

RECOMMENDATION SYSTEM FOR MOVIES

Mr. Sumit Ghadi*¹, Miss. Rashi Chaubal*², Miss. Aditi Pawar*³, Mr. Govind Agrawal*⁴

Prof. Rahul P. More*⁵

*^{1,2,3,4}Student, Department Of Computer Engineering, Zeal College Of Engineering And Research,
Pune, India

*⁵Professor, Department Of Computer Engineering, Zeal College Of Engineering And Research,
Pune, India

ABSTRACT

Due to the digital era growing so rapidly much of digital content is produced among which movies is topmost. There are movies of every genre and sometimes it becomes difficult to find movies based on our favorite genre. Recommendation systems are among most popular application of data science. This Movie Recommendation System will be based on Machine learning model. There is procedure of certain steps to be followed to recommend and display the movies of genre accordingly. Labeled Data is use for this purpose. Data set from 2 sources IMDb and TMDb are used and provided to the module. Data preprocessing is done before applying precondition. Naive Bayes approach which serves to be popular algorithm for classifying text is used. In this paper, a content-based movie recommender system is built that asks the sequential user predilections in user modeling and predicts the ideal movies [1]. The more interactive approach of proposed system with user regarding genre preference method outperforms the existing systems.

Keywords: Movie Recommendation, Genre, Dataset, Content Based.

I. INTRODUCTION

In this digital era data production is growing rapidly in number of fields, every second 1.7MB of data is produced so it can be overwhelming for user to find what he required. Data in which is in such volumes turns out to be erratic. Also, without apposite preprocessing the statistics, gets wasted. In such cases, users need to have multiple runs for getting desired data before getting the required one to address this issue researchers found recommendation systems. A recommendation system arrange for similar information to the users by considering their preferences (older & present). The main advantages of recommendation system are it has reduced the overhead of market research a user needs to do before buying, watching, ordering on online platforms. Data is strained and personalized depending on user requests. With rapidly increasing records on the Internet, recommendation systems are becoming more necessary because to their efficacy in providing statistics in a short time period. Recommender systems have been proving their importance in fields ranging from music, movies, news, to many more. Today many of renowned organizational like YouTube, LinkedIn, and Amazon have implemented recommendation systems for rewarding customer requirements. YouTube recommendation engine works by suggesting user next videos based on his current as well as past interest. Amazon recommendation schemes work such that it suggests allied items that the customers can buy. Netflix's recommendation engine works by considering the types of shows that a client watches and provides recommendations. Basically, Recommendation systems can be roughly classified into three categories, namely Content Based, Collaborative Filtering, and Hybrid approach [2]. A content-based recommendation system works in such a way that it takes into consideration user's past behavior and finds a particular pattern in them which eventually helps system to recommend the user for future. This paper recommends for a content-based recommendation which uses genre relationship. The dataset used for this purpose is IMDb and TMDb dataset which in total consisted of 50 records. The system asks user for genre rating which helps the system to work accordingly and suggest the user with movies. Content-based filtering models analyzes likings given by a particular operator, and try to build a model around this data. Content-based systems are broadly divided into three methods—wrapper methods, filter methods, and embedded methods [2].

II. METHODOLOGY

A. System Architecture

A recommendation system is a subclass of information filtering system that helps in predicting the "preference" a user would like based on user's past behavior. The dataset considered for this model is from IMDb and TMDb respectively. Initially the dataset Consist of various genres, a user when redirected to login page is asked for rating of 5 genres.

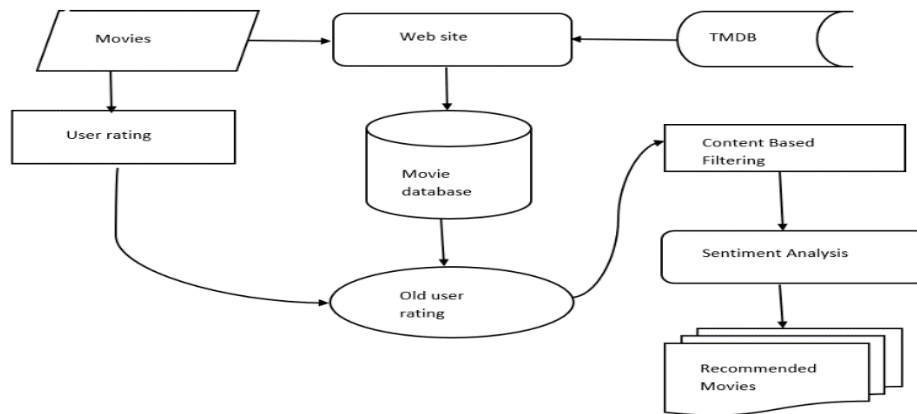


Figure 1: System Architecture

The 5 genres are after user has provided with genres which are stored in .csv file. Further we find 3 genres which has maximum value. The indexes for those 3 genres having maximum value are found. Then the product of those 3 genres which serve to be maximum of all genres in a particular list is calculated. As of now indexes for those 3 genres are found out, we get the product of those indexes in book1final.csv file. Book1final.csv dataset consist of genre's first name, description link etc. In such a way we get the product of such genres which user most like from our main dataset. For convenience all these csv files are copied into temporary list. As product of genres of indexes is obtained, the list of sorted in descending order containing Information about main dataset. Lastly the movies which has maximum product is displayed on top of list.

The Algorithm for system is as follows: -

Step 1: We get the genres of users in a csv file.

Step 2: Then we find 3 genres which has maximum value.

Step 3: We store indexes of those 3 maxima in separate list which is stored in temporary file.

Step 4: We calculate the product of those 3 genres which are maximum of all the genres present in list for particular user.

Step 5: As we know the indexes of max 3 genres, now we try to get product of those indexes in our main dataset which has name book1final.csv.

Step 6: book1final.csv is dataset in which we get genres first then name, description, image, link so on and so forth.

Step 7: From the reference of step 5 above, we get the product of such genres which user most like from our main dataset. All the .csv files are copied to temporary list for our calculations.

Step 8: As we get the product of genres of indexes which we get in Step no.5 we sort the list in descending order which contains all the information of our main dataset.

Step 9: The logic for sort in reverse is that we get the movie which has maximum product. It should be displayed on top of the list and which is most suitable for the user.

III. MODELING AND ANALYSIS

YouTube recommendation engine works by suggesting user next videos based on his current as well as past interest. Amazon recommendation schemes work such that it suggests allied items that the customers can buy. Netflix's recommendation engine works by considering the types of shows that a client watches and provides recommendations. A content-based recommendation system works in such a way that it takes into consideration user's past behaviour and finds a particular pattern in them which eventually helps system to recommend the user for future. This paper recommends for a content-based recommendation which uses genre

relationship. The dataset used for this purpose is IMDb and TMDb dataset which in total consisted of 50 records. The system asks user for genre rating which helps the system to work accordingly and suggest the user with movies. Content-based filtering models analyses likings given by a particular operator, and try to build a model around this data. Content-based systems are broadly divided into three methods—wrapper methods, filter methods, and embedded methods.

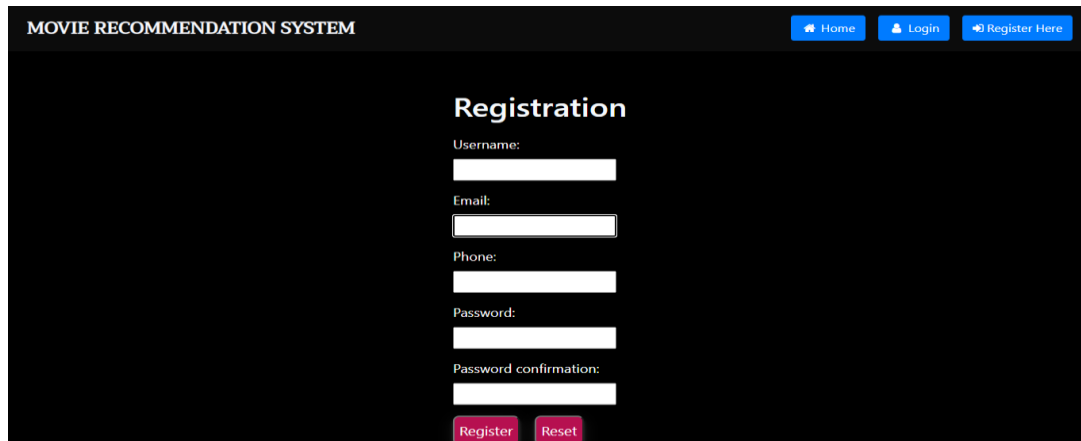


Figure 2: Registration page.

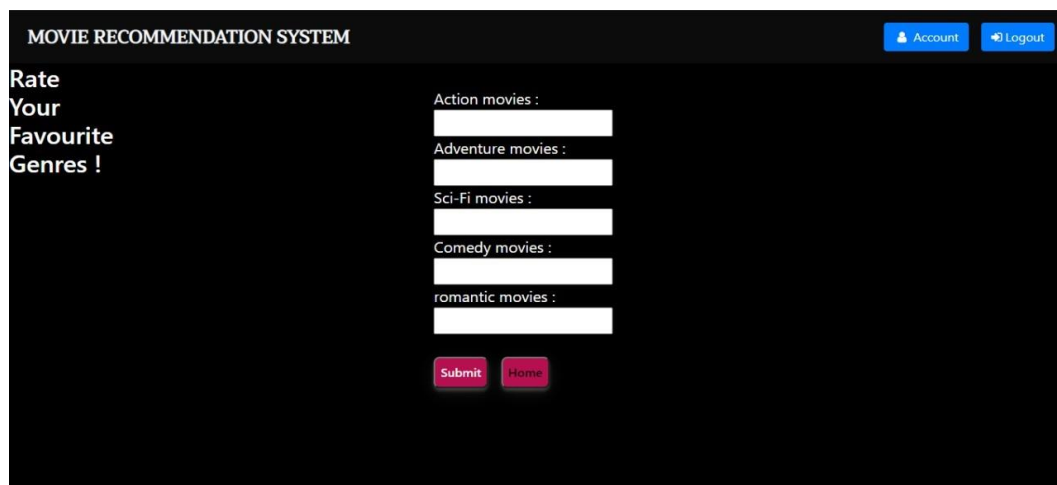


Figure 3: Rating page

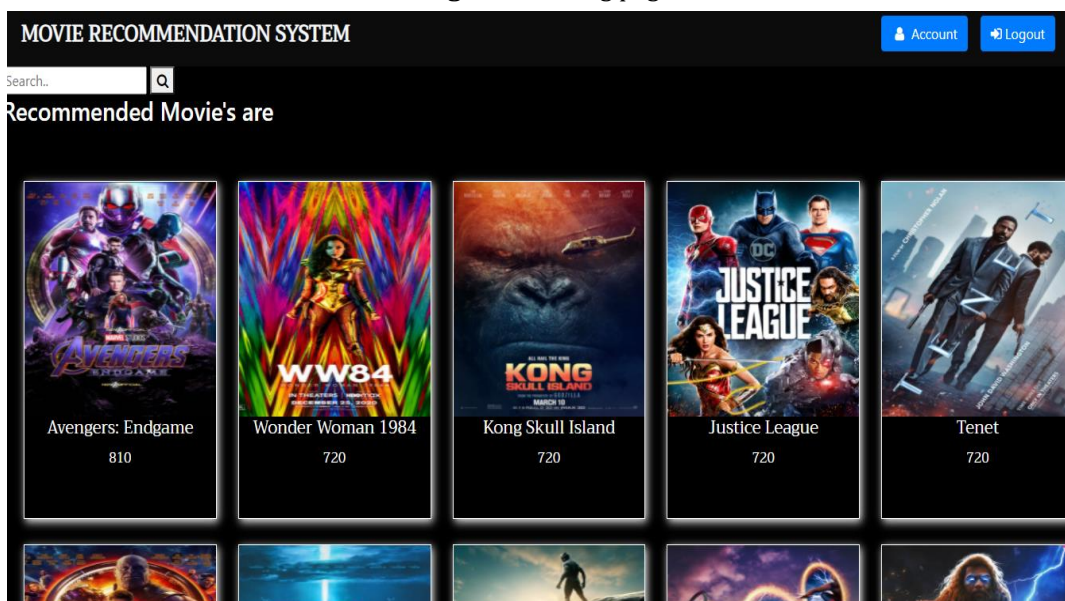


Figure 4: Recommendation page.

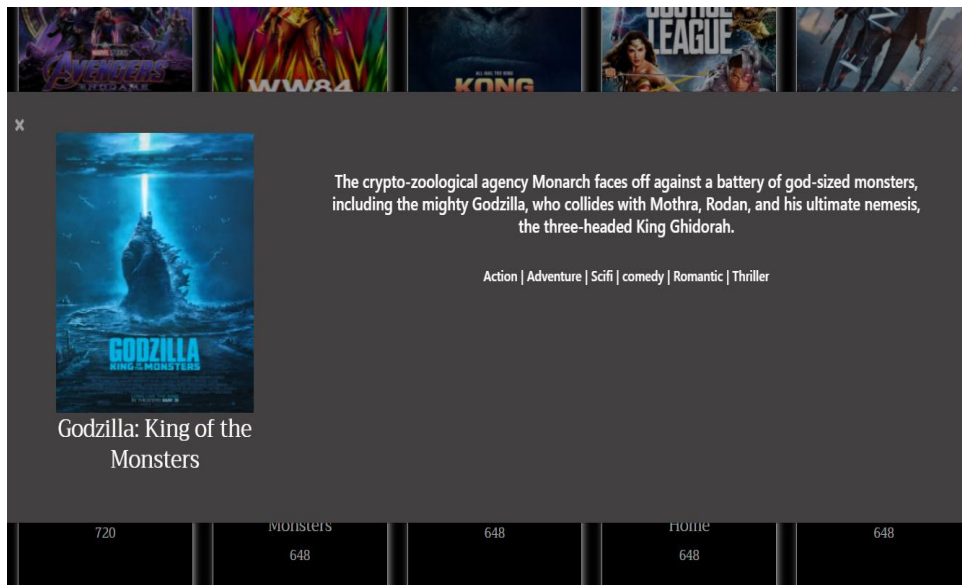


Figure 5: Information page

IV. RESULTS AND DISCUSSION

Recommendation systems collect customer data and auto analyze this to generate customized and personalized recommendations for customers. Recommendation systems are necessity for today in every field because of rapidly increasing data. Thus, this recommendation system will suggest user for movie's based on their interest which he/she would like to watch, thus reducing the overhead of searching for one in busy schedule thus providing with more customized and personalized experience.

V. CONCLUSION

This recommendation system is meant to provide user with similar movies which he/she would like to watch based on the past behavior. Thus, this recommendation system has eventually reduced the overhead of market research needed to be done by a user before watching any movie. This system can further be enhanced by adding more genres and updating genres rating. Also, the implementation of group-based recommendation can be done. A hybrid system can be implemented considering the benefits from content based and collaborative filtering models.

ACKNOWLEDGEMENT

We would like to thank Prof. Rahul More for his cooperation and enthusiasm. Without him, research to this paper and project would not have been possible. At times discussion for further processing was needed, we would like to thank him for his valuable time and thus helping us out at times we were stuck with some issues.

VI. REFERENCES

- [1] Bagher Rahimpour Cami, Hamid Hassanpour and Hoda Mashayekhi, "A content-based movie recommender system based on temporal user preferences", 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), Shahrood, Iran, 2017.
- [2] SRS Reddy, Sravani Nalluri, Subramanyam Kuniseti, S. Ashok and B. Venkatesh, "Content-Based Movie Recommendation System Using Genre Correlation", Smart Innovation, Systems and Technologies at Singapore, 2018.