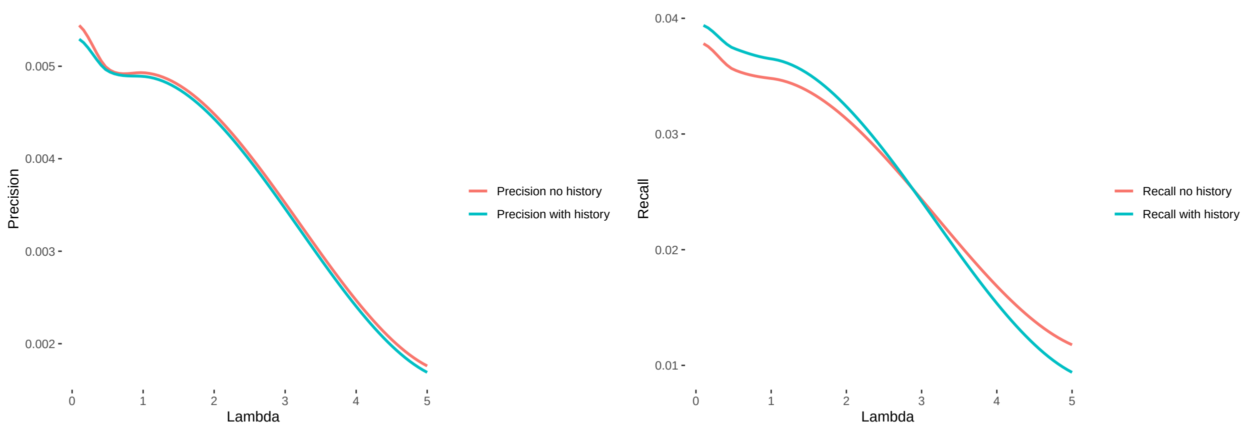

Figure 6: Tuning of the degree of Regularization

The tuning of the ALS models indicate that the best prediction performances are obtained with 50 latent factors and the degree of regularization set to 0.1. The predicting results for the model using these hyperparameters are shown in Table 6 below. Table 6 includes the accuracy and quality performance of the ALS model with and without user information included, and the difference in performance of the model with user information included compared to the model with user information excluded. Table 6 shows that, in a list of 20 recommendations, the model with user history included has a precision of 0.007 and recall of 0.0445 and the model with user history excluded a precision of 0.0013 and recall of 0.0084. The model with user history included is thus more accurate in predicting relevant articles. This is in line with Burke (2007) among others that claim that more user information leads to more complex user profiles and more relevant recommendations.

Both models perform comparable for the quality measures with the model with user history included performing marginally better. Altough the model with user history included recommends more novel items, the serendipity score is similar for both models which means that most novel items are not relevant recommendations.

Table 6: ALS Recommendation Accuracy

	ALS User History excluded	ALS User History included	Difference
TP	0.0264	0.1392	427.27%
FP	3.7662	19.8608	427.34%
FN	0.5421	2.8588	427.36%
Precision	0.0013	0.0070	438.46%
Recall	0.0084	0.0445	429.76%
Novelty	0.4038	0.8204	103.17%
Coverage	0.5727	0.5746	0.33%
Serendipity	0.0107	0.1142	967.29%

4.1.3 Funk SVD with Explicit Data

The models using Funk SVD and explicit data are tuned secondly and the results are shown in Figure 7 and 8. Unlike the models using ALS, these models show different results for when user history is included and excluded. The models without user information approach an negative exponential function where all additional latent factors under 30 total have a negative impact on prediction precision and recall and all additional factors over 30 a positive impact. The models with user history included show a slight positive trend where additional latent factors contribute to better prediction precision and recall.

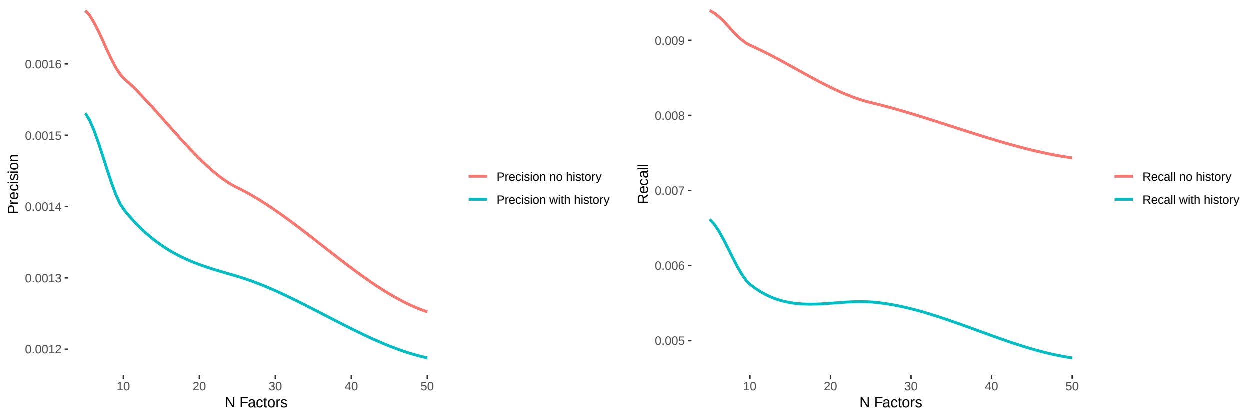

Figure 7: Tuning of the number of Latent Factors

The results for tuning the regularization term for the Funk SVD models are similar to those of the ALS models where there is a downwards trend for higher regularization terms. This indicates that Funk SVD models also relatively quickly suffer from overfitting. Different to the ALS models is that the difference between the history including and excluding model is larger. In addition, model precision shows that one model does not strictly outperform the other one. Models without user information show higher prediction precision and recall on average for all levels of regularization.

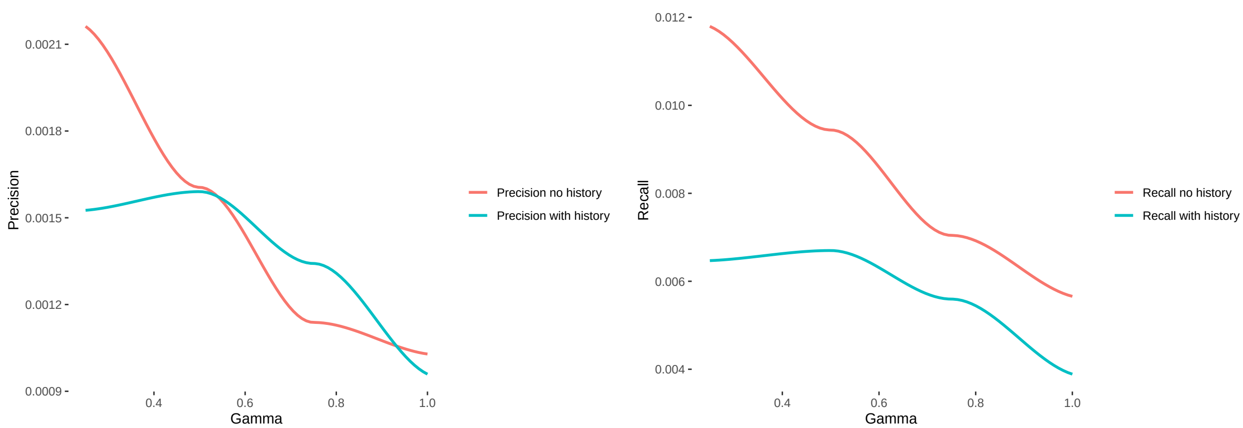

Figure 8: Tuning of the degree of Regularization

The tuning results indicate that the model has, on average, the best predicting performance with the number of latent factors set to 10 and the degree of regularization set to 0.5. The results of the models using these hyperparameters are shown in Table 7 in the same way as the ALS models. Table 7 shows that both precision and recall are higher for the model that excludes user history which is contradictory to the results for the ALS model with implicit data. This means that the model with user history excluded is the more accurate model. Even though it is less accurate, the model with user history included performs better for quality measures, showing higher novelty and coverage than the model with user history excluded.

Table 7: Funk SVD Recommendation Accuracy

	Funk SVD User History excluded	Funk SVD User History included	Difference
TP	0.0358	0.0179	-50.00%
FP	19.9642	19.9821	0.09%
FN	3.6759	3.0596	-16.77%
Precision	0.0018	0.0009	-50.00%
Recall	0.0096	0.0041	-57.29%
Novelty	0.2001	0.4370	118.39%
Coverage	0.4284	0.6093	42.23%
Serendipity	0.0072	0.0078	8.33%

The best performing Funk SVD model has a precision and recall of 0.0018 and 0.0096 compared to 0.0070 and 0.0445 of the best ALS model. This means that the ALS model outperforms the Funk SVD model in situations without limitations. When comparing the models with limited user history, the Funk SVD model is slightly more accurate and the ALS model provides higher quality recommendations. Although, these results are not fully in line with the null hypothesis which states that CF models using explicit data provides the best performance, they do provide interesting insights regarding recommender performance with limited user data.

Figure 9 below shows the number of recommendations made by all models per category. This is used to further investigate the topical coverage of both recommendation models and to what extent they suffer from popularity bias. It shows that the categories “Sports”, “News” and “Food and Drinks” are the most recommended categories with differences between the two methods. Figure 9 also confirms that the ALS models produce recommendations with higher topical coverage, which is in line with the prediction results. Compared to the number of mentions of each category shown in Figure 3, “Sports” is relatively more recommended than it is mentioned. In addition, the lifestyle and finance categories are relatively infrequently recommended compared to how many times the article is mentioned. This indicates that these models are not significantly affected by popularity bias.

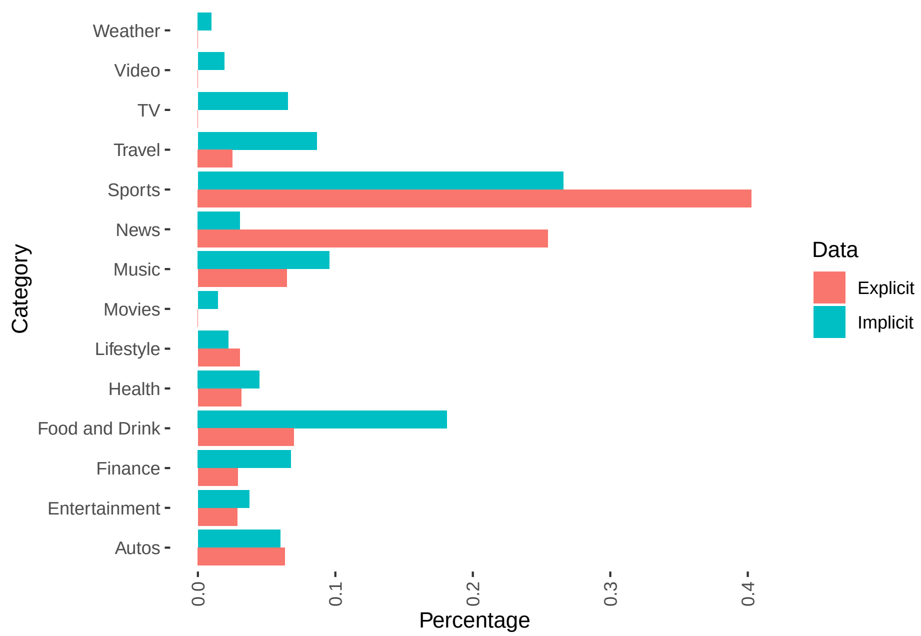

Figure 9: Number of Recommended Articles per Category for Collaborative Filtering Recommendation

4.2 Content-Based recommendation

Recommendations using a CB approach are done using TF-IDF and cosine similarity as described in the methods section. The results of this approach are shown in Table 7 in the same way as the CF models. The results in Table 7 show that the model with user history included has higher precision but lower recall than the model with user history excluded. In addition to this, the model with user history excluded has higher topical coverage in its recommendations than the model with use history included. Both models return the same ratio of novel recommendations. However, the serendipity is 0 for both models which holds that the correct recommendations were not novel. This is in line with the hypothesis which states that CB recommenders suffer more from the popularity bias. Considering all results, it is indicated that the model with user history excluded performs relatively better than the model with user history included.

Table 8: Content-Based Recommendation Accuracy

	User History excluded	User History included	Difference
TP	0.0618	0.0816	32.04%
FP	19.6916	19.6873	-0.02%
FN	4.0135	8.0689	101.04%
Precision	0.0031	0.0040	29.03%
Recall	0.0141	0.0092	-34.75%
Coverage	0.5727	0.3918	-31.59%
Novelty	0.4038	0.4009	-0.72%
Serendipity	0.0000	0.0000	0.00%

When comparing the CB result to those of CF, it is seen that that the precision for the best possible CB model is higher, and the recall is lower than that of the best possible ALS model. This holds that the CB models make relatively less false positive recommendations at the cost of not finding all relevant items. With regard to the quality measures, all CF models outperform the CB models which means that the CF model produces higher quality recommendations.

The number of recommended articles per category is also calculated for the CB recommendations and presented in Figure 10. It shows, similarly to the CF recommendations, that “Sports”, “News” and “Food and Drinks” are the most recommended categories. Compared to the CF recommendations, the recommendations are more evenly distributed across all categories. This means that the CB model suffers less from popularity bias compared to the CF models which is in line with the null hypothesis.

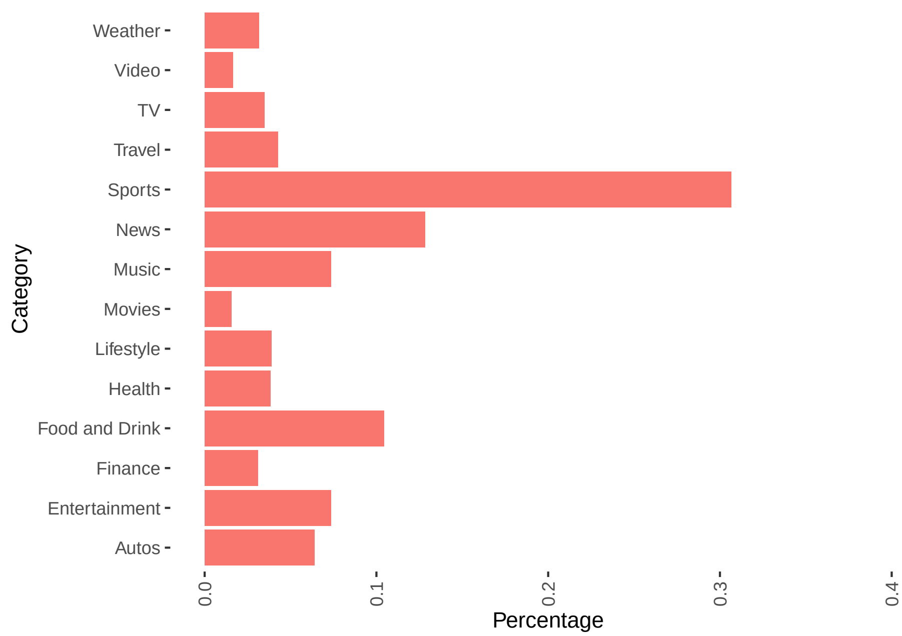

Figure 10: Number of Recommended Articles per Category for Content-Based Recommendation

5 Discussion of the Results

The purpose of this research is to gain better understanding of how restrictions in obtaining user information affect recommender systems with news recommendation in specific. The motivation for this research problem is the increase in data information collection restrictions in the form of the GDPR among others. The effect of these restrictions is investigated by testing several recommendation approaches on both accuracy and quality. The results from this approach and the implication that they have are discussed in this section.

5.1 Accuracy

The prediction accuracy of the most accurate CF model using implicit data is higher than that of the most accurate CF models using explicit data. This is contrary to the corresponding null hypothesis which states that using explicit data should result in more complex user profiles and therefore more accurate predictions. Although these findings are not in line with those of Zheng et al. (2018), they are with those of Feng et al. (2020) who find improved accuracy for news recommendation when using implicit feedback instead of explicit. A possible explanation for these results is that transforming the data gives too much weight to the positive data points which skews the estimation of latent factors. This theory is supported by the relatively large increase in false positive predictions made by the models using explicit data compared to the models with implicit data. These results indicate that adding context by transforming implicit data into explicit data yields a disadvantageous effect on prediction accuracy and is therefore not a useful technique.

The prediction of the most accurate CF model is also higher than that of the most accurate CB model. This provides some evidence in favor of the null hypothesis which states that CB models are outperformed by CF models. Overall, the null hypothesis is rejected as CF models with explicit data do not provide more accurate recommendations than models using implicit data which in not in line with the null hypothesis. Evidence from these results indicates that the most accurate approach for news recommendation is CF with implicit data. This indicates that user information improves the accuracy of news recommendations. A practical implication of this is that the collection user information remains relatively important for online news platforms to ensure performance, although it can be done implicitly which is easier than explicitly.

In the work of F. Wu et al. (2020), a benchmark precision and recall of 0.048 and 0.018 were obtained when applying ALS on a dataset for movies. This research found a precision and recall of 0.007 and 0.045 when applying ALS which is significantly lower. The precision and recall were also lower than the benchmark for the SVD model where F. Wu et al. (2020) obtained 0.091 and 0.033 compared to the 0.001 and 0.009 in this research. The lower accuracy can be explained by the difference in recommended item. Where news articles can have a relatively rapid decay time, that of movies is relatively slower and thus stay relevant for longer. For example: where a year-old movie can still be considered recent, a week-old news article can already be outdated and thus irrelevant. This is quantitative evidence for the claim of Mitchelstein and Boczkowski (2009) who describe the importance of recency in news articles. As a consequence, online news platforms should provide recommendations that are recent enough to be relevant.

5.2 Quality

Similarly to the results for prediction accuracy, CF models outperform CB models for prediction quality. The best performing models for both explicit and implicit data outperform the best CB model for quality measures. The null hypothesis states that CF models on average return higher quality recommendations than CB models. These findings are thus in line with the null hypothesis, and it is therefore not rejected. Quality measures were added because they are often neglected even though they are important for providing a good user experience. Apart for the Funk SVD model, the results for prediction accuracy do not point to a different outcome than the results for prediction quality. This brings into question how useful these measures are. However, even though the quality measures provide the same results in model selection, it still gives relevant insight on how the recommendation model performs that accuracy measures do not. An example of this is the measure for serendipity which is not possible to derive from prediction accuracy alone and has a significant impact on the user experience according to Gemmis et al. (2015) and Ziarani and Ravanmehr (2021). This sort of information is useful for online platforms to publish the type of content desired by their audience without falling victim to click-bait or duplication as described by Raza and Ding (2022).

Despite scoring lower for coverage compared to CF models, the recommendations for CB models are more equally distributed between all the categories. This indicates that CB recommendations are less affected by how many times an article is read in the training data or how many articles there are per category. This evidence that CB recommenders are less prone to popularity bias than CF recommenders which is in accordance with the work of Liu, Dolan, and Pedersen (2010). News publishers who have a high variety in popularity between their articles can favor CB recommenders over CF recommenders to prevent popular articles becoming overly represented in the recommendations.

5.3 Data Sparsity

On most aspects, CF models that include user history perform significantly better than the ones without user information. The ALS model with implicit data performs significantly better with than without user information on both prediction accuracy and quality. These results can be explained by how CF models perform better when more user information is available as this allows them to make more complex user profiles to estimate preferences as described in Burke (2007). The opposite can be seen for the Funk SVD models with explicit data. Although prediction quality is also lower, the accuracy is slightly higher when user history is excluded. A possible explanation for this is that by leveraging additional information from the explicit data, the model with user history overfits on the training data which results in lesser prediction performance. This result gives evidence that transforming data from implicit to explicit could be a viable technique to improve prediction accuracy when limited user information is available. This result is in line with the findings of Zheng et al. (2018) who achieve improved prediction accuracy by augmenting implicit feedback. As a result, transforming implicit data to explicit can improve news article recommendation accuracy and can therefore be a useful technique in specific situations like when limited user information is available.

Similarly to the Funk SVD model, the CB model that excludes user information performs better than the one that includes it for both accuracy and quality. This can be explained by how CB recommenders perform worse for users with a varying taste as mentioned by Liu, Dolan, and Pedersen (2010). By only considering the current user information, and therefore current taste, CB models only find articles that fit the current preferences of the user. These results are in line with the null hypothesis which states that CB approaches are more resilient to data sparsity. Based on these findings, CB are the best option for news platforms when they are unable to collect sufficient user information. Although the performance for both accuracy and quality is not as good as for the best possible CF method, CB methods are still able to provide relatively good recommendations without it.

5.4 Concluding

The main research question was aimed at finding out how relevant news recommendations are made by recommender systems in situations with limited access to user information. The findings above show that recommender systems without user information are less succesful in recommending relevant news articles compared to systems that have access to user information. The recommendation quality is also lower for recommender systems without access to user information where especially novelty and serendipity suffer from a lack of user information. However, the findings of this research show that it is still possible to provide relevant recommendations with less user information. By employing CB recommendation, online news recommendation can still provide relevant recommendation whilst complying with GDPR guidelines or catering to privacy-conscious users.

6 Conclusion

In this last section, the process of this research and the results are summarized. The limitations of this research are also discussed and used to make recommendations for future work.

6.1 Summary and Key Findings

The transition from offline to online has changed the way news is consumed and published. Recommenders are essential for users to cope with the amount supply and find what they are looking for in a timely matter. However, recommender systems depend on user input to provide relevant recommendations and new regulations are straining the collection of user information. The goal of this research is to investigate the recommendation of news articles by recommender systems with limited user information available. Data from MIND consisting of user impressions on the Yahoo News website was used to investigate this matter. The analysis was done by performing ALS and Funk SVD as CF methods and TF-IDF with cosine similarity as CB method. For the Funk SVD model, the implicit data was transformed to explicit data to try creating more context. For each method, a model with user information and one without it was made. The difference in performance between the two gives insight into the effect of limited user information. Performance was measured in accuracy in precision and recall and in quality in coverage, novelty and serendipity.

Results show that CF methods are the most accurate with ALS with implicit data the most accurate approach. This is not in line with the null hypothesis that states that explicit data leverages more context and therefore provides more accurate recommendations. This finding underlines the importance of user information in making recommendations since the methods that use user information perform better than those that use item characteristics.

Next to higher accuracy, the results also show that the recommendation from CF score higher on quality which is in line with the null hypothesis. Quality measures were also evaluated in the analysis because they are often neglected in similar analyses even though they provide useful insight into the performance of the recommendations. These results do not provide evidence that including these measures leads to improved model selection. However, these measures provide useful insight that cannot be derived by only the accuracy measures and help avoid in avoiding duplication of content and click-baiting.

Regarding resilience to data sparsity, CB models perform relatively best when limited user information is available, also in line with the null hypothesis. CB is also the only method that shows an increase in performance rather than a decrease when excluding user information. This result indicates that it is still possible to make relevant recommendations with limited user information available.

Considering all findings from the analysis, it can be concluded that recommender system can be effective in situations without user information, albeit to a lesser extent than when user information is available. The results show that although user information remains vital for making relevant recommendations, the dependency on it can be reduced by applying CB methods. However, this also means that This is a somewhat unexplored part of the field and therefore requires future studies to validate these findings. On the existing research, this work validates earlier findings from F. Wu et al. (2020) and Zheng et al. (2018) by finding comparable results with the same approaches. Additionally this work extends on the work of Raza and Ding (2022) by finding similar results to F. Wu et al. (2020) whilst also exploring qualitative, beyond accuracy measures. By doing so, it also builds on the work of Gemmis et al. (2015) and Ziarani and Ravanmehr (2021) which investigates serendipity in recommender systems.

6.2 Limitations and Future Work

This research knows a few limitations and shortcomings. This section discusses them and provides solutions to be applied in future research.

This first limitation is related to recency. As described by Mitchelstein and Boczkowski (2009), recency plays an important role for news recommendation as online news has broken the twice-a-day news cycle. In this research, publishing time has not been considered in making recommendations as recency is less of an important factor for other categories than news. By excluding news articles for recommendations that are not recent, possible relevant recommendations made by the recommender system. For example, when a user reads a movie review for a new film which he likes, a recommendation for a movie review for a similar, but older film would be excluded while this could still be a relevant suggestion. In future research, recency can be taken into account by only considering news articles in the news or weather category. As recency plays an significant role for these categories, accuracy can be improved by incorporating recency in the model.

The second limitation is related to the use of hybrid methods. Schein et al. (2002) describes how all methods suffer from a form of the cold-start problem. Burke (2007) and Thorat, Goudar, and Barve (2015) describe in his overview of recommender systems methods that this can be overcome by combining multiple methods into a hybrid method. In this recommender system hybrid, the strengths of one component are used to overcome the weaknesses of the other. As an example, when limited user information is available, a hybrid recommender system consisting of a CF and CB method would in first instance be more based on the results from the CB recommender. Hybrid systems are not included in this research as they fall outside of its scope. The aim was to find out what method is most resilient to data sparsity. According to the work of Burke (2007) and Thorat, Goudar, and Barve (2015), applying hybrid models would most likely result in the best result and would therefore not give insight into what component contributes to creating relevant recommendations. Future research can overcome this by solely focusing on hybrids which consists of all possible component combinations. This holds that hybrid models are created using two CF methods, two CB methods or any combination in between.

The third and last limitation identified in this research is related to the methods used for the CB recommendation. For this part, TF-IDF is applied to article titles in combination with cosine similarity to make recommendations which consist of articles with similar titles. As a title never fully captures the complete context of an article, a lot of context is not taken into account in this recommending approach. This can be overcome by applying the same approach to the abstract text instead of the title. This is not done in this research as this would significantly increase the computational costs whilst most likely not providing the same level of benefits in the form of additional insight. A second limitation related to the CB recommendation is the method itself. More elaborate and complex models for CB recommending have been developed and presented in the works of F. Wu et al. (2020) and Zheng et al. (2018) among others. One of the most contribution to the developments they made is the use of neural nets in their recommender systems to derive more latent context from articles which is used to create a better understanding of how an article relates to other articles. These methods are not applied in this research for the same reasons as for not using abstract text.

Literature List

Adomavicius, Gediminas, and Alexander Tuzhilin. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions." *IEEE Transactions on Knowledge and Data Engineering* 17: 734–49.

Aguirre, Elizabeth, Dominik Mahr, Dhruv Grewal, Ko De Ruyter, and Martin Wetzels. 2015. "Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness." *Journal of Retailing* 91 (1): 34–49.

Aguirre, Elizabeth, Anne L Roggeveen, Dhruv Grewal, and Martin Wetzels. 2016. "The Personalization-Privacy Paradox: Implications for New Media." *Journal of Consumer Marketing*.

Allan, Stuart. 2006. *EBOOK: Online News: Journalism and the Internet*. McGraw-Hill Education (UK).

Balabanović, Marko, and Yoav Shoham. 1997. "Fab: Content-Based, Collaborative Recommendation." *Communications of the ACM* 40 (3): 66–72.

Blanchett Neheli, Nicole. 2018. "News by Numbers: The Evolution of Analytics in Journalism." *Digital Journalism* 6 (8): 1041–51.

Bucy, Erik P. 2003. "The Interactivity Paradox: Closer to the News but Confused." In *Media Access*, 67–92. Routledge.

Burke, Robin. 2007. "Hybrid Web Recommender Systems." *The Adaptive Web: Methods and Strategies of Web Personalization*, 377–408.

Carrascal, Juan Pablo, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo De Oliveira. 2013. "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online." In *Proceedings of the 22nd International Conference on World Wide Web*, 189–200.

European Commission. 2016. "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

Feng, Chong, Muzammil Khan, Arif Ur Rahman, and Arshad Ahmad. 2020. "News Recommendation Systems-Accomplishments, Challenges & Future Directions." *IEEE Access* 8: 16702–25.

Fischer, Gerhard. 2001. "User Modeling in Human–Computer Interaction." *User Modeling and User-Adapted Interaction* 11: 65–86.

Flaxman, Seth, Sharad Goel, and Justin M Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (S1): 298–320.

Forman-Katz, Naomi, and Katerina Eva Matsa. 2022. "News Platform Fact Sheet." *Pew Research Center's Journalism Project*. Pew Research Center. https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/.

Funk, Simon. 2006. "Netflix Update: Try This at Home."

Gemmis, Marco De, Pasquale Lops, Giovanni Semeraro, and Cataldo Musto. 2015. "An Investigation on the Serendipity Problem in Recommender Systems." *Information Processing & Management* 51: 695–717.

Goldberg, David, David Nichols, Brian M Oki, and Douglas Terry. 1992. "Using Collaborative Filtering to Weave an Information Tapestry." *Communications of the ACM* 35 (12): 61–70.

Hahsler, Michael. 2022. "Recommenderlab: An r Framework for Developing and Testing Recommendation Algorithms." *arXiv Preprint arXiv:2205.12371*.

Han, Jiawei, Micheline Kamber, Jian Pei, et al. 2012. "Getting to Know Your Data." In *Data Mining*, 2:39–82.

Han, Jungkyu, and Hayato Yamana. 2017. "A Survey on Recommendation Methods Beyond Accuracy." *IEICE TRANSACTIONS on Information and Systems* 100 (12): 2931–44.

Harcup, Tony, and Deirdre O'neill. 2017. "What Is News? News Values Revisited (Again)." *Journalism Studies* 18 (12): 1470–88.

Hastie, Trevor, Rahul Mazumder, Jason D Lee, and Reza Zadeh. 2015. "Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares." *The Journal of Machine Learning Research* 16: 3367–3402.

Hildebrandt, Mireille. 2022. "The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems." *Frontiers in Artificial Intelligence*, 73.

Hu, Xuehui, and Nishanth Sastry. 2019. "Characterising Third Party Cookie Usage in the EU After GDPR." In *Proceedings of the 10th ACM Conference on Web Science*, 137–41.

Hu, Yifan, Yehuda Koren, and Chris Volinsky. 2008. "Collaborative Filtering for Implicit Feedback Datasets." In *2008 Eighth IEEE International Conference on Data Mining*, 263–72.

Ilievski, Ilija, and Sujoy Roy. 2013. "Personalized News Recommendation Based on Implicit Feedback." In *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*, 10–15.

Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender Systems: An Introduction*. Cambridge University Press.

Karimi, Mozhgan, Dietmar Jannach, and Michael Jugovac. 2018. "News Recommender Systems–Survey and Roads Ahead." *Information Processing & Management* 54 (6): 1203–27.

Koren, Yehuda, Robert Bell, and Chris Volinsky. 2009. "Matrix Factorization Techniques for Recommender Systems." *Computer* 42 (8): 30–37.

Lawson-Borders, Gracie. 2006. *Media Organizations and Convergence: Case Studies of Media Convergence Pioneers*. Routledge.

Liu, Jiahui, Peter Dolan, and Elin Rø{}nby Pedersen. 2010. "Personalized News Recommendation Based on Click Behavior." In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, 31–40. Association for Computing Machinery. https://doi.org/10.1145/1719970.1719976.

Lü, Linyuan, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. 2012. "Recommender Systems." *Physics Reports* 519 (1): 1–49.

Madden, Mary. 2014. "Public Perceptions of Privacy and Security in the Post-Snowden Era."

Maddodi, Srivatsa et al. 2019. "NETFLIX Bigdata Analytics-the Emergence of Data Driven Recommendation." *Srivatsa Maddodi, & Krishna Prasad, K.(2019). Netflix Bigdata Analytics-The Emergence of Data Driven Recommendation. International Journal of Case Studies in Business, IT, and Education (IJCSBE)* 3 (2): 41–51.

Malhotra, Naresh K, Sung S Kim, and James Agarwal. 2004. "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model." *Information Systems Research* 15 (4): 336–55.

Mitchelstein, Eugenia, and Pablo J Boczkowski. 2009. "Between Tradition and Change: A Review of Recent Research on Online News Production." *Journalism* 10 (5): 562–86.

Montgomery, Alan L, and Michael D Smith. 2009. "Prospects for Personalization on the Internet." *Journal of Interactive Marketing* 23 (2): 130–37.

Mooney, Raymond J, and Loriene Roy. 2000. "Content-Based Book Recommending Using Learning for Text Categorization." In *Proceedings of the Fifth ACM Conference on Digital Libraries*, 195–204.

Nguyen, Luong Vuong, Tri-Hai Nguyen, Jason J Jung, and David Camacho. 2021. "Extending Collaborative Filtering Recommendation Using Word Embedding: A Hybrid Approach." *Concurrency and Computation: Practice and Experience*, e6232.

Plisson, Joël, Nada Lavrac, Dunja Mladenic, et al. 2004. "A Rule Based Approach to Word Lemmatization." In *Proceedings of IS*, 3:83–86.

Raza, Shaina, and Chen Ding. 2022. "News Recommender System: A Review of Recent Progress, Challenges, and Opportunities." *Artificial Intelligence Review*, 1–52.

Resnick, Paul, and Hal R Varian. 1997. "Recommender Systems." *Communications of the ACM* 40 (3): 56–58.

Ricci, Francesco, Lior Rokach, and Bracha Shapira. 2010. "Introduction to Recommender Systems Handbook." In *Recommender Systems Handbook*, 1–35. Springer.

Sanchez-Rola, Iskander, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. "Can i Opt Out yet? Gdpr and the Global Illusion of Cookie Control." In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, 340–51.

Schedl, Markus, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. "Current Challenges and Visions in Music Recommender Systems Research." *International Journal of Multimedia Information Retrieval* 7: 95–116.

Schein, Andrew I, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. 2002. "Methods and Metrics for Cold-Start Recommendations." In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 253–60.

Schwartz, Barry, and Barry Schwartz. 2004. "The Paradox of Choice: Why More Is Less." In. Ecco New York.

Shoemaker, Pamela J, Tim P Vos, and Stephen D Reese. 2009. "Journalists as Gatekeepers." In *The Handbook of Journalism Studies*, 93–107. Routledge.

Smith, Brent, and Greg Linden. 2017. "Two Decades of Recommender Systems at Amazon. Com." *Ieee Internet Computing* 21 (3): 12–18.

Stocking, Galen, and Maya Khuzam. 2023. "Digital News Fact Sheet." *Pew Research Center's Journalism Project*. Pew Research Center. https://www.pewresearch.org/journalism/fact-sheet/digital-news/#economics.

Sundar, S Shyam, and Sampada S Marathe. 2010. "Personalization Versus Customization: The Importance of Agency, Privacy, and Power Usage." *Human Communication Research* 36 (3): 298–322.

Tankard, Colin. 2016. "What the GDPR Means for Businesses." *Network Security* 2016 (6): 5–8.

Thorat, Poonam B, Rajeshwari M Goudar, and Sunita Barve. 2015. "Survey on Collaborative Filtering, Content-Based Filtering and Hybrid Recommendation System." *International Journal of Computer Applications* 110: 31–36.

Trstenjak, Bruno, Sasa Mikac, and Dzenana Donko. 2014. "KNN with TF-IDF Based Framework for Text Categorization." *Procedia Engineering* 69: 1356–64.

Verhoef, Peter C, Werner J Reinartz, and Manfred Krafft. 2010. "Customer Engagement as a New Perspective in Customer Management." *Journal of Service Research* 13 (3): 247–52.

Wang, Xuejian, Lantao Yu, Kan Ren, Guanyu Tao, Weinan Zhang, Yong Yu, and Jun Wang. 2017. "Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration." In *Proceedings of the 23rd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2051–59.

Wu, Chuhan, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. "Neural News Recommendation with Multi-Head Self-Attention." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6389–94.

Wu, Fangzhao, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, et al. 2020. "Mind: A Large-Scale Dataset for News Recommendation." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3597–3606.

Zerlang, Jesper. 2017. "GDPR: A Milestone in Convergence for Cyber-Security and Compliance." *Network Security* 2017 (6): 8–11.

Zhang, Shuai, Lina Yao, Aixin Sun, and Yi Tay. 2019. "Deep Learning Based Recommender System: A Survey and New Perspectives." *ACM Computing Surveys (CSUR)* 52: 1–38.

Zheng, Guanjie, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. "DRN: A Deep Reinforcement Learning Framework for News Recommendation." In *Proceedings of the 2018 World Wide Web Conference*, 167–76. International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3178876.3185994.

Ziarani, Reza Jafari, and Reza Ravanmehr. 2021. "Serendipity in Recommender Systems: A Systematic Literature Review." *Journal of Computer Science and Technology* 36: 375–96.