



IJIRCCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Movie Recommendation Website Using Machine Learning

Dr. P L Rajarajeswari¹, M Swathi Meenakshi², R Swetha³, R Yamuna⁴, R Yazhini⁵

Associate Professor, Dept of Computer Science and Engineering, Saranathan College of Engineering,

Trichy, Tamil Nadu, India¹

Student, Dept of Computer Science and Engineering, Saranathan College of Engineering,

Trichy, Tamil Nadu, India²⁻⁵

ABSTRACT: The Movie Recommendation website is a personalized way to find the required content over the internet. This is a search engine designed for all language movies and it also provides the user with the basic information about the movie, recommendation and classification of movie reviews. It alters the existing system using the content based filtering approach. Similar movies are recommended using parameters such as movie actors and genre. This Recommendation website helps the user with recommended movies based on their preferences, thus increase the options for selecting a movie for the users. Content-based filters will provide accurate results compared to various types such as shared filters, based on user reviews (values), and it will suggest movies we have not seen at this point. Sentiment Analysis has been used to analyze the user sentiments in the reviews given to a particular movie. To find the similarity among the movies, cosine-similarity has been used. Movie review classification is done by using Random Forest Classifier, a classification algorithm. Tmdbv3 API is used to gather the data from “The Movie” dataset. The proposed system shows an improvement in the accuracy, quality and scalability of the movie recommendation website than the existing one. The proposed system recommends the top-n recommendation list of movies to users based on their interest and preferences.

KEYWORDS: Content-Based Filtering, Cosine Similarity, Random Forest Classifier

I. INTRODUCTION

A recommendation system is a subclass of information filtering system that seeks to predict the rating or the preferences a user might give to an item. In simple words, it is an algorithm that suggests relevant items to users. A recommendation process is a form of filtering information a challenging program to do important things for the user, and make recommendations based on user priorities. A recommendation system is a model used for information filtering where it tries to predict the preferences of a user and provide suggestions based on these preferences. These systems have become increasingly popular nowadays and are widely used today in areas such as movies, music, books, videos, restaurants, food, places and other utilities. These systems collect information about a user's preferences and behavior, and then use this information to improve their suggestions in the future. This project deals with the development of a search engine for all language movies. Instead of searching about a particular movie across all over the internet, “The Movie Recommendation Website” deals with all the searching more efficiently and returns the relevant information about the particular movie.

II. RELATED WORK

[1] Lina Chen, Tianqi Zhou, Jian Shen, “Move Recommendation System Employing the User-Based CF in Cloud Computing”, 2017. System proposed in this work uses the concepts of Collaborative Filtering. It focuses on a way to provide with efficient and effective algorithm for moving pictures. The proposed model uses MapReduce Framework from java to handle large datasets. By exploitation of MapReduce framework, high efficiency and reliability can be achieved for the proposed model.

[2] Unnathi Bhandary, hing-She (Mike) Wu, “Movie Recommendation using Collaborative Filtering”, ICSESS-2018. Collaborative Filtering with Apache Mahout is the main approach for recommendation in this paper. Collaborative filtering approach takes into account the user's behaviour pattern and preferences. Prediction is done based on the

similarity among different users. Machine learning libraries are imported and used with the help of apache mahout. The whole system is enforced exploitation of the mixture of Collaborative filtering and Apache Mahout.

[3] Shreya Agarwal, “An Improved Approach for Movie Recommendation System”. Traditional method has become outdated for recommendation. Hence, the Hybrid method which is the latest version for specifically improving the quality of movie recommendation system. The proposed system involves positive and negative features of both the concepts of Collaborative filtering method and Content based filtering. Negative features can overcome by using each other’s positive features. Implementation is done with the help of Genetic Methodology and Support Vector Machine (SVM) algorithm as its classifier. Thanks to the hybrid method, healthy improvement in the working of the entire system is identified.

[4] Mihhail Matskin, Chang Gao, “Content-Based Recommendation System for Movie Webiste”, 2016. Specialized recommender system for the movie website is proposed in this paper. Textual meta data is extracted and analysed and they’re found to be distinctive and diverse which makes it the main difference from other recommender systems for movies. After Analysis, similarity is identified and movies are recommended. An additional feature for setting weight in the textual metadata to find similarity, is also suggested by this proposed model.

[5] Jeffrey Lund, yiu-Kai N G, Dept. of Computer Science, “Movie Recommendation using Deep Learning Approach”, Brigham Young University Provo, USA. Collaborative filtering using auto encoders are used for this proposed system. The Movie Lens dataset is used for this approach to train the proposed model. Movie ratings for particular user is predicted based on ratings from different users. The proposed system uses neural network model to recommend the movies to the user along with the usage of Collaborative Filtering method. Additionally, the system makes use of regularization to reduce recommendation errors.

[6] Vallari Manaci, Anjali Diwate, Priyanka Korade, Anita Seanthi, “MoView Engine: An Open- Source Movie Recommender”, ITM Web of Conference, ICACC- 2020. The main concept involved in this recommendation system is the Hybrid method which improves the quality of the system. Hybrid method uses combination of both content-based and collaborative approach. It uses one hot encoding followed by similarity matrix generation. Then DNN algorithm with SoftMax Activation function is performed on the matrix generated to provide the recommended list of movies.

[7] Patrick Adolf Telnoni, Reza Budiaawan, Mutia Qana’a, “Comparison of Machine Learning Classification Method on Text based Case in Twitter”, Telkom University, Bandung, Indonesia, IEEE-2019. This paper deals with how the classifier algorithm works for text based classification. It helps to analyse the strength and uses of various classification and regression algorithms. It briefly explains about how the algorithm works for different scenarios.

[8] Angshuman Paul, Dipti Prasad Mukherjee, Prasun Das, Adhinandhan Gangopadhyay, Appa Rao Chintha and Saurach Kundu, “Improved Random Forest Classifier”, IEEE- 2018. In this paper, an advanced study is conducted on Random forest algorithm such that how the performance is increased and evaluated based on the internal parameters is discussed and it helps in improving the performance of the classification by limiting number of trees, iteratively removing the unwanted features.

III. PROPOSED SYSTEM

This project uses content-based filters that will provide accurate results compared to various types such as shared filters, based on user review. This algorithm attempts to guess what a user may like based on their activity. The accuracy obtained using the content based filtering algorithm is 98.84%. This system overcomes the cold launch problem. All language movies are recommended in this system. This is a Content based Recommendation System which uses Sentiment Analysis and Similarity of Cosine to recommend movies.

IV. ARCHITECTURE DIAGRAM

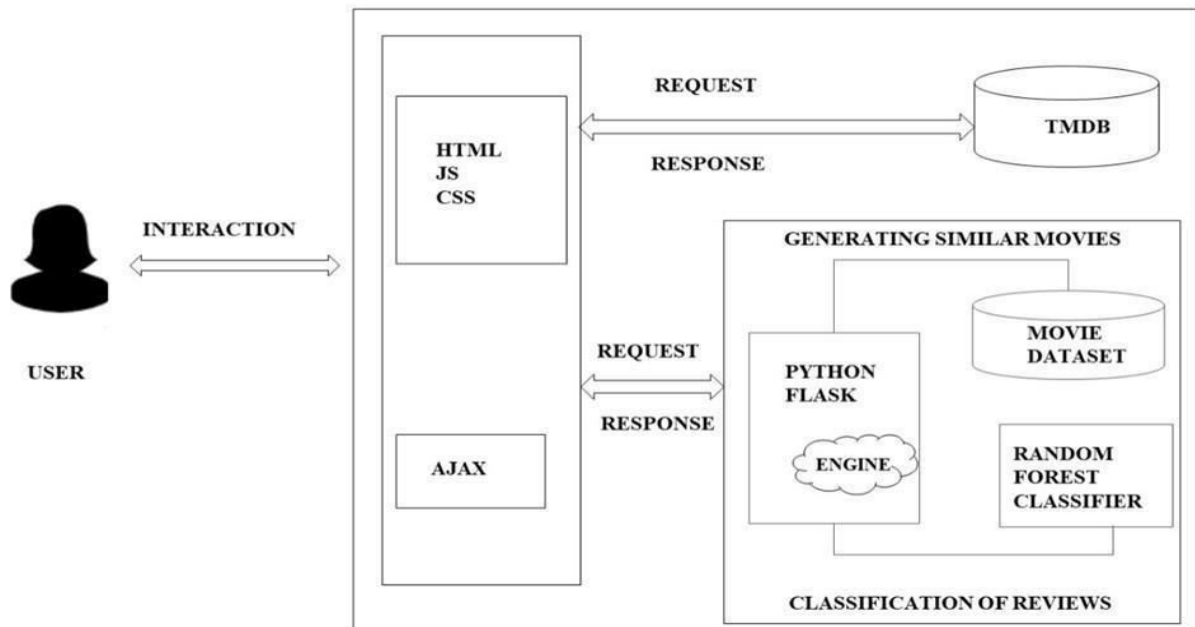


Fig. 1 Architecture Diagram

V. PROJECT MODULES

5.1 Data Collection

Movie tricks (title, category, genre, runtime, rating, poster, etc) are extracted using an API via TMDB. The dataset used is The Movies Dataset. This dataset contains metadata for all 45,000 movies in the MovieLens dataset. The dataset is collected from the data science community - Kaggle. The Movies Dataset - <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>.

5.2 Web scraping from Wikipedia

Web Removal is a method of copying where certain data is collected and copied on the web, usually uploaded to a local site or spread sheet, for later retrieval or analysis. The unavailability of data in The Movies Dataset from year 2018 to 2022 creates a reason to remove the features of movies and their characters of this age on Wikipedia. The reviews of the movies in the IMDB website are web scraped and pre-processed to perform the sentiment analysis.

5.3 Data pre-processing

Here, the raw data has been pre-processed in four files to make the dataset into suitable format. Data pre-processing includes data cleaning, filling the missing values and changing the null values as unknown. After pre-processing the dataset, the movies of all languages which are released from the year 1916 to 2017 are obtained.

5.4 Data Processing

After generating the datasets, the dataset was prepared using the “pandas” library in python to create different data frames. The datasets will be divided into different groups and frames which is required for the next step, that is, transferring the data to the frontend using Flask (A python web framework). The dataset consist of multiple list as a dictionary which can be passed to the html files so that it can be processed easily and the order of information will be preserved.

5.5 Sentiment Analysis

Sentiment analysis, a Natural Language Processing (NLP) technique is an automated mining of user generated opinionated text data such as reviews, comments and feedback. Here, Tfidf vectorizer is used to convert text into price representations as 0 or 1 which is used by separator and fit_transform to convert all available information into vector. After converting into required format, labeled reviews are displayed as POSITIVE (1) or NEGATIVE (0). The Random Forest Classifier, a supervised learning algorithm is used to process text voting display the classified movie reviews with prediction which also gives the accuracy of 98.84% .

5.6 User Registration and Login

The sign up page and login page are created using HTML ,CSS and php. Then, the phpMyAdmin in the Wamp server is used to create the “movie” database and the “regtb” table. The connection to the MySQL database is done by importing mysql.connector in main.py file. After the sanitization and validation of user inputs, the records will be inserted into the ” regtb” table, so that the user information is registered. When the user reloads the website, he/she will be redirected to the login page. Then, the user can login to the website, using the valid email and password.

5.7 Movie Recommendation

The movie information has been displayed on the Client Side using AJAX. With Ajax, web applications can send and retrieve data from a server in sync. Using Ajax, python data can be displayed on the client side in the form of Request and Feedback. To do this, the TMDB API has been used. The top n movies are recommended by taking the similarity measures of actors and genres of the movies.

VI. RESULT

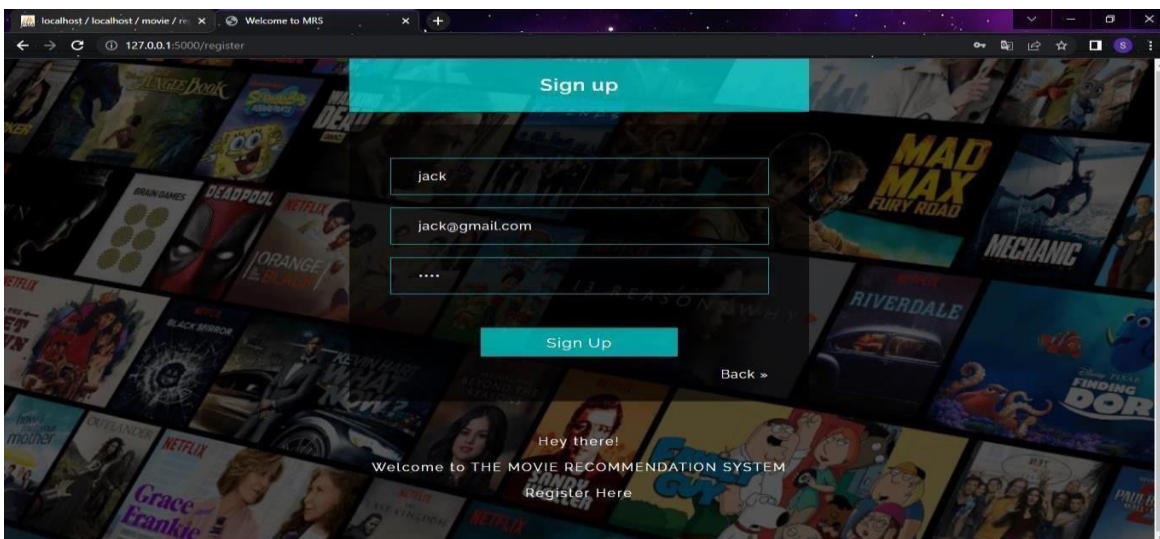


Fig. 2 Registration Page

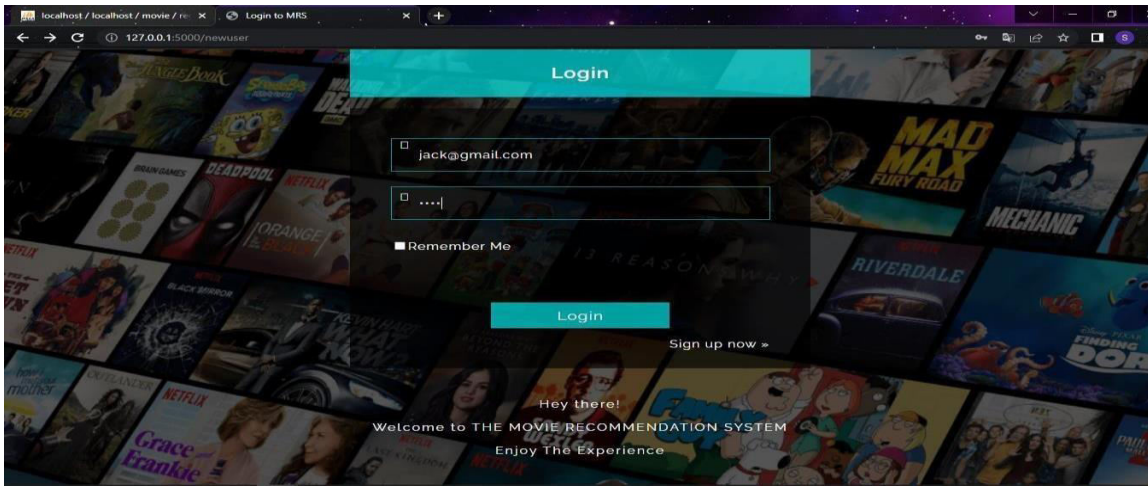


Fig.3 Login Page

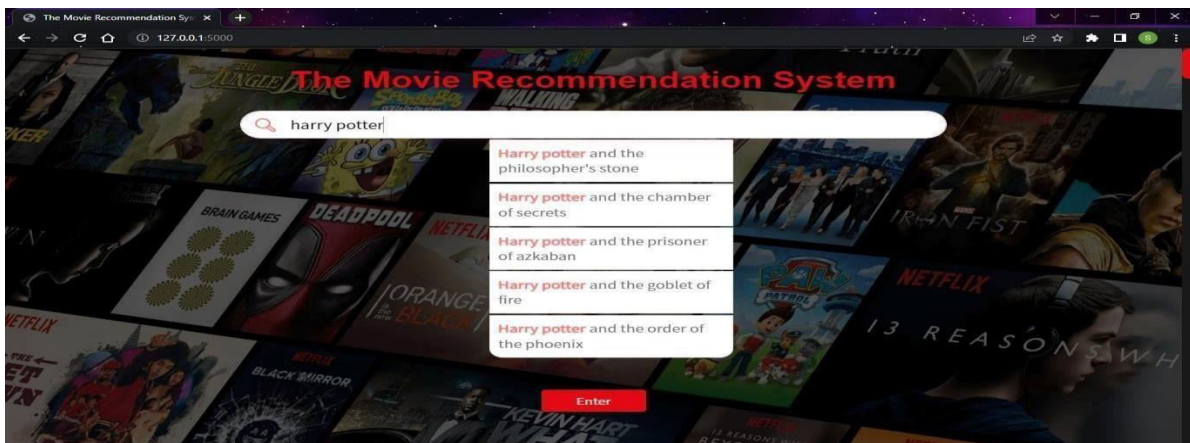


Fig. 4 Input Page

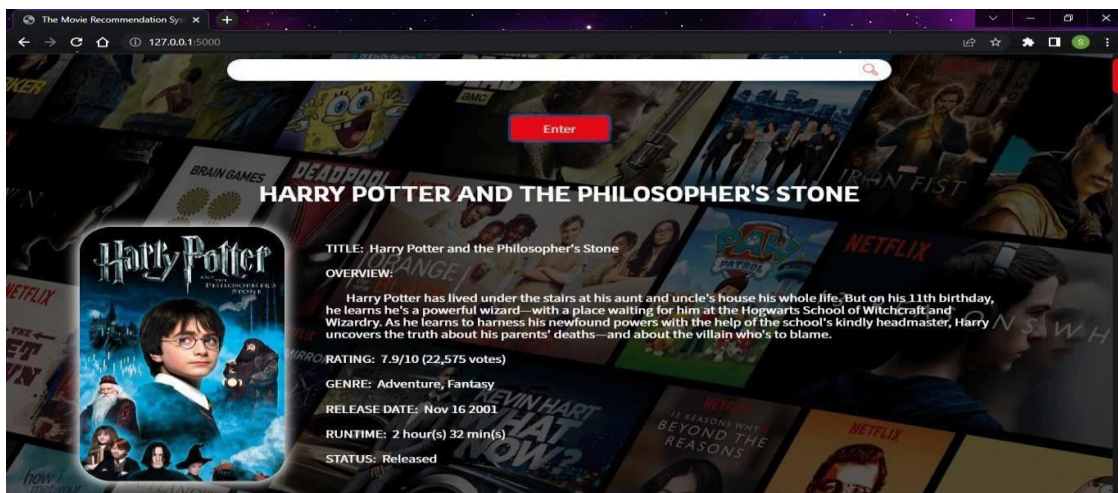


Fig. 5 Movie Information

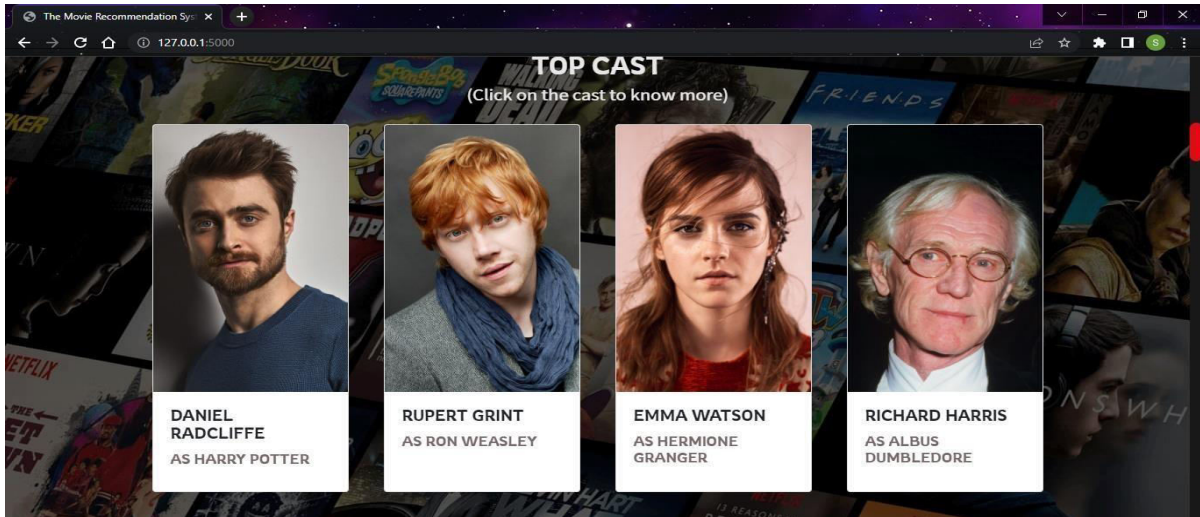


Fig. 6 Cast Information

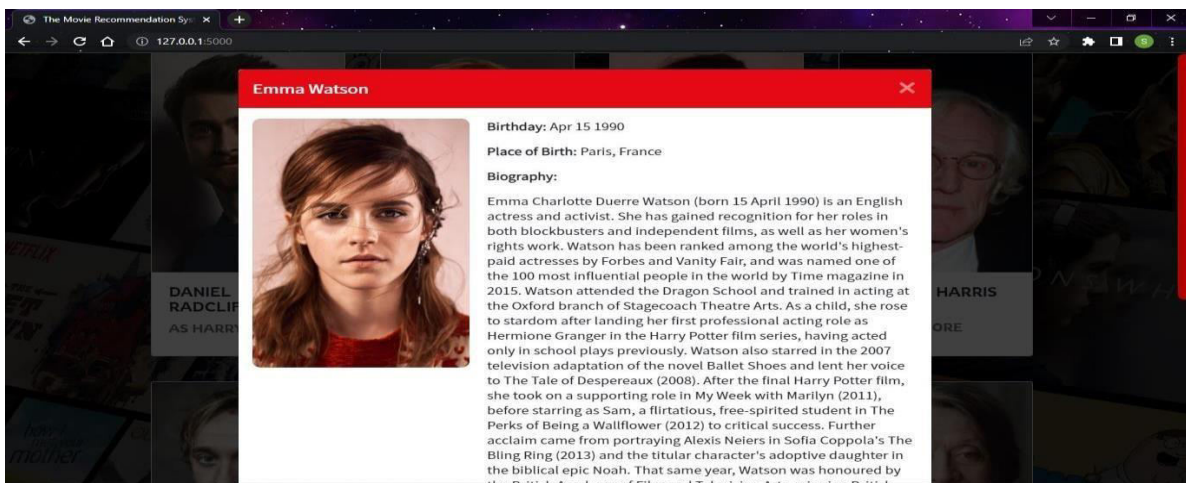


Fig. 7 Actor Information

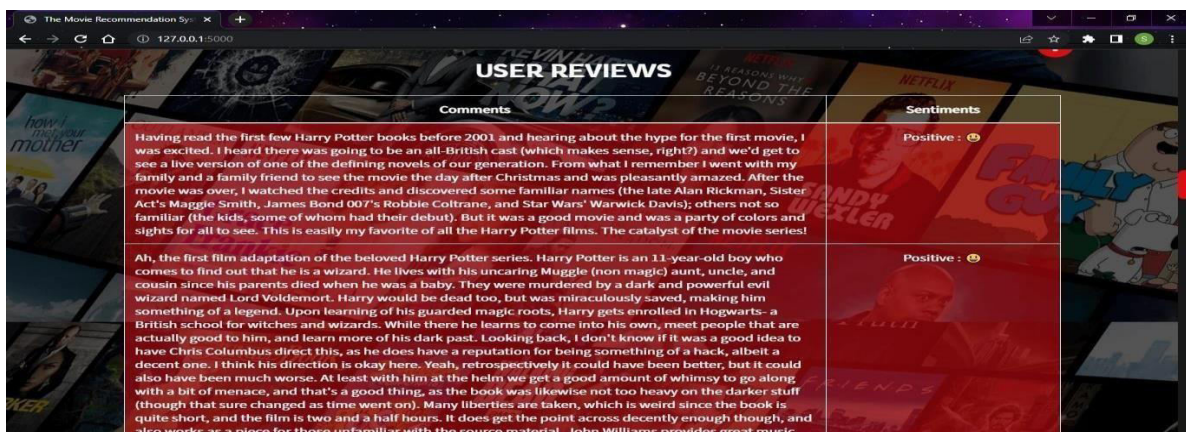


Fig.8 Review Classification

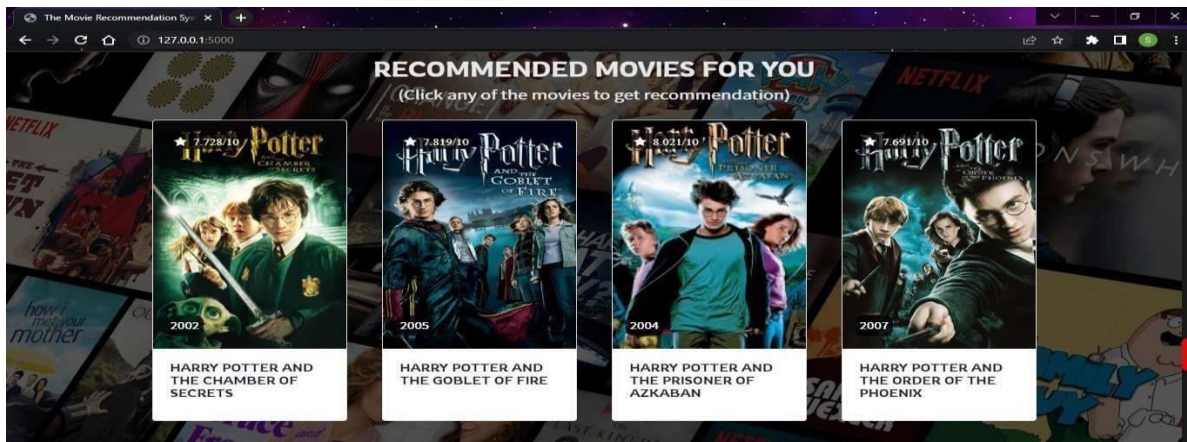


Fig. 9 Recommended Movies

VII. FUTURE ENHANCEMENTS

The future work to consider is that improving the performance of recommending movies with considerably large movie dataset using the enhanced concepts of deep neural networks with soft max activation function. Classifier output are predefined (class label) in current system, our future work is to consider the concept of Sentimental Analysis using Convolution Neural Network with a wide variety of consideration such as sarcastic statement, emoticons and much more. The user - friendly and more simplistic user interface will also be considered for the users and assuring to achieve still better performance and efficiency of the entire search engine. The suggestion process of extracting relevant data from the user's media platforms, preferences will be improved. More information about emotional tone of the user from different social media platforms will be extracted for the further improvement of the recommendation system using sentiment analysis. In future, there is enough scope for research to explore emotion information about users and integrate them in recommendation system.

VIII. CONCLUSION

The Movie Recommendation Website uses metadata such as cast, genre, plot, year and actors. Our engine is capable of giving information about the movies of all languages. The movie recommendation website recommends movies based on actors acted and genres of the movie. It is also capable of recommending similar movie, based on user search track. After conducting series of analysis and testing it is found that most individual saves up the time using The Movie Recommendation Website than searching every single piece of information about the particular movie in the internet.

REFERENCES

1. Angshuman Paul, Dipti Prasad Mukherjee, Prasun Das, Adhinandhan Gangopadhyay, Appa Rao Chintha and Saurach Kundu, (2018) "Improved Random Forest Classifier".
2. C. N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, (2005) "Improving Recommendation Lists through topic Diversification", WWW '05: Proceedings of the 14th international, May.
3. Dharmendra Pathak, Sandeep Matharia, Murthy, Dept. of Information Technology, (2013) "ORBIT: Hybrid Movie Recommendation Engine", International Conference on Emerging Trends in Computer, Communication and Nano Technology (ICECCN).
4. G. Linden, B. Smith, and J. York, (2003) "Amazon Recommendations: Item to- item Collaborative Filtering".
5. Jeffrey Lund, yiu-Kai N G, Dept. of Computer Science, (2018) "Movie Recommendation using Deep Learning Approach".
6. K. S. Jones, (1972) "A Statistical Interpretation of term Specificity and its Application in Retrieval".
7. Lina Chen, Tianqi Zhou, Jian Shen, (2017) "Movie Recommendation System Employing the User- Based CF in Cloud Computing".
8. Michael Hashler, (2011) "Recommender Lab: A Framework for Developing and Testing Recommendation Algorithms".



9. MihhailMatskin, Chang Gao, (2016) "Content-Based Recommendation System for Movie Website".
10. Mukesh Kumar Kharita, Atul Kumar, Pardeep Singh (2018) "Item - Based Collaborative Filtering in Movie Recommendation in Real time".
11. Patrick Adolf Telsoni, Reza Budiaawan, Mutia Qana'a, (2019) " Comparison of Machine Learning Classification Method on Text- based 77 Case in Twitter ".
12. Qaiser and R. Ali, (2018) "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents".
13. R. Bell, Y. Koren, and C. Volinsky, (2007) "Modeling relationships at Multiple scales to improve Accuracy of large Recommender Systems".
14. Rujhan Singla, Saamarth Gupta, Anirudh Gupta, Dinesh Kumar Viswakarma, Dept. of Information Technology, (2020) "FLEX: A Content Based Movie Recommender".
15. Shreya Agarwal, (2017) "An Improved Approach for Movie Recommendation System".
16. Unnathi Bhandary, Hing-She (Mike) Wu, (2018) "Movie Recommendation using Collaborative Filtering".
17. Vallari Manaci, Anjali Diwate, Priyanka Korade, Anita Seanthi, (2020) "MoView Engine: An Open-Source Movie Recommender".

BIOGRAPHY

Dr. P. L. Rajarajeswari received B.E degree in the department of Computer Science and Engineering from Bharathidasan University, Saranathan College of Engineering, Trichy, India in the year 2003, M.E. degree in Computer Science and Engineering from Anna University, VLB Janakiammal College of Engineering and Technology, Coimbatore- India in the year 2010 and received her Ph.D. degree in Information and Communication Engineering from Anna University, Chennai, Tamil Nadu, India in the year 2017. Her Academic experience in teaching field is around 16 years. She had published around 20 research papers indexed in SCI, Scopus, WEB of science and UGC care approved journals. She has also published 1 patents. She is an active life time member of ISTE. Her area of interest include Wireless Sensor Networks, Wireless Communication and IOT. She had also delivered many guest lectures to the students of various Engineering colleges. Presently, she is designated as Associate Professor in the Department of Computer Science and Engineering in Saranathan College of Engineering, Trichy.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details