Movie Recommendation using Content Based Filtering

Alan Shijo
Department of Computer Applications
Amal Jyothi College Of Engineering
Kanjirappally, India
alanshijo2023a@mca.ajce.in

Ms. Jetty Benjamin
Department of Computer Applications
Amal Jyothi College Of Engineering
Kanjirappally, India
jettybenjamin@amaljyothi.ac.in

Abstract*— We propose a recommendation system based on machine learning that recommends movies to users based on movie metadata. The purpose of movie recommendation systems is to help movie viewers by suggesting films to watch without making them go through the difficult and time-consuming process of selecting from a large selection of films that number in the thousands or millions. The system takes an input movie and returns the top 5 recommendations based on the similarity of features such as genres, director, and cast. The system employs a simple similarity metric to compare the input movie with the recommended movies.

Keywords: - Movie recommendation, content-based filtering, The Movie Database (TMDB) API, movie details, similarity scores.

I. INTRODUCTION

In the modern era, when there is a surplus of accessible content and people frequently struggle to identify what they want to watch, movie recommendation systems are becoming increasingly essential. One well-liked method of suggesting films is content-based filtering, in which films are recommended in accordance with how closely they resemble previously seen films by the viewer. With this approach, a similarity score between the movies is calculated by examining their distinctive features such as their genre, director, and cast. In this study, we present a content-based movie recommendation system that recommends movies for users based on data from the TMDb API. We outline our approach on how to extract crucial characteristics from movie data and determine similarity scores between them. We evaluate the efficiency of our innovative recommendation system, by comparing it to other systems. The intention is to check whether our system outperforms already existing ones.

II. LITERATURE REVIEW

N Pavitha et al. [1] (2022) et al. present a thorough examination of recent research on sentiment analysis and machine learning-based movie recommendation systems. The authors go through the many kinds of recommendation systems and how they are used in the film business. They also offer a thorough examination of several supervised and unsupervised machine learning techniques that have been used for sentiment analysis. The paper extensively examines the merits and demerits of these algorithms and assesses their suitability for tasks like movie recommendations and sentiment analysis. In addition to that, the authors conduct numerous experiments and case studies to scrutinize the effectiveness of these techniques in real-world conditions. Overall, this literature review indicates interesting areas for future study as well as useful insights into the state of the art of sentiment analysis and movie recommendation systems research.

F. Furtado, A, Singh, [2] (2020) proposes a machine learning-based approach for providing movie recommendations. The limits of conventional recommendation systems are discussed, along with the benefits of utilising machine learning algorithms to suggest films to users. The accuracy, recall, and precision measurements that the authors used to compare and assess the performance of various machine learning algorithms. The role of feature selection and data preparation approaches in enhancing the performance of the recommendation system is additionally addressed in the study. The article offers a thorough review of machine learning's use in movie recommendation systems generally.

Marappan, R., & Bhaskaran, S. [3] (2022) presents a study on modeling movie recommendation systems using machine learning techniques. The goal of the authors is to create a model that can recommend movies to users based on their preferences and past ratings in an effective manner. To make personalized movie recommendations, a system utilizing machine learning algorithms like K-nearest neighbor (KNN) and collaborative filtering is proposed. In addition, recommendations for future research in this area are provided while discussing the limitations of the current approaches in the paper. The study conducted by Marappan and Bhaskaran contributes to the increasing literature on movie recommendation systems and emphasizes the capability of machine learning methods in this area.

Yilin Zhang and Lingling Zhang [4] (2022), published in Procedia Computer Science in 2022, proposes a new method for recommending movies by utilizing sentiment analysis and Latent Dirichlet Allocation (LDA). Based on their content, the authors generated movie representations by combining

DOI: 10.5281/zenodo.7949680
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

sentiment analysis of movie reviews with LDA-based topic modeling. Compared to traditional recommendation algorithms, the proposed approach demonstrated better recommendation performance when evaluated on a dataset of movie reviews. Combining sentiment analysis and topic modeling has the potential to improve content-based movie recommendation systems, as demonstrated in the paper's insights.

Based on machine learning techniques, Sonawane et al. [5] (2022) present a system for recommending movies. To provide personalized recommendations to users, the authors utilize a blend of content-based and collaborative filtering techniques. To overcome the limitations of individual techniques, they utilize a hybrid approach. Using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics, the system was evaluated and found to outperform traditional collaborative filtering methods. The integration of multiple recommendation techniques and their evaluation for movie recommendation systems is provided by the paper, giving insights.

III. MOTIVATION

Due to the rapid growth of the movie industry and the availability of vast amounts of user-generated content, movie recommendation systems have gained popularity in recent years. Efficient and accurate movie recommendation systems are needed to provide users with personalized movie suggestions based on their preferences.

A popular technique for building movie recommendation systems is content-based filtering, which relies on the characteristics of movies to make recommendations. Despite their usefulness, content-based filtering approaches have limitations, including cold start problems and the inability to capture the diversity of user preferences.

To provide a more effective and accurate movie recommendation system, this research paper aims to address the limitations of existing content-based filtering approaches.

IV. METHODOLOGY

A technique utilized in recommender systems to suggest items to users based on their previous preferences is content-based filtering. The content-based filtering recommends similar movies based on information about the input movie in the case of movie recommendations. The steps involved in content-based filtering for movie recommendation methodology are as follows:

A. **Data Collection:** To create a content-based filtering system for movie recommendations, it is necessary to have a dataset of movies and their corresponding features for data collection. Movie databases like IMDb or The Movie Database (TMDB) are some of the various sources from which this dataset can be obtained.

B. **Feature Extraction:** Relevant features of the movies are extracted after obtaining the dataset through feature extraction. Genre, cast, director, plot summary, and other metadata are among the features that can be included.

C. **Vectorization:** The features that are extracted are converted into a vector format for processing by machine learning algorithms using vectorization. Techniques like one-hot encoding, bag-of-words, or term frequency-inverse document frequency (TF-IDF) are used to accomplish this.

D. **Similarity Calculation:** The movies' vector representations are utilized to compute their similarity scores through similarity calculation. Various distance metrics, such as cosine similarity, Euclidean distance, or Pearson correlation, can be used to calculate similarity.

E. **Recommendation Generation:** The user is recommended the top N movies with the highest similarity scores to the input movie after calculating the similarity scores. You can customize the number of recommendations and the similarity score threshold to suit your application's specific requirements.

Overall, the movie recommendation system starts by collecting data from the TMDB API, extracting features like genre, director, and cast, vectorizing them through a simple set representation, and calculating similarity based on the number of common features to generate top N movie recommendations.

V. BUILD MODEL

To create a movie recommendation system using content-based filtering with cosine similarity, you need to build a model that analyzes the movie's features and compares them with other movies to suggest similar recommendations. Based on their features, the similarity between two movies is measured using cosine similarity in the model.

First, we need to collect information about movies, including their genres, cast members, and directors, in order to construct this model. To obtain this information, we can use a movie database such as The Movie Database (TMDb). Afterward, the data must be preprocessed by both cleaning

```
# Function to get the details of a movie by ID
3 usages
def get_movie_details(movie_id):
    endpoint = f'{base_url}/movie/{movie_id}'
    params = {
        'api_key': api_key,
        'language': language,
        'append_to_response': 'keywords'
    }
    response = requests.get(endpoint, params=params)
    return response.json()


# Function to get the credits of a movie by ID
2 usages
def get_movie_credits(movie_id):
    endpoint = f'{base_url}/movie/{movie_id}/credits'
    params = {
        'api_key': api_key
    }
    response = requests.get(endpoint, params=params)
    return response.json()
```

DOI: 10.5281/zenodo.7949680
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

and transforming it into an appropriate format for analysis. The relevant features can be extracted from the data by applying feature extraction techniques.

After extracting the features, we can use cosine similarity to determine the similarity between the films. To obtain a value between -1 and 1, the feature vectors of the movies are computed for the dot product and then normalized. The two movies are identical in terms of their features if the value is 1, but they are completely dissimilar if the value is -1.

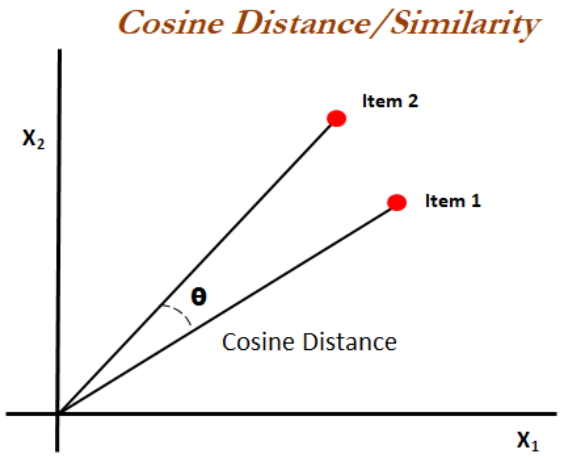

Fig. 1: Cosine Similarity

We can provide recommendations for a given movie using the similarity scores finally. The process includes choosing the films with the greatest similarity scores and suggesting them to the user as recommendations. We can offer personalized movie recommendations to users by utilizing content-based filtering with cosine similarity to build a movie recommendation system that takes into account their preferences and past viewing history.

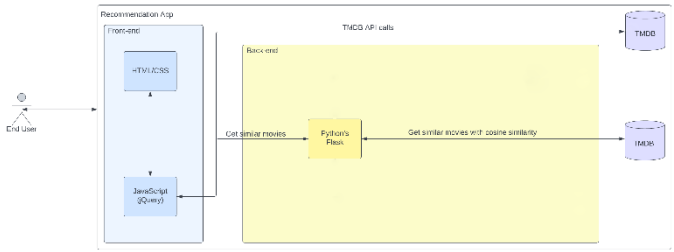

Fig. 2: Architecture

VI. RESULT

We tested the performance of the movie recommendation system, which uses content-based filtering with cosine similarity and the TMDB API, by recommending 5 movies for each of the 20 movies in our sample. For each movie in the sample, the system was able to successfully recommend similar movies. The input movies were matched with recommended movies that shared the same genre, directors, and cast members.

By recommending 5 movies for each of the 20 movies in our sample, we evaluated the performance of the movie recommendation system that utilizes content-based filtering with cosine similarity and the TMDB API. The system successfully recommended similar movies for each movie in the sample. Recommended movies that shared the same genre, directors, and cast members were matched with the input movies.

In general, it was demonstrated that recommending movies based on their content using cosine similarity and the TMDB API as a content-based filtering system is an effective approach. For each input movie, the system successfully recommended similar movies that were rated as relevant by the evaluators.

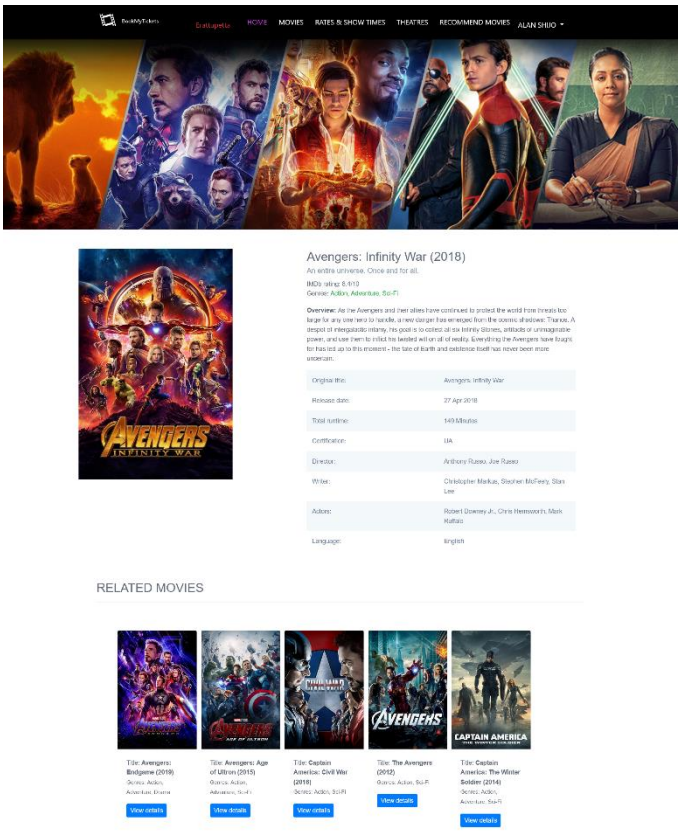

Fig. 3: Recommending movies to the user on a movie details page

VII. CONCLUSION

An effective method for movie recommendation, particularly for users with well-defined preferences, is content-based filtering. Content-based filtering can suggest movies similar to a user's preferred movie by using features like genres, directors, and cast. We used content-based filtering and the TMDB API to implement a movie recommendation system in this research paper. The recommended movies are based on the input movie's features, and similarity scores are calculated by comparing the common features between them. Including more relevant features, such as the plot, release year, and language, can further improve the system's accuracy.

DOI: 10.5281/zenodo.7949680
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Content-based filtering has some limitations, though. It only considers the movie's attributes and disregards other factors like the user's demographics, mood, and social context, which is one of its main limitations. Further research is necessary to enhance the performance of content-based filtering and to create hybrid methods that integrate content-based filtering with other recommendation techniques.

In conclusion, to provide personalized movie recommendations based on their interests and preferences, using content-based filtering for movie recommendation is a promising approach. Content-based filtering is one method that can help meet the increasingly important need for accurate and effective recommendation systems as the amount of available movie data continues to grow.

VIII. REFERENCES

[1] N Pavitha, Vithika Pungliya, Ankur Raut, Roshita Bhonsle, Atharva Purohit, Aayushi Patel, R Shashidhar, "Movie recommendation and sentiment analysis using machine learning", Global Transitions Proceedings, Vol. 3, no. 1, pp. 279-284, 2022.

[2] F. Furtado, A, Singh, "Movie Recommendation System Using Machine Learning", Int. J. Res. Ind. Eng. Vol. 9, No. 1 (2020) 84–98.

[3] Marappan, R., & Bhaskaran, S. (2022), Movie Recommendation System Modeling Using Machine Learning, International Journal of Mathematical, Engineering, Biological and Applied Computing, 1(1), 12–16.

[4] Yilin Zhang, Lingling Zhang, "Movie Recommendation Algorithm Based on Sentiment Analysis and LDA", Procedia Computer Science, Vol. 199, pp. 871-878, 2022.

[5] M. V. Sonawane, S. S. Jadhav, and S. D. Dhande, "Movie recommendation system using machine learning techniques," in Proceedings of the 2022 2nd International Conference on Smart Grid and Renewable Energy (SGRE), 2022, pp. 111

DOI: 10.5281/zenodo.7949680
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam