The Content-Based Recommender System for predicting the target audience and popularity of movies

[1] Mrs.K.Vasumathi, [2] Dr.S.Selvakani, [3] D.Logakrishna
[1]Assistant Professor, [2]Assistant Professor and Head, [3]PG Research Scholar
[1,2,3] PG Department of Computer Science,
[1,2,3] Government Arts and Science College, Arakkonam – 631051.

Abstract - Obtaining the current price prediction for a company's stock or other commercial tool can be achieved by using previous data sets. Due to the potential for substantial financial gains, individuals are often attracted to investing in stock markets and exchanges. Predicting stock prices is a challenging task as it is affected by various factors, such as political conditions, global economy, and the company's financial performance. In this project, we propose to gather input data, eliminate unwanted null values, perform feature engineering, and apply Decision Tree and Linear Regression models for training purposes. The stock market is a dynamic, unpredictable, and non-linear environment, making the prediction of stock prices a challenging issue.

Keywords - Prediction, Machine Learning, Data Set, Stock Price prediction, Decision tree algorithm, and linear regression.

I. INTRODUCTION

Stock price prediction involves predicting the current value of a product in the market. The two primary methods used for stock price prediction are analyzing historical data from the past 12 months and forecasting for the next 12 months. Sequential data processing techniques such as decision trees and linear regression are commonly used for this purpose. There are three different ways to predict stock prices: regular price, least possible price, and best possible price. The primary functions of stock price prediction remain historical analysis and future forecasting.

Processing sequential data for financial time series prediction is a challenging task due to their noisy, non-stationary and chaotic nature. However, with the rapid development of information technologies, machine learning techniques inspired by human brain that consist of inter-connected neurons have been proposed to improve prediction performance. While traditional methods such as linear regression and Autoregressive Moving Average (ARMA) have been used in the past, recent studies have shown that machine learning techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) have significantly improved prediction efficiencies by 60-86 percent.

To reduce variance in the model, Random Forest and Support Vector Machines are used to predict stock values using machine learning algorithms. The Stock Closing Price Prediction using Machine Learning Techniques is capable of performing both regression and classification tasks by combining multiple decision trees to determine the final output. Investors can benefit from the analysis of future conditions of the market provided by stock market predictions. However, predicting stock prices is a challenging task as it depends on various factors including political conditions, global economy, company's financial reports and performance, etc. Therefore, the outputs generated from different traditional and machine learning methods can help investors avoid risks and improve performance.

Future studies in this area are planned to focus on extracting various uncorrelated features from the raw data and studying multivariate structures. The ultimate goal of this project is to improve the quality of the stock market predictions by using the stock value.

II. REVIEW OF LITERATURE

Lalita Sharma, Anju Gera [1] says that In many applications such as the World Wide Web, the process of recommendation plays a crucial role. The primary aim of this paper is to highlight the challenges associated with the techniques used to generate recommendations. The techniques can be broadly categorized into three categories: Collaborative Filtering, Content-Based, and Hybrid Recommendations. By providing an overview of these challenges, we can improve the quality of recommendations by developing new approaches and methods that can be used for research and practice in this field. This can act as a pathway to further advancements in the area of recommendation techniques.

- The Collaborative Filtering (CF) problem involves predicting the rating of an item that has not yet been rated. Various approaches are used to calculate similarities between items and users to make this prediction.
- A large database containing the items to be recommended and their features is known as the Item Profile.
- Users provide information about their preferences to the system, and based on this information, the system creates a profile for the users by combining it with the item information.

- The system recommends appropriate items to the user based on the information available in their profile.
- One approach to recommendation is to implement Collaborative Filtering (CF) and Content-Based (CB) techniques separately and then combine their predictions.

Neelam Singh, Shivanshi Tripathi, Devesh Pratap Singh [2] says that The rapid advancement of technology and the widespread use of social media platforms such as Facebook, Instagram, and Twitter have revolutionized the way people express their opinions, observations, and sentiments about products and services. As a result, a vast amount of data is being produced and accumulated. Recommendation systems are gaining momentum in extracting insights from this data to make decisions that can be presented in various statistical and graphical formats. These systems have been successful in predicting or suggesting products, including food, movies, restaurants, and more.

a. Recommender systems based on collaborative filtering are known for their flexibility as they can be utilized across various domains and even generate recommendations that span different domains if implemented correctly.
b. Collaborative fissltering is particularly effective when applied to a large user base.
c. Collaborative filtering engines can address the issue of the "filter bubble" by allowing users to discover and connect with new subspaces within the item space.Limitations of Collaborative Filtering-Based Recommender Systems:

Mohd Abdul Hameed, Omar Al Jadaan, S. Ramachandram [3] says Collaborative filtering is the most popular technique for making recommendations. These systems make predictions about a user's preferences for products or services by studying past relationships between users and items with similar preferences and tastes. This article examines different aspects of collaborative filtering recommendation systems and classifies them based on their approach to computing similarity. Every buyer and seller try to predict the stock market price movements to get maximum profits and minimum losses.

- User and item profiles are created by collaborative filtering algorithms instead of storing large amounts of term frequency data for each user and document.
- Content analysis is not necessary for collaborative filtering algorithms.
- Collaborative filtering algorithms do not require language development.

Badrul Sarwar, George Karypis [4] says that The task of providing personalized recommendations for information, products, or services during a live interaction is accomplished by applying knowledge discovery techniques through recommender systems. The k-nearest neighbor collaborative filtering-based systems, in particular, are experiencing significant success on the Web. Statistical Aural metrics evaluate the aura y of a system by comparing the numerical recommendation s ores against the actual user ratings for the user-it.

- Sparsity. In practice, many commercial recommender systems are used to evaluate large item sets (e.g., Amazon. recommends books and CDnow.com recommends music albums).
- Salability. Nearest neighbor algorithms require computation that grows with both the number of users and the number of items.

Yehuda Koren and Robert Bell [5] says that the collaborative filtering (CF) approach to recommendation systems has gained significant attention and advancement in recent times. Its popularity has been fueled by its central role in the recently concluded Netflix competition. This chapter provides an overview of the recent developments in this field. It describes matrix factorization techniques, which have become the preferred choice for implementing CF, along with recent advancements. The decision of how to split the timeline into bins should balance the desire to achieve finer resolution (hence, smaller bins) with the need for enough ratings per bin (hence, larger bins).

- In our implementation, each bin corresponds to roughly ten consecutive weeks of data, leading to 30 bins spanning all days in the dataset.
- A day t is associated with an integer Bin(t) (a number between 1 and 30 in our data), such that the movie bias is split into a stationary part and a time changing part

Joseph A. Konstan, and John Riedl [6] says that community of people to predict a person's affinity for items or information. The system shares ratings between like-minded individuals. However, current recommender systems lack transparency as they are black boxes that offer no insight into how recommendations are generated.

Explanations offer transparency by revealing the reasoning and data that underpins As per, Joseph Automated collaborative filtering (ACF) systems use recorded interests of a recommendation. It represented some advantage of Justification. User understanding of the reasoning behind a recommendation, so that he may decide how much confidence to place in that recommendation.

- User Involvement. User involvement in the recommendation process, allowing the user to add his knowledge and inference skills to the complete decision process.
- Education. Education of the user as to the processes used in generating a recommendation, so that he may better understand the strengths and limitations of the system.
- Acceptance. Greater acceptance of the recommender system as a decision aide, since its limits and strengths are fully visible and its suggestions are justified

III. PROPOSED METHODOLOGY

1. Acquire movie data and movie intrinsic features from TMDb dataset and computes similar movie using a content-based movie recommendation system.

2. Use similar movie information and voting data from the IMDb beta set. Predict the movie popularity using the Deep learning approach.

3. Compute target audience prediction using fuzzy c means.

(1)Movie Recommendation Module

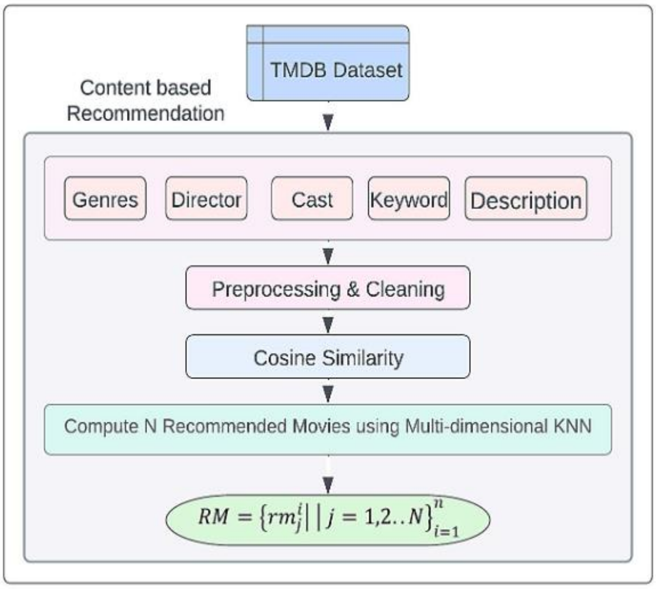

Figure 1. Popularity box plot of Animation Movies

IV. EXPERIMENTAL RESULTS

The objective of our study is to propose an expert system that could help the movie maker execute necessary changes if needed at the appropriate time. Our system can food cost the level of popularity of the upcoming movie before the production has started for the earliest stage of the production and with significant accuracy. About system focused not only on the popularity of the upcoming movie but also on the movie's popularity among all age groups

Table 1. Comedy movie's target audience

Movie Title	Genres	Popularity			
		Junior	Teenage	Mid age	Senior
Euloge	Comedy Drama	54.8	69.53	70.95	73.3
Death At a Funeral	Comedy Drama	52.3	67.8	70.68	73.22
And So It Goes	Comedy Drama Romance	63.45	83.27	85.33	88.06

Algorithm 2 Movie popularity prediction using deep learning.
Input:
Recommended Movie details RM = {rmij|j = 1, 2, . . . , N }n i=1 Movie Rating Database Mr
Output:
Prediction of Popularity Class (0, 1, 2, 3, 4, 5) for movie mi
1. for(Each movie mi ∈ M)
2. for(Rating r = 1, 2, . . . , 10)
3. V ir = N∑ k=1vir,k ,

4. where vir,k = no. of vote with rating r of kth recommended movie of mi
5. V i = {V ir |r = 1, 2, . . . , 10 voting details of each movie mi
6. V = {V i}n i=1 voting details of all movie M
7. R = {Ri j|j = 1, 2, . . . , N }n i=1, where Rij = rating of jth recommended movie of mi
8. Create Dataset X = V ∪ R
9. Y = {yi|yi ∈ (0, 1, 2, 3, 4, 5)}n i=1
10. X = Xtrain ∪ Xtest
11. Y = Y_train ∪ Y_test
12. Y ^_test = CNN(X_train,Y_train,X_test)
13. Return Y ^_test

- The system would use user rating by each age group and take the number of vote details from each group to each recommended movie—system analysis of all the voting and rating information from each group of all the recommended movies.
- Ultimately, the module would predict the popularity of the upcoming movie for each age group. The target audience prediction module takes input from the first module output, and that also takes the IMDb rating data set as input.
- Movie recommendation module produces similar movies
- (rmij | j = 1, 2, . . . , N) of a given movie mi. We have used a set of recommended movies of all movie RM = ∪n i=1{rmij | j = 1, 2, . . . , N } present in the data set.

We have also taken the IMDb rating data set, including voting and rating information of all movies present in the data set. The proposed system divides the audience into four groups according to age demography. The proposed system forecast how much preferable the movie would be for each group for an upcoming movie. Which group would prefer the film most and which group would not like the movie. Table 2 presents the viewer's age groups.

TABLE 2. Viewer's age group

Age	Group	Group name
0-17	Gr-1	Junior
18-29	Gr-2	Teenage
30-44	Gr-3	Mid-age
45+	Gr-4	Senior

V. Conclusion

A substantial amount of financing is consumed in every box office movie. However, most movies fail to achieve success. Earlier, the most significant number of works have been done on post-production or post-release forecast. The estimate does not influence as the investor has already consumed their funds on the film production. The pre-production or early production stage forecast needs high accuracy and the best time to ensure investment. The objective of our study is to propose an expert system that could help the movie maker execute necessary changes if needed at the appropriate time. Our system can food cost the level of popularity of the upcoming movie before the production has started for the earliest stage of the production and with significant accuracy. About system focused not only on the popularity of the upcoming movie but also on the movie's popularity among all age groups. Movie Maker can estimate the target audience and assess how the different audience groups would respond to the upcoming movie.

VI. References

[1] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, ''Item-based collaborative filtering recommendation algorithms,'' in Proc. 10th Int. Conf. World Wide Web (WWW), 2001, pp. 285–295.
[2] J. L. Herlocker, J. A. Konstan, and J. Riedl, ''Explaining collaborative filtering recommendations,'' in Proc. ACM Conf. Computes. supported Cooperat. Work (CSCW), 2000, pp. 241–250.
[3] L. Sharma and A. Gera, ''A survey of recommendation system: Research challenges,'' Int. J. Eng. Trends Technol., vol. 4, no. 5, pp. 1989–1992, 2013.
[4] M. A. Hameed, O. Al Jadaan, and S. Ramachandram, ''Collaborative filtering based recommendation system: A survey,'' Int. J. Comput. Sci. Eng., vol. 4, no. 5, p. 859, 2012.
[5] M. J. Pazzani, ''A framework for collaborative, content-based and demographic filtering,'' Artif. Intell. Rev., vol. 13, no. 5, pp. 393–408, 1999
[6] P. Nagarnaik and A. Thomas, ''Survey on recommendation system methods,'' in Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS), Feb. 2015, pp. 1603–1608.
[7] Y. Koren and R. Bell, ''Advances in collaborative filtering,'' in Recommender Systems Handbook. Boston, MA, USA: Springer, 2015, pp. 77–118.