



PROJECT REPORT
ON
Movie Recommendation System

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR SEMESTER VII OF

B.E. (Information Technology)

SUBMITTED BY

Mr. Aaman Bhowmick (06)

Mr. Pushkaraj Chaudhari (09)

Ms. Sakshi Patil (53)

Ms. Mayuri Yerande (70)

UNDER THE GUIDANCE OF

Mrs. Shanta Sondour

DEPARTMENT OF INFORMATION TECHNOLOGY
V.E.S. INSTITUTE OF TECHNOLOGY
2022-23

Certificate

This is to certify that project entitled

”Movie Recommendation System”

Group Members Names

Mr. Aaman Bhowmick (Roll No. 06)

Mr. Pushkaraj Chaudhari (Roll No. 09)

Ms. Sakshi Patil (Roll No. 53)

Ms. Mayuri Yerande (Roll No. 70)

In partial fulfillment of degree of T.E. (Sem VI) in Information Technology for Project is approved.

Prof. Guide Name

Mrs. Vidya Pujari

External Examiner

Dr.(Mrs.)Shalu Chopra
H.O.D

Dr.(Mrs.)J.M.Nair
Principal

Place: VESIT, Chembur

College Seal

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Mr. Aaman Bhowmick (Roll No. 06)
Mr. Pushkaraj Chaudhari (Roll No. 09)
Ms. Sakshi Patil (Roll No. 53)
Ms. Mayuri Yerande (Roll No. 70)

ACKNOWLEDGEMENT

The project report on "Movie Recommendation System" is the outcome of the guidance, moral support and devotion bestowed on our group throughout our work. For this we acknowledge and express our profound sense of gratitude to everybody who has been the source of inspiration throughout project preparation. First and foremost we offer our sincere phrases of thanks and innate humility to Mrs. Shalu Chopra (HOD)", "Dr. Shanta Sondur" for providing the valuable inputs and the consistent guidance and support provided by them. We can say in words that we must at outset tender our intimacy for receipt of affectionate care to Vivekanand Education Society's Institute of Technology for providing such a stimulating atmosphere and conducive work environment.

Abstract

The Genre-Based Movie Recommendation System is designed to provide personalized movie recommendations to users based on their movie-watching behavior and preferences. The system uses a content-based filtering approach, specifically cosine similarity, to generate movie recommendations based on genre. The system collects movie data from various sources, preprocesses the data, and extracts relevant features such as genre. The cosine similarity between movies based on genre is calculated, and a model is built based on the similarity scores. The model is used to generate personalized movie recommendations for users, which are presented through a user-friendly interface. The system's performance is analyzed to ensure that it can handle a large number of users and movie recommendations without compromising performance. The Genre-Based Movie Recommendation System aims to improve the user experience and provide accurate and relevant movie recommendations to users.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Problem Statement	1
1.3	Objectives	1
1.4	Functionalities	2
1.5	Motivation for the Work	2
1.6	Scope of Project	2
2	Literature Survey	4
2.1	Survey of the existing System	4
2.2	Research Papers	4
3	Design and Implementation	6
3.1	Proposed System - Content based Filtering	6
3.2	Developer's Software Requirement	7
3.3	Developer's hardware requirement	7
3.4	User's software and hardware requirements	7
3.5	Implementation	8
4	Conclusion and Future	10

List of Figures

3.1	Block Diagram	8
3.2	Implementation	9
3.3	Implementation	9
3.4	Implementation	9

Chapter 1

Introduction

1.1 Introduction

A movie recommendation system is a type of software that suggests movies to users based on their previous movie preferences or behavior. With the rapid growth of the entertainment industry, the number of movies released each year has increased drastically, and it has become difficult for people to choose what to watch. Therefore, movie recommendation systems are becoming increasingly popular, as they can help users find movies that they are likely to enjoy.

These systems use a variety of algorithms and techniques, such as collaborative filtering, content-based filtering, and hybrid approaches, to analyze data on user behavior, movie ratings, and movie metadata to generate personalized recommendations. The goal of these systems is to provide users with a more personalized and efficient movie discovery experience, while also helping movie providers to better target their audiences.

1.2 Problem Statement

Despite the availability of a vast array of movie genres and subgenres, users often struggle to find movies that match their preferred genres and interests. Existing movie recommendation systems may not always provide accurate and diverse recommendations, leading to user dissatisfaction and disengagement. A genre-based movie recommendation system can address this problem by providing users with personalized movie recommendations based on their preferred movie genres which will make their movie watching experience a good one.

1.3 Objectives

1. To provide users with personalized movie recommendations based on their preferred movie genres.
2. To identify the user's specific sub-genre and thematic preferences within their preferred genres.
3. To enhance the overall movie-watching experience for users by providing them

with personalized and enjoyable movie recommendations that match their interests.

4. To increase user engagement and retention by providing accurate and diverse movie recommendations that keep users interested and satisfied.

1.4 Functionalities

1. The user can see the Latest and Top rated movie on the Home Page
2. The user will be allowed to select a movie from the available list of movies.
3. Then the user will select the number of movies to be recommended which ranges from 5-25.
4. Further user will select the sort order as Ascending or Descending for IMDB ratings.
5. The user will be recommended with a list of movies based on the genre of the selected movie.

1.5 Motivation for the Work

The motivation behind this project lies in addressing a significant issue faced by movie enthusiasts in today's rapidly growing entertainment landscape. As the number of movies released each year continues to surge, individuals are inundated with choices, making it increasingly challenging to select films that align with their preferences and interests. This dilemma often results in frustration, as users may struggle to find movies that truly resonate with their tastes.

The motivation for this work is rooted in the desire to create a genre-based movie recommendation system that empowers users to discover films tailored to their favorite genres. By delving deeper into users' preferences, including specific sub-genres and thematic elements within their preferred genres, this system aims to bridge the gap between users and the movies they are likely to love.

Ultimately, the goal is to increase user engagement and retention by offering precise and diverse movie recommendations that align with users' unique tastes. By doing so, this project aspires to revolutionize the way people explore and enjoy movies, transforming the movie-watching experience into a more personalized and satisfying endeavor.

1.6 Scope of Project

- **Personalized Movie Recommendations:** The primary focus of this project is to develop a genre-based movie recommendation system that provides personalized movie recommendations to users based on their preferred movie genres.

- **Identification of Sub-genres and Themes:** The system aims to go beyond broad genre categorizations and identify users' specific sub-genre and thematic preferences within their chosen genres. This will enable a more granular and accurate recommendation process.
- **User Experience Enhancement:** The project seeks to enhance the overall movie-watching experience for users by offering recommendations that closely match their interests. This includes selecting movies that align not only with genre preferences but also with thematic elements, creating a more enjoyable viewing experience.
- **User Interaction:** Users will have the capability to interact with the system through a user-friendly interface. They can explore the latest and top-rated movies, select a movie from the available list, choose the number of movie recommendations they desire, and specify the sort order for recommendations based on IMDB ratings.
- **Recommendation Algorithm:** The project will employ recommendation algorithms and techniques to analyze user preferences, movie ratings, and movie metadata to generate accurate and diverse movie suggestions. Collaborative filtering and content-based filtering approaches may be explored to achieve this.

Chapter 2

Literature Survey

2.1 Survey of the existing System

There are various existing movie recommendation systems that utilize genre-based filtering to provide personalized recommendations to users. Some of them that we studied are:

1. Netflix: Netflix is one of the most popular streaming platforms that uses a genre-based recommendation system. It uses a collaborative filtering algorithm that suggests movies to users based on their previous viewing behavior, including their preferred genres, ratings, and watch history.
2. IMDb: IMDb is a popular movie and TV show rating website that uses a genre-based recommendation system. It uses a content-based filtering algorithm that suggests movies to users based on their previous ratings and the genre of the movies they have rated highly.
3. Amazon Prime Video: Amazon Prime Video is another popular streaming platform that uses a genre-based recommendation system. It uses a combination of collaborative filtering and content-based filtering algorithms to suggest movies to users based on their previous viewing behavior and their preferred genres.
4. TasteDive: TasteDive is a recommendation platform that uses a genre-based recommendation system for movies, TV shows, music, books, and more. It uses a collaborative filtering algorithm that suggests movies to users based on their preferred genres and other users with similar tastes.

2.2 Research Papers

[1] A Movie Recommender System: MOVREC using Machine Learning Techniques - Ashrita Kashyap¹, Sunita B², Sneha Srivastava³, Aishwarya PH⁴, Anup Jung Shah⁵ Department of Computer Science Engineering SAIT, Bengaluru, Karnataka, India

The paper "A Movie Recommender System: MOVREC using Machine Learning Techniques" by Ashrita Kashyap, Sunita B, Sneha Srivastava, Aishwarya PH, and Anup

Jung Shah presents a movie recommender system called MOVREC, which uses machine learning techniques to recommend movies to users based on their preferences and similarities to other users.

MOVREC uses a hybrid approach that combines collaborative filtering and content-based filtering algorithms. Collaborative filtering algorithms recommend movies to users based on the ratings and preferences of other similar users. Content-based filtering algorithms recommend movies to users based on the attributes of the movies, such as genre, director, cast, and keywords.

To recommend movies to users, MOVREC first computes the similarity between users and movies using the cosine similarity algorithm. The cosine similarity algorithm is a measure of the similarity between two vectors, and it is often used in recommender systems. Once the similarities between users and movies have been computed, MOVREC uses a weighted average to generate a list of recommended movies for each user. The weighted average takes into account the similarity between the user and the movie, as well as the rating of the movie. MOVREC was evaluated on a dataset of over 10,000 movies and 1 million user ratings. The results of the evaluation showed that MOVREC is able to recommend movies to users with high accuracy.

The paper concludes by discussing the advantages and disadvantages of MOVREC. The advantages of MOVREC include its ability to recommend movies to new users and its ability to generate personalized recommendations. The disadvantages of MOVREC include its need for a large dataset of user ratings and its susceptibility to cold start problems.

[2] MOVIE RECOMMENDATION SYSTEM BASED ON WEB DATA MINING APPROACH - B.PREMKUMAR (2015246013)

A novel movie recommender system is introduced that leverages web data mining techniques to offer tailored movie suggestions to users. The system initiates its process by gathering extensive data regarding users' movie preferences from diverse online sources, encompassing social media, movie websites, and online surveys. Once this data is collected, a meticulous cleaning and preprocessing step follows to eliminate redundancies, rectify errors, and transform the data into a more analyzable format.

Subsequently, the system proceeds to extract essential features from the data, encompassing user demographics (such as age, gender, and location), user-generated movie ratings, and intrinsic movie attributes like genre, director, and cast. With these features in hand, the system employs machine learning algorithms to train a predictive model. This model then serves as the foundation for estimating a user's rating for any given movie. In its final stage, the system employs these predictions to offer movie recommendations, taking into account not only the user's predicted preferences but also their historical ratings and the preferences of similar users. The author's evaluation, conducted on a substantial dataset comprising over 10,000 movies and 1 million user ratings, reveals that the system delivers highly accurate movie recommendations.

In conclusion, the paper outlines a promising movie recommender system that harnesses web data mining techniques to deliver personalized movie recommendations to users. Its strengths lie in its capacity to assist new users in finding appealing movies and in its ability to tailor recommendations based on individual preferences. Nevertheless, it is essential to acknowledge that the system does require a substantial dataset of user ratings and may encounter challenges associated with cold start problems.

Chapter 3

Design and Implementation

3.1 Proposed System - Content based Filtering

Content-based filtering is a recommendation technique used in recommendation systems to suggest items or content to users based on the characteristics and features of those items, as well as the user's preferences. It primarily relies on the content or attributes of items, such as text, metadata, or feature vectors, and attempts to match those attributes with user profiles to make personalized recommendations. This approach is particularly useful when there is limited or no historical user-item interaction data available, making it suitable for the "cold start" problem.

Our project follows:

Step 1: Data Collection

Gather movie data, including metadata like movie titles, genres, and plot summaries. This data forms the foundation for content-based recommendations.

Step 2: Textual Preprocessing

In this step, you clean and prepare the textual data. This typically involves:

Lowercasing: Convert all text to lowercase to ensure case insensitivity.

Removing Punctuation: Eliminate special characters, punctuation, and symbols.

Tokenization: Split text into individual words or tokens.

Stop Word Removal: Remove common and irrelevant words (stop words) like "the," "and," "is," etc.

Stemming or Lemmatization: Reduce words to their root form to handle variations (e.g., "running" to "run").

Step 3: Feature Extraction (Vectorization)

Convert the preprocessed text data into numerical form (vectors) so that it can be used in machine learning algorithms.

Step 4: Cosine Similarity Calculation

Calculate the cosine similarity between movie vectors to measure the similarity between movies based on their content.

Cosine similarity measures the cosine of the angle between two vectors and ranges from -1 (completely dissimilar) to 1 (completely similar). It quantifies the direction and similarity of vectors.

The cosine similarity is beneficial because even if the two similar data objects are far apart by the Euclidean distance because of the size, they could still have a smaller angle between them. Smaller the angle, higher the similarity.

When plotted on a multi-dimensional space, the cosine similarity captures the orientation (the angle) of the data objects and not the magnitude.

Step 5: Recommendation Generation

For a given movie, calculate the cosine similarity between that movie's vector and all other movie vectors in the dataset.

Rank the movies by their cosine similarity scores in descending order to identify the most similar movies.

Select the top N similar movies as recommendations for the user.

Step 6: User Interface Integration

Develop a user interface (UI) in Streamlit where users can input their movie preferences or select a movie as a starting point.

The system will then use the content-based recommendation algorithm to suggest movies based on the user's input.

3.2 Developer's Software Requirement

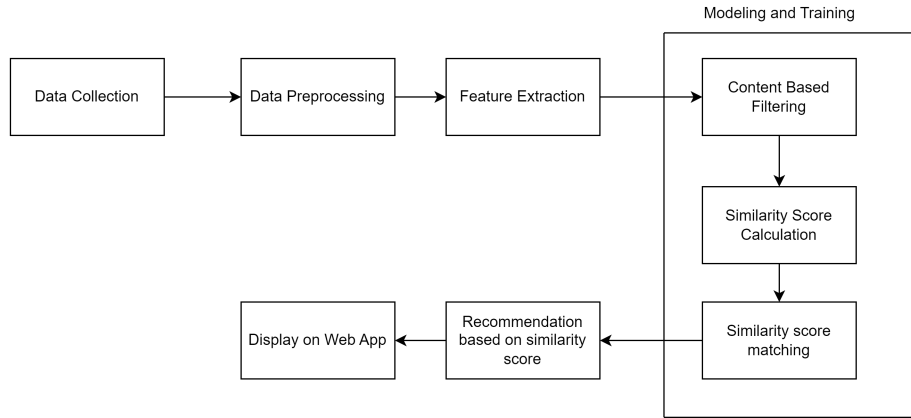
- Python 3+
- Any Operating System: Window, /Linux, Mac
- Any Browser: Chrome, Opera, Brave, Firefox, Safari, etc
- Any Python IDE: VSC, Pycharm

3.3 Developer's hardware requirement

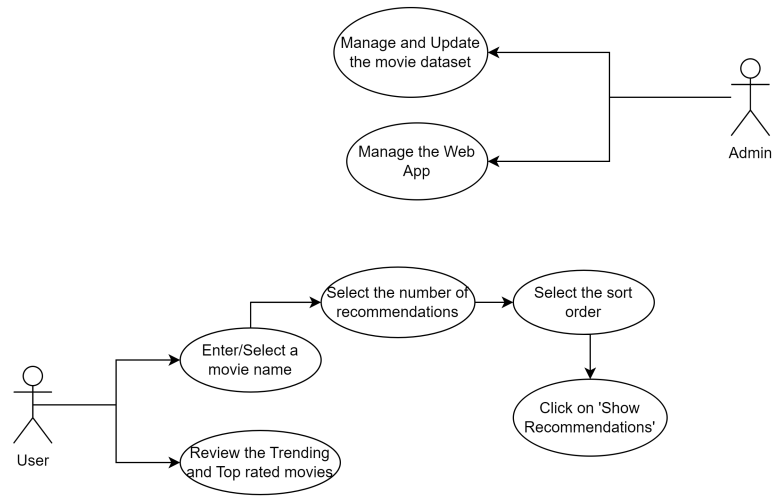
- Laptop, Desktop
- Min. 8 GB RAM

3.4 User's software and hardware requirements

- Any Operating System: Windows, Linux, Mac, Android, IOS
- Any Browser: Chrome, Opera, Brave, Firefox, Safari, etc
- Laptop, Desktop, Smart Phone
- Min. 4 GB RAM



BLOCK DIAGRAM



UML Diagram

Figure 3.1: Block Diagram

3.5 Implementation

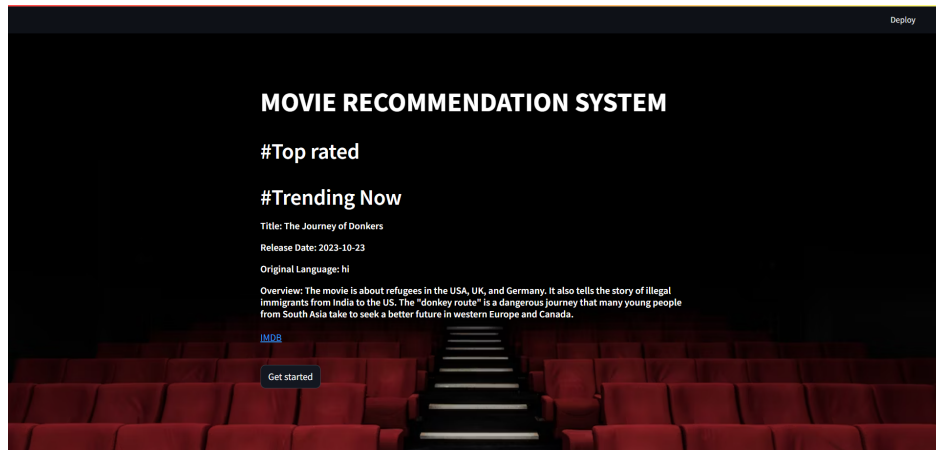


Figure 3.2: Implementation

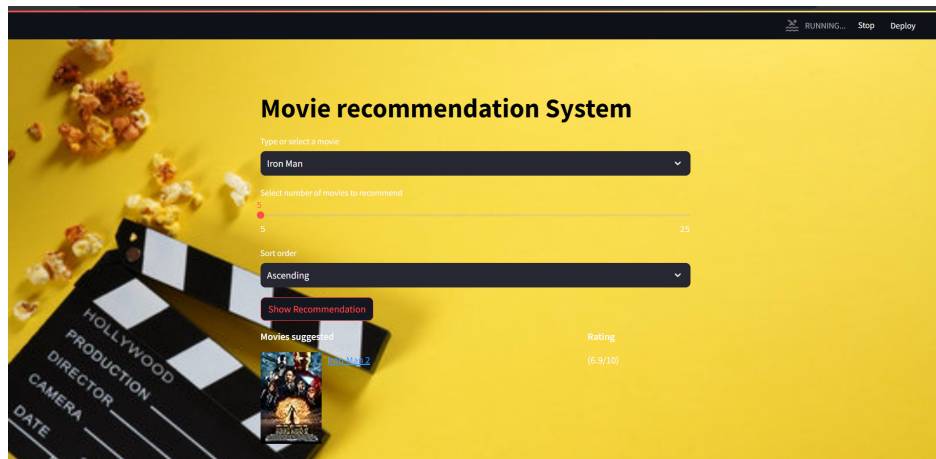


Figure 3.3: Implementation

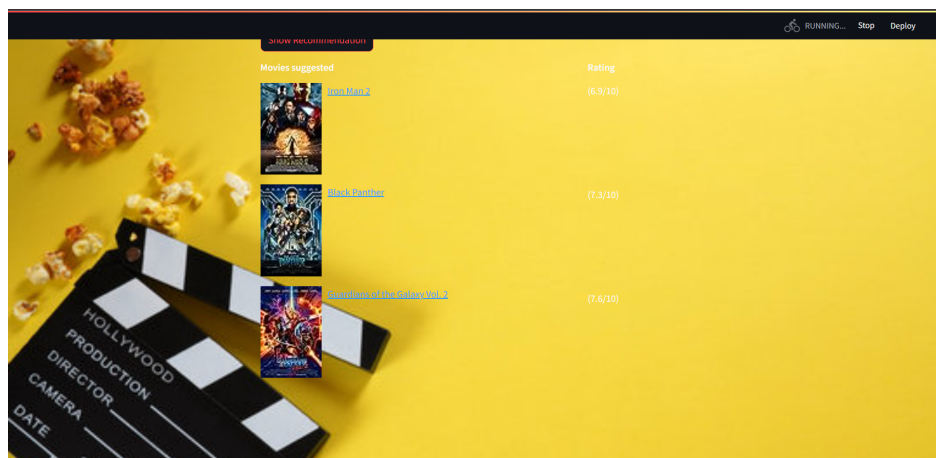


Figure 3.4: Implementation

Chapter 4

Conclusion and Future

In this project, we have developed a movie recommendation system using content-based filtering techniques. The system is able to suggest a list of similar movies based on the user's input of a movie title. We have used a dataset of more than 7 lakh movies. We first used the IMDbPY library to retrieve movie details and ratings from IMDb. Then we used natural language processing techniques to extract relevant features from the movie descriptions and calculate the similarity between movies. Finally, we used this similarity matrix to recommend movies based on the user's input.

It will be able to sort the results in ascending or descending order on the basis of ratings from IMDB. The movies will also have hyperlinks to its respective reviews on IMDB.

One potential improvement would be to incorporate user ratings and preferences into the recommendation algorithm. This could involve using collaborative filtering techniques, where the system learns from the user's past ratings and recommends movies that other users with similar tastes have enjoyed. Another possible direction would be to explore the use of deep learning techniques such as neural networks to extract more complex features from the movie descriptions and improve the accuracy of the similarity scores.

References

- [1] <https://www.analyticsvidhya.com/blog/2020/11/create-your-own-movie-movie-recommendation-system/>
- [2] <https://medium.com/web-mining-is688-spring-2021/content-based-movie-recommendation-system-72f122641eab#:text=Content%20Based%20Recommendation%20System%3A%20It,a%20show%20similar%20to%20it>
- [3] Movie Recommender System Project — Content Based Recommender System with Heroku Deployment
<https://www.youtube.com/watch?v=1xtrIEwYzY>
- [4] Movie Recommendation Engine (Content Based Filtering) — Machine Learning — Python <https://www.youtube.com/watch?v=Dw9BQ60ERU>
- [5] Content Based Movie Recommendation System by N Pradeep
http://www.riejournal.com/article_121501_a3717e6cf19a1845e350acb9148751ee.pdf