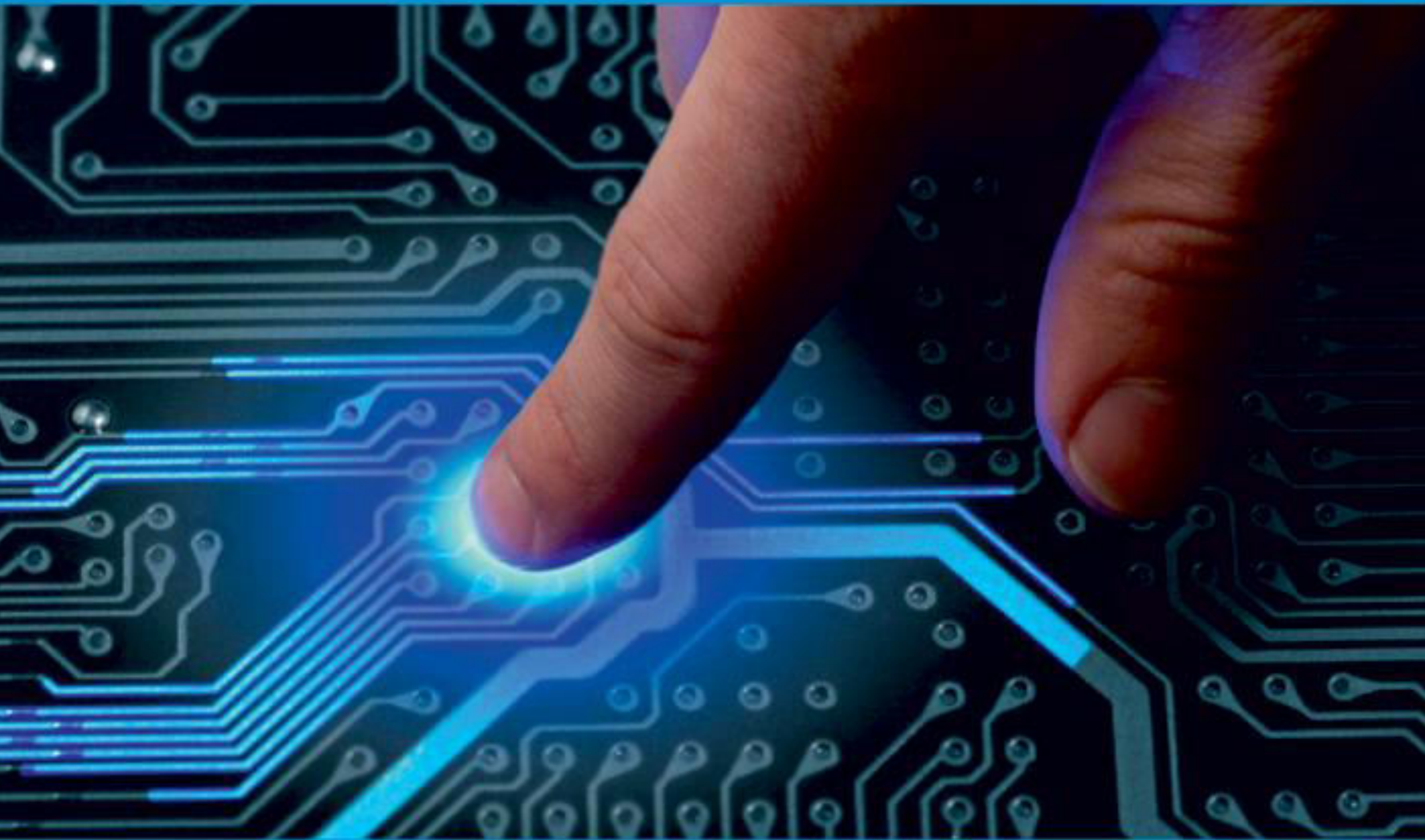
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

Recommendation and Sentiment Analysis System for Movies

Asmita Patil[1], Anuja Bhosale[2], Prachi Ingale[3], Santosh Tamboli[4]

[1]B.E. Student, Dept. of I.T., Vidyalankar Institute of Technology, University of Mumbai, India

[2]B.E. Student, Dept. of I.T., Vidyalankar Institute of Technology, University of Mumbai, India

[3]B.E. Student, Dept. of I.T., Vidyalankar Institute of Technology, University of Mumbai, India

[4]Assistant Professor, Dept. of I.T., Vidyalankar Institute of Technology, University of Mumbai, India

ABSTRACT: Recommendation systems are the systems that recommend different things like movies, clothing, accessories, food items, etc. The main aim of the movie recommendation system is to recommend. movies to users. according to their interests. It will help users save time while browsing the internet looking for movies from the thousands of already existing ones. This paper describes the Movie Recommendation system using Cosine. Similarity. Various factors like genres, directors, and actors/actresses are considered while recommending the movies. A movie's watchability will also be decided by the reviews of the movie as a user will choose to watch a movie that is preferred by most people rather than one which is disliked. Analyzing these reviews or sentiments as positive and negative will help the users.

KEYWORDS: Recommendation, Sentiment Analysis, Reviews,Cosine Similarity.

I. INTRODUCTION

Due to the abundance of information and the rate of information flow over the internet, there is a lot of confusion created among the users related to the usefulness of the information. Even on YouTube, when a user wants to watch a video of a particular concept, generally, there are a lot of videos available out there. Now, since the results are ranked appropriately, there may not be many issues but what if the results were not ranked appropriately? Well, in that case, we would probably spend a lot of time finding the best possible video which suits us and satisfies our needs. This recommendation results when you search for something on a website. Next time, when a user visits a particular website, without even searching, sometimes the system can show recommendations that users might like. So, the job of a recommendation system is to suggest the most relevant items to the user. Recommendation systems are used on YouTube for video recommendation and Amazon and Flipkart for product recommendation. Whatever users search on such websites, there is a system that monitors the user behavior and then ultimately suggests things/items with which you are highly likely to engage.

A lot of movie-related datasets are already available on Kaggle and other websites. We need to filter the data to consume it because generally, we are not interested in everything available to us. Recommendation systems are extensively used today and have applications in multiple industries like e-commerce, retail, banking, entertainment, etc. The most common approaches to implementing recommendation systems are Content-based Filtering (CBF), Collaborative Filtering (CF), and Hybrid Filtering (HF)[1].This paper describes the Recommendation using Content-based Filtering (CBF).

The process of recommendation will be as follows- the user will enter a movie and the system will suggest recommended movies and display reviews classified as positive and negative. A movie review enables a user to understand the overall idea or rating of that movie and make the decision- whether to watch it or not. A movie's success or failure depends on its reviews given by users. Therefore, a major challenge is to classify it into positive or negative reviews. This can be achieved using Sentiment Analysis.

Sentiment Analysis also called opinion mining, identifies, and categorizes the opinions and analyzes the sentiments for a particular topic as positive or negative. The process of sentiment analysis includes tokenization, word filtering, negation handling, stemming, and classification. Stemming is the process of removing prefixes and affixes to convert the word into its stem or root form[2] . The movie Recommendation is performed using the Cosine Similarity algorithm and the Support Vector Classifier algorithm is used to perform the Sentiment Analysis on the reviews of the movie chosen.

The paper structure is as follows: Section II describes the Related Work, Section III discusses the proposed System, Section IV describes the Experimental Results, and Section V discusses the Conclusion.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

II. RELATED WORK

In this Section, various existing Movie Recommendation Systems are studied. Similarly, their features and limitations are also discussed.

In [3] movie reviews are classified using various machine learning models. They have performed tokenization to convert the input string into a word vector, stemming to extract the root of the words, selection of important words by feature selection, and a classification model is used to analyze the sentiments. The model is trained on the real dataset using the TMDB reviews dataset in which reviews are already classified as positive or negative. They evaluated and compared the model with eight different machine-learning classification algorithms. The research paper showed that the accuracy of random forests is highest as compared to other algorithms.

Theresearcher in[4] used both types of filtering methodscontent-based and collaborative-based filtering. Movies are suggested based on previously watched movies. When a user watches a movie, he/she will give feedback in the form of a good smiley(positive) or a bad smiley(negative). Based on these user opinions, movies will be suggested to the user. The user first needs to login into the system and the user's behavioral characteristics will be stored in the database and the system will suggest movies based on it. They have used KNN collaborative Filtering algorithm.

In[5]to recommend a movie, three types of filtering methods can be used. They are Content-based filtering, Collaborative filtering, and Hybrid filtering. Content-based filtering is used to analyze the content of movie based and give recommendations. In Collaborative filtering, the history of the user and other similar users' history is used to recommend movies by finding out the similarity. They have used Cosine similarity for a recommendation of movies and Naïve Bayes(NB) and Support Vector Machine(SVM) classifiers for the sentiment of reviews. To find out the similarity, vector which is taken as data objects are defined in product space. In higher similarity, the distance is smaller, and in lower similarity the distance is large. This research paper included the comparison between NB (Naïve Bayes) and SVM (Support Vector Machine) depending on various factors. The accuracy of SVM was 98.63% and the Accuracy of NB was 97.33%.

In [6]data from YouTube is used to recommend movies. After watching a movie trailer on youtube, the user will give feedback in the form of likes and comments. You tube's data is integrated with movie ratings and sentiment of movie reviews to recommend movies to the user. They have used deep neural networks and multi-layer perceptron models for better performance.

In[7] Author used a collaborative filtering technique for Movie Recommendation System. Collaborative filtering consisted of two major approaches: the user-based approach and the Item-based approach. Further, in this filtering methodology, discussed how machine learning algorithms can be implemented for movie recommendation and also to predict the ratings of the movies which are not rated and sort the movies as per the user's preference.

In [8]Author used Hybrid Movie Recommendation System using sentiment analysis. It used the concept of opinion mining, and this system helped the user by giving recommendations of movies based on the choice of the user, watch history, and the ratings provided by the user after watching the movie. It allowed the user to easily search the movie without wasting a lot of time scrolling and searching keywords etc.

In [9]Author a content-based filtering technique is used to recommending movies. In collaborative-filtering-based recommendation systems, there is a comparison between the users based on their watching patterns. This required a proper understanding of the user's preferences and watching patterns because if a user never watches a particular movie, it will not be added to the list of the movies recommended to other users as it was not liked or watched by any user. So, this problem is solved using content-based filtering as this system will not involve different users but instead, a recommendation is made for a single user. In this paper, the recommendation is implemented based on various parameters like genre, movie length, director, and cast.

In [10] Author useda content-based filtering technique implemented using the machine learning algorithm KNN and Cosine similarity. It gave proper weightage to the preferences of the users and hence helped in recommending appropriate movies.

In[11]recommendation system helped the user make choices. Various recommendation systems are based on collaborative filtering and content-based filtering but these systems have various limitations-Storing movie data, and storing user history. Hence, a hybrid approach is used in this research paper that employed both content-based and collaborative-based systems. This system used tweets from microblogging sites for sentiment analysis. This is used to understand the latest trends and sentiments of users.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

III. PROPOSED ALGORITHM

In this section, the implementation of the Movie Recommendation and Sentient Analysis system has been discussed. The diagrams shown in the Fig. 1 and Fig. 2 explain the methodology that has been used in the project. The proposed system consists of two components- Movie Recommendation and Sentiment Analysis. The system will also show movies sorted by popularity, rating, and release dates.

A. *Data pre-processing*

For Movie Recommendations, the dataset is prepared from Kaggle and Wikipedia. The dataset gathered from Kaggle is processed in the following steps:
Step 1. After exploring the data, the following columns- Movie Id, Movie Title, Genres, Director, and Actors (Top 3) are considered relevant for a movie.
Step 2. Values for Genres and Cast are present in dictionaries. Lambda functions are used to extract these values.
Step 3. Rows having null values are removed. All these data attributes are converted into Lower Case.

Data is collected from Wikipedia for Movie Recommendations in the following steps:
Step 1. Pandas Library is used to collect data using web scrapping from Wikipedia. This data includes the movie's title, director, and cast.
Step 2. TMDB (The Movie Database) API is used to collect the genres of these movies.
Step 3. Basic Pre-processing similar to the operations performed on the dataset obtained from Kaggle is implemented here as well.
Step 4. The data from both sources are combined for recommending movies.

Sentiment Analysis:
The dataset for Sentiment Analysis is obtained from Kaggle for the training model. Following pre-processing operations are performed which are Removing punctuation, Removing Links, Removing Digits, and Converting text to lowercase.

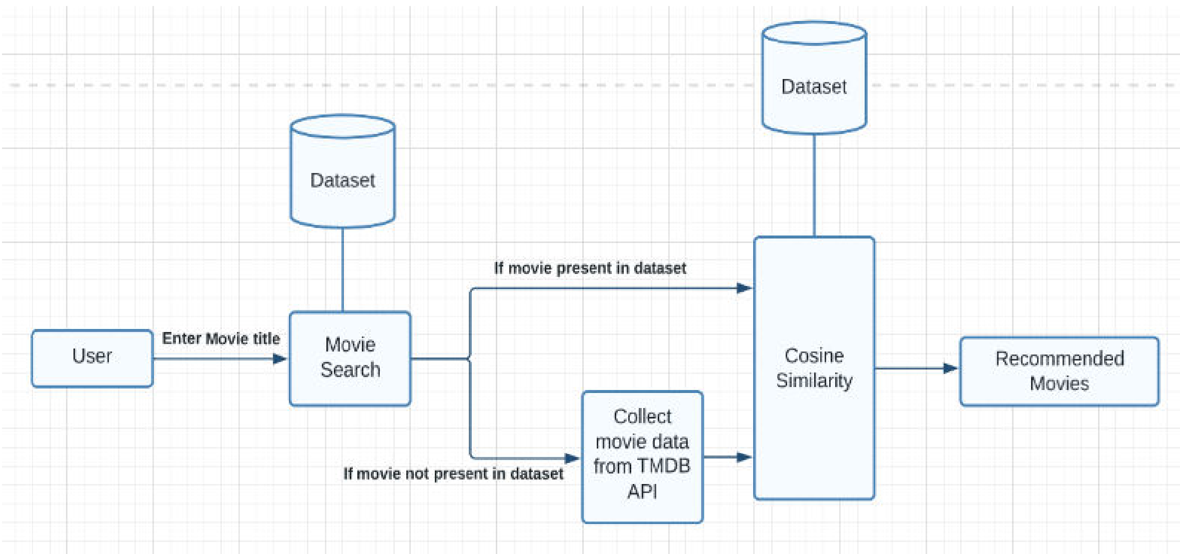


Fig.1 Flow chart of the Movie Recommendation System.

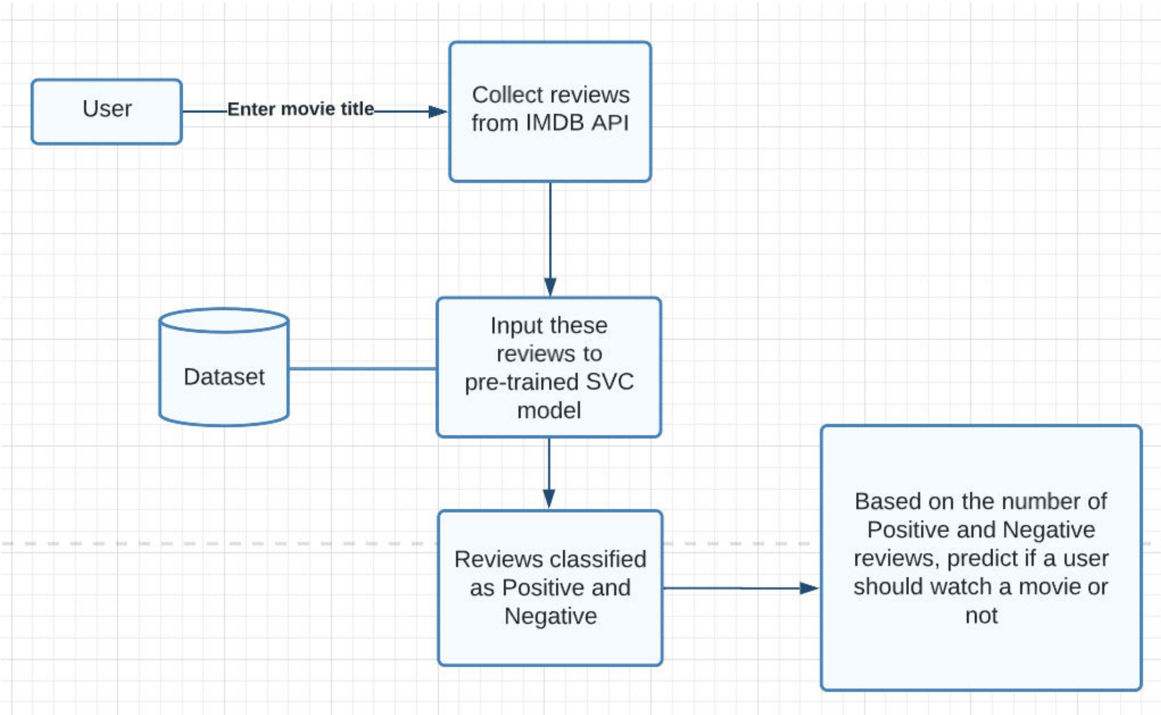


Fig.2 Flow chart of the Sentiment Analysis for Movie Recommendation.

B. *Movie Recommendation:*

The Genre, Cast, and Director fields of the movies dataset are combined into a single field for training the model. Count Vectorizer is used to convert text to numerical data. Machine Learning models need numeric inputs for producing results.

The Cosine Similarity algorithm from the Sklearn library is trained on the numeric data vectors that are considered as the data objects in data sets in cosine similarity. If the distance is smaller, the similarity will be more, but if the distance is greater, the similarity will be less. Cosine similarity is a measure to find how similar data objects are. It is mathematically expressed as the cosine of the angle between two vectors in multidimensional space.

$$Cos\,\theta = \frac{\vec{a}, \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_{1}^{n} a_i b_i}{\sqrt{\sum_{1}^{n} a_i^2} \sqrt{\sum_{1}^{n} b_i^2}} \quad (1)$$

The angle between two vectors determines its direction and is measured in 'θ'. This angle θ can be calculated by using Eq. (1). When θ = 0 °, the 'x' and' y' vectors overlap and prove to be similar. When θ = 90 °, the 'x' and ' y' vectors are therefore dissimilar.

The cosine similarity model is further pickled to be used later while implementing the User Interface.

C. *Sentiment Analysis:*

The dataset obtained from Kaggle is split into train and test datasets. Tf-idf vectorizer is used to convert these reviews into a numerical representation. IMDB API is used to collect reviews for a movie. Support Vector Classifier (SVC) is used to pickle the model. It operates by searching hyperplanes in N-dimensional space that are used to classify the data points. Also, based on the number of positive and negative reviews, users will be suggested whether he/she should watch the movie or not.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

IV. EXPERIMENTAL RESULTS

The results shown below in table 1 are calculated after training the data using different algorithms for sentiment analysis on reviews.

Machine Learning Algorithms	Accuracy Score
Support Vector Classifier	89.67
Naive Bayes	86.48
Logistic Regression	89.36
Decision Tree	71.03
K-Nearest Neighbor	76.14

TABLE 1. Accuracy of different machine learning algorithms for sentiment analysis

From the above Table 1 on comparing different machine learning algorithm's accuracy, we get to know that Support Vector Classifier (SVC) has the greatest accuracy. Hence, we have used Support Vector Classifier to implement sentiment analysis.

Below are some Screenshots of the original work:

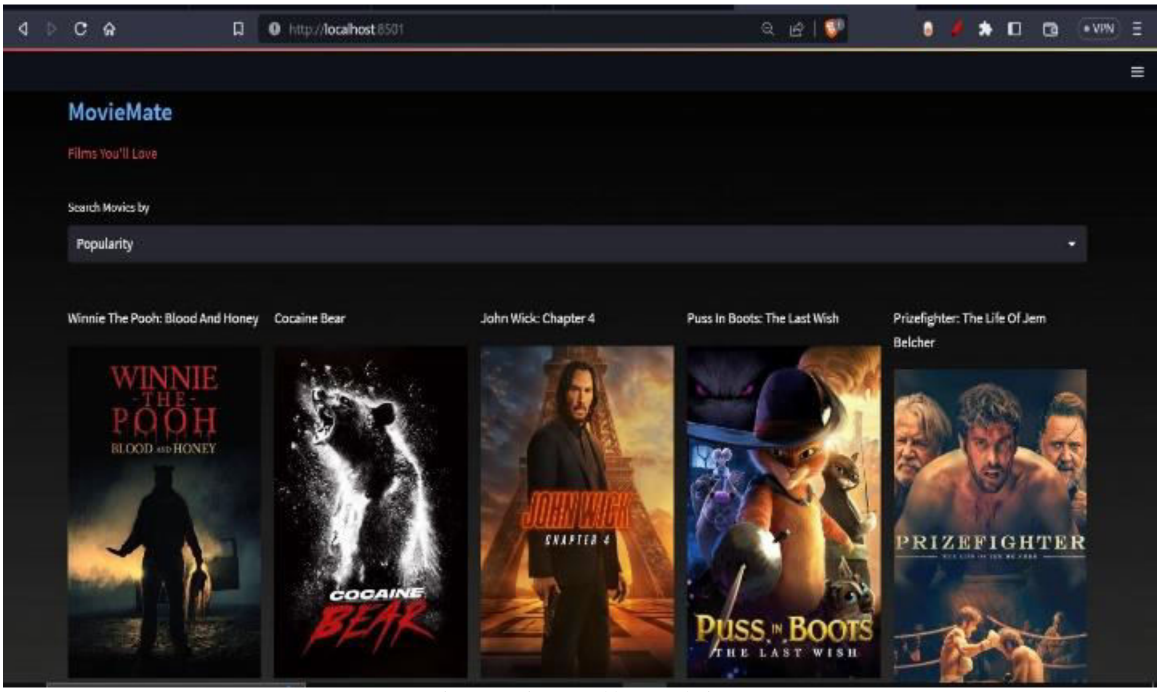


Fig. 3 Movies sorted by Popularity

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

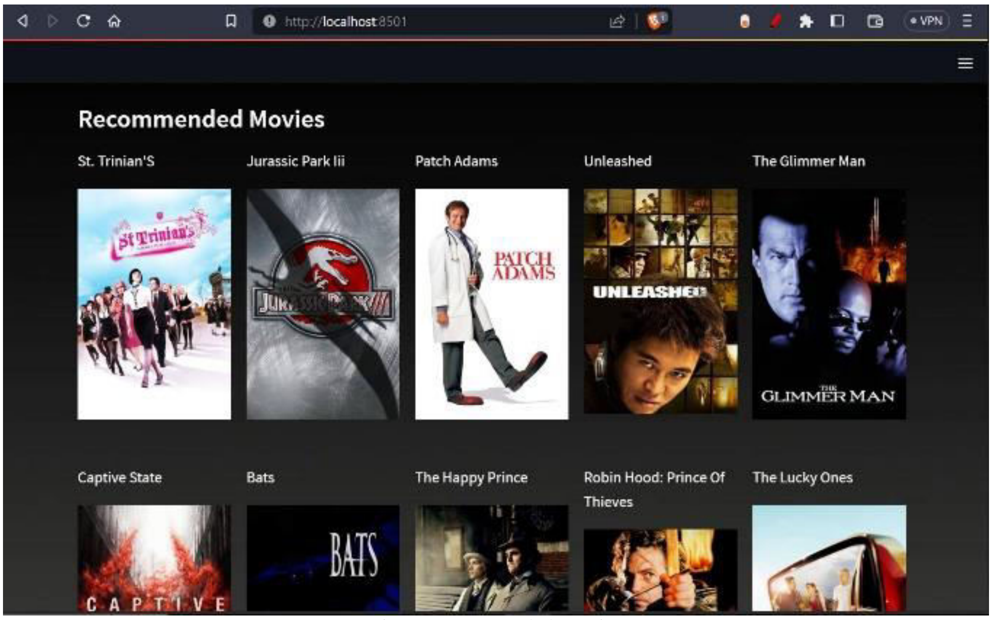


Fig. 4 Recommended Movies

Fig. 5 Movie Information

Fig. 6 Sentiment Analysis of Movie Reviews

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 3, March 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1103104 |

V. CONCLUSION AND FUTURE WORK

Thus, we have implemented a movie recommendation and sentiment analysis system which not only recommends movies to the users but also shows the reviews. We have used Cosine Similarity to implement the Recommendation system and Support vector Classifier to implement Sentiment Analysis. The user interface is easy to use and simple. The system will also show movies sorted by popularity, rating, and release dates.

For future work implement the system with more accurate recommendations customized to user preferences. Expand the dataset to give more widespread recommendations incorporating more languages. Improve the accuracy of the model used for sentiment analysis.

REFERENCES

[1] Y. Kulkarni, K. Bhat, P. Narke, P. Prince, and P. Tekade, “A Survey of Content-Based Movie Recommendation System with Sentiment Analysis Based on Machine Learning,” *International Journal of Scientific Research in Engineering and Management*, 2022, [Online]. Available: www.ijsrem.com

[2] G. Singh and R. Singh Boparai, “A Survey on Movie Recommendation System,” *International Journal of Advanced Trends in Computer Applications (IJATCA)*, vol. 3, no. 7, pp. 11–14, 2016, [Online]. Available: www.ijatca.com

[3] M. Yasen and S. Tedmori, “Movies Reviews Sentiment Analysis and Classification.”

[4] A. Chauhan, D. Nagar, and P. Chaudhary, “Movie recommender system using sentiment analysis,” in *Proceedings of International Conference on Innovative Practices in Technology and Management, ICIPTM 2021*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 190–193. doi: 10.1109/ICIPTM52218.2021.9388340.

[5] N. Pavitha*et al.*, “Movie recommendation and sentiment analysis using machine learning,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 279–284, Jun. 2022, doi: 10.1016/j.gltp.2022.03.012.

[6] D. Roy and C. Ding, “Movie Recommendation using YouTube Movie Trailer Data as the Side Information,” in *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 275–279. doi: 10.1109/ASONAM49781.2020.9381349.

[7] U. Thakker, R. Patel, and M. Shah, “A comprehensive analysis on movie recommendation system employing collaborative filtering,” *Multimed Tools Appl*, vol. 80, no. 19, pp. 28647–28672, Aug. 2021, doi: 10.1007/s11042-021-10965-2.

[8] R. Lavanya and B. Bharathi, “Systematic analysis of Movie Recommendation System through Sentiment Analysis,” in *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 614–620. doi: 10.1109/ICAIS50930.2021.9395854.

[9] “Movie recommendation using Sentiment Analysis and Ajax.” [Online]. Available: www.ijariie.com

[10] R. H. Singh, S. Maurya, T. Tripathi, T. Narula, and G. Srivastav, “Movie Recommendation System using Cosine Similarity and KNN,” *Int J Eng Adv Technol*, vol. 9, no. 5, pp. 556–559, Jun. 2020, doi: 10.35940/ijeat.E9666.069520.

[11] S. Kumar, K. De, and P. P. Roy, “Movie Recommendation System Using Sentiment Analysis from Microblogging Data,” *IEEE Trans Comput Soc Syst*, vol. 7, no. 4, pp. 915–923, Aug. 2020, doi: 10.1109/TCSS.2020.2993585.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details