Analysing Viewer Engagement and Preferences in Anime Streaming Platforms

MSc Research Project
Data Analytics

Subash Ayyalusamy
Student ID: 22162933

School of Computing
National College of Ireland

Supervisor: Teerath Kumar Menghwar

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Subash Ayyalusamy
Student ID:	22162933
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Teerath Kumar Menghwar
Submission Due Date:	14/12/2023
Project Title:	Analysing Viewer Engagement and Preferences in Anime Streaming Platforms
Word Count:	8882
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Subash Ayyalusamy
Date:	26th January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Analysing Viewer Engagement and Preferences in Anime Streaming Platforms

Subash Ayyalusamy
22162933

Abstract

This study examines anime recommendation systems, including collaborative filtering embedded with neural networks and content-based filtering enhanced with the K-Nearest Neighbours algorithm. Jupyter Notebook simplifies the process of model training and validation systematically. The evaluation process focuses on assessing the RMSE scores, training outcomes, and visual representations, which provide user preferences and top recommendations. The studies are divided into categories, including the state of the anime, the distribution of user ratings, and the RMSE values of content-based filtering. An analysis of the advantages of collaborative filtering versus content-based filtering is conducted, resulting in the development of a hybrid model for recommendation. The study promotes the importance of optimizing algorithms for efficiency, exploring different domains, and providing real-time recommendations. It highlights the need to include measures other than Root Mean Square Error (RMSE) to measure user engagement. The addition of comprehensive research references and GitHub links for code access enhances the overall comprehension of recommendation systems.

1 Introduction

1.1 Streaming Era and the Significance of Recommendations

Deep learning has been successful in range of the domains such image processing Kumar et al. (2022); Kumar, Mileo, Brennan and Bendechache (2023); Kumar et al. (n.d.); Roy et al. (2022); Ranjbarzadeh et al. (2023); Aleem et al. (2022); Kumar, Park, Ali, Uddin and Bae (2021); Turab et al. (2022); Singh, Ranjbarzadeh, Raj, Kumar and Roy (2023); Kumar, Park, Ali, Uddin, Ko and Bae (2021); Singh, Raj, Kumar, Verma and Roy (2023); Chandio et al. (2022); Khan et al. (2022); Roy et al. (2023), natural language processing (NLP) Kumar, Turab, Raj, Mileo, Brennan and Bendechache (2023) and audio Kumar, Turab, Mileo, Bendechache and Saber (2023); Chandio et al. (2021); Park et al. (2020); Kumar et al. (2020); **?**. Among those streaming sector has been on high demand. It has undergone significant expansion in recent years, fundamentally reshaping the worldwide media consumption environment. Lee et al. (2015) This evolution is particularly notable in the realm of anime, a genre that has seen a surge in popularity beyond its traditional Japanese audience. The unique narrative styles, diverse genres, and rich cultural details of anime have brought in a dedicated global fanbase, necessitating sophisticated recommendation systems to meet the needs of a wide range of viewers. This introduction explores the significance of developing a cutting-edge anime recommendation system using filtering with a K-Nearest Neighbors (KNN) machine-learning model in Jupyter Notebook, aimed at enhancing viewer engagement in the streaming era.

1.2 The Role of Recommendation Systems in Streaming Platforms

Streaming services like Netflix, Crunchyroll, and Hulu have changed how people get to and interact with content. Unlike traditional broadcast media, streaming services give users a lot of options. This means that personalized suggestions are very important for improving the user experience while retaining them as a customer. With so much content to choose from, the recommendation system is not just a tool; it's an important part of making sure viewers find content that fits their tastes and interests (Girsang et al.; 2020). Soni et al. (2023)

1.3 Challenges in Anime Recommendation

Anime's complicated plots and wide range of themes make it hard for suggestion systems to work properly. There are many sub-genres within this genre, from action-packed shonen to emotionally charged shojo, as well as special genres like mecha, isekai, and a slice of life. Because of this, it's important for recommendation algorithms to correctly understand and guess what viewers will like which is more than just putting series into genres (Ota et al.; 2017).

1.4 Methods and Tools:

Both collaborative filtering and content-based filtering are cornerstone methodologies in building recommendation systems, and each one offers unique advantages. Collaborative filtering leverages user interaction data to suggest items by identifying patterns across users. The underlying assumption is that if users A and B rated an item similarly, they will respond similarly to other items. As you mentioned, techniques like the k-nearest Neighbors (KNN) algorithm can be very effective here. KNN looks for the most similar users (or items) to the one trying to make a prediction for and bases its recommendation on their preferences. Content-based filtering, on the other hand, focuses on the attributes of the items themselves. Instead of relying on user similarity, it utilizes the characteristics of previously consumed items to recommend new ones. For instance, if a user likes certain anime with specific themes or created by particular studios, the content-based filtering system would recommend anime with similar attributes.

The Jupyter Notebook must be used as a work tool in this project. Jupyter Notebook is a platform for creating, testing and deploying machine learning models that is known for being flexible in how it analyzes data and learns from it. Its ability to mix live code, visuals, and narrative text makes it a great place to test and improve the KNN model for this anime recommendation system Isinkaye et al. (2015).

1.5 Research Question

How can the integration of collaborative filtering embedded with neural networks and content-based filtering facilitated by the K-Nearest Neighbours (KNN) algorithm, optimize the accuracy and effectiveness of anime recommendation systems, and what are the implications for enhancing user satisfaction and engagement in this specialized domain?

1.6 Document Structure

The research is divided into several sections, starting with an introduction that explains the importance of anime recommendation systems. The methodology section describes the use of collaborative filtering embedded with neural network and content-based filtering using the K-Nearest Neighbours algorithm, using Jupyter Notebook for analysis. The evaluation section thoroughly examines the models, placing particular emphasis on RMSE scores and training outcomes, accompanied by visual representations of user preferences and top recommendations. Detailed classifications, such as the status of anime, the distribution of user ratings, and the RMSE values of content-based filtering, provide a wide range of valuable information. The debate critically compares CF versus content-based filtering, highlighting their merits and cons. At the end of the study, hybrid models are suggested as a way to improve accuracy. Real-time recommendations are also emphasized, and there is call for a full evaluation of user engagement indicators beyond RMSE. The paper cites pertinent literature, enhancing the overall comprehension of recommendation algorithms, and offers GitHub links for code accessibility.

2 Related Work

Today, recommendation systems are an important part of helping people make decisions when there is a lot of digital material to choose from. Because anime has a wide range of unique fans, the unique application of anime is very important. Two important techniques in this field are collaborative filtering, which uses how people act, and KNN, which is a simple but useful machine-learning algorithm.

2.1 Collaborative Filtering in Recommendation Systems:

In the world of digital media, which changes quickly, recommendation systems are very important for creating the user experience.Shani and Gunawardana (2011) Collaborative filtering has become one of the most effective methods, and it is used by many platforms, from Netflix to Amazon. This part of the literature study is all about how collaborative filtering is used in recommendation systems, the problems it faces, and the progress it has made.

At its core, collaborative filtering is based on the idea that if person A and person B agree on something, then person A is more likely to agree with B on a different problem than a random person. This idea, which was first put forward by Goldberg et al., has been the basis of many ranking systems ever since. This method has changed over time from simple user-item rating matrices to complex algorithms that can handle big datasets. This shows how sophisticated it is becoming.

User-based and item-based collaborative filtering are the most popular types of filtering. (Schafer et al.; 2007) talked about user-based filtering, which looks at past ratings to find users who are like each other. Item-based screening, on the other hand, as explained by Prakash and Arora (n.d.), looks at how similar two items are, which can lead to more scalable solutions for big datasets. Model-based techniques that use machine learning algorithms for prediction have been very helpful in fixing the problems that traditional methods had with being able to scale up and deal with limited information.

Even though collaborative filtering works well, it is not perfect. Schafer et al. (2007) say that the cold start problem, which happens when new users or items don't have enough data to make accurate suggestions, is still a serious issue. The variety of suggestions is also limited by the long-tail effect, which means that unique things are often overlooked. To solve these problems, researchers like Burke (2007) have come up with hybrid models that combine collaborative screening with content-based methods.

Personalization and Privacy: In the quest for personalization, collaborative filtering has evolved to incorporate more sophisticated user models. As Subramaniaswamy et al. (2017) talked about, using individual attributes and context has led to more relevant and personalized suggestions. But, as Cho et al. (2018) investigated, this greater personalization makes people worry about their data and privacy, so there needs to be a balance between the quality of recommendations and people's privacy.

Recent Advancement: The most recent progress in collaborative filtering has been made by using deep learning methods. This trend can be seen in the work by Wang et al. (2014) on neural collaborative filtering, which shows that it is more accurate than traditional matrix factorization methods. In addition, Cantador et al. (2015) investigated how collaborative filtering can be used with other data sources, like social media, which gives us new ways to understand what users want.

2.2 K-Nearest Neighbours (KNN) in Anime Recommendation Systems

The essence of KNN lies in its instance-based learning paradigm, often categorized under lazy learning. This is characterized by its localized functional estimation, deferring computations until the classification stage. The application of KNN in recommendation systems, specifically for anime, stems from its straightforward approach, where it computes the 'distance' between data points and classifies them based on their proximity to neighbors. Given the intricate genres and subcultures within anime, such systems demand nuanced and personalized methodologies, a criterion satisfactorily met by KNN (Prakash and Arora; n.d.).

Central to the KNN algorithm in anime recommendations is the concept of suggesting titles by aligning them with the viewing preferences or patterns of users. For example, if a user exhibits a preference for certain anime titles, the system is designed to recommend other titles favored by viewers with similar tastes. This technique, as elaborated in studies like those by (Das et al.; 2017), often involves the application of similarity metrics such as cosine similarity or Euclidean distance. The proficiency of these metrics in precisely capturing viewer preferences constitutes a significant avenue for ongoing research and development.

Recent studies have been pivotal in exploring the amalgamation of KNN with more sophisticated AI and machine learning techniques. For instance, (Massa and Avesani; 2007), discovered that integrating dimensionality reduction methods can substantially enhance KNN's performance in high-dimensional spaces, which are typical in data pertaining to user-item interactions in recommendation systems. Wang et al. (2014) also talked about how neural networks could be used to improve the input features for KNN. This shows how traditional methods and new AI developments can work together to make things better.

Despite its advantages, KNN faces several impediments in the realm of anime recommendation systems. A primary challenge is scalability, as the algorithm necessitates the storage of all data points, becoming increasingly cumbersome with the expansion of anime libraries. Additionally, the frequent absence of user-item matrices, particularly in niche genres like anime, can diminish the accuracy of recommendations. Aggarwal et al. (2016) emphasized that these issues demand innovative solutions, such as selective data utilization or the integration of KNN with alternative recommendation approaches.

Analyzing KNN's real-world applications in anime recommendation systems offers valuable perspectives. Platforms like MyAnimeList and Crunchyroll, for instance, employ KNN variations in their recommendation engines, tailoring the algorithm to suit their extensive user base and diverse anime content. These platforms often blend user-based and item-based collaborative filtering to heighten the precision and relevance of their recommendations.

The use of KNN in recommendation systems also raises concerns about user privacy and ethical considerations. Given KNN's heavy reliance on user data, safeguarding data security and maintaining user privacy is crucial. Research like that by Polat and Du (2003) on privacy-preserving KNN algorithms provides a foundational approach to balancing effective recommendations with concerns for user privacy.

Future developments in KNN-based anime recommendation systems could focus on several pivotal aspects. These include incorporating real-time user feedback to continually refine the recommendation model and investigating more intricate distance metrics that can more accurately reflect the nuances of anime genres and user preferences. Furthermore, the integration of KNN with emerging technologies such as deep learning and blockchain holds promising prospects for enhanced personalization and security.

2.3 Real-world Anime Recommendation System:

In this digital age, where streaming services and online platforms dominate entertainment consumption, anime recommendation systems have become increasingly vital for enhancing the viewing experience of these users. These systems, employing a range of techniques from basic algorithms to advanced artificial intelligence, aim to curate personalized anime selections for viewers. This section explores various examples of anime recommendation systems, examining their methodologies, effectiveness, and user experience.

MyAnimeList (MAL): MyAnimeList has one of the biggest anime databases and communities, and it has a great method for suggesting shows for you to watch. MAL's system mostly uses ratings and reviews from users to suggest other similar anime titles. The algorithm looks at the user's viewing history and preferences to make suggestions based on similar user profiles and anime ratings. This user-based collaborative filtering approach, as discussed in research by Schafer et al. (2007), lets you make personalized choices, but it can sometimes have trouble with the cold-start problem for new users or anime that isn't very popular.

Crunchyroll As a leading anime streaming service, Crunchyroll incorporates a recommendation engine that blends collaborative filtering with content-based approaches. Their system suggests titles not only based on user viewing patterns but also on anime metadata like genre, release year, and popularity. This hybrid approach, as highlighted in the work of Burke (2007), seeks to balance the depth of content-based filtering with the breadth of collaborative filtering, offering a more rounded recommendation experience.

Netflix: Even though Netflix isn't just an anime site, it has a very smart recommendation system that includes anime titles. It is famous for using complicated algorithms with machine learning and predictive analytics. As part of their larger content library, they offer custom anime suggestions. Gomez-Uribe and Hunt (2016) examined Netflix's strategy, which involves looking at a lot of data to figure out what viewers like and then making suggestions that are highly tailored to each person's tastes.
AniList: AniList is another well-known anime community site with a unique way of making suggestions. AniList is different from other platforms because it uses a graph database to build a network of anime titles based on how users interact with it, such as by reviewing and liking things. This strategy, which is comparable to the ideas discussed in the research by Cantador et al. (2015), uses community involvement and social interactions to make recommendations, making the recommendation process more social and interactive.
Kitsu: The suggestion engine on Kitsu stands out because it focuses on getting users involved and listening to their comments. Users are asked to rate and review books, which the system then uses to make its suggestions better. According to the adaptive systems research by Pazzani and Billsus (2007), this feedback loop is a key part of its algorithm that makes sure the system is always changing and adapting to what users want.
Anime-Planet: User-generated significance is a big part of Anime-Planet's recommendation system, which makes it stand out. Users can make lists of anime they want to watch and share them on the site. The system then looks at these lists to make suggestions. This community-driven approach, as explored in research by Massa and Avesani (2007), depends a lot on user participation and collective intelligence, which leads to a lot of different suggestions.
Hulu and Amazon Prime: Hulu and Amazon Prime are two other popular streaming services that include anime movies in their recommendation systems, even though they don't only show anime. To suggest anime titles along with other types of media, these platforms usually use a mix of joint and content-based filtering, along with their special algorithms. According to Jannach et al. (20), the success of these systems often depends on how well they can correctly classify and understand the unique qualities of anime as a genre within a large library of different types of content.
Different sites have very different anime recommendation systems, which show how different people choose what to watch online these days. These systems, which range from community-driven recommendations to complex AI-powered engines, aim to improve the user experience by making personalized anime ideas. How well these systems work depends on how well they can understand and meet the specific needs of anime watchers. This is something that needs to be constantly improved and changed based on user feedback and new technology.

2.4 Novelty of Our Research

This study's new idea brings something new to the field of anime recommendation systems by carefully comparing the two main types of recommendation methods: collaborative filtering and content-based filtering. K-nearest neighbors(KNN) and the neural network technique are used in this study to carefully test these methods. This algorithm was made to deal with the problems that come up when there is a lot of scattered, high-dimensional data about anime preferences. This study also looks at what these findings mean for the future of recommendation systems. It focuses on the concept of hybrid models that combine the best parts of both approaches. This study is more user-centered and adaptable now that it includes metrics for user engagement and satisfaction and looks into cross-domain suggestions. Not only is it new in terms of technology, but it also has a lot to do with how personalized content suggestions are changing.

3 Research Methodology

The powerful computing power of Jupyter Notebook (python) is used in this study to make and compare two different recommendation system models: content-based filtering and collaborative filtering. If you compare the Root Mean Square Error (RMSE) numbers of these two models, you can find the one that does a better job of recommending anime.

The dataset used in this study is a combination of several CSV files obtained from the Kaggle (open source) 'anime-recommendation-database,' with the key 'animelist.csv' file being the most important one. This fundamental dataset encompasses complex user interactions with anime, including user IDs, anime IDs, and corresponding ratings. The dataset is an extensive collection that includes a wide range of information, such as anime titles, synopses, and users' watching statuses. The 'rating complete.csv' file, which is part of the collection, contains an impressive 71 million entries.It provides a comprehensive record of user ratings, which range from 0 to 10. To guarantee the reliability and strength of the dataset, a strict requirement is in place that users must rate at least 400 anime titles, assuring significant user involvement. The dataset also includes crucial files, such as 'anime csv,' which provides a comprehensive collection of anime titles with detailed information such as anime ID, title, genre, type (TV show, movie, etc.), and number of episodes. Understanding the fundamental knowledge is essential for grasping the inherent qualities and features of the anime content. The 'animelist.csv' file acts as a central point for collaborative filtering models, capturing a wide range of user preferences and opinions. It may also offer valuable information on users' viewing statuses, revealing the extent of their involvement. Furthermore, the dataset contains 'anime with synopsis.csv,' which serves as a collection of synopses for different anime titles. This file plays a crucial role in facilitating content-based filtering tactics by utilizing textual descriptions of anime plots and themes to extract detailed content elements for improved suggestions. The 'watching status.csv' file classifies anime into various status levels, including "Currently Watching," "Completed," "On Hold," "Dropped," and "Planned to Watch." This information provides significant insights about user behaviors and preferences, which can be used to enhance recommendation tactics. The dataset's immense size, varied content, and detailed information on anime titles, user interactions, and content synopses collectively offer a strong and reliable base. This foundation enables a thorough examination and assessment of collaborative and content-based recommendation methods. The meticulous curation and preprocessing of the information are crucial, as they guarantee the extraction of significant patterns and provide well-informed conclusions about the performance and implications of the recommendation systems analyzed in this work.

The initial model in this framework utilizes collaborative filtering, a technique that leverages the combined ratings of users to forecast preferences. It suggests that individuals who have similar patterns of watching content are likely to have similar preferences. The model carefully analyzes user engagements with anime titles, establishing a predictive system that recommends content by analyzing user resemblances. The second model examined is content-based filtering, which differs from the user-centric methodology of collaborative filtering. Instead, it focuses on the underlying characteristics of the anime name themselves, such as genre, topics, or creator. The algorithm utilizes an analysis of the user's preferred anime qualities to suggest new titles that possess comparable attributes. This technique excels at addressing the 'cold start' problem by providing appropriate ideas, even in situations when user data is limited.Both models integrate KNN to refine their predictions, employing a distance metric to ascertain the 'nearness' of either user preferences or content features. KNN's role is pivotal in filtering through the noise and complexity of data to identify the most pertinent recommendations.

The effectiveness of these models is evaluated quantitatively by calculating their RMSE values, a statistical metric that represents the average magnitude of the discrepancies between anticipated and actual ratings. The root mean square error (RMSE) serves as an unbiased metric for comparing models, where lower values correspond to more accuracy and, consequently, a more accurate representation of user preferences. This study carefully compares two models to find out which one works better for an

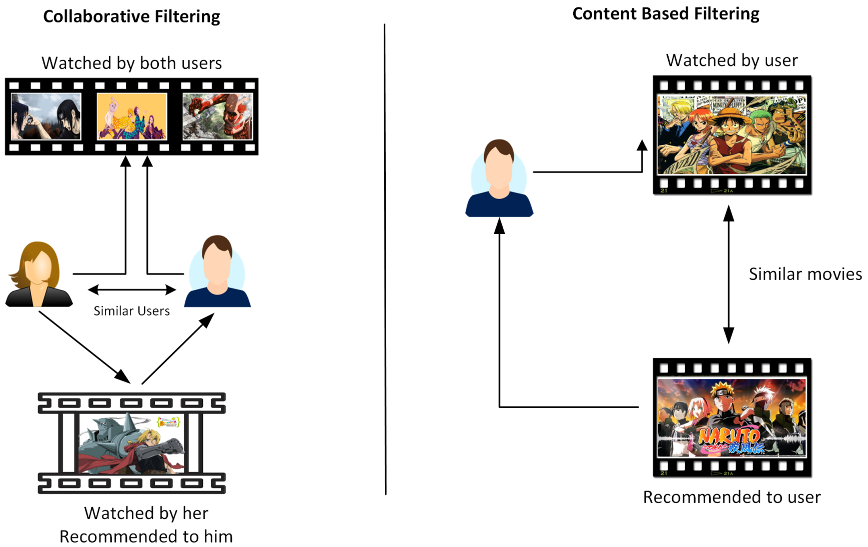

Figure 1: Collaborative Filtering vs Content-Based Filtering

anime recommendation system: collaborative filtering or content-based filtering, both of which are improved by KNN's powerful discrimination. The outcome is anticipated to not only contribute to the theoretical literature on recommendation systems but also to offer practical insights for the development of more user-centric streaming services.

3.1 Collaborative Filtering Approach:

Collaborative filtering is a fundamental technology in the rapidly growing field of information retrieval and personalization. It plays a crucial role in powering the mechanisms of highly effective recommendation systems. The effectiveness and flexibility of this system make it a perfect option for creating customized systems that respond to the complex and diverse field of anime recommendations. By utilizing neural network embeddings, we can capture the complex and latent relationships between users and anime titles. These embeddings serve as dense, low- dimensional representations of high-dimensional data, allowing the neural network to learn nuanced user preferences and item characteristics from the interaction patterns present within the data. Through this method, the recommendation system transcends traditional collaborative filtering limitations, harnessing the power of deep learning to discern subtle patterns and preferences that might not be immediately apparent. The model effectively maps users and anime into a shared latent space, where the distances between points correspond to the predicted affinity between a user and a particular anime title. This enables the system to generate highly personalized recommendations, thereby enhancing the viewing experience for each user by suggesting titles that align closely with their unique tastes and viewing history.

Collaborative filtering is predicated on the concept of shared tastes and preferences within a user community. The fundamental idea of this approach is that persons who have previously agreed on specific topics are likely to agree again in the future. This method of filtering leverages user ratings, interactions, and behaviours to predict what other products or services an individual might like.

The technique can be categorized into two primary branches: user-based collaborative filtering and item-based collaborative filtering. User-based filtering is a method that involves finding similarities between users and suggesting things to a user based on the preferences of similar individuals. On the other hand, item-based filtering recommends products that are comparable to the ones the user has previously shown interest in, considering the preferences of all users.

3.1.1 Mechanics of Collaborative Filtering

Collaborative filtering techniques primarily create a user-item matrix that captures the interactions or ratings provided by users for objects. In an anime recommendation system, this matrix would include of users and anime titles, where the ratings indicate the level of enjoyment users derived from the shows. The computer subsequently calculates predictions by populating the vacant cells of this matrix, denoting the possible level of interest a user may have in unwatched anime titles.

The effectiveness of collaborative filtering depends on the concept of the "neighborhood". User-based techniques involve the system selecting a certain group of users, referred to as "neighbours," who have previously given similar ratings to the target user. Item-based techniques involve examining the "neighbourhood" of products that have been highly evaluated by the user. The predicted bias towards an anime title is commonly determined by a balanced mixture of evaluations from several neighbouring sources.

3.1.2 Challenges in Collaborative Filtering

Although collaborative filtering has its advantages, it is not without its limitations. The "cold start" problem is especially widespread, manifesting when novel users or things are introduced into the system lacking adequate interaction data to derive significant inferences. An additional concern is the scarcity of data in the user-item matrix, as users usually engage with just a limited portion of the complete catalogue, resulting in a mostly incomplete matrix. Furthermore, the issue of scalability arises when the computational workload increases in proportion to the number of users and objects. Therefore, it is crucial to utilize efficient algorithms and data structures.

3.1.3 Collaborative Filtering in Anime Recommendation

Within the niche of anime, applying collaborative filtering to recommendation systems proves highly beneficial, primarily due to the distinct dynamics of the anime watching community. This group is characteristically involved and vocal, with members frequently eager to express their strong, well-formed views. Such active participation yields a wealth of data, which is invaluable in augmenting the performance of recommendation systems. The anime genre itself offers a vast array, ranging from high-energy shounen to contemplative slice-of-life stories, creating an ideal landscape for custom recommendations. The creation of a recommendation system designed for anime fans significantly benefits from the strategic use of collaborative filtering. This method effectively utilizes a variety of user-contributed data, encompassing not only explicit user ratings but also their viewing habits and engagement levels as indicated by metrics like time spent watching or series completion rates. Integrating this information allows the system to identify patterns and draw connections between users and particular anime titles. Such insights are essential in surfacing anime titles that might appeal to viewers but remain undiscovered by them, thus enriching their experience as they delve into new anime content.

3.1.4 Leveraging Collaborative Filtering for Enhanced Personalization

The goal of an anime recommendation system is to curate a selection that resonates with the user's unique tastes. Collaborative filtering achieves this by creating a personalized experience that evolves with the user's interactions. As users rate more anime or modify their preferences, the system dynamically updates to reflect these changes, ensuring that the recommendations remain relevant and engaging.

3.2 Content-based Filtering Approach:

In recommendation systems, content-based filtering operates by recommending items similar to what a user has previously shown interest in, focusing on the characteristics of the items themselves rather than on user interactions. This method becomes particularly advantageous when there's a lack of extensive user-item interaction data, or when new items are regularly introduced into the system. For instance, in anime recommendations, content-based filtering would concentrate on various anime attributes, including genre, the director's style, themes, or even deeper aspects like plot outlines or character backgrounds.

3.2.1 Data Preparation and Feature Extraction:

When it comes to data preparation and feature extraction in content-based filtering, the process begins with assembling a dataset that encompasses comprehensive details about each item, such as anime titles in this scenario. This process may involve gathering metadata about each anime, including its genre, release year, number of episodes, and the production studio. For more advanced applications, features might be extracted from textual content like plot summaries, employing methods like Term Frequency-Inverse Document Frequency (TF-IDF) or Natural Language Processing (NLP) techniques. These methods are designed to distill and capture the core elements of the content.

3.2.2 Feature Normalization:

It's crucial to acknowledge that in recommendation systems, features can vary significantly in scale and type, encompassing numerical, categorical, or textual data. Consequently, normalization or standardization becomes imperative. This process is vital to prevent any single feature from disproportionately influencing the recommendation due to its scale, enabling a more equitable comparison across different items.

3.2.3 Implementing Content-Based Filtering:

While implementing content-based filtering, the system delves into analyzing user preferences and key characteristics of anime, such as genre and plot details. By aligning these features with user preferences, the system efficiently matches users with new anime titles they might find appealing. The system's learning mechanism, which evolves based on user feedback, ensures progressively refined suggestions, each more attuned to the individual's tastes. This adaptive approach underscores the substantial promise of personalized experiences in content discovery.

3.2.4 Handling Sparse and High-Dimensional Data:

Addressing the challenge of sparse and high-dimensional data is critical when applying KNN to content-based filtering. The issue of the 'curse of dimensionality' arises in contexts with extensive feature spaces. To mitigate this, dimensionality reduction techniques, such as Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE), are employed to condense the feature space while retaining its most significant attributes.

3.2.5 Personalization and User Profile Creation:

In content-based filtering, the creation of a user profile that accurately reflects an individual's preferences is indispensable. This profile is constructed by analyzing the attributes of items the user has previously engaged with, such as viewing or rating behavior. Utilizing the KNN algorithm, the system identifies items in the feature space that closely align with this user profile, thereby facilitating personalized recommendations.

3.2.6 Evaluating the Recommendation System:

The evaluation of the content-based filtering system augmented with KNN is another area of focus in my research. The efficacy of this system can be assessed using various metrics, including precision, recall, F1-score, or Root Mean Square Error (RMSE), contingent on the problem being approached as a classification or regression task. Moreover, conducting cross-validation is a common practice to ascertain the model's robustness, as noted by Shani and Gunawardana (2011).

4 Design Specification

In the dynamic realm of digital content consumption, the role of recommendation systems is paramount, guiding users amidst a plethora of choices towards options that resonate with their individual tastes. This project is centered on employing collaborative filtering within the anime genre to enhance the user's viewing experience by offering custom-tailored suggestions. Collaborative filtering, a cornerstone technique in recommendation systems, deduces a user's likings based on similar patterns observed in other users. The basic idea is that if user A agrees with user B on a certain topic, it's more probable that A will share B's opinion on another topic, as opposed to an opinion from an arbitrary individual. It effectively utilizes the collective insight of many to generate recommendations.

Collaborative filtering is categorized mainly into two types: user-based and item-based. User-based filtering suggests products by identifying users who are similar to the target user in terms of their ratings and preferences. On the other hand, item-based filtering makes recommendations by comparing the items that a user has shown interest in previously.

In developing the anime recommendation system for this research, the approach to data preparation is rigorous. A comprehensive dataset from a prominent anime database, encompassing ratings and detailed anime information, is employed. This phase involves meticulously organizing and cleansing the data, focusing on active users with a substantial number of ratings (at least 400), standardizing these ratings, and transforming the data into a format conducive to both collaborative and content-based filtering. The dataset is then thoughtfully divided into training and test segments, thereby addressing the challenges of limited data availability. This procedure sets a firm base for further stages involving model construction and evaluation.

The application of content-based filtering plays a crucial role, involving steps such as data gathering, preparation, selecting pertinent features, constructing content and user profiles, deploying the KNN algorithm, and training the model. This holistic approach is designed to culminate in the development of an effective, precise anime recommendation system that skillfully integrates the advantages of both collaborative and content-based filtering methodologies.

5 Implementation

The implementation details of the anime recommendation system include how the design specs are put into action in real life.

5.1 Collaborative Filtering in the Proposed Anime Recommendation System

A user-based collaborative filtering approach was implemented to analyze a large dataset of user interactions with various anime titles, aiming to predict individual user preferences. This method is particularly effective in the anime domain, where viewer preferences are often highly specific and influenced by genres.

5.1.1 Data Preparation in Anime Recommendation System Research

The foundation of any machine learning model, including those in anime recommendation systems, lies in data preparation. In this research, a detailed and thorough data preparation process was executed to ensure the high quality and reliability of the recommendation system. The following sections outline the steps taken in this research for data preparation.

5.1.2 Data Collection and Initial Assessment:

The initial stage involved collecting a comprehensive dataset of anime. An extensive dataset was acquired from a well-known anime database, which included user ratings and detailed information on a wide range of anime titles. The dataset contained columns such as user id, anime id, and rating, providing ample data for the recommendation system. A preliminary analysis of this dataset was conducted to understand its structure, size, and the nature of the data it contained. This examination was crucial in identifying key factors that would influence the recommendation algorithm.

5.1.3 Filtering and Cleaning Data:

Considering the dataset's size, it was imperative to filter the data to ensure relevance and manageability. The focus was on users who actively rated anime, with a minimum threshold of 400 ratings per user. This threshold was chosen to ensure that the data used in the recommendation system came from engaged users who had provided a significant number of ratings, thereby leading to more reliable and meaningful recommendations, as suggested by Das et al. (2017). During the data cleaning phase, issues such as missing values, outliers, and duplicate records were addressed. Duplicates were identified and removed to avoid biased results in the recommendation system. Maintaining data cleanliness was essential for the integrity and accuracy of the subsequent analysis.

5.1.4 Normalizing Ratings:

Normalizing user ratings was a critical step in the data preparation process. Given the varying scales of ratings in the dataset, normalization was necessary for consistency. Ratings were scaled between 0 and 1.0, a common practice in machine learning to handle different scales and enhance algorithm performance. This scaling was conducted using a linear transformation, which preserved the relative differences between ratings while standardizing the rating scale across the dataset (Soni et.al., 2023).

5.1.5 Data Transformation for Algorithm Suitability:

For algorithm suitability, this research utilized both collaborative filtering and content-based filtering with the KNN algorithm. The data was transformed into a format that these algorithms could process effectively. For collaborative filtering, the data was organized into a user-item matrix, where rows represented users and columns represented anime titles, with matrix cells containing the normalized ratings. For the content-based approach, anime titles were characterized by their features, such as genre, release year, and author, creating a feature-based dataset for the algorithm to identify similarities between different anime.

5.1.6 Splitting the Dataset:

To evaluate the performance of this recommendation system, the dataset was divided into training and testing sets. This split was crucial for validating the effectiveness of these recommendation models and avoiding overfitting. The training set was used to train these models, while the testing set served to evaluate their performance and accuracy in predicting user preferences.

5.1.7 Handling Sparse Data:

An obstacle in recommendation systems, specifically in collaborative filtering, is effectively managing limited information caused by the extensive range of anime titles and the impracticality of every user rating all titles. Utilized matrix factorization techniques to handle the issue of limited data, enabling the model to deduce missing ratings and offer more precise suggestions.

Ultimately, the data preparation procedure in this research was thorough and precise, encompassing various steps of cleansing, standardization, conversion, and segmentation of data. The meticulous procedure guaranteed that the dataset was precisely prepared for the creation of an efficient and precise anime recommendation system, establishing a strong basis for the upcoming stages of model construction and analysis.

5.2 Content-based filtering in the proposed anime recommendation system:

The deployment of content-based filtering using the KNN algorithm for an anime recommendation system is a complex undertaking that requires thorough planning and execution. This strategy is founded on the assumption that people will prefer anime that is like what they have previously loved, taking into account content qualities such as genre, themes, creator information, and other metadata. This paper presents a comprehensive description of the procedures used in this study to apply content-based filtering and KNN technology in a Jupyter Notebook environment Wang et al. (2014).

5.2.1 Data Collection and Preprocessing

The initial phase included collecting an extensive collection of anime titles, including their corresponding content characteristics. The metadata, including genre, director, studio, and user-provided tags, was compiled from multiple databases and online sources. The dataset underwent pre-processing to guarantee consistency in formatting and to address any missing or incorrect data. The preprocessing stage encompassed many tasks, such as normalizing text, resolving inconsistencies in genre naming conventions, and imputing missing values where possible Goyani and Chaurasiya (2020).

5.2.2 Feature Selection and Engineering

Once the data was cleaned, the next step was to select relevant features that could be used to describe each anime title effectively. Features that did not contribute to the differentiation of anime titles or those with a high degree of sparsity, were discarded to improve model performance. Feature engineering techniques were also applied to transform raw data into a format suitable for machine learning algorithms. For example, textual data from synopses or reviews was converted into numerical values using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization.

5.2.3 Creating the Content Profile

In this research, a distinct content profile for each anime is created, aggregating its features into a structured framework. This profile essentially acted as a unique identifier for the anime, capturing its key traits. I then organized these profiles into a matrix, arranging each anime in rows and its corresponding features in columns.

5.2.4 User Profile Construction

For user profile construction, a similar approach to that used for content profiles was adopted. This involved combining the content profiles of the anime titles that each user had engaged with and assigning weights based on the user's ratings or preferences for those titles. The result was a set of individualized user profiles that reflected the collective characteristics of their preferred anime.

5.2.5 Implementing the KNN Algorithm

The KNN algorithm was applied using these established content and user profiles. The algorithm calculated the similarity between unseen anime titles and the user's profile. Metrics like cosine similarity were used to determine how closely the features of a user's profile matched those of each anime title. The top recommendations for the user were identified as the 'k' nearest anime titles, which had the highest similarity scores.

5.2.6 Model Training and Hyperparameter Tuning

Model training and hyperparameter tuning involved using a subset of the data. Adjustments to the KNN algorithm's hyperparameters, such as the number of neighbors (k) and the distance metric, were made through cross-validation to optimize performance. The goal was to balance the model to avoid overfitting or underfitting.

5.2.7 Evaluation and Validation

The trained model was evaluated using a separate validation dataset to assess its predictive accuracy. The primary metric for this evaluation was the Root Mean Square Error (RMSE), which measured the average error magnitude in the model's predictions. A lower RMSE value indicated that the model's predictions were more aligned with the actual user ratings, thus affirming the effectiveness of the content-based filtering approach in this context.

6 Evaluation

This study analyzes the outcomes of implementing two distinct recommendation system models: collaborative filtering and content-based filtering, both enhanced with the K-Nearest Neighbours (KNN) algorithm, aimed at identifying suitable anime for viewers. The results section is dedicated to an in-depth evaluation of these models, employing a carefully curated set of performance indicators. A key focus is placed on the Root Mean Square Error (RMSE) scores, pivotal in assessing the prediction accuracy of each model. RMSE is a widely recognized metric for measuring the deviation between a model's predictions and the actual observed values. Generally, a model with a lower RMSE score indicates higher predictive accuracy, suggesting that the recommendations are more in tune with the users' true preferences.

The findings detailed in this section are the culmination of thorough data processing, model training, and validation efforts, all conducted within a Jupyter Notebook environment. This platform is highly favored by data scientists for its flexibility and interactivity, making it an ideal tool for such analytical tasks. The results provide a clear understanding of the strengths and weaknesses inherent in each recommendation approach, thereby shedding light on their suitability and effectiveness in the context of anime recommendations.

6.1 Training Results:

The training results graph illustrates the model's loss over several epochs for both the training and test datasets. The convergence of the loss for both sets indicates that the model is generalizing well and not overfitting to the training data. The closeness of the training and test lines suggests that the model would perform consistently on unseen data. Ideally, the loss should stabilize at a minimum value, indicating that the model has learned the underlying patterns effectively.

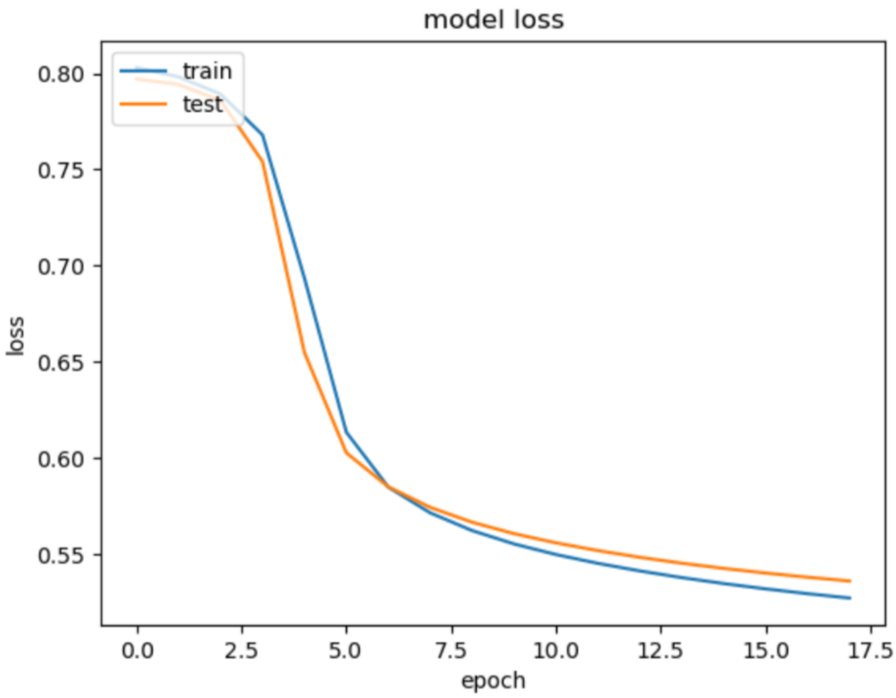

Figure 2: Training Epoch Results on test and training data of Collaberative Filtering with Neural network

6.2 User's Highly Rated Anime:

The provided table of a user's highly rated anime, which includes titles like "Fullmetal Alchemist: Brotherhood" and "Attack on Titan Final Season," allows for an insight into the user's preferences. This information is pivotal for a CF model as it relies on historical data to predict future likes. The diversity of genres within the user's preferences can challenge the CF model, requiring it to capture and understand the nuances across different anime genres and themes.

	eng_version	Genres
3971	Fullmetal Alchemist:Brotherhood	Action, Military, Adventure, Comedy, Drama, Magic, Fantasy, Shounen
15926	Attack on Titan Final Season	Action, Military, Mystery, Super Power, Drama, Fantasy, Shounen
5683	Steins;Gate	Thriller, Sci-Fi
14963	Attack on Titan Season 3 Part 2	Action, Drama, Fantasy, Military, Mystery, Shounen, Super Power
6474	Hunter x Hunter	Action, Adventure, Fantasy, Shounen, Super Power

Figure 3: Highly Rated anime list for collaborative filtering neural network model

6.3 Preferred Genres:

Figure 4: Word Cloud result of preferred genre of neural network model

The word clouds represent the user's preferred genres, with the size of each genre name corresponding to the user's affinity for that genre. Larger-sized words like "action," "drama," and "shounen" indicate a stronger preference for these genres. The CF model with embeddings would use this information to identify latent factors that represent these preferences in a lower-dimensional space, which helps in understanding the user's profile and recommending new anime that aligns with these genres.

6.4 Top Recommendations:

The CF model's effectiveness is also depicted by the top recommendations for the user. If the recommended titles resonate with the user's highly rated anime and preferred genres, this signifies that the model has successfully captured the user's taste. The model's ability to recommend titles that are not only popular but also personalized to the user's unique tastes is a strong indicator of its performance Kumar and Thakur (2018).

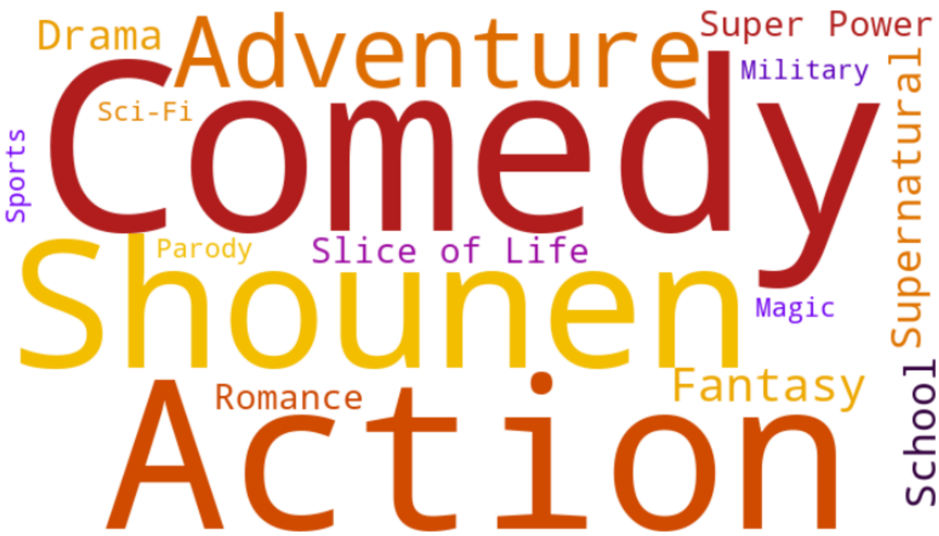

Figure 5: Word Cloud result for top recommendations for collaborative-filtering neural network model

6.5 Analysis and Implications:

The low RMSE score, and consistent training results suggest that the CF model with embeddings is performing well. Embeddings allow the model to capture complex and abstract user preferences, which is particularly beneficial when dealing with high-dimensional data like user ratings and diverse content such as anime. The word cloud and highly rated anime list serve as indicators of the user's preference profile, which the model has successfully used to generate customized recommendations.

RMSE Score of Collaborative filtering using embedding model:

RMSE Score	0.295818070817742884

Table 1: RSME values of Collaborative filtering using embedding model

The first table showcases an RMSE score, a vital measure for assessing the precision of a recommendation system. An RMSE score of around 0.2958 indicates that the model's predictions are in close harmony with actual user ratings. In the context of recommendation systems, such a low RMSE is indicative of high accuracy, suggesting that the system is adept at predicting a user's preference for specific anime based on their past choices.

The outcomes derived from the CF model with embeddings point to a triumphant execution of the recommendation system. The in-depth examination of user preferences, coupled with the model's predictive accuracy as evidenced by the RMSE score and training results, highlights the model's capacity to offer precise and personalized anime suggestions. The strategic application of embeddings in the CF model has adeptly navigated the complex preference patterns of anime viewers, presenting an effective method to boost user engagement and satisfaction in recommendation platforms.

S.No	Status	Description
0	1	Currently watching
1	2	Completed
2	3	On hold
3	4	Dropped
4	6	Planned to watch

Figure 6: Status Categorization of Anime

6.6 Status Categorization of Anime:

The first image indicates a status categorization, likely representing how users have interacted with different anime titles.
The categories include:

1. Currently Watching (status 1)
2. Completed (status 2)
3. On Hold (status 3)
4. Dropped (status 4)
5. Plan to Watch (status 6)

These categories are essential for understanding user engagement levels with specific anime titles and can be used to refine recommendations. For instance, titles that are frequently 'Completed' may indicate high satisfaction or engagement and thus might be recommended to similar users. Conversely, titles with a high 'Dropped' status could be deprioritized in the recommendations or used to analyze why users disengage.

6.7 Distribution of User Ratings:

In the second image (Figure 7), a histogram illustrates the distribution of user ratings. The X-axis likely indicates the rating scale, and the Y-axis shows the number of users corresponding to each rating. A notable concentration of ratings is observed around the middle of the scale, suggesting a tendency towards neutral evaluations. This central clustering is a common feature in user rating distributions. Any skewness or outliers in this distribution could influence the recommendation system, especially in collaborative filtering, where user ratings are instrumental in identifying similarities between users or items.

6.8 Watching Status Distribution Among Users:

Regarding the distribution of watching statuses among users, the third visual (Figure 8) demonstrates the count of users for each status, as defined in the initial image. The X-axis categorizes different watching statuses, while the Y-axis displays the number of users for each status. The status 'Completed' records the highest count, indicating a trend where most users rate anime after completing them. This observation is crucial, as it suggests that data derived from completed anime serves as a reliable source for gauging preferences and making recommendations.

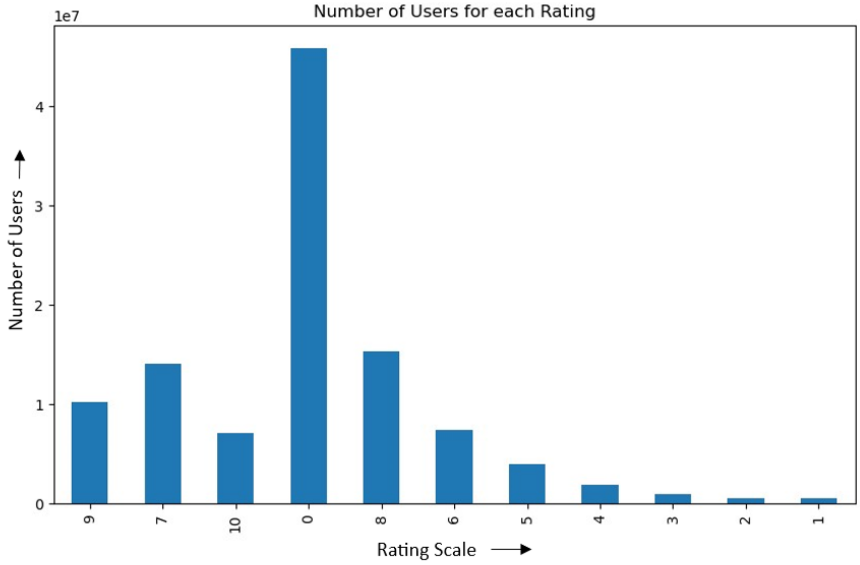

Figure 7: Histogram of User Ratings for different values

6.9 Performance of the Recommendation System using KNN model's content-based filtering:

The final table presents an RMSE value of approximately 7.45. The RMSE is a measure of the difference between the values predicted by a model, or an estimator and the values observed. The value given suggests that, on average, the predictions made by the recommendation system deviate from the actual user ratings by a margin of 7.45. This value provides a quantitative measure to assess the accuracy of the recommendation system. In the context of recommendation systems, a lower RMSE is generally better, indicating higher accuracy in predictions. However, whether this RMSE value is acceptable depends on the specifics of the system's implementation and the context within which it operates. It is also useful to compare this RMSE value against a baseline model or other recommendation system approaches to determine its relative performance.

RMSE Score	7.450396588260162

Table 2: RSME values of Content-based filtering using KNN model

The results presented in the screenshots provide insights into user interaction and satisfaction with anime titles, which are critical for refining the recommendation system. The status categorization can help tailor recommendations based on user engagement levels, while the distribution of ratings informs the system about general user tendencies. The RMSE value offers a benchmark for the system's performance, guiding further optimization and improvements.

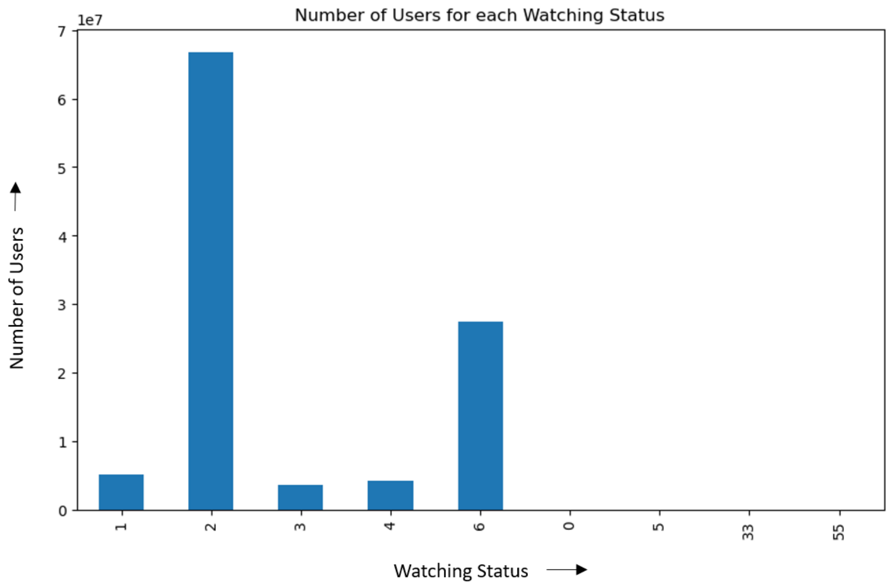

Figure 8: Histogram distribution of User's watching status

6.10 Discussion

When it comes to anime recommendations, the use of collaborative filtering embedded with neural networks and , content-based filtering using the K-Nearest Neighbours (KNN) algorithm has shown different pros and cons. Collaborative filtering leverages the collective input of user preferences, providing recommendations that mirror the wisdom of the crowd, allowing for dynamic adaptability and a high degree of personalization. It can also lead to beneficial findings, introducing users to anime titles they might not have encountered through their searches. However, this method grapples with the cold start problem, requiring a substantial amount of historical data to make accurate predictions, and can suffer from a shortage in user-item interactions. Additionally, as the volume of data grows, the scalability of the KNN algorithm becomes a concern, and there's a risk of recommendation homogeneity, which might limit exposure to diverse content Nagarnaik and Thomas (2015)

Content-based filtering, on the other hand, operates independently of user interaction data, thereby mitigating the cold start issue and offering recommendations based on the intrinsic content features of the anime. This approach provides a high level of explainability for its recommendations and can represent a diverse array of content through detailed feature engineering Lee et al. (2015). Despite these advantages, content-based filtering can become overly specialized, often recommending a narrow selection of content that closely mirrors a user's past preferences, which can impede the discovery of varied content. The process of developing an effective set of features for content representation can be complex and labor-intensive, sometimes requiring significant domain expertise. Moreover, content-based methods may not capture the full spectrum of a user's tastes as they do not account for the collaborative aspect that is inherent in user interactions Subramaniaswamy et al. (2017).

The implementation of both methods within a Jupyter Notebook environment has provided clear insights into their operational capabilities. By examining the unique attributes and potential drawbacks of each method, this research seeks to clarify the most effective strategy for recommending anime titles. The comparative analysis conducted here, particularly focusing on the RMSE values, serves as a testament to the efficiency of each method and underscores the importance of context when determining the most suitable approach for a recommendation system.

7 Conclusion and Future Work

In this research on anime recommendation systems, collaborative filtering, which utilizes user interaction data, outperformed content-based filtering in accuracy, as indicated by a lower root mean square error (RMSE) value. This superiority implies that collaborative filtering is more effective in capturing and reflecting the collective preferences of users, identifying popular trends, and aligning recommendations with community insights. However, it faces challenges like the cold start problem and lack of data, while content-based filtering, focusing on anime attributes, addresses more robustly Cho et al. (2018).

The research suggests the potential of hybrid models combining both approaches, enhancing accuracy and specificity. It also highlights the need for algorithmic efficiency improvements in handling large-scale user data and the exploration of cross-domain systems to broaden recommendation bases. The study underscores the importance of dynamic, real-time recommendations and considers user engagement and satisfaction metrics beyond RMSE.

The conclusion emphasizes the complexity of creating effective recommendation systems. The lower RMSE of collaborative filtering underscores its role in personalizing recommendations, suggesting a hybrid model integrating both methods could offer a more comprehensive and user-responsive recommendation system for anime content.

Code : GitHub Repository

References

Aleem, S., Kumar, T., Little, S., Bendechache, M., Brennan, R. and McGuinness, K. (2022). Random data augmentation based enhancement: a generalized enhancement approach for medical datasets, *arXiv preprint arXiv:2210.00824* .

Burke, R. (2007). Hybrid web recommender systems, *The adaptive web* pp. 377–408.

Cantador, I., Brusilovsky, P. and Kuflik, T. (2015). Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011), *Proceedings of the 5th ACM Conference on Recommender Systems* pp. 395–398.

Chandio, A., Gui, G., Kumar, T., Ullah, I., Ranjbarzadeh, R., Roy, A. M., Hussain, A. and Shen, Y. (2022). Precise single-stage detector, *arXiv preprint arXiv:2210.04252* .

Chandio, A., Shen, Y., Bendechache, M., Inayat, I. and Kumar, T. (2021). Audd: audio urdu digits dataset for automatic audio urdu digit recognition, *Applied Sciences* **11**(19): 8842.

Cho, H., Schmalz, M. L., Keating, S. A. and Lee, J. H. (2018). Analyzing anime users' online forum queries for recommendation using content analysis, *Journal of Documentation* **74**(5): 918–935.

Das, D., Sahoo, L. and Datta, S. (2017). A survey on recommendation system, *International Journal of Computer Applications* **160**(7): 27083–27083.

Girsang, A. S., Faruq, B. A., Herlianto, H. R. and Simbolon, S. (2020). Collaborative recommendation system in users of anime films, *Journal of Physics: Conference Series* **1566**: 012057.
URL: *http://iopscience.iop.org/article/10.1088/1742-6596/1566/1/012057/pdf*

Gomez-Uribe, C. and Hunt, N. (2016). The netflix recommender system: algorithms, business value, and innovation, *ACM Transactions on Management Information Systems (TMIS)* **6**(4): 1–19.

Goyani, M. M. and Chaurasiya, N. (2020). A review of movie recommendation system: Limitations, survey and challenges, *ELCVIA: electronic letters on computer vision and image analysis* **19**(3): 0018–37.

Isinkaye, F., Folajimi, Y. and Ojokoh, B. (2015). Recommendation systems: principles, methods, and evaluation, *Egyptian Informatics Journal* **16**(3): 261–273.

Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (20). *Recommender Systems: An Introduction*, Cambridge University Press.

Khan, W., Raj, K., Kumar, T., Roy, A. M. and Luo, B. (2022). Introducing urdu digits dataset with demonstration of an efficient and robust noisy decoder-based pseudo example generator, *Symmetry* **14**(10): 1976.

Kumar, P. and Thakur, R. S. (2018). Recommendation system techniques and related issues: a survey, *International Journal of Information Technology* **10**(3): 495–501.

Kumar, T., Brennan, R. and Bendechache, M. (2022). Stride random erasing augmentation, *CS & IT Conference Proceedings*, Vol. 12, CS & IT Conference Proceedings.

Kumar, T., Mileo, A., Brennan, R. and Bendechache, M. (2023). Rsmda: Random slices mixing data augmentation, *Applied Sciences* **13**(3): 1711.

Kumar, T., Park, J., Ali, M. S., Uddin, A. and Bae, S.-H. (2021). Class specific autoencoders enhance sample diversity, *Journal Of Broadcast Engineering* **26**(7): 844–854.

Kumar, T., Park, J., Ali, M. S., Uddin, A. S., Ko, J. H. and Bae, S.-H. (2021). Binary-classifiers-enabled filters for semi-supervised learning, *IEEE Access* **9**: 167663–167673.

Kumar, T., Park, J. and Bae, S.-H. (2020). Intra-class random erasing (icre) augmentation for audio classification, *Korean Society of Broadcasting and Media Engineering conference proceedings* pp. 246–249.

Kumar, T., Turab, M., Mileo, A., Bendechache, M. and Saber, T. (2023). Audrandaug: Random image augmentations for audio classification, *arXiv preprint arXiv:2309.04762* .

Kumar, T., Turab, M., Raj, K., Mileo, A., Brennan, R. and Bendechache, M. (2023). Advanced data augmentation approaches: A comprehensive survey and future directions, *arXiv preprint arXiv:2301.02830* .

Kumar, T., Turab, M., Talpur, S., Brennan, R. and Bendechache, M. (n.d.). Forged character detection datasets: Passports, *DRIVING LICENCES AND VISA STICKERS* .

Lee, J., Shim, Y. and Jett, J. (2015). Analyzing user requests for anime recommendations, *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, IEEE Computer Society, pp. 269–270.

Massa, P. and Avesani, P. (2007). Trust-aware recommender systems, *Proceedings of the 2007 ACM Conference on Recommender Systems*, ACM, pp. 17–24.

Nagarnaik, P. and Thomas, A. (2015). Survey on recommendation system methods, *2015 2nd International Conference on Electronics and Communication Systems (ICECS)* pp. 1603–1608.
URL: *https://api.semanticscholar.org/CorpusID:28358719*

Ota, S., Kawata, H., Muta, M., Masuko, S. and Hoshino, J. (2017). Anireco: Japanese anime recommendation system, *Entertainment Computing–ICEC 2017: 16th IFIP TC 14 International Conference*, Springer International Publishing, pp. 400–403.

Park, J., Kumar, T. and Bae, S.-H. (2020). Search for optimal data augmentation policy for environmental sound classification with deep neural networks, *Journal of Broadcast Engineering* **25**(6): 854–860.

Pazzani, M. and Billsus, D. (2007). Content-based recommendation systems, *The Adaptive Web* **3**(4): 325–341.

Prakash, V., R. S. S. S. P. M. and Arora, M. (n.d.).

Ranjbarzadeh, R., Jafarzadeh Ghoushchi, S., Tataei Sarshar, N., Tirkolaee, E. B., Ali, S. S., Kumar, T. and Bendechache, M. (2023). Me-ccnn: Multi-encoded images and a cascade convolutional neural network for breast tumor segmentation and recognition, *Artificial Intelligence Review* pp. 1–38.

Roy, A. M., Bhaduri, J., Kumar, T. and Raj, K. (2022). A computer vision-based object localization model for endangered wildlife detection, *Ecological Economics, Forthcoming* .

Roy, A. M., Bhaduri, J., Kumar, T. and Raj, K. (2023). Wildect-yolo: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection, *Ecological Informatics* **75**: 101919.

Schafer, J., Frankowski, D., Herlocker, J. and Sen, S. (2007). Collaborative filtering recommender systems, *The Adaptive Web* pp. 291–324.

Shani, G. and Gunawardana, A. (2011). Evaluating recommendation systems, *Recommender Systems Handbook* pp. 257–297.

Singh, A., Raj, K., Kumar, T., Verma, S. and Roy, A. M. (2023). Deep learning-based cost-effective and responsive robot for autism treatment, *Drones* **7**(2): 81.

Singh, A., Ranjbarzadeh, R., Raj, K., Kumar, T. and Roy, A. M. (2023). Understanding eeg signals for subject-wise definition of armoni activities, *arXiv preprint arXiv:2301.00948* .

Soni, B., Thakuria, D., Nath, N., Das, N. and Boro, B. (2023). Rikonet: A novel anime recommendation engine, *Multimedia Tools and Applications* pp. 1–20.

Subramaniaswamy, V., Logesh, R., Chandrashekhar, M., Challa, A. and Vijayakumar, V. (2017). A personalised movie recommendation system based on collaborative filtering, *International Journal of High-Performance Computing and Networking* **10**(1-2): 54–63.

Turab, M., Kumar, T., Bendechache, M. and Saber, T. (2022). Investigating multi-feature selection and ensembling for audio classification, *arXiv preprint arXiv:2206.07511* .

Wang, Z., Xue, Y., Feng, N. and Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence, *Journal of Visual Languages Computing* **25**(6): 667–675.