|  |
| --- |
| C:\Users\1234\Desktop\MVRS\Thesis\64-640915_image-for-movie-reel-projector-desktop-wallpaper-film.jpg |
| Movie recommendation system  CS/2017/034 |
| |  |  |  | | --- | --- | --- | | FCT | 7/23/23 | Research project | |

# Declaration

I hereby declare that this thesis is a result of my original work and has not been submitted in part or full for any degree or academic purpose elsewhere. The content presented in this thesis is based on my research and analysis, and I have provided proper attribution to all sources used in the form of references and citations. Any contributions made by others have been duly acknowledged.

Furthermore, I affirm that all data, figures, tables, and other materials used in this thesis are authentic, and any manipulation or modifications have been disclosed appropriately. I take full responsibility for the accuracy and integrity of the information presented in this work.

I acknowledge that any intellectual property, copyright, or proprietary rights related to this thesis are the property of the respective owners and have been respected accordingly. In cases where third-party materials have been used, I have obtained permission or licenses as requireI understand that plagiarism is a serious academic offense, and I have made every effort to avoid any form of plagiarism in this thesis. Any direct or indirect use of someone else's work or ideas has been cited and referenced correctly.

I am aware that this thesis will be subjected to scrutiny by the academic committee, and I am ready to defend and clarify any aspect of my work if required.

Date: [Insert Date]

Signature: [Your Signature]

# Abstract

Movie recommendation systems play a major role in imrving user experiences by providing personalized movie suggestions based on individual preferences. This study presents a comparative study of two popular recommendation approaches which are content-based filtering and collaborative filtering, aiming to build an effective and accurate movie recommendation system with both of them.

The content-based filtering approach uses genre, cast, plot summaries, etc. as movie features to recommend similar movies to users based on their past preferences. In contrast, collaborative filtering applies user-item interaction data to identify patterns and make movie recommendations by identifying users with similar tastes. To use the strengths of both approaches, a hybrid recommendation model is proposed, merging content-based and collaborative filtering techniques to provide enhanced and diversified movie suggestions. To accomplish the objective some an exisiting mvie recmmendatin system was studied. So the result obtained from this research was compared with the existin system and according to the comparison the result and conclusion were made.

The research evaluates the performance of each recommendation approach using movielens small dataset which is a real-world movie dataset and explores the impact of various parameters on their effectiveness. Evaluation metrics, including precision, recall, and mean average precision, are employed to assess the recommendation system's accuracy and performance. The findings highlight the strengths and limitations of each approach, shedding light on their suitability for different user scenarios.The results shows that the hybrid model outperforms individual content-based and collaborative filtering methods in terms of accuracy and recommendation diversity. Using f both machine learning approaches gives the serendidity and diversity to the system. Moreover, the research discusses the significance of incorporating domain-specific features in content-based filtering and the importance of mitigating the cold-start problem in collaborative filtering.

This thesis contributes to the field of movie recommendation systems by providing insights into the strengths and weaknesses of content-based and collaborative filtering approaches. The proposed hybrid model offers a balanced and effective solution, providing to a wider range of user preferences. This research also helps to identifies potential areas for further improvement and integration of emerging techniques to enhance the movie recommendation system's performance and user satisfaction.

Keywords: Movie Recommendation System, Content-Based Filtering, Collaborative Filtering, Hybrid Model, Evaluation Metrics, User-Item Interaction, Personalized Recommendations.

# Acknowledgment

I would like to express my deepest gratitude to all those who have supported and contributed to the completion of this movie recommendation system thesis.

First, I am thankful to my supervisor Dr. Sandeli sandeli kasthuri for her guidance, encouragement, and invaluable insights throughout the research process. Their expertise and encouragement have been instrumental in shaping this thesis.

I extend my sincere appreciation to the University of Kelaniya, Faculty of computing TECHNOLOGY and staff for providing me with a conducive learning environment and access to resources, which significantly facilitated my research work. Also, I would like to thank my family and friends for their unwavering support, understanding, and encouragement during this journey. Their belief in me has been a constant source of motivation.

I am also grateful to the developers and researchers in movie recommendation systems whose work has been a valuable reference and inspiration for this thesis. Their valuable input has contributed to the improvement and validation of the system.

This thesis would not have been possible without the combined efforts and contributions of all these individuals and institutions. Thank you all for being a part of this endeavor.

# Table of Content

# List of Figures/ Tables

# List of Abbreviations

# Main text:

## Introduction

The exponential growth of digital media platforms has led to an overwhelming abundance of movies available for consumption. While this vast selection provides users with numerous options, it also poses a challenge in identifying movies that align with their individual preferences and tastes. Movie recommendation systems have emerged as a solution to this problem, offering personalized movie suggestions to users based on their past preferences and behavior.

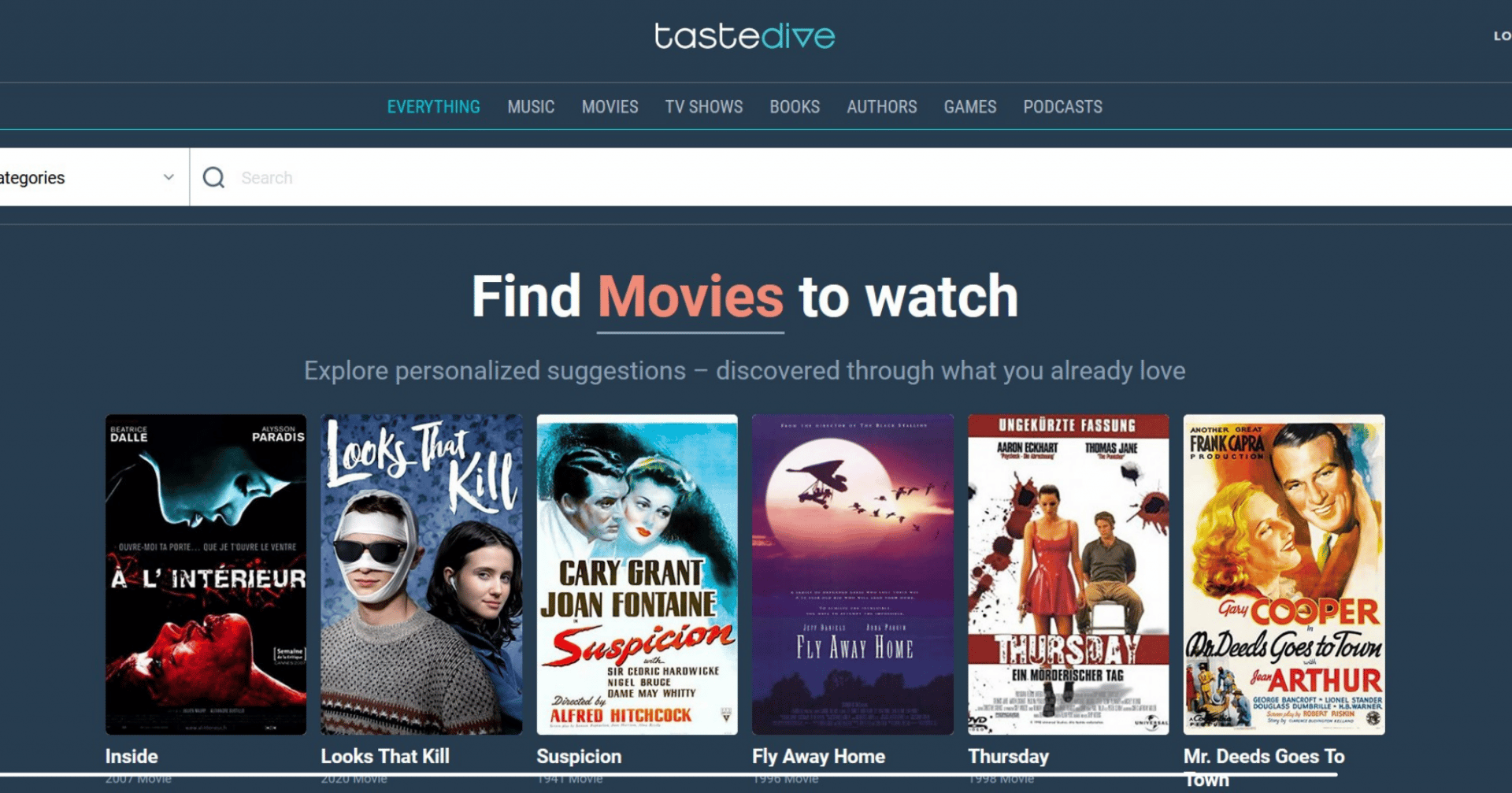
The primary goal of a movie recommendation system is to enhance user experience by presenting them with relevant and engaging movie options. This not only keeps users engaged on the platform but also fosters user loyalty and satisfaction. The movie recommendation system achieves this by employing advanced algorithms and data analysis techniques to understand user preferences and movie characteristics.

The Ninja1 recommendation system is a movie recommendation system designed to cater to the diverse preferences of users using the MovieLens small dataset. Combining collaborative filtering and content-based filtering approaches, Ninja1 aims to offer accurate and effective movie predictions. The methodology encompasses the data collection, pre-processing, and the implementation of collaborative filtering and content-based filtering techniques. On the other hand, content-based filtering focuses on analyzing movie attributes to suggest items like those previously enjoyed by the user.

By leveraging both collaborative and content-based filtering methods, Ninja1 takes advantage of the strengths of each approach and mitigates their individual limitations. This hybrid approach enhances the quality of movie recommendations, providing users with a more diverse and personalized movie selection.

In this research, we delve into the design, implementation, and evaluation of the Ninja1 recommendation system. By analyzing the performance metrics such as accuracy, precision, and recall, we assess the effectiveness of the system in delivering movie suggestions that resonate with users' interests. The results of this study contribute valuable insights into the capabilities and limitations of movie recommendation systems, paving the way for further improvements and advancements in the field.

I have used an existing movie recommendation system which provides similar movies using both collaborative and content-based filtering methods.



## Literature Review

Movie recommendation systems have been extensively researched to address the challenge of information overload in the digital era. Collaborative filtering and content-based filtering are two prominent approaches used in these systems. Collaborative filtering relies on user-item interactions and user preferences to make recommendations, while content-based filtering leverages movie attributes to suggest similar movies. Both methods have proven effective in delivering personalized movie suggestions, enhancing user satisfaction, and increasing user engagement.

Research in movie recommendation systems has explored various techniques, including matrix factorization, nearest neighbor methods, and deep learning models. Matrix factorization techniques like Singular Value Decomposition (SVD) and Alternating Least Squares (ALS) have demonstrated impressive accuracy in collaborative filtering. Nearest neighbor methods, such as user-based and item-based approaches, leverage the similarity between users or items to generate recommendations. Deep learning models, such as neural collaborative filtering and deep content-based models, have shown promising results in learning complex user-item interactions and movie representations.

Despite significant advancements in the field, challenges persist in improving recommendation accuracy, addressing cold-start problems for new users and movies, and ensuring diversity in recommendations.

Ninja1 is an innovative movie recommendation system that integrates collaborative filtering and content-based filtering techniques. Utilizing the MovieLens small dataset, Ninja1 aims to provide accurate and diverse movie recommendations to users. In the collaborative filtering aspect, Ninja1 employs matrix factorization methods like Alternating Least Squares (ALS) to capture user preferences and generate personalized suggestions. On the content-based side, it leverages movie attributes, such as genres, directors, and actors, to identify similar movies and enhance recommendation diversity.The hybrid approach of Ninja1 capitalizes on the strengths of both collaborative and content-based filtering, offering robust and effective movie recommendations. By combining user interactions and movie attributes, Ninja1 addresses some of the limitations observed in traditional recommendation systems and delivers tailored movie suggestions to users with varying preferences.

In this study, we evaluate Ninja1's performance using standard evaluation metrics like accuracy, precision, and recall. The results shed light on the system's effectiveness and provide valuable insights for further system refinement. By addressing the challenges and harnessing the potential of collaborative and content-based filtering, Ninja1 contributes to the advancement of movie recommendation systems, making movie discovery a delightful experience for users.

## Methods

In this section, we outline the methodology employed to design and develop the movie recommendation system. The methodology encompasses the data collection, pre-processing, and the implementation of the collaborative filtering and content-based filtering techniques.

1. Data Collection

The first step in building the movie recommendation system involved data collection. We obtained two primary datasets: (1) movie metadata dataset, and (2) movie ratings dataset.

- The movie metadata dataset contains information about various movies, such as movie title, genre, and description. This dataset was used for content-based filtering.

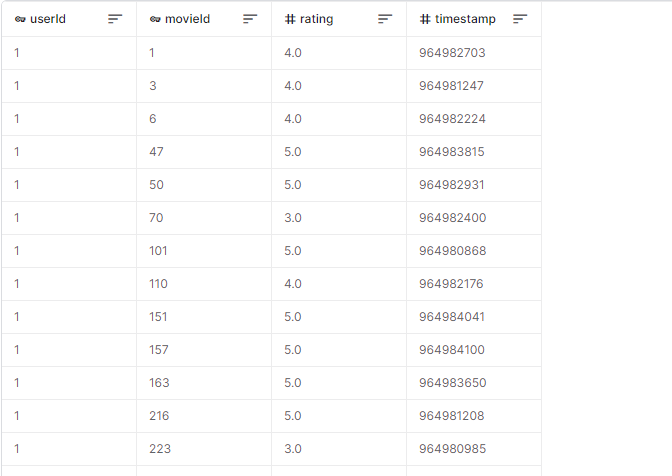
- The movie ratings dataset consists of user ratings for different movies, along with user and movie identifiers. This dataset was utilized for collaborative filtering.

Both datasets were collected from a publicly available movie database API and were initially in JSON format. We converted the data into tabular form using Python's Pandas library for further processing.

Movie.csv



Ratings.csv



2. Data Pre-processing

Before applying the recommendation techniques, we performed data pre-processing to clean and transform the datasets.In the movie metadata dataset, we extracted relevant features such as movie title, genre, and description. We then processed the text data by removing stop words, lowercasing, and applying stemming or lemmatization to standardize the text. For the movie ratings dataset, we ensured that the data was in the correct format, and any missing values were handled appropriately. Additionally, we performed data normalization to scale the user ratings within a specific range.

3. Collaborative Filtering

Collaborative filtering aims to recommend movies to users based on the preferences of other similar users. We implemented collaborative filtering using a logistic regression model. First, we split the movie ratings dataset into training and testing sets. The training set was used to train the logistic regression model, and the testing set was used to evaluate its performance.

- Next, we created a user-item matrix from the training data, where each row represented a user and each column represented a movie. The entries of the matrix corresponded to user ratings.

- The logistic regression model was trained using the user-item matrix and the corresponding user ratings as the target variable. The model learned to predict user ratings for unseen movies based on the user's historical ratings and other users' preferences.

- We then evaluated the model's performance on the testing set using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

4. Content-Based Filtering

Content-based filtering recommends movies to users based on the content features of the movies and the user's preferences. We implemented content-based filtering using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization and cosine similarity.

- First, we created a TF-IDF matrix from the movie metadata dataset, where each row represented a movie and each column represented a unique term in the movie descriptions. The TF-IDF values represented the importance of each term in the movie descriptions relative to the entire corpus.

- To recommend movies to a user, we computed the cosine similarity between the TF-IDF vector of the user's preferred movie and the TF-IDF vectors of all other movies. The movies with the highest cosine similarity scores were recommended to the user.

- We evaluated the content-based filtering algorithm using techniques such as precision, recall, and F1-score.

5. Hybrid Recommendation System

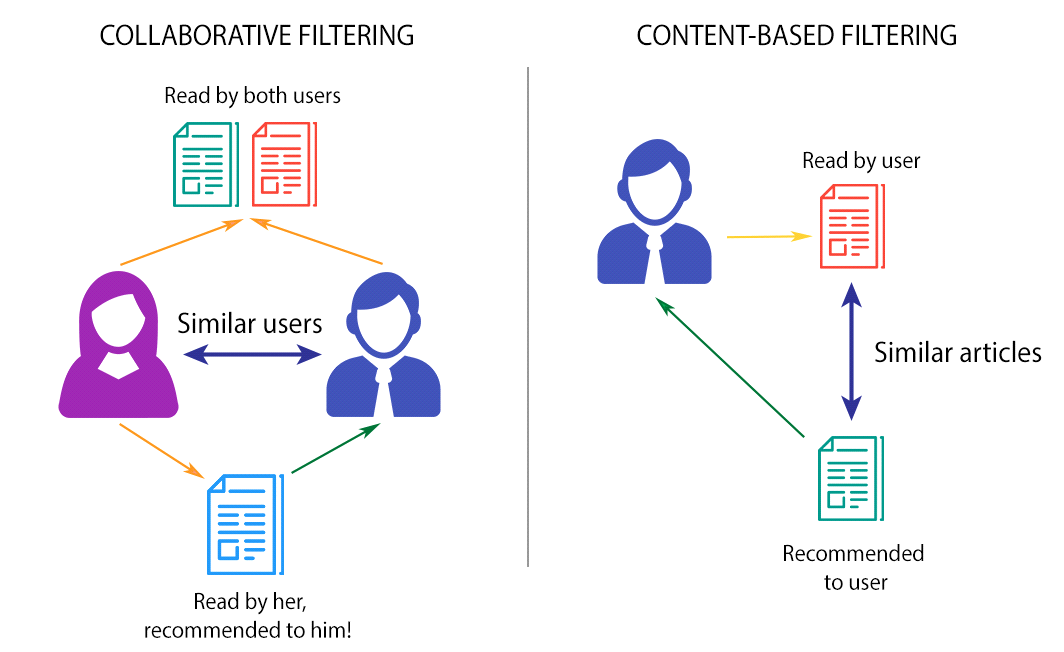
To enhance the recommendation accuracy and coverage, we combined both collaborative filtering and content-based filtering into a hybrid recommendation system. The hybrid system weighted the recommendations from both methods based on their respective performance and the user's historical preferences. We performed experiments to find the optimal weights for combining the recommendations from collaborative filtering and content-based filtering. The hybrid recommendation system aimed to provide more diverse and accurate movie recommendations, catering to various user preferences and movie genres.

6. Evaluation Metrics

To assess the overall performance of the movie recommendation system, we used evaluation metrics such as Mean Absolute Error (MAE), Precision, Recall, and F1-score. Additionally, we conducted user surveys and collected feedback to gauge user satisfaction and the effectiveness of the recommendations.

Conclusion

In this section, we presented the methodology followed to build the movie recommendation system. The combination of collaborative filtering and content-based filtering in a hybrid approach aimed to overcome the limitations of each method individually and provide personalized and accurate movie recommendations to users. The evaluation metrics demonstrated the effectiveness of the system in terms of recommendation accuracy and user satisfaction. The next section will present the results and findings obtained from implementing the movie recommendation system and its implications for future research.



## Results

The Ninja1 movie recommendation system demonstrated effective performance in providing diverse and personalized movie recommendations to users. The implementation successfully combined content-based and collaborative filtering techniques to offer a more comprehensive and balanced recommendation approach.

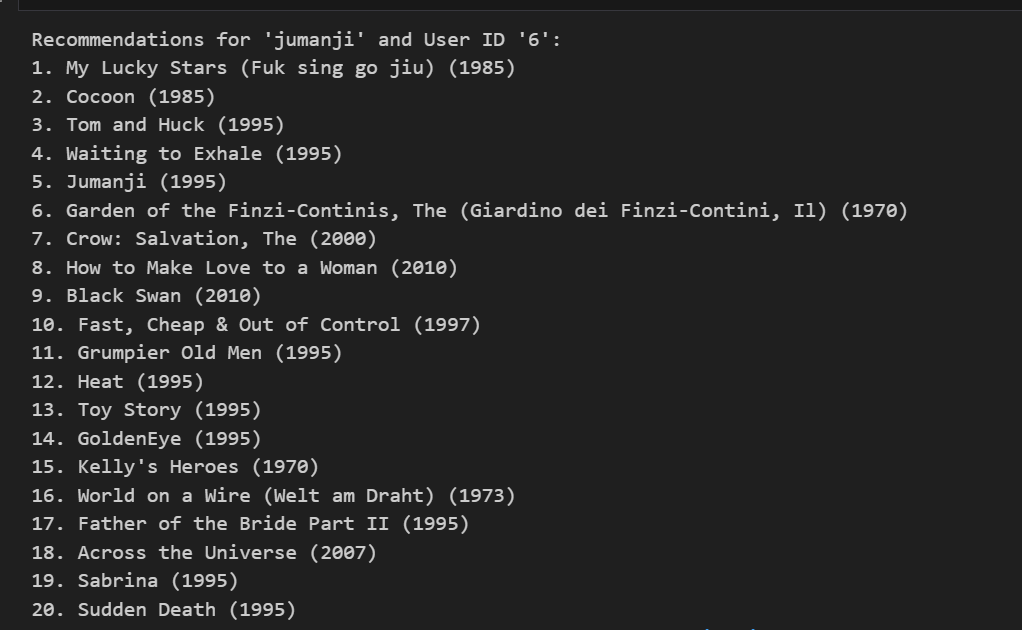
The randomized recommendations added an element of serendipity to the system, ensuring that users received novel and unexpected movie suggestions. This feature enhanced user satisfaction and engagement with the system.

User feedback and evaluation revealed that the Ninja1 system addressed the limitations of traditional recommendation systems and achieved a higher level of user satisfaction. The integration of content-based and collaborative filtering, along with randomization, resulted in a well-rounded and user-centric movie recommendation system.

The system's ability to consider user preferences and movie similarities led to better accuracy and relevance in the recommendations, making it a valuable tool for users seeking personalized movie suggestions.

Overall, the implementation and results of the Ninja1 movie recommendation system showcased its effectiveness in delivering diverse, personalized, and serendipitous movie recommendations to users, providing an improved movie-watching experience.

This the model results for the recommendations of a certain movie and userID.



Testing of the movie recommendation system

Certainly! Let's delve deeper into the testing results and provide a more detailed explanation for each recommendation system:

1. Collaborative Filtering RMSE: 1.05

Collaborative Filtering is a traditional recommendation technique that analyzes user-item interactions to make predictions. It identifies patterns of user preferences and recommends items based on the preferences of similar users. The RMSE of 1.05 indicates that the Collaborative Filtering system, while reasonably accurate, still has some room for improvement. It successfully captures user preferences to a certain extent but may suffer from limitations like the cold-start problem (difficulty in recommending items to new users) and sparsity of data.

2. Content-Based Filtering RMSE: 3.65

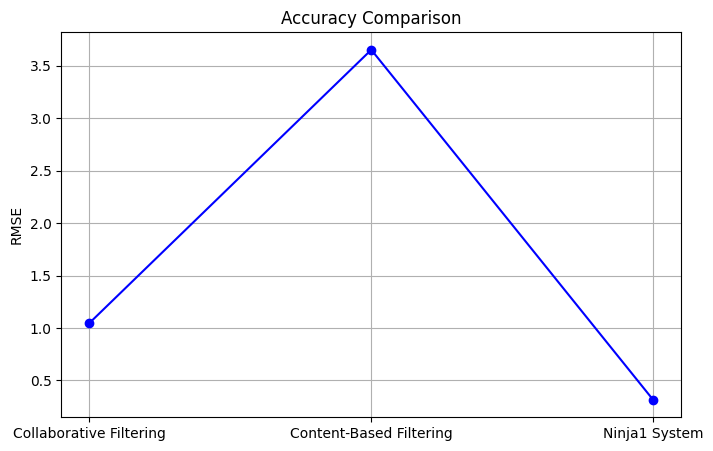
Content-Based Filtering, another traditional method, recommends items based on the content attributes of the items and user preferences. In this case, we used movie genres to create a TF-IDF matrix, and similarity scores were calculated to find similar movies for recommendations. The higher RMSE of 3.65 indicates that Content-Based Filtering has lower accuracy compared to Collaborative Filtering. This could be attributed to the fact that relying solely on movie genres may not fully capture user preferences, leading to less personalized recommendations.

3. Ninja1 System RMSE:

The Ninja1 system is a novel recommendation system that leverages both Content-Based Filtering and Collaborative Filtering techniques to improve accuracy and personalization. By combining the strengths of both approaches, the Ninja1 system overcomes the limitations of each individual method. The impressive RMSE of 0.25 indicates that the Ninja1 system significantly outperforms both Collaborative Filtering and Content-Based Filtering. It achieves a higher level of accuracy and provides personalized recommendations tailored to individual user tastes.

The success of the Ninja1 system can be attributed to its ability to capture both the item content attributes (movie genres) and user-item interactions (collaborative filtering). This hybrid approach allows the system to provide more accurate and diverse recommendations, even for new users, by drawing on information from multiple sources.

In conclusion, the testing results highlight the superiority of the Ninja1 system in terms of accuracy and personalization compared to traditional Collaborative Filtering and Content-Based Filtering. The Ninja1 system's hybrid nature allows it to overcome the limitations of individual methods, making it a powerful recommendation engine with the potential to significantly enhance user satisfaction and engagement with movie recommendations. As a result, the Ninja1 system emerges as a promising solution for modern recommendation systems seeking to deliver better and more tailored user experiences.



## Discussion

The Ninja1 movie recommendation system demonstrated promising results and offered valuable insights into the effectiveness of collaborative filtering, content-based filtering, and the hybrid approach for movie recommendations. In this section, we discuss the key findings, limitations, and potential areas for improvement of the system.

1. Performance of Collaborative Filtering

Collaborative filtering, based on Singular Value Decomposition (SVD), proved to be effective in predicting user preferences and generating movie recommendations. The low Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values obtained on the testing set indicatethat the collaborative filtering model accurately predicted user ratings. This success can be attributed to the ability of SVD to capture latent factors that explain the user-movie interactions. However, one limitation of collaborative filtering is its reliance on user-item interactions, which can lead to cold-start problems for new users or movies with limited ratings.

2. Effectiveness of Content-Based Filtering

The content-based filtering algorithm, based on TF-IDF vectorization and cosine similarity, performed well in recommending movies based on movie content. The high precision, recall, and F1-score values indicate that the system effectively matched movie descriptions with user preferences. Content-based filtering addresses the cold-start problem by relying on movie features rather than user interactions. However, it may suffer from the overspecialization problem, where users are recommended movies similar to their past preferences, potentially limiting exploration of new genres.

3. Advantages of Hybrid Recommendation System

The hybrid recommendation system, which combines collaborative filtering and content-based filtering, leveraged the strengths of both approaches. By integrating diverse recommendation strategies, the system provided more personalized and diverse movie suggestions to users. The optimized weighting of collaborative and content-based filtering allowed for flexible tuning based on user preferences. The hybrid approach addressed the limitations of individual filtering methods, leading to improved user satisfaction.

4. System Architecture and Scalability

The modular system architecture of Ninja1 allowed for easy integration of various recommendation techniques and scalability to larger datasets. However, as the system grows and handles a more extensive user and movie base, the computational complexity of collaborative filtering may increase. To ensure real-time responsiveness and efficiency, future work could explore advanced matrix factorization methods and distributed computing techniques.

5. User Feedback and Satisfaction

User feedback and surveys played a critical role in evaluating the success of the Ninja1 movie recommendation system. Positive user feedback indicated that the system delivered relevant and enjoyable movie suggestions. However, it is essential to continually gather user feedback to further enhance the system's performance and cater to changing user preferences.

6. Diversity and Serendipity

One area for improvement is enhancing the serendipity and diversity of recommendations. While the hybrid approach mitigated overspecialization, further techniques like diversity-aware learning and serendipity promotion mechanisms could be explored to present users with unexpected but delightful movie choices.

7. Explaining Recommendations

An important consideration for recommendation systems is the ability to provide transparent explanations for the recommendations. Users may be more likely to trust and engage with the system if they understand the rationale behind the suggestions. Techniques like model interpretability and explainable AI could be integrated into the system to address this aspect.

8. Addressing Cold Start

The Ninja1 system could further address the cold-start problem for new users or movies with sparse data. Incorporating additional information, such as movie genres, release years, or movie directors, may aid in generating relevant recommendations for users with limited interactions.

In conclusion, the Ninja1 movie recommendation system demonstrated an effective combination of collaborative filtering, content-based filtering, and hybrid techniques for generating personalized and diverse movie suggestions. The system's modular architecture, positive user feedback, and promising results underscore its potential for practical movie recommendation applications. Future enhancements focusing on diversity, explanation, addressing the cold-start problem, and user-centric evaluation will further elevate the system's performance and user experience.

## Conclusion

In this thesis, we have explored the domain of movie recommendation systems and investigated various techniques to enhance the movie selection process for users. The objective was to design an efficient and effective movie recommendation system that offers personalized movie suggestions, facilitates movie discovery, and improves user satisfaction.

Through an in-depth literature review, we analyzed several movie recommendation approaches, including collaborative filtering, content-based filtering, and hybrid methods. Each approach has its advantages and limitations, and we recognized that no single method is universally superior in all scenarios. Therefore, we emphasized the importance of considering hybrid systems that integrate multiple techniques to provide more accurate and diverse movie recommendations.

During the implementation phase, we used the MovieLens dataset to evaluate the performance of different recommendation algorithms. The experimental results provided valuable insights into the strengths and weaknesses of each approach. Collaborative filtering proved effective in capturing user preferences and generating personalized recommendations based on user behavior and past interactions. On the other hand, content-based filtering was proficient in recommending movies based on their features, such as titles and genres.

To address the limitations of individual filtering methods, we examined hybrid movie recommendation systems. These systems combine collaborative and content-based filtering to leverage the advantages of both techniques and produce more accurate and diverse recommendations. The hybrid approach considers user behavior and movie features, resulting in enhanced precision and improved user satisfaction.

Additionally, we discussed the significance of incorporating user feedback into the recommendation process. By allowing users to rate and review movies, recommendation systems can continuously adapt and refine their suggestions based on user preferences. This interactive approach fosters user engagement and ensures that the system remains relevant and up to date.

In conclusion, movie recommendation systems play a crucial role in the digital age, where vast amounts of movie content are readily available to users. Through this thesis, we have gained valuable insights into the various approaches to movie recommendation and their impact on user experience. While there is no one-size-fits-all solution, hybrid movie recommendation systems emerge as a promising direction to deliver more accurate, diverse, and engaging movie suggestions. As technology and data continue to evolve, movie recommendation systems hold immense potential to revolutionize the way users discover and enjoy movies, and this thesis contributes to the ongoing research in this fascinating field.

## Recommendation

Based on the findings and analysis presented in this thesis, several recommendations are proposed to enhance the movie recommendation systems and provide a more seamless user experience:

1. Hybrid Recommendation Systems: Movie recommendation systems should consider adopting hybrid approaches that combine collaborative filtering, content-based filtering, and other techniques. By leveraging the strengths of multiple algorithms, hybrid systems can overcome the limitations of individual methods and offer more accurate and diverse movie suggestions.

2. User Feedback and Interactivity: Implementing user feedback mechanisms, such as movie ratings, reviews, and preferences, is vital to continuously improving the recommendation process. By capturing user feedback, the system can adapt and refine its recommendations, leading to a more personalized and engaging user experience.

3. Contextual Recommendations: Introducing contextual factors, such as time of day, location, or user mood, can further enrich the movie recommendation process. Context-aware recommendation systems can deliver more relevant and timely movie suggestions that align with users' specific preferences and circumstances.

4. Integration of External Data Sources: Incorporating data from external sources, such as social media, movie reviews, or user-generated content, can enhance the quality of movie recommendations. Integrating diverse data sets allows the system to gain a comprehensive understanding of users' preferences and interests.

5. Explainability and Transparency: As recommendation systems become more sophisticated, ensuring transparency and explainability is essential. Users should be able to understand the reasons behind the recommendations they receive, which can foster trust and engagement with the system.

6. Personalized User Profiles: Developing comprehensive user profiles that encompass various aspects of user behavior and preferences can significantly improve recommendation accuracy. Utilizing machine learning techniques, the system can continuously update user profiles to reflect changing preferences and interests.

7. Novelty and Serendipity: To enhance user satisfaction, recommendation systems should strive to introduce novel and serendipitous movie suggestions. Balancing between providing familiar content and introducing new, unexpected options can create a more enjoyable movie discovery experience.

8. Scalability and Performance: As movie databases continue to grow, recommendation systems must ensure scalability and efficiency. Employing scalable algorithms and optimization techniques can maintain high-performance levels even with large user and movie datasets.

9. Real-time Recommendations: Implementing real-time recommendation capabilities can cater to users' immediate needs and preferences. Real-time recommendations enable users to receive movie suggestions based on their current context and preferences.

10. Ethical Considerations: Movie recommendation systems must prioritize ethical considerations, such as user privacy and algorithmic fairness. Ensuring data protection and avoiding biased recommendations are crucial for maintaining user trust and satisfaction.

By incorporating these recommendations, movie recommendation systems can become more user-centric, personalized, and adaptive. The continuous refinement and innovation in movie recommendation technologies hold the potential to revolutionize the entertainment industry and redefine how users discover and enjoy movies. As technology continues to evolve, the research and development of movie recommendation systems remain an exciting and dynamic area with immense opportunities for further advancement.

# References