Shodhshauryam, International Scientific Refereed Research Journal
Available online at : www.shisrrj.com
© 2024 SHISRRJ | Volume 7 | Issue 2
ISSN : 2581-6306
doi : https://doi.org/10.32628/SHISRRJ

Movies Recommendation System Using Hybrid Filtering and Naive Bayes Algorithm

B. Chandini[1] , B. Rupadevi [2]

[1]Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

[2]Associate Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

Article Info

Publication Issue :
March-April-2024
Volume 7, Issue 2

Page Number : 90-98

Article History
Received : 15 March 2024
Published : 30 March 2024

ABSTRACT

There is a growing need to develop a strong sentiment analysis model that categorizes movie evaluations because customer opinions can improve product efficiency and because reviews determine whether a movie succeeds or fails. Tokenization is used in this paper to convert the input string into a vector of words; stemming is used to extract word roots; feature selection is used to extract key words; and classification is applied to categorize reviews as positive or negative. We construct this model just by utilizing content based collaborative filtering and for the sentiment analysis we are going to use the Naive Bayes algorithm, Collaborative Filtering Algorithms, Content-Based Filtering Algorithms, Hybrid Filtering Algorithms to predict.

Keywords: Machine Learning, Naive Bayes, Natural Language Processing (NLP), Sentiment Analysis

I. INTRODUCTION

In today's digital age, where streaming platforms offer an overwhelming abundance of movies and TV shows, users often struggle to navigate through the sea of content to find something that truly resonates with them. The sheer volume of options can lead to decision fatigue and dissatisfaction, highlighting the need for innovative solutions in content discovery. A Movie Recommendation System Coupled with Sentiment Analysis emerges as a viable strategy to deal with this issue. By harnessing the power of advanced algorithms and machine learning techniques, this system aims to analyze user preferences and emotional states to deliver personalized movie recommendations.

At its core, the system collects and analyzes user data, including viewing history, ratings, and even textual reviews, to gain insights into individual preferences. By integrating sentiment analysis, it goes beyond mere ratings to understand how users feel about specific movies, genres, or themes. This deeper understanding enables the system to offer recommendations that not only align with users' tastes but also cater to their current emotional states.

II. EXISTING AND PROPOSED SYSTEMS

A. Existing systems

We can't identify any kind movie rating or app rating etc. we don't know which movie is the best movie at present we blindly follows based on analyses of the of people either movie is good or bad based on that which people uses to follow it.

1. Collaborative Filtering Component:

The system employs traditional collaborative filtering by analyzing user behavior, preferences, and historical data.

2. Item-Based Collaborative Filtering: The system also incorporates item-based collaborative filtering to identify movies that are similar based on user interactions.

3. Content-Based Filtering Component:

The system utilizes content-based filtering, considering features such as genre, actors, directors, and keywords associated with movies. This method helps in recommending movies based on the intrinsic characteristics of the films that match a user's preferences.

4. Sentiment Analysis Component:

Sentiment analysis is applied to user reviews and comments associated with movies. Natural Language Processing techniques help in determining the sentiment expressed in textual content, extracting opinions, and gauging user satisfaction.

Positive sentiments from user reviews contribute to the recommendation strength for a particular movie. Additionally, sentiment analysis helps in understanding the nuanced preferences of users, such as their preference for uplifting or emotionally charged movies

B. Proposed system

we are going to implement a recommendation system using content based collaborative filtering where we will get all the recommendation system along with movie details like name, rating, genre and the movie image as well and the star cast details with their biography as well and the sentiment analysis of the reviews has been done. To get all the details we are using the "The movie dB".

Advantages:

- **High accuracy.**
- **Time Saving.**
- **High reliability.**
- **Low complexities.**

III. LITERATURE SURVEY

[1]Xiaoyuan Su and Taghi M. Khooftaar: "A survey of collaborative filtering techniques". This survey provides an extensive overview of collaborative filtering techniques, which are widely used in recommendation systems. It discusses various collaborative filtering algorithms, including memory-based and model-based approaches, highlighting their strengths, weaknesses, and applications.

[2]: Bing Liu, "Sentiment analysis and opinion mining". Liu's work provides a comprehensive review of sentiment analysis and opinion mining techniques. It covers various aspects of sentiment analysis, including text classification, feature extraction, and sentiment lexicons, offering insights into the challenges and opportunities in this field.

[3]: Paolo Cremonesi, Franca Garzotto , Alessandro Vassena, "Content-based recommendation systems:

State of the art and trends". With an emphasis on the state of the art and new developments in content-based recommendation systems, this article provides a thorough study of the field. It goes over the ideas underlying content-based filtering as well as new developments in evaluation techniques and recommendation algorithms.

[4]: Gediminas Adomavicius, Alexander Tuzhilin, "Hybrid Recommender Systems: Survey and Experiments". Adomavicius and Tuzhilin provide a comprehensive survey to improve accuracy and coverage. The paper discusses different hybridization strategies, along with experimental evaluations and real-world applications.

[5]: Jin-Woo Park, Jianfeng Wang, Ruochen Jiang, Hyung-Jeong Yang "Understanding and Predicting Movie Box Office Revenue: An Integrated Approach". This study explores the prediction of movie box office revenue using an integrated method that blends conventional predictors, social media data, and sentiment analysis. It illustrates how useful sentiment analysis is in understanding audience reactions and predicting box office performance.

IV. METHODLOGY

4.1 Data Collection: Gather movie-related data from various sources such as IMDb, TMDb (The Movie Database), or other movie databases. Obtain user ratings and reviews from platforms like IMDb, Rotten Tomatoes, or user-generated reviews. Collect metadata including movie titles, genres, release years, directors, actors, ratings, and user reviews.

4.2 Data Preprocessing:

Clean the collected data by handling missing values, duplicates, and inconsistent formats. Normalize textual data such as movie titles, genres, and user reviews by removing special characters, performing tokenization.

Extract features from textual data for sentiment analysis, such as sentiment scores or sentiment labels (positive, negative, neutral).

4.3 Sentiment Analysis: Perform sentiment analysis on user reviews to determine the sentiment polarity (positive, negative, neutral) associated with each movie.

4.4 **Model Training and Evaluation:** Train machine learning models or deep learning models on historical user-item interactions and movie attributes.

Validate the recommendation system using techniques like cross-validation or split validation to ensure robustness and generalization.

4.5 DataSets used

There are several datasets that researchers and practitioners often use for movie recommendation systems, especially when incorporating hybrid filtering and sentiment analysis. Here are some popular datasets:

4.5.1 Credits Dataset:

User Ratings: This dataset contains user ratings for various movies. It includes information like user ID, movie ID, and rating given by the user to the movie.User Preferences: This dataset captures user preferences, such as genres they like, actors they prefer, directors they follow, etc. It helps in understanding user tastes and preferences

	A	B	C
1	cast	crew	id
2	[{'cast_id':	[{'credit_ic	862
3	[{'cast_id':	[{'credit_ic	8844
4	[{'cast_id':	[{'credit_ic	15602
5	[{'cast_id':	[{'credit_ic	31357
6	[{'cast_id':	[{'credit_ic	11862
7	[{'cast_id':	[{'credit_ic	949
8	[{'cast_id':	[{'credit_ic	11860
9	[{'cast_id':	[{'credit_ic	45325
10	[{'cast_id':	[{'credit_ic	9091
11	[{'cast_id':	[{'credit_ic	710
12	[{'cast_id':	[{'credit_ic	9087
13	[{'cast_id':	[{'credit_ic	12110
14	[{'cast_id':	[{'credit_ic	21032
15	[{'cast_id':	[{'credit_ic	10858
16	[{'cast_id':	[{'credit_ic	1408
17	[{'cast_id':	[{'credit_ic	524
18	[{'cast_id':	[{'credit_ic	4584
19	[{'cast_id':	[{'credit_ic	5
20	[{'cast_id':	[{'credit_ic	9273
21	[{'cast_id':	[{'credit_ic	11517
22	[{'cast_id':	[{'credit_ic	8012
23	[{'cast_id':	[{'credit_ic	1710
24	[{'cast_id':	[{'credit_ic	9691
25	[{'cast_id':	[{'credit_ic	12665
26	[{'cast_id':	[{'credit_ic	451
27	[{'cast_id':	[{'credit_ic	16420
28	[{'cast_id':	[{'credit_ic	9263
29	[{'cast_id':	[{'credit_ic	17015
30	[{'cast_id':	[{'credit_ic	902
31	[{'cast_id':	[{'credit_ic	37557

Figure 1. Credit dataset for analysis purpose

4.5.2 METADATA Dataset:

Information about movies like genres, release year, director, actors, etc., is crucial for collaborative filtering. It helps in finding similarities between movies based on their attributes.

i. director_name
ii. actor_1_name
iii. actor_2-name
iv. actor_3_name
v. genres
vi. movie_title
vii. comb

These are the attributes that are selected to predict the model.

4.5.3 Review Dataset

For a review dataset plays a crucial role in understanding user sentiments towards movies. Here's an explanation of the review dataset and its significance in such a project: Review Data Set Content: User Reviews: This dataset contains textual reviews written by users after watching movies. It includes information such as the user ID, movie ID, review text, and possibly a sentiment score (positive/negative).Critic Reviews: In addition to user reviews, you can also include reviews from professional critics or platforms like Rotten Tomatoes or IMDb. These reviews often come with ratings or qualitative assessments of the movie's quality.

Here are the some reviews based upon the user choice.

1 I loved the Da Vinci Code, but now I want something better and different!..

1 I thought the Da Vinci Code was a pretty good boo

V. EXPERIMENTAL SETUP

Obtain a diverse dataset of movies including metadata like genres, ratings, user reviews, and sentiment scores. Ensure the dataset covers a wide range of genres, release years, and popularity levels to provide a comprehensive testing ground. Preprocess textual data such as user reviews by tokenizing, lemmatizing, and removing stop words. Perform sentiment analysis on user reviews to extract sentiment scores (positive, negative, neutral) for each review. Extract features from the dataset such as movie genres, ratings.

5.1 Algorithms used

5.1.1 NAÏVE BAYES

Naive Bayes algorithm can be incorporated into movie recommendation systems using hybrid filtering techniques, combining collaborative filtering (CF) and content-based filtering (CBF). Sentiment analysis can further enhance the system by analyzing user reviews or ratings to understand their sentiments towards movies.Here's a general approach to using Naive Bayes for movie recommendation with hybrid filtering and sentiment analysis

Data Collection: Gather a dataset that includes information about movies (e.g., genre, actors, directors, release year) and user data (e.g., ratings, reviews, preferences).Preprocessing: Clean and preprocess the data, handling missing values and irrelevant information. Convert text data (e.g., reviews) into a format suitable for sentiment analysis, such as tokenization, stop word removal, and stemming/lemmatization.

Feature Extraction: For content-based filtering, extract features from movies (e.g., genre vectors, actor/director information). For sentiment analysis, extract sentiment-related features from user reviews or ratings.

Training Naive Bayes Classifier: Use the preprocessed data to train a Naive Bayes classifier for sentiment analysis. This classifier can predict the sentiment (positive, negative, neutral) of user reviews.

Hybrid Filtering: Combine collaborative filtering and content-based filtering approaches. Collaborative filtering recommends movies based on user-item interactions and similarities among users or items. Use the Naive Bayes sentiment classifier to filter the recommendations based on sentiment. For example, if a user has a preference for positive sentiment movies, filter out negative sentiment movies from the recommendations.

Recommendation Generation: Based on the hybrid filtering results and sentiment analysis, generate personalized movie recommendations for each user. Consider diversity in recommendations to provide a balanced list of movies across genres, sentiment, and popularity.

Evaluation: Evaluate the performance of the recommendation system using metrics such as precision, recall, F1-score, and user satisfaction surveys.

Iterative Improvement: Continuously update and refine the recommendation system based on user feedback and new data to improve accuracy and relevance of recommendations.

By combining Naive Bayes sentiment analysis with hybrid filtering techniques, movie recommendation systems can provide more accurate and personalized recommendations tailored to individual user preferences and sentiments towards movies.

5.1.2 Sentimental Analysis

The inclusion of sentiment analysis algorithms in movie recommendation systems enhances the accuracy and personalization of recommendations by considering users' emotional responses to movies. Here's how sentiment analysis contributes to defining a movie recommendation system using hybrid filtering and sentimental analysis

Understanding User Preferences: Sentiment analysis helps in understanding user preferences by analyzing their sentiments expressed in reviews or ratings. Positive sentiment indicates a liking for certain aspects of movies, such as plot, acting, or visuals, while negative sentiment highlights dislikes or areas of improvement.

Enhancing Collaborative Filtering: In hybrid filtering, sentiment analysis complements collaborative filtering by providing insights into users' emotional responses to movies. It helps in refining recommendations by considering not just user-item interactions but also the sentiment associated with those interactions. For example, it can prioritize recommending movies with positive sentiment from users who typically enjoy similar genres or themes.

Improving Content-Based Filtering: Sentiment analysis can also enhance content-based filtering by incorporating sentiment-related features of movies. For instance, sentiment analysis can categorize movies as having predominantly positive, negative, or neutral sentiments based on user reviews. This information can be used to recommend movies that align with a user's sentiment preferences.

Personalized Recommendations: By combining sentiment analysis with hybrid filtering techniques, movie recommendation systems can generate highly personalized recommendations. For example, if a user consistently expresses positive sentiment towards romantic comedies, the system can prioritize recommending similar movies with positive sentiments from other users who share similar tastes.

Filtering Out Irrelevant Recommendations: Sentiment analysis helps in filtering out irrelevant recommendations that may have negative sentiment, ensuring that users receive recommendations that align with their emotional responses and preferences.

Enhancing User Experience: By considering sentiment in recommendations, movie recommendation systems can enhance the overall user experience by providing more relevant and enjoyable movie suggestions. This can lead to increased user satisfaction and engagement with the platform.

Overall, sentiment analysis plays a crucial role in refining movie recommendations in hybrid filtering systems, ensuring that recommendations not only match users' preferences based on past interactions but also align with their emotional responses and sentiments towards movies.

VI. IMPLEMENTATION

It requires some steps to be followed.

Register: user will register with the user details

Login: User will login using email and password.

Search movie Name: User will search movie name and it will give some recommendation and displays the cast, reviews etc.

Ticket Bookings: In ticket bookings we can book the ticket and do the payment.

My Bookings: User can view all his ticket bookings in My bookings.

Logout: Finally, Logout from the application

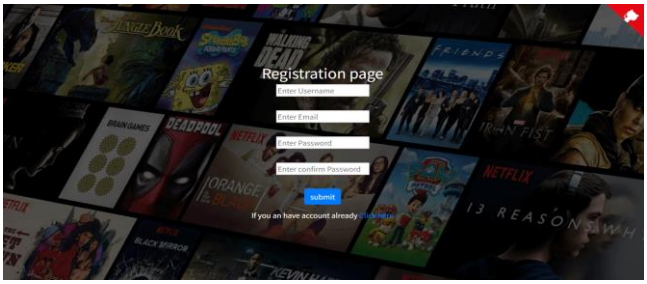


Fig 2 Registration page

Fig 2, First, the user need to register by entering the following details like username, Email, Password, at last they need to confirm the password.

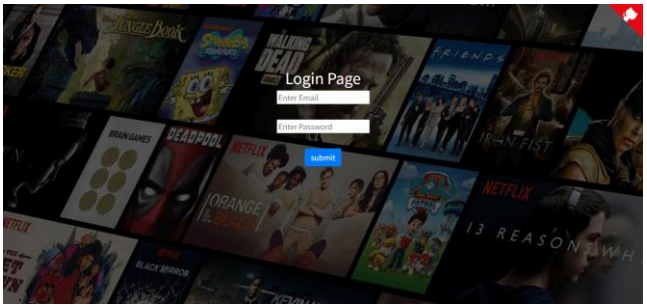


Fig 3 Login page

Fig 3 tells that After registration the user needs to login with respective email id or user name and password and they need to confirm the password. After login in to the website the following image will be displayed

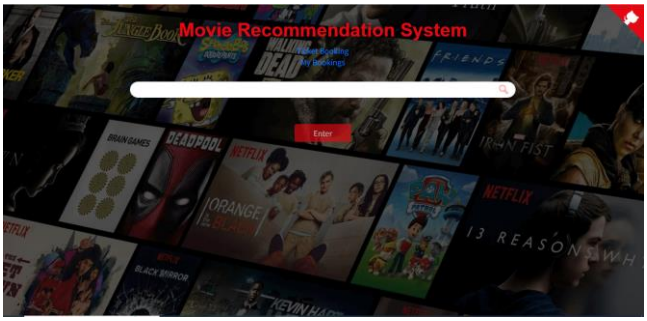


Fig 4 Ticket Booking Page

Fig 4 tells the user has to book the tickets. First the user has to select the movie. Based on the user preference is going to recommendation all the movies. And then after selecting the movies they need to enter how many tickets they want to book and then next column its will go for calculating the price, and then the user needs to enter card number and cvv number and expire date and then submit. It will load the page for some time.

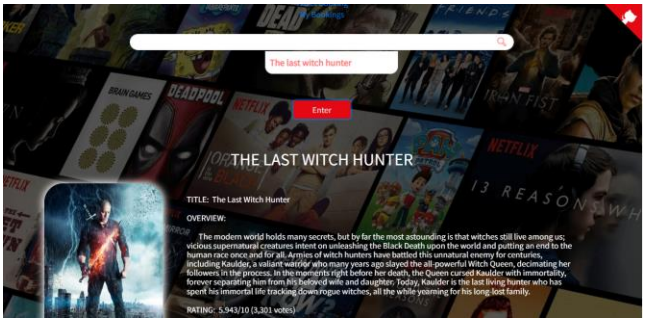
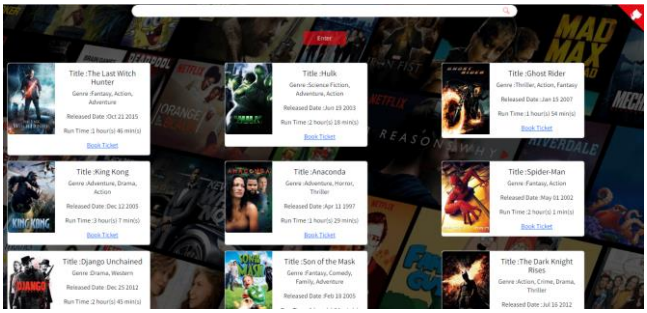


Fig 5 Search for an movie

Fig 5 Tells that the user will go for selecting an movie



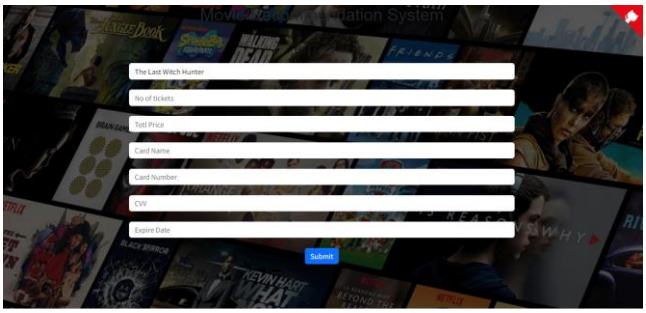


Fig 6 Payment

Fig 6 Tells that the user will go booking the ticket by entering the Details

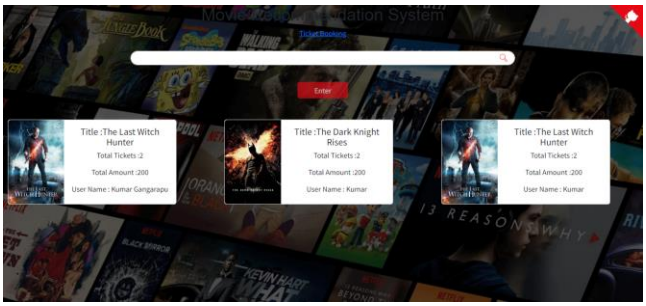


Fig 7 My Bookings

Fig 7 Displays that what are the movies the user has been Booked

VII. CONCLUSION

In conclusion, the development of a Movie Recommendation System coupled with Sentiment Analysis presents a promising solution to the challenges of content overload and user dissatisfaction in the digital entertainment landscape. By leveraging advanced algorithms and techniques, such as collaborative filtering, content-based filtering, and sentiment analysis, this system aims to enhance content discovery and user satisfaction..

VIII. FUTURE ENHANCEMENT

Enhanced Contextual Understanding: Integrating contextual information such as user location, time of day, and device type to provide more relevant recommendations tailored to the user's current context.

Incorporation of Social Influence: Leveraging social network data and influence modeling techniques to incorporate the impact of social interactions and recommendations from friends or influencers.

Continuous Learning and Adaptation: Implementing mechanisms for the system to continuously learn from user feedback and adapt its recommendations over time, ensuring they remain accurate and reflective of changing preferences.

Explainability and Transparency: Enhancing the transparency of recommendation algorithms by providing users with explanations or insights into why certain recommendations are made, fostering trust and understanding.

Interactive Interfaces: Developing interactive interfaces that allow users to provide real-time feedback and refine their preferences, enabling a more dynamic and engaging recommendation experience.

IX. REFERENCES

[1]. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. IEEE Computer Society, 42(8), 30-37.

[2]. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[3]. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.

[4]. Park, J. W., Wang, J., Jiang, R., & Yang, H. J. (2012). Understanding and predicting movie box office revenue: An integrated approach. Journal of Marketing Research, 49(4), 532-544.

[5]. Cremonesi, P., Garzotto, F., & Vassena, A. (2011). A survey of content-based recommender systems. In Recommender systems handbook (pp. 73-105). Springer, Boston, MA.

[6]. Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009, 4.

[7]. Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58.

[8]. Hosseinmardi, H., Zamani, H., Shahriari, H. R., & Alhajj, R. (2015). Sentiment analysis algorithms and applications: A survey. Knowledge-Based Systems, 89, 14-37.

[9]. Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434).

[10]. Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). Recommender systems: An introduction (Vol. 40). Cambridge University Press.