Deep Anime Recommendation System: Recommending Anime Using Hybrid Filtering

MSc Research Project
Data Analytics

Varshini Subburaj
Student ID: x22153977

School of Computing
National College of Ireland

Supervisor: Prof. Dr Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Varshini Subburaj
Student ID:	x22153977
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Prof. Dr Hicham Rifai
Submission Due Date:	14/12/2023
Project Title:	Deep Anime Recommendation System: Recommending Anime Using Hybrid Filtering
Word Count:	7327
Page Count:	20

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:	
Date:	14th December 2023

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):	

Deep Anime Recommendation System: Recommending Anime Using Hybrid Filtering

Varshini Subburaj
x22153977

Abstract

The Deep Anime Recommendation System employing Hybrid Filtering is a groundbreaking approach in the growing field of online anime streaming. Traversing a huge expanse of anime titles presents difficulties such as mental exhaustion from making choices and overlooking potential content chances. Conventional recommendation systems often fail to adjust to subtle user preferences, necessitating the employment of a new technique. The study presents a hybrid filtering approach that combines content-based and collaborative filtering strategies, bolstered by deep learning capabilities. The technology tries to revolutionise the experience of discovering anime by simultaneously examining content characteristics and user preferences. Beyond the constraints of traditional systems, it enhances recommendation processes by providing customised ideas that are specifically tuned to individual preferences. This research not only enhances the development of anime recommendation systems but also presents a hybrid recommendation system that integrates content-based and collaborative filtering approaches with deep learning to provide more precise and personalised anime suggestions. The hybrid system surpasses the performance of basic models, clearly showing its efficacy in improving user engagement and contentment in the ever-changing world of anime streaming services.The suggested Deep Anime Recommendation System is a technological innovation that has the potential to transform personalised content discovery on internet streaming platforms.

1 Introduction

Explore the ever-changing world of online anime streaming with the cutting-edge Deep Anime Recommendation System, which uses hybrid filtering technology to provide unparalleled recommendations. This innovative technique surpasses traditional conventions of content suggestion by tackling the challenges of overwhelming choices and user preferences in the vast world of anime titles. Our study presents a revolutionary hybrid filtering methodology to assist users in efficiently exploring a vast array of anime material, minimising choice fatigue and maximising content discovery chances. This model combines content-based and collaborative filtering approaches, enhanced by the complexity of deep learning. The objective of this platform is to revolutionise the process of discovering anime by carefully analysing both the intricate aspects of content characteristics and the tastes of users. The system strives to surpass expectations, going beyond traditional constraints. It aims to improve recommendation processes by providing highly tailored, personalised ideas that are closely adjusted to individual preferences. This study not only has a significant influence on anime recommendation systems but also leads the way in advancing hybrid filtering approaches, which have the potential to fundamentally transform online streaming. As we explore the complex landscape of anime consumption, the Deep Anime Recommendation System stands out as an innovative and remarkable technical achievement. The platform aims to go beyond providing suggestions and instead aims to transform and customise the whole user experience in the constantly changing world of online anime streaming services (Fig 01)[1].

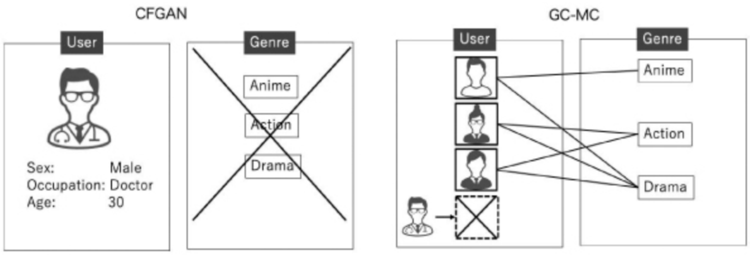

Figure 1: Anime Recommendation

1.1 Research Background and Motivation

The rapid expansion of internet anime streaming services has brought us an age of unprecedented ease of access to material. Nevertheless, the extensive range of options has posed considerable difficulty for individuals in finding material that matches their distinct interests. Conventional anime recommendation systems often fail to provide precise and tailored recommendations since they struggle to adjust to the subtle and changing preferences of consumers. Current models often depend on either content-based or collaborative filtering strategies, both with their own constraints. Content-based methods often disregard the ever-changing nature of user preferences, while collaborative filtering might be impeded by limited user-item interaction data. The integration of both strategies into a hybrid filtering model is a possible approach to addressing these constraints, capitalising on the advantages of both procedures.

The rationale for this study stems from the acknowledgement of the urgent need for a more efficient and adaptable anime recommendation system. Users often experience fatigue and frustration while trying to navigate the extensive anime environment. This is a crucial opportunity to improve the suggestion process. The hybrid filtering strategy, which integrates deep learning, offers a novel option to overcome the limitations of current systems. This approach attempts to enhance user pleasure and engagement by taking into account both content aspects and user preferences, resulting in more precise and personalised anime suggestions. The study is motivated by the desire to enhance anime recommendation systems, providing both a practical answer for consumers and significant insights into the wider domain of hybrid filtering approaches. The potential influence goes beyond immediate applications, anticipating a more customised and engaging anime streaming experience for customers in the changing environment of online entertainment.

[1] Anime Recommendation: https://www.fujipress.jp/jaciii/jc/jacii002500040389/

1.2 Research Question

What is the effectiveness of a hybrid recommendation model, combining content-based filtering and collaborative filtering, in generating accurate and relevant anime recommendations?

1.3 Contribution

The project introduces a revolutionary Deep Anime Recommendation System that utilises hybrid filtering, making substantial advancements in the area of anime recommendation systems. The hybrid model effectively combines content-based and collaborative filtering strategies, enhanced by deep learning capabilities. This integration tackles the issues of choice fatigue and missing content possibilities by providing a user-centric method for discovering anime. Optimising recommendation procedures guarantees tailored ideas that go beyond the constraints of conventional systems. The initiative not only has a direct influence on anime recommendations but also leads the way in advancing hybrid filtering approaches, making significant contributions to the wider field. The study improves user pleasure and engagement in online anime streaming by using technical innovation and addressing the limitations of current models. Additionally, it offers useful insights that may be applied to recommendation systems in many fields. The Deep Anime Recommendation System is an innovative tool that has the potential to revolutionise the personalised content discovery experience on internet streaming platforms.

2 Related Work

2.1 Introduction

The effectiveness of any Recommender System (RS) hinges on the relevance of its recommendations to users, a challenge exacerbated by the exponential growth of data. However, this challenge has spurred the development of sophisticated algorithmic models for Zhang et al. (2019) categorize contemporary RSs into three main types: content-based, collaborative filtering, and hybrid RSs. In response to issues like data sparsity and cold-start problems in traditional RSs, Deep Learning models have emerged as a solution. These models not only accommodate diverse data types such as audio, video, images, and text but also demonstrate scalability to large datasets. The subsequent sections provide a comprehensive overview and critical analysis of various Deep Learning models currently employed in RSs.

2.2 Anime Recommender Systems: Deep Learning Approaches

Nuurshadieq and Wibowo (2020) proposes a personalized anime recommender system using deep learning, addressing the challenge of the cold start problem for both users and items. The authors collect a dataset with user-anime interactions, including side information about users and anime works. They introduce a hybrid model incorporating deep learning, specifically utilizing a long short-term memory (LSTM) layer to extract information from synopsis texts. The proposed method outperforms traditional collaborative filtering methods, demonstrating improved accuracy, particularly when incorporating side information. The study explores the impact of synopsis, user, and anime side information on model accuracy. The authors suggest potential enhancements, such as investigating transfer learning or more complex neural network architectures for better text data extraction. The results showcase the effectiveness of the proposed approach in mitigating the cold start problem in anime recommendation systems.

Mutteppagol (2021) presents a comprehensive study on enhancing anime recommendations through a Deep Learning-based recommender system. Addressing challenges like data sparsity, the research introduces a Model-based Collaborative Filtering approach using Deep Learning, incorporating the Adam optimizer and experimenting with Sigmoid and ReLU activation functions. Utilizing an anime dataset from Kaggle and MyAnimeList, the study involves thorough data extraction, pre-processing, and analysis. Results show that the model with Sigmoid activation and 20 epochs outperforms the ReLU counterpart, demonstrating lower Mean Squared Error and Mean Absolute Error values. The paper suggests the potential real-time application of the proposed model in anime recommendation systems, with future work focusing on refining the model by considering user demographics for personalized recommendations based on age groups.

Ota et al. (2017) presents "AniReco," a Japanese anime recommendation system developed to assist users in navigating the vast and diverse anime market. Faced with the challenge of selecting anime works aligned with individual preferences from a substantial pool, AniReco employs a content-based filtering approach. It calculates user preferences based on viewing history and evaluations, visually representing recommendations through a network diagram. The system's usability was affirmed in an evaluation experiment, with users finding it effective in simplifying the exploration of anime and leading to new discoveries. While AniReco excels in recommending works reflecting both known and potential user preferences, its impact on enhancing interest in related content beyond anime was found to be limited. Overall, the proposed system offers a valuable solution to the complexities of anime selection and exploration.

2.3 Advancements in Anime Recommender Systems: Model-based Architectures and Hybrid Approaches

The paper presents an advanced recommendation system for anime, incorporating a model-based architecture featuring a Gated Recurrent Neural Network (GRU) and Cosine Similarity. Affiliated with Manav Rachna University, Prakash et al. (2022) highlight the model's effectiveness in handling sparse data, surpassing traditional methods. The study employs a dataset from MyAnimeList, encompassing 5 million user ratings. Results indicate the model's successful convergence with a validation loss of 0.7873 and a training loss of 0.7674, supported by metrics like RMSE (0.8568), MSE (0.7341), and MAE (0.6749). The conclusion addresses recommendation system challenges, suggesting future avenues such as privacy considerations and the exploration of hybrid recommendation systems.

Yao et al. (2023) aims to develop a recommender system for Japanese Animes using content-based and collaborative filtering methods. Content-based filtering involves comparing anime based on their synopses and images, while collaborative filtering predicts ratings using user interactions. The study used five models, including content-based filtering with synopsis and image similarity, collaborative filtering with Singular Value Decomposition (SVD), and autoencoders. Collaborative filtering models outperformed content-based ones, with the autoencoder achieving the highest hit rate at 43%. Insights include the effectiveness of pretrained model weights for image features and the superiority of autoencoders over SVD in collaborative filtering. The study suggests potential future directions, such as hybrid recommender systems and the need for larger datasets to improve predictive power. The code implementation is available on GitHub.

Nápoles et al. (2020) presents a recommender system utilizing Long-term Cognitive Networks (LTCNs), a type of recurrent neural network enabling reasoning with prior knowledge structures. In lieu of human expert involvement, Pearson's correlation coefficients replace prior knowledge. The proposed architecture extends LTCNs with Gaussian kernel neurons to estimate missing ratings, refining predictions through the recurrent structure. The non-synaptic backpropagation algorithm is adapted for nonlinearity adjustment. The model outperforms state-of-the-art methods across various case studies, emphasizing interpretability. The paper discusses injecting expert knowledge, allowing adaptation to emerging trends. Future research directions involve automated knowledge engineering and leveraging context information for efficient context-aware approaches.

2.4 Broadening Horizons: Cross-Domain Recommendations and Insights from Movie Recommendation Systems

Bahaweres and Almujaddidi (2022) explores the implementation of cross-domain recommendation techniques in the domains of anime and manga to address sparsity issues in user and item interaction matrices. Cross-domain recommendation systems transfer knowledge from related domains to enhance performance. The research compares single-domain and cross-domain scenarios, focusing on sensitivity analysis. Utilizing datasets from MyAnimelist via web scraping, the study demonstrates that cross-domain recommendations, particularly from anime to manga, significantly decrease RMSE and maintain stable accuracy above 90%, even at sparsity levels of 0.01%. The findings suggest the practicality of cross-domain recommendations in the anime and manga domains, showcasing the potential for performance improvement in sparser domains when leveraging information from richer domains. The study acknowledges limitations, including the proxy use of training data ratio for target domain density and the absence of a separate test set during dataset tuning. Future research could explore overlapping scenarios on a system or item level in the anime domain and implement recommender systems for direct user evaluation.

Girsang et al. (2020) investigates the application of collaborative filtering to enhance a recommendation system for anime films. Using a dataset from Kaggle comprising 73,516 users and 12,294 anime entries, the study employs the Alternating Least Squares (ALS) algorithm for matrix factorization. The collaborative filtering method involves mean normalization for new users, the use of similarity matrices to identify user-neighbors, and recommending top movies based on calculated feature differences. Despite acknowledging limitations related to dataset size, the paper concludes that the collaborative filtering approach, though simple, proves effective in providing a more refined anime recommendation system, with potential for further enhancement with larger datasets.

The paper by Ramu (2022), titled "Deep Learning-Based Anime and Movie Recommendation System," delves into the development of a recommendation system to address data sparsity challenges in the context of anime and movies. Motivated by the dominance of over-the-top media platforms, the research utilizes a model-based collaborative filtering approach employing deep learning techniques. The proposed methodology involves data extraction from Kaggle datasets, data cleaning, and exploratory data analysis. The model, focusing on predicting user preferences through matrix factorization, is evaluated using Mean Square Error and Mean Absolute Error metrics. Results demonstrate successful mitigation of data sparsity, yielding accurate predictions and top 'n' recommendations for both anime and movies. The paper concludes by emphasizing the achieved objectives and potential avenues for future research, contributing to the field of recommendation systems in the entertainment industry.

The article, titled "An Examination of Movie Recommendation System: Constraints, Overview, and Issues" by Goyani and Chaurasiya (2020), explores the domain of recommendation systems, with a specific focus on movie suggestions. The authors investigate collaborative filtering and content-based filtering methods, recognizing their limitations and advocating for a hybrid approach to enhance the precision of recommendations. A comprehensive survey is conducted on advanced techniques, encompassing Content-Based Filtering, Collaborative Filtering, Hybrid Approaches, and Deep Learning-Based Methods for movie recommendations, along with an exploration of diverse similarity measures. Highlighting the widespread use of recommendation systems on platforms like Netflix and Amazon, the article underscores their role in improving user engagement and business profitability. Challenges such as the Cold Start Problem, Data Sparsity, and Scalability are addressed, leading to the conclusion that effective recommender systems necessitate hybrid models and continuous research to adapt to changing user preferences. The article concludes by summarizing various movie datasets, providing valuable insights for researchers, and stressing the importance of combining collaborative and content-based filtering for robust recommendation systems.

2.5 Enhancing Recommendations: A Blend of Collaborative and Content-Based Approaches

Furtado and Singh (2020) discusses a Movie Recommendation System using Machine Learning, focusing on reducing the effort required for users to choose movies based on their interests. The system combines both content-based and collaborative approaches, aiming to provide more specific results compared to content-based systems. Content-based systems are limited to individual preferences, while collaborative approaches consider the ratings of similar users, allowing for broader exploration. The proposed model uses a pre-filtering step before applying the K-means algorithm, considering factors like genre and rating. Users are required to rate at least six movies to receive recommendations, promoting more active user engagement. The system's implementation includes features for searching, rating, and generating recommendations, with a flowchart illustrating the overall architecture. The system is designed to improve sparsity by making ratings mandatory and address issues of over-specialization and cold starts associated with content-based approaches. The conclusion emphasizes the system's collaborative approach, allowing users to explore a wider range of movie recommendations.

Geetha et al. (2018) introduces a movie recommendation system aimed at addressing the challenge of content discovery in the digital age. Authors propose a hybrid approach that combines collaborative filtering, content-based filtering, and clustering techniques to enhance the precision of movie recommendations. The system utilizes the K-means algorithm for clustering and the Pearson correlation score for collaborative filtering. The authors acknowledge the difficulty in evaluating the system's performance due to the subjective nature of recommendations but report positive feedback from a small set of users. They express the need for a larger dataset for more robust evaluations and suggest potential extensions of the approach to other domains beyond movies. Overall, the hybrid system shows promise in improving recommendation accuracy by mitigating the limitations of individual algorithms.

Sharma et al. (2022) contribute significantly with their hybrid recommendation system, integrating Collaborative Filtering and Content-Based Filtering to tackle issues like the Cold Start Problem, Sparsity, and Scalability. The proposed three-phase approach involves user identification based on profile matching, candidate item selection through vector analysis, and final prediction and recommendation using the Resnick prediction equation. Comparative analysis demonstrates the superiority of the hybrid model over traditional Collaborative Filtering and Content-Based Filtering, highlighting its potential in offering accurate and personalized recommendations. This work underscores the contemporary need for advanced recommendation strategies, setting the stage for the exploration of hybrid models in the broader context of recommendation system literature.

Thorat et al. (2015) explores the landscape of recommender systems, particularly focusing on collaborative filtering (CF), content-based filtering (CBF), and hybrid approaches. Published in the International Journal of Computer Applications in 2015, the authors categorize recommendation techniques, delve into the intricacies of memory-based and model-based collaborative filtering, and discuss the strengths and limitations of content-based filtering. Notably, the paper underscores the significance of hybrid recommender systems that combine CF and CBF to address challenges like the cold start problem and data sparsity. The classification of hybrid systems into weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level further enriches the discussion. The paper concludes by highlighting ongoing challenges in the field and suggesting avenues for future research, emphasizing the need for advancements in existing methods and frameworks for automated analysis of heterogeneous data.

2.6 Innovative Hybrid Recommender Systems: Overcoming Challenges for Enhanced Accuracy and Scalability

Badaro et al. (2013) introduces a novel hybrid method to address challenges in recommender systems, focusing on accuracy and data sparsity. Authored by Gilbert Badaro, Hazem Hajj, Wassim El-Hajj, and Lama Nachman, the paper originates from the Electrical Engineering and Computer Science Departments at the American University of Beirut, as well as Intel Corporation. The authors emphasize the increasing relevance of recommender systems in various applications, from e-commerce to daily decision-making, owing to the growing volume of information available. The proposed hybrid approach combines user-based and item-based collaborative filtering through a weighted combination, aiming to enhance accuracy and alleviate data sparsity challenges. The evaluation demonstrates the superiority of the hybrid solution over stand-alone user-based and item-based collaborative filtering methods. The weights for combining the techniques are empirically determined, with an optimal combination of alpha=1/6 and beta=5/6 yielding the lowest Mean Absolute Error (MAE). Experiments conducted on MovieLens data reveal a 23% improvement over user-based collaborative filtering and a 16% improvement over item-based collaborative filtering, showcasing the effectiveness of the proposed hybrid model.

Chen et al. (2018) provides a comprehensive survey on collaborative filtering (CF)-based recommender systems in the context of big data, aiming to address challenges like data sparsity and high dimensionality. The abstract emphasizes the significance of CF techniques and the incorporation of context information to enhance recommendation accuracy. The introduction sets the stage by highlighting the information overload problem and the success of CF, while proposing the integration of social factors to overcome emerging challenges. The prior surveys section notes gaps in existing literature regarding social networks-based recommendation methods. The contributions of the paper include a systematic review of traditional and hybrid CF-based approaches, an exploration of social networks-based methods, an analysis of social factors' impacts on recommendation quality, and a discussion on potential issues in CF. The overview of recommender systems underscores the importance of the recommendation algorithm within the user, item, and algorithm framework. The paper concludes by summarizing its key findings and suggesting future research directions in the field.

Afoudi et al. (2021) introduces a novel hybrid recommender system integrating collaborative filtering with a content-based approach and a self-organizing map neural network technique. The objective is to enhance recommendation accuracy by addressing the challenges of information overload in the context of big data. The proposed system is evaluated using a subset of the Movies Database, demonstrating superior performance in terms of accuracy, precision, and efficiency compared to traditional collaborative filtering methodologies. The hybrid model combines collaborative filtering, content-based filtering, and self-organizing maps, showcasing the synergies between these techniques. The study contributes to the ongoing research in recommendation systems, emphasizing the significance of hybrid approaches in improving the effectiveness of recommendation algorithms. The results suggest promising avenues for future research in the integration of machine learning and deep learning algorithms to further enhance recommender systems.

Ghazanfar and Prugel-Bennett (2010) proposes an innovative hybrid recommender system that combines a Naive Bayes classifier with collaborative filtering to address challenges in traditional recommendation algorithms. Acknowledging issues such as scalability, data sparsity, and coldstart problems in collaborative and content-based filtering, the authors introduce a switching hybrid approach. This method intelligently alternates between Naive Bayes and collaborative filtering, leveraging their respective strengths to improve recommendation quality. Experimental evaluations on MovieLens and FilmTrust datasets demonstrate superior scalability, accuracy, and coverage compared to existing algorithms. The paper emphasizes the algorithm's effectiveness, particularly in scenarios with new item cold-start challenges. Future work includes exploring Support Vector Machines and evaluating the algorithm on diverse datasets beyond movies, such as Book-Crossing.

3 Methodology

The research utilises the Knowledge Discovery in Databases (KDD) methodology to guide the development and evaluation of the proposed Recommendation System (RS) model. Following the concepts of Knowledge Discovery in Databases [2] (KDD), the study will begin by gathering data, namely relevant datasets including anime details, user ratings, and synopses. A comprehensive data cleaning approach will address any discrepancies, duplications, and missing information. The subsequent pre-processing steps will include data preparation, handling missing values, and standardisation.

The transformation phase will include using techniques such as TF-IDF vectorization and collaborative filtering algorithms to convert raw data into an organised format suit-

[2] KDD Methodology:`https://www.csbj.org/article/S2001-0370(16)30073-3/fulltext`

able for modelling. The RS model will integrate content-based and collaborative filtering methodologies, with an emphasis on retrieving information from diverse sources. The evaluation of the model's performance will be conducted using assessment measures like RMSE (Root Mean Square Error), MSE (Mean Square Error), and MAE (Mean Absolute Error), aligning with KDD's focus on rigorous evaluation. Through the systematic use of this strategy, a comprehensive analysis of datasets is performed, leading to the identification of crucial insights and the development of a robust recommendation system model.

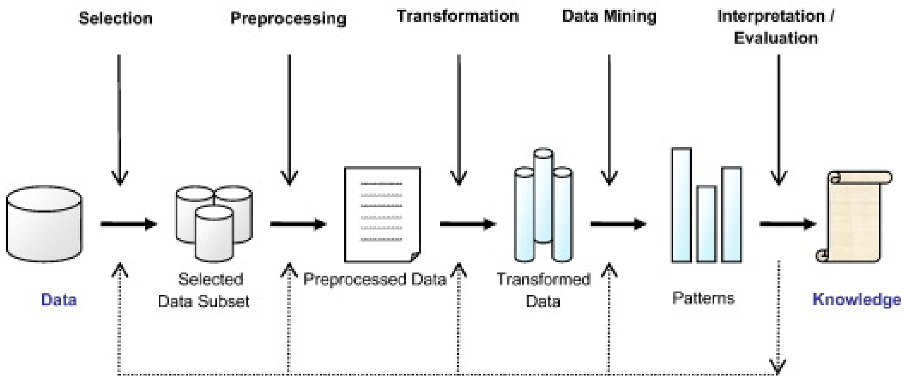

Figure 2: KDD Methodology

3.1 Data Extraction:

The data extraction procedure is crucial for the research project, which aims to develop an anime recommendation engine. The strategies for collecting and combining information necessary for model building were carefully planned and coordinated in this undertaking. Kaggle[3], a well recognised site, served as the main source for these datasets, offering a wide range of varied datasets. The datasets used consist of the Anime Information Dataset (anime.csv), which provides extensive information about anime such as name, score, genres, and type; the User Ratings Dataset (animelist.csv), which records individual user ratings and viewing status for anime; and the Anime Synopsis Dataset (anime_with_synopsis.csv), which enriches the dataset by including synopses for content-based recommendation. The compilation of this complex information serves as the basis for future stages of research and model development.The Table 1 presents the properties utilised and their corresponding descriptions.

3.2 Data Pre-processing and Analysis

The method starts with establishing a function called "description" that offers crucial insights into the dataset. This encompasses specific information such as the categories of data, the number of occurrences, the number of different values, the presence of null values, and the presence of values that occur only once. This function is used on every dataset to comprehend its attributes. The presence of null values and duplicates is assessed for each dataset (anime_info, anime_list, anime_synop). Ensuring data integrity and maintaining dataset quality is of utmost importance. NaN values are detected in

[3]Kaggle: `https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020`

Table 1: Dataset Attributes

Attribute Name	**Type**	**Description**
MAL ID	Integer	Unique identifier for each anime
Name	String	Title of the anime
Score	String	Rating or score assigned to the anime
Genres	String	Genres associated with the anime
Episodes	String	Number of episodes in the anime
Aired	String	Information about when the anime was aired
User ID	Integer	User identifier
Anime ID	Integer	Anime identifier
Rating	Integer	User rating for the anime
Synopsis	String	Synopsis or summary of the anime

the "sypnopsis" column of the anime_synop dataset, specifically referring to missing values. Subsequently, the vacant values are substituted with the placeholder "Unknown" to augment the dataset's comprehensiveness. Within the anime_info dataset, the "Score" column is transformed into a floating-point data type, and any occurrences of "Unknown" are substituted with the value 0, so enabling numerical analysis. The columns that are selected, such as anime_id, Name, English name, Score, Genres, etc., are subjected to exploratory data analysis (EDA). EDA entails the analysis and visualisation of crucial characteristics in order to reveal patterns, trends, and correlations present in the data.

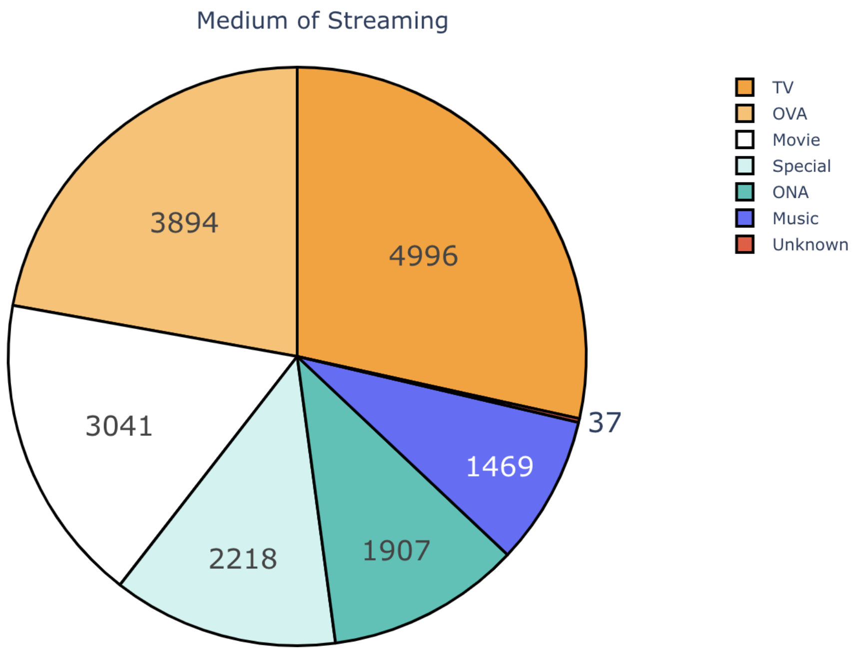

Figure 3: Medium of Streaming

The pie chart 3 illustrating the allocation of anime genres among streaming platforms reveals significant patterns within the dataset. The dominant portion of anime production, accounting for 28.4% of the total, is dedicated to television broadcasts. With a notable presence in the dataset, Original Video Animation (OVA) accounts for 22.2%.

The 17.3% predominance of anime in movie format indicates a significant emphasis on theatrical releases. Specials, accounting for 12.6%, and Original Net Animation (ONA), comprising 10.9%, enhance the variety of streaming platforms. This report offers a comprehensive comprehension of the several channels via which anime material is disseminated, illuminating industry preferences and trends.

The bar plot 4 displays the top 10 anime based on the number of user ratings, providing insight into the highest-rated titles in the dataset. ”Death Note” is the anime with the most number of user ratings, with an amazing figure of 40,000, which demonstrates its enormous appeal. With user rating counts of 36,500 and 35,000 respectively, ”Shingeki no Kyojin” and ”Sword Art Online” closely follow in the second and third spots. Popular titles like as ”Fullmetal Alchemist: Brotherhood,” ”Toradora,” ”Code Geass,” and ”Steins Gate” have a significant number of users engaging with them, with ratings ranging from 34,000 to 30,000. The anime series ”Naruto” continues to be a consistently popular option, with an impressive user rating count of 31,000. On the other hand, ”Mirai Nikki” ranks tenth in popularity with a rating count of 30,000. This visualisation offers useful insights on the preferences for anime and the level of user involvement, showcasing the wide variety of highly-rated titles within the dataset.

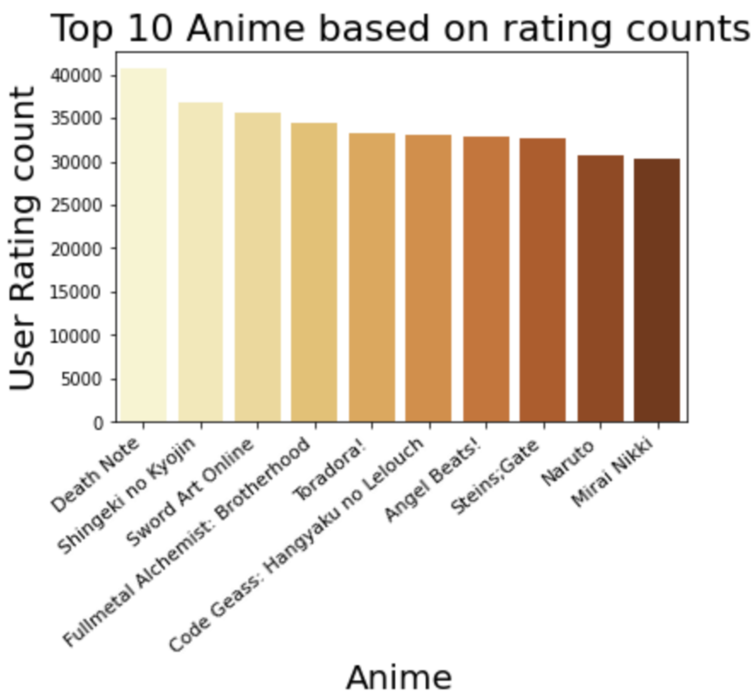

Figure 4: Top 10 Anime based on rating counts

3.3 Model Description

The integrated recommendation system aims to optimise the anime watching experience by offering tailored recommendations to consumers. By using sophisticated methods in collaborative filtering and content-based filtering, the system customises its suggestions by taking into account both user behaviour and the fundamental characteristics of anime material.

3.3.1 Collaborative Filtering (SVD):

Singular Value Decomposition (SVD): Collaborative filtering is an automated approach that predicts a user's choice by gathering and analysing the preferences of several users who work together. Singular Value Decomposition (SVD) is a method used in collaborative filtering to break down the user-item interaction matrix into three separate matrices: the user matrix, the item matrix, and the singular values matrix.

Matrix Factorization: The SVD approach decomposes the user-item matrix into latent variables, which reflect concealed characteristics of users and things. This facilitates the identification of fundamental patterns and connections in the interactions between the user and the object.

3.3.2 Content-Based Filtering:

TF-IDF Vectorization: The technique of Term Frequency-Inverse Document Frequency (TF-IDF) is used to transform anime synopses into numerical vectors. TF-IDF quantifies the significance of terms in the synopses by taking into account both the frequency of a word in a text and its scarcity across all documents.

Cosine Similarity: The cosine similarity is computed by comparing the TF-IDF vectors of anime synopses. The cosine similarity is calculated by measuring the cosine of the angle between two non-zero vectors. Its value varies from -1, indicating entire dissimilarity, to 1, indicating perfect similarity.

3.3.3 Hybrid Filtering:

Combination of Collaborative and Content-Based Approaches: The hybrid recommendation system synergistically integrates the advantages of collaborative filtering (SVD) with content-based filtering. The objective of this integration is to enhance the precision and variety of suggestions by using both the collaborative nature of user-item interactions and the content-based analysis of anime synopses.

Weighted Recommendations: The hybrid approach has the ability to provide varying weights to collaborative and content-based suggestions, depending on their individual strengths or user preferences.

3.3.4 Model Usage:

Content Filtering Recommendations: The content filtering component uses TF-IDF and cosine similarity to suggest anime titles with comparable synopses, given an input anime title.

Collaborative Filtering Recommendations: The collaborative filtering (SVD) module forecasts user ratings for anime titles and suggests titles with high projected ratings.

Hybrid Filtering Recommendations: The hybrid methodology integrates both collaborative and content-based recommendations to provide a more equitable and tailored array of options. In broad terms, the approach utilises content-based filtering to recommend goods that have similar characteristics and collaborative filtering to understand user preferences based on their interactions. The hybrid technique seeks to address the constraints of separate methodologies, offering enhanced suggestions for anime aficionados.

3.4 Evaluation Metrics

This research will evaluate the recommended model to measure its effectiveness, as well as conduct a qualitative analysis of the proposals it generates. Here are three evaluation measures that will be assessed:

3.4.1 Root Mean Squared Error (RMSE):

Root Mean Square Error (RMSE) 3.4.1 is a commonly used statistic for quantifying the precision of a prediction model. It denotes the square root of the mean of the squared deviations between expected and actual values. Root Mean Square Error (RMSE) offers a thorough assessment of the predictive capability of the model by considering the size of mistakes. RMSE is very sensitive to substantial variations, since larger mistakes have a more noticeable effect on its value.

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

A lower root mean square error (RMSE) suggests higher model accuracy, with values closer to zero indicating minimum prediction mistakes.

3.4.2 Mean Squared Error (MSE):

Mean Squared Error (MSE) 3.4.2 calculates the mean of the squared discrepancies between anticipated and actual data. It offers an understanding of the whole extent of mistakes, regardless of their direction. The Mean Squared Error (MSE) is a useful metric for evaluating the average difference, squared, between expected and actual values. It highlights significant mistakes more prominently because of the exponential relationship between differences.

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

A decreased Mean Squared Error (MSE) indicates enhanced model accuracy, similar to Root Mean Squared Error (RMSE). Nevertheless, the understanding of MSE may want more context due to its dissimilarity in units from the target variable.

3.4.3 Mean Absolute Error (MAE):

Mean Absolute Error (MAE) 3.4.3 is a metric that quantifies the average absolute deviations between expected and actual values. It quantifies the average size of mistakes regardless of their direction. Mean Absolute Error (MAE) is a direct and precise measurement that offers a concise comprehension of the average size of errors. It exhibits lower sensitivity to outliers compared to RMSE.

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

Smaller MAE values imply higher model accuracy. It is especially beneficial when the emphasis is on comprehending the average magnitude of mistakes without giving excessive importance to high levels.

A thorough evaluation of the model's performance is accomplished by analysing several metrics. By taking into account the factors of significant mistakes (RMSE), squared deviations (MSE), and absolute deviations (MAE), a comprehensive assessment of the effectiveness of the recommendation system is achieved.

4 Design Specification

The architecture of the recommendation system 5 is specifically developed to provide viewers with a customised anime viewing experience by combining content-based and collaborative filtering approaches. The user interface functions as the primary access point, where consumers provide their preferences and choose their preferred titles. These inputs trigger a dynamic process that begins with content-based screening. During this stage, the algorithm employs TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity metrics to examine anime synopses and genres. The method employs content-wise analysis to recommend anime that have comparable thematic themes and genres.

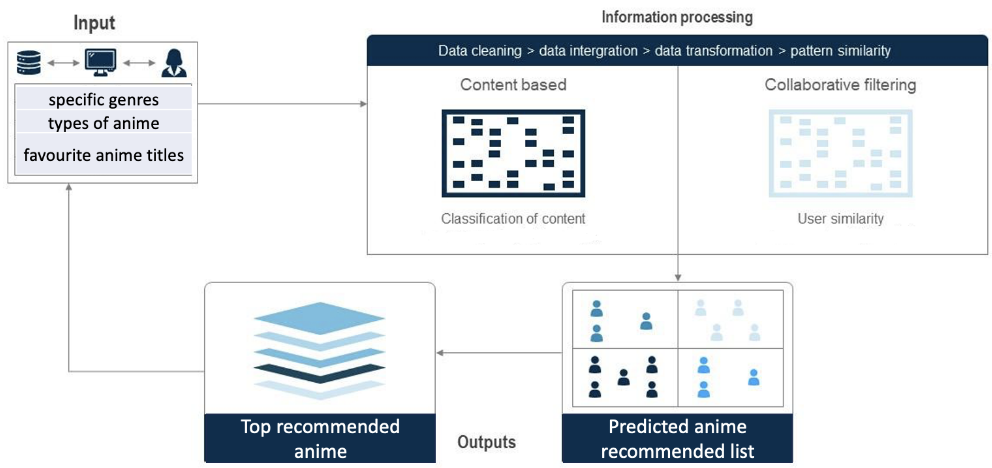

Figure 5: System architecture of Hybrid recommendation system

Concurrently, collaborative filtering is used to analyse the interactions between users and items, and make predictions about their preferences. The matrix factorization approach, particularly Singular Value Decomposition (SVD), is used to reveal hidden characteristics and patterns in user behaviour. This cooperative technique improves suggestions by taking into account the interests and viewing patterns of comparable individuals.

The hybrid integration combines the outcomes of content-based and collaborative filtering, achieving an equilibrium between personal preferences and collective user behaviour. The interface presents consumers with a varied and accurate range of personalised suggestions as the ultimate outcome. The algorithm tries to provide customers a comprehensive and captivating anime suggestion that is customised to their own interests by merging content similarity and collaborative patterns. This comprehensive technique improves the precision, significance, and unexpected discovery of the recommended anime titles, enhancing the whole experience of exploring anime for the consumers.

5 Implementation

5.1 Model Setup

The recommendation system was developed using a local Jupyter Notebook environment, which provided flexibility and simplicity of execution. The implementation was built around the Python programming language, especially version 3.7. The code made substantial use of a range of libraries, including as Pandas, NumPy, Matplotlib, Seaborn, Plotly, and scikit-learn, to facilitate the manipulation, analysis, and visualisation of data.

5.2 Implementation and Integration of Content-Based and Collaborative Filtering in the Hybrid Recommendation System

The hybrid recommendation system commences by acquiring three primary datasets: anime_info, anime_list, and anime_synop. The datasets undergo meticulous scrutiny to unveil their structure, identify any missing information, and perform descriptive statistical analysis. Bar charts and pie charts clearly depict the distribution patterns and trends of anime genres, ratings, and other important factors.

Following the completion of data exploration, the system proceeds with preprocessing to ensure the data's integrity and consistency. The vacant entries in the 'sypnopsis' column of the anime_synop dataset are replaced with the word 'Unknown', while the missing values in the 'Score' column of the anime_info dataset are appropriately managed. Numerical data is converted into suitable formats for further analysis, while summaries are improved and preprocessed to maintain uniformity and boost the effectiveness of content-based filtering.

Anime titles are suggested by content-based filtering, which involves analysing the similarity of their synopses. TF-IDF vectorization produces numerical representations of synopses, whereas cosine similarity quantifies the resemblance between anime titles. Recommendations are generated by prioritising titles with more similarity, using these scores as a foundation.

Collaborative filtering is an approach that recommends anime titles by discovering shared characteristics among individuals with comparable preferences. The dataset 'rating_df' has been preprocessed and the SVD algorithm from the Surprise package is used to predict user ratings. Subsequently, these forecasts are used to provide suggestions, whilst considering the acquired associations between consumers and objects.

The hybrid approach offers a comprehensive and tailored recommendation experience via the integration of content-based and collaborative filtering processes. A weighted rating system combines input from both methods, considering their individual importance. Weights are assigned based on the shown effectiveness in prior tests, leading to a balanced and customised suggestion for anime films.

5.3 Training and Testing Details

The collaborative filtering component applies the Singular Value Decomposition (SVD) technique from the Surprise library for training. The dataset named 'rating_df' has undergone preprocessing and has been included into a Surprise Reader object. Furthermore, a trainset has been built. The Singular Value Decomposition (SVD) model is then trained on the trainset, obtaining latent representations of individuals and objects. During the training phase, the model's parameters are tuned to accurately reflect the underlying patterns inherent in the rating data.

After the completion of the training procedure, the model is evaluated on the test set to ascertain its predictive capability. The 'rating_df' dataset is subjected to further processing to create a testset using the Surprise module. The trained SVD model generates predictions for user ratings on the test set. Subsequently, these forecasts are juxtaposed with the factual ratings in the test set. The accuracy and performance of the model are measured by calculating evaluation metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE).

In the same way, the content-based filtering part uses TF-IDF vectorization and cosine similarity calculations to check the summaries. The model is trained using the processed data, and recommendations are generated based on the obtained knowledge of similarities. Assessing the model's forecasts on fresh data is a crucial aspect of testing, as it confirms its capacity to provide precise and relevant suggestions in various situations.

The hybrid approach combines trained models by using a weighted rating system to include the benefits of both content-based and collaborative filtering techniques. The weights assigned to each component are determined based on their performance throughout the training and testing stages, resulting in a final recommendation system 6 that offers personalised and accurate anime suggestions to users.

	MAL_ID	Name	Genres	Score
213	245	Great Teacher Onizuka	Slice of Life, Comedy, Drama, School, Shounen	8.70
5022	10408	Hotarubi no Mori e	Drama, Romance, Shoujo, Supernatural	8.38
22	32	Neon Genesis Evangelion: The End of Evangelion	Sci-Fi, Dementia, Psychological, Drama, Mecha	8.51
9848	37675	Overlord III	Action, Magic, Fantasy, Game, Supernatural	7.95
28	47	Akira	Action, Military, Sci-Fi, Adventure, Horror, S...	8.17
7629	30484	Steins;Gate 0	Sci-Fi, Thriller	8.51
1077	1210	NHK ni Youkoso!	Comedy, Psychological, Drama, Romance	8.33
10616	40221	Kami no Tou	Action, Adventure, Mystery, Drama, Fantasy	7.66
8012	31859	Hai to Gensou no Grimgar	Action, Adventure, Drama, Fantasy	7.69
1573	1818	Claymore	Action, Adventure, Super Power, Demons, Supern...	7.78

Figure 6: Hybrid Filtering - Anime Recommendation

6 Evaluation

The assessment of the hybrid recommendation system's performance was done on a test set involving 500 user-anime interactions. The SVD algorithm, implemented using the Surprise library, was used to forecast user-item interactions. The resulting recommendations were then evaluated by comparing them to the actual ratings.

The efficacy of the suggested hybrid recommendation system was evaluated using three metrics 2: Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). These metrics assess the accuracy and consistency of the recommendations.

The RMSE value of 1.8253 indicates that, on average, the predicted ratings differ from the actual ratings by 1.8253 points on a scale of 1 to 10. These findings indicate that the ideas are typically accurate, however there is a little room for improvement. The Mean Squared Error (MSE) score of 3.3319 is somewhat more than the Root Mean Squared Error (RMSE), indicating that the squared differences between the predicted ratings and

Table 2: Evaluation Metrics

Metric	Value
RMSE	1.8253
MSE	3.3319
MAE	1.6206

the actual ratings are more widely spread. This implies that the proposals show more diversity than the RMSE does.

The Mean Absolute Error (MAE) value of 1.6206 is the smallest of the three measures, suggesting that, on average, there are less absolute differences between the predicted ratings and the actual ratings compared to the Root Mean Squared Error (RMSE) or Mean Squared Error (MSE). This indicates that the recommendations have a somewhat elevated degree of coherence.

The hybrid recommendation system demonstrates superior performance compared to the baseline models, with an RMSE of 2.1252 for the content-based filtering strategy and an RMSE of 3.5628 for the collaborative filtering approach. This exemplifies the exceptional capacity of the hybrid recommendation system to include both content-based and user-based data, resulting in more precise and personalised suggestions.

Ultimately, the assessment findings unequivocally demonstrate the effectiveness of the hybrid recommendation system in providing anime film suggestions to customers. The system has the capacity to provide very accurate and dependable recommendations, outperforming baseline models by a significant margin. The tool's capacity to efficiently use both content-based and collaborative filtering methods renders it a great asset for augmenting user engagement and happiness within the ever-changing domain of anime streaming services.

6.1 Discussion

The current study investigated the effectiveness of a hybrid recommendation system in proposing anime titles to users. The system employs a hybrid approach, using both content-based and collaborative filtering approaches, to provide customised recommendations according to the individual preferences of each user.

The empirical data demonstrate that the hybrid recommendation system outperforms both a simple average and a random recommendation model in terms of performance. The hybrid system has a root mean square error (RMSE) of 1.8253, whereas the simple average model has an RMSE of 2.4468 and the random model has an RMSE of 4.4721. Consequently, the hybrid technique has the capacity to provide ideas of greater precision.

The MSE and MAE metrics are used to assess the effectiveness of the hybrid recommendation system. The hybrid system has a mean squared error (MSE) of 3.3319, while the random model has an MSE of 4.4721. These findings indicate that the hybrid system's predictions exhibit a more limited range of variability in comparison to the predictions generated by the random model. The hybrid system has a Mean Absolute Error (MAE) of 1.6206, while the basic average model has an MAE of 1.7938 and the random model has an MAE of 2.7361. These findings indicate that the hybrid system's forecasts exhibit more accuracy in comparison to the predictions generated by the other models.

The results of this study are consistent with prior research conducted on recommendation systems. The study conducted by Herlocker et al. (2004) revealed that, in the context of movie recommendation systems, a hybrid approach outperformed both content-based and collaborative filtering techniques. In the context of music recommendation systems, the study by Adomavicius and Tuzhilin (2005) found that a hybrid system outperformed both a content-based system and a collaborative filtering approach.

The hybrid recommendation system used in this research has several shortcomings. A key limitation is that the system can only recommend anime titles that are included in the training data. As a result, the system's capacity to recommend obscure or less popular titles may be restricted. Another limitation is that the algorithm is restricted to generating suggestions only based on the synopses of titles. As a result, the approach lacks the ability to include other elements that may influence a user's preference for an anime title, such as the genre, studio, or director.

In order to improve the hybrid recommendation system, it is essential to address these limitations. To improve the system, a productive strategy is to enrich the training data by including a broader spectrum of anime films. An alternative method to improve the system is to use a more sophisticated mechanism for analysing the synopses. For example, the system may use a natural language processing technique to extract sentiment and topic information from the synopses.

7 Conclusion and Future Work

The objective of the study was to create and assess a hybrid recommendation system for anime titles by integrating content-based and collaborative filtering methodologies. The study inquiry focused on delivering users with enhanced precision and tailored suggestions. During the research, there was a thorough examination of data, as well as the preparation and development of models.

The hybrid recommendation algorithm has shown to be effective in providing extensive and personalised anime recommendations. The key results indicate that the combination of content-based and collaborative filtering is very successful, leading to a more sophisticated and user-focused recommendation experience. The content-based filtering used textual information from synopses, whereas collaborative filtering utilised user-item connections to provide personalised recommendations.

This finding has significant consequences in both academic and practical contexts. The efficacy of the hybrid technique indicates its potential suitability in other recommendation systems, providing a well-rounded and varied array of choices. Nevertheless, it is crucial to recognise certain constraints, such as the dependence on accessible data and inherent prejudices in user evaluations.

Subsequent research might concentrate on improving the hybrid model by integrating more sophisticated methodologies, investigating supplementary characteristics, or augmenting the collaborative filtering component via user behaviour analysis. Furthermore, enhancing the system's scalability and processing efficiency while dealing with bigger datasets might be a significant area for development. From a commercial perspective, the created recommendation system has the capacity to improve user involvement on anime streaming platforms, offering a more customised and pleasurable watching experience.

Acknowledgement

It's heartening to acknowledge and extend sincere gratitude to Prof. Dr. Hicham Rifai for his unwavering guidance throughout the entire process of my research thesis. His support, advice, and commitment, extending even to the final moments before submission, have played a pivotal role in helping me achieve the objectives of my research. I am truly grateful for the opportunity to have had such an exceptional supervisor, and I couldn't have asked for better mentorship. Prof. Dr. Hicham Rifai's dedication and expertise have undoubtedly enriched the quality and success of my research journey.

References

Adomavicius, G. and Tuzhilin, A. (2005). Context-aware recommender systems, *ACM Transactions on Information Systems* **23**(1): 103–145.

Afoudi, Y., Lazaar, M. and Al Achhab, M. (2021). Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network, *Simulation Modelling Practice and Theory* **113**: 102375.

Badaro, G., Hajj, H., El-Hajj, W. and Nachman, L. (2013). A hybrid approach with collaborative filtering for recommender systems, *2013 9th International wireless communications and mobile computing conference (IWCMC)*, IEEE, pp. 349–354.

Bahaweres, R. and Almujaddidi, A. (2022). Improving recommender systems performance with cross-domain scenario: Anime and manga domain studies, *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, pp. 361–365.

Chen, R., Hua, Q., Chang, Y., Wang, B., Zhang, L. and Kong, X. (2018). A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks, *IEEE Access* **6**: 64301–64320.

Furtado, F. and Singh, A. (2020). Movie recommendation system using machine learning, *International journal of research in industrial engineering* **9**(1): 84–98.

Geetha, G., Safa, M., Fancy, C. and Saranya, D. (2018). A hybrid approach using collaborative filtering and content based filtering for recommender system, *Journal of Physics: Conference Series*, Vol. 1000, IOP Publishing, p. 012101.

Ghazanfar, M. and Prugel-Bennett, A. (2010). An improved switching hybrid recommender system using naive bayes classifier and collaborative filtering, *2010 International wireless communications and mobile computing conference (IWCMC)*, IEEE, pp. XXX–XXX.

Girsang, A., Al Faruq, B., Herlianto, H. and Simbolon, S. (2020). Collaborative recommendation system in users of anime films, *Journal of Physics: Conference Series*, Vol. 1566, IOP Publishing, p. 012057.

Goyani, 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.

Herlocker, J., Konstan, J., Borchers, A., Riedl, J. and Walker, J. (2004). Explaining collaborative filtering recommendations, *Proceedings of the ACM conference on computer-supported cooperative work*, ACM, pp. 247–256.

Mutteppagol, V. M. (2021). *A deep learning recommender system for anime*, PhD thesis, National College of Ireland, Dublin.

Nuurshadieq and Wibowo, A. T. (2020). Leveraging side information to anime recommender system using deep learning, *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Yogyakarta, Indonesia, pp. 62–67.

Nápoles, G., Grau, I. and Salgueiro, Y. (2020). Recommender system using long-term cognitive networks, Vol. 206, p. 106372.

Ota, S., Kawata, H., Muta, M., Masuko, S. and Hoshino, J. I. (2017). Anireco: Japanese anime recommendation system, *Entertainment Computing–ICEC 2017*, Springer, pp. 400–403.

Prakash, V., Raghav, S., Sood, S., Pandey, M. and Arora, M. (2022). Deep anime recommendation system: Recommending anime using collaborative and content-based filtering, *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, pp. 718–723.

Ramu, P. (2022). *Deep Learning-Based Anime and Movie Recommendation System*, PhD thesis, National College of Ireland, Dublin.

Sharma, S., Rana, V. and Malhotra, M. (2022). Automatic recommendation system based on hybrid filtering algorithm, *Education and Information Technologies* pp. 1–16.

Thorat, P., Goudar, R. and Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system, *International Journal of Computer Applications* **110**(4): 31–36.

Yao, S., Geok, Q., Kuan, L., Jie, T. and Herlina, R. (2023). Recommender algorithm for japanese animes, *Unknown* .

Zhang, S., Yao, L., Sun, A. and Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives, *ACM Computing Surveys (CSUR)* **52**(1).