Movie Popularity and Target Audience Prediction Using the Content-Based Recommender System

Madasu V Bhgya Sree[1], Suthari Vamshi[2], Eppalapally Saikiran[3], Najeema Afrin[4]

[1,2,3]Department of CSE, CMR Technical Campus, Telangana, India
[4]Assistant Professor, Department of CSE, CMR Technical Campus, Telangana, India

Abstract: The movie is an important part of our daily entertainment. For the worldwide movie industry, this is a highly thriving and substantial sector that attracts attention from people of all ages. In a recent study, it has been observed that only a small number of films are successful. The film production industry's stakeholders have been extremely stressed by uncertainty in the sector There is a growing belief among filmmakers and researchers that certain expert systems must be established to predict the movie's success probability in advance of its production with sufficient accuracy. In order to anticipate film popularity at the final production stage, a large amount of research has been carried out. We need to predict at the early stages of film production and provide specific information on upcoming movies, so that movie makers can estimate their future films and make necessary changes. The study suggests that a content Based Recommendation System for movies, CBRS using Preliminary Film features such as genre, cast, director, keyword and movie description, should be developed. We have created a new set of features and proposed the Random Forest deep learning (DL) model for building an multiclass movie popularity prediction system by using RS data, film ratings and voting information from similar movies. We've also made proposals for a system of predicting the popularity of an upcoming film according to different audience groups. The audience group was divided into four age groups: Junior, Senior, Middle and Elderly. The publicly available internet movie database IMDb and the film database HTTP(TMDb) data were used in this study. We have implemented a multiclass classification model that has yielded 96.8% accuracy which is superior to all the reference models. The potential of predictive and prescriptive data analysis within IT systems to underpin industry decisions has been highlighted in this study.

Keywords: Content-based, recommendation system, expert systems, deep learning, audience groups.

I.INTRODUCTION

The world's movie industry is fast becoming a lucrative source of revenue, while multi billion dollars have been invested in this sector. There are many people involved in this sector, and massive investment is needed for both qualitative and quantitative reasons. The United States and Canada's overall box office revenue reached $11.32 billion for the year 2019. Indeed, few numbers of movies have been a success in the real world. The need for some expert systems which predict the film's chances of success and lead production with correct accuracy is constantly felt by producers and researchers. There is a great deal of diversity in the film industry. In making a film, the sign can't number of parameters from different dimensions is used. It is an extremely difficult job to represent the success of a new film or its degree of success. Prior to this, a number of studies have been carried out on the post production or after release forecast. However, this is not in the investor's best interest because their funds have already been used to produce a film. The early production stage and pre-production prediction with satisfying accuracy have been beneficial to secure investment. A forecast made soon after the cast, director, and storyline have been nalized would assist the investor in making a facial decision. After a rigorous study, we have seen signicant research on movie hit prediction before the official release. Predictions performed shortly before or following, the official release (the last stage in lm production) may have additional data to use and produce a more precise prediction. Early-stage (production) forecast of movie success is the most beneficial. At the earliest stage of production, very little work was done to forecast film success. There is no significant accuracy in the early stage fore-cast of previous works. Maximum of the works are performed only to focus success probability of the upcoming movie. Some of them classify the problem into a binary problem (hit/ flop), and in some work, they classify the problem into a multiclass problem. Most of the time, movie producers just create a new film to

target specific audiences or groups. Audience age is one of the essential criteria for the target audience. Some movies are being produced in an effort to reach a younger audience. Some movies are aimed at young audiences, sometimes targeting a middle or elderly audience, and some films are geared towards all of them. Suppose we're able to predict whether or not the upcoming film will be popular with the target audience group. At the earliest stage of production, film producers would be in a position to benefit if we were able to monitor the influence of an upcoming movie on all age groups. In this way, the Movie Maker could make some necessary modifications if needed. The movie hit fore- casting and target audience prediction of the upcoming movie at the early stage of the movie production are interrelated and meaningful. The risk for the film industry could be reduced as a result of this work.

Our research problem proposed a system to predict moviesuccess at an early stage of movie production and performs movie Target audience prediction. The CB movie recommendation system has been used for both of these works. Three important parts of our research work can also be integrated. A movie recommendation system is the first module of our framework, as described below. We have considered only five essential features of the movie like genre, cast, director, keywords and movie description as the feature sets to build the recommendation system. At a very early stage in production, all of these basic features are available. A set of comparable films is included in the proposed recommendation system based on a forthcoming film. The second module only accepts similar movies from the first module, uses movie rating and voting information of similar movies, and creates a novel feature set. Next, we have proposed a RF model and use the newly created data set to predict the movie's popularity. We have divided the popularity of a movie into six classes super-duper hit (SDH), super hit (SH), hit (H), above average (AA), average (A) and flop (F). We'll be building a module for estimation of target audience at the third and final stage. We've divided the audience into four age groups: junior, teenage, middle aged and senior. In order to take into account, the movie rating and voting information, we used the same movie set from the first module and created a new set of features for each age group. Using the new data set,we built a model using fuzzy c means and

cosine similarity to estimate the popularity of the upcoming movie among allage groups.

This study has made the following main contributions:

1. Using a Recommendation System to predict the forthcoming popularity of films before their initial production phase is among the most significant research activities conducted in recent studies.

2. Proposed a model based on the fuzzy c mean and cosine distance to estimate the popularity of an upcoming film among different age groups.

Hence in this work, contented based recommended system movie popularity is presented. The rest of the work is arranged as follows: The section II describes the literature survey. The section III presents deep learning based content-based recommender system. The section IV evaluates the result analysis. Finally, the conclusion is provided in section V.

II.LITERATURE REVIEW

[1] 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.

A recommender framework may be a set of instruments for data recovery. By taking into consideration the explicit and implicit preferences and behavior of users, this improves their ability to access information and recommends products and services which are consistent with user tastes. The e commerce field has become a very popular area for recommendation systems. Today, there is a great deal of diversity in information available on the internet and it's difficult for end users to find out what they need. For users who are continuously adjusted to their changing taste, recommender systems provide a specific view of the user. Despite the fact that a large number of recommendation methods have been developed in other areas, recommender systems continue to encounter problems and difficulties which limit their accuracy. It summarizes the main challenges and problems that need to be solved in implementing recommender schemes, summarizing recent research achievements and recommendations on how to address them.. Moreover, we present an evaluation method used to assess the performance of recommender systems that goes beyond this. [2] N.

Das, S. Borra, N. Dey, and S. Borah, ``Social networking in web-based movie recommendation system,'' in *Social Networks Science: Design, Implementation, Security, and Challenges*. Cham, Switzerland: Springer, 2018, pp. 25_45.

Most online stores today use movie recommendation systems as a standard practice. Based on their search history and known preferences, movie recommendation systems make predictions as to the user's response. The selection of items is normally performed on the basis of their characteristics and content, or in collaboration with user ratings using smart algorithm features such as classification and grouping. It is essential to predict accurately what customers or users will be likely to encounter when I visit them, because it benefits both service providers and clients. The Chapter presents the evolution, basic concepts, classification, classic and novel models, requirements, similarity controls, assessment methods, challenges, impact of social networks on film recommendation systems as well as their future.

[3] P. Nagarnaik and A. Thomas, ``Survey on recommendation system methods,'' in *Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS)*, Feb. 2015, pp. 1603_1608.

The way users and websites communicate with each other has changed over the last years thanks to recommendation systems. To identify users' interests, the recommendation system sorts through a large number of data and makes it easier to find information. A number of methods have been applied for this purpose Collaborative filtering (CF) is a way to automatically predict customers' interest based on the collection of data from other customers, and many collaborative base algorithms are used for this purpose. Among the many patterns of pattern detection that are capable of handling large data sets, CHARM algorithm is one not supported by any prior association mining algorithms which do not have a large dataset. The paper covers a range of techniques in recommendation systems and offers a new system for delivering effective recommendations on web pages, such as combining the collaboration filtering method with CHARM algorithms that combine pattern discovery methods like clustering or association rule mining.

III.METHODOLOGY

This research study aimed to develop a model that will predict movie popularity and its age-wise preference using movie recommendations. Our objective is to classify the movie popularity among the six classes {SDH, SH, H, AA, A, F} at theearly movie production stage. Next, our objective is to find the movie's target audience and determine its influence on audience groups. Regroup the audience into four age demography (Junior, Teenage, Mid-Age, Senior). Our final output of the system will be age-wise movie popularity prediction.

In this study, we used a content-based movie recommendation system to find out a similar movie. In the next step, we use the voting information and rating of each recommended movies. All these data are used to train the 1-D RF deep learning model. The output of the RF model is the classification of the film among six classes. We predict the scale of popularity of the movie.
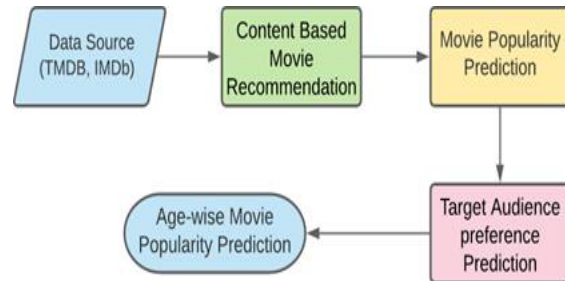

Fig 1. Framework workflow.

Our third module of the system takes recommended movie information and age-wise voting information. We have grouped the voting information into four age groups. We useFuzzy C-Mean to calculate the movie preference for each age group.

The framework of our job has three significant steps, which are listed below Fig. 1

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

2. Utilize comparative motion picture data and voting information from the IMDb beta set.

3.Compute target audience prediction using fuzzy c means.

4.We have introduced a new data set containing voting and rating information of the recommended movies, used topredict the movie popularity class.

5.We have proposed new parameters called Global centroid for each age group.

6.We have also presented a new approach for estimating the interest or popularity of and upcoming movie amongdistinct age groups.

The proposed system mainly has three modules:
- DATASET DESCRIPTION
- DATA PROCESSING
- MOVIE POPULARITY LABELLING

The proposed system has three major interrelated modules. The first module is a content-based movie recommendation system model, which produces 10 most similar movies of all the movies listed in the data set. The second module isa movie hit prediction module. The output of the first moduleis the input of the second module. Next, the third module groups the audience according to their age and predict the most suitable target audience group for each movie. Finally, the whole system provides an age-wise movie popularity prediction of the upcoming movie. The overall process flowdiagram has shown in figure 2 which is also the complete system architecture.
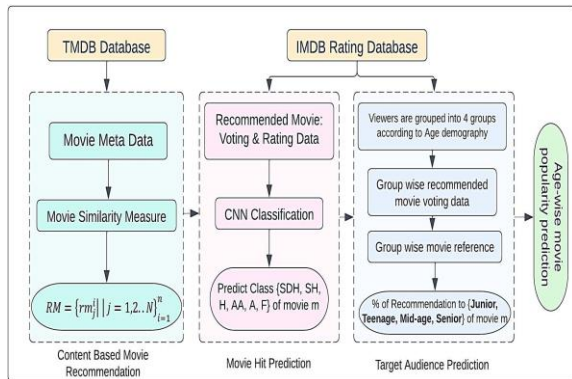
Fig 2: System Architecture

Content-based filter used for finding a similar movie. Whichuses movie attributes to find out the similarity between the two movies. The second module's purpose in this research problem is to build a movie hit prediction system. This module accepts the earlier module's output. The third and final module forecasts the target audience's preference according to the age of demography. To predict the upcoming movie's target audience, we have considered all the recommended movies delivered by our first module. In this paper, we are creating a content-based recommender system which takes the content as input and gives the output of popular movies that relates to the study and also predict the outcome of our content i.e. hit or flop. Using the Movie content, we predict the movie's success or degree of success in a multiclass way. This research helps us to predict or analyse the success rate of the upcoming movie with high accuracy.

IV.EXPERIMENTAL RESULTS

In this section of paper, we are going to implement the content-based recommender system based on deep learning. The experimental results and analysis are evaluated here.
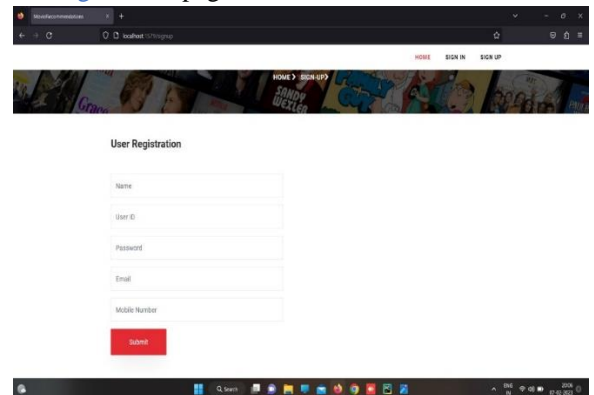

Fig 3: Web page of movie recommendations
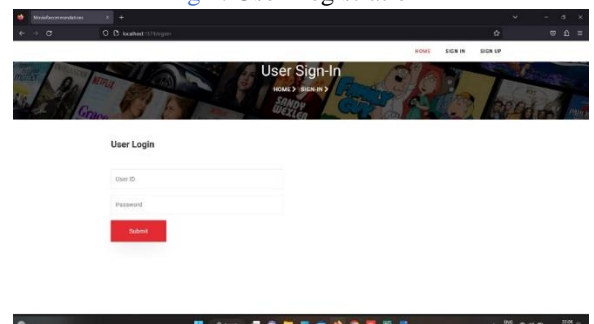

Fig 4: User Registration

Fig 5: User Login page
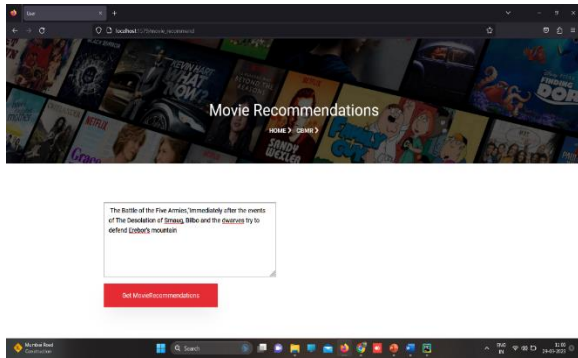

Fig 6: User homepage
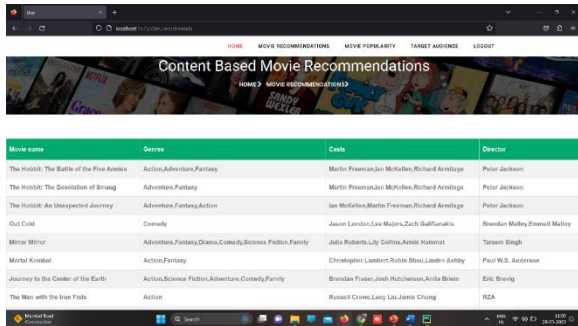
Fig 7: Data entering in the system

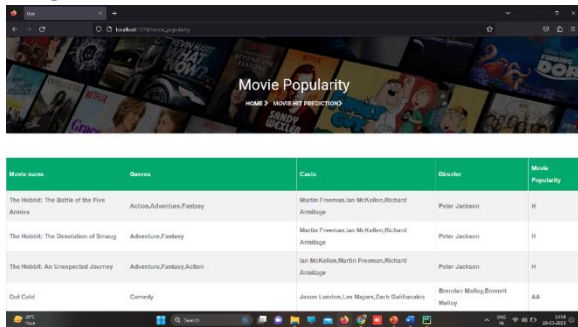Fig 8: Content based recommendation of movies
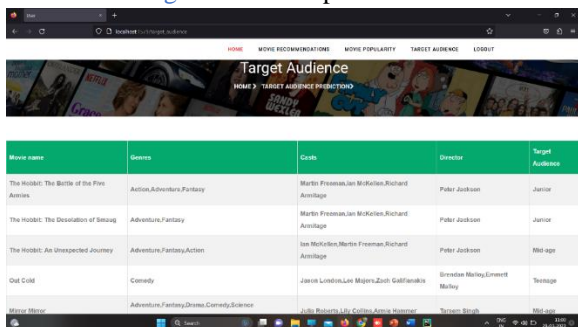

Fig 9: Movie hit prediction

Fig 10: Target audience prediction

V.CONCLUSION

Each film at the Box Office encompasses a huge sum of financing to spend. But the larger part of films do not turn out to be fruitful. Prior, the foremost critical number of works have been done on post-production

or post-release figure. The appraise does not impact as the financial specialist has as of now devoured their stores on the film generation. The pre-production or early master- duction arrange estimate needs tall precision and the most excellent time to guarantee speculation. The objective of our study is to propose an master framework that might offer assistance the motion picture producer execute essential changes on the off chance that required at the suitable time. Our framework can nourishment fetched the level of ubiquity of the up and coming motion picture some time recently the generation has begun for the most punctual organize of the generation and with noteworthy precision. Around framework centered not as it were on the notoriety of the up and coming motion picture but moreover on the movie's notoriety among all age bunches. Motion picture Producer can gauge the target group of onlookers and evaluate how the diverse gathering of people bunches would react to the up and coming motion picture. Encourage, our target is to construct a strong framework appropriate to all motion picture businesses. We have utilized the final hundred a long time (1915-2016) of motion picture information from TMDb and the IMDb database. Our approach to centered motion picture ubiquity and finding out the target gathering of people of an up and coming motion picture is exceptionally much one of a kind. In our approach, we have utilized a suggestion framework to discover comparative motion pictures from a given motion picture and utilize comparative motion pictures for determining purposes. Besides, it has been challenging to at the same time utilize to isolated the database (TMDb and IMDb). The measure of the TMDb information set is 4803, and the estimate of the IMDb rating information set is 85855. Since, we are utilizing both the dataset, the measure of the combined information set comes down to 23332 as it were. Within the future, interactive media data such as audio and video data can be utilized to attain superior comes about; besides, a blurb of an up and coming film can be utilized. With a estimation investigation on the Social Media information, it is conceivable to analyze past prepare tickets. It'll be valuable for moviemakers to know approximately later preparing that the film industry has experienced in relation to advertise desires. It would be conceivable to partition the group of onlookers by age, demography or calling. This way you'll be able target your following film and advance it much more effortlessly.

VI. FUTURE ENHANCEMENT

The proposed system is an fabulous instrument for the motion picture industry. In future work, interactive media information like sound and video information may well be consolidated additionally, and the blurb of the upcoming movie might be utilized for superior comes about. Later prepare tickets may be analyzed utilizing estimation examination of the social media information. Data with respect to later preparing on the showcase desire from the motion picture industry will be advantageous for the motion picture producers. The group of onlooker's bunches can be separated concurring to age and concurring to the demography or calling of the group of onlookers. That will be much simpler for focusing on and advancing an up and coming motion picture.

VII. ACKNOWLEDGEMENT

REFERENCE

[1] 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.

[2] N. Das, S. Borra, N. Dey, and S. Borah, ''Social networking in web based movie recommendation system,'' in *Social Networks Science: Design, Implementation, Security, and Challenges*. Cham, Switzerland: Springer, 2018, pp. 25–45.

[3] P. Nagarnaik and A. Thomas, ''Survey on recommendation system meth- ods,'' in *Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS)*, Feb. 2015, pp. 1603–1608.

[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] 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.

[6] Y. Koren and R. Bell, ''Advances in collaborative filtering,'' in *Rec- ommender Systems Handbook*. Boston, MA, USA: Springer, 2015, pp. 77–118.

[7] J. L. Herlocker, J. A. Konstan, and J. Riedl, ''Explaining collaborative fil- tering recommendations,'' in *Proc. ACM Conf. Comput. supported Coop- erat. Work (CSCW)*, 2000, pp. 241–250.

[8] M. J. Pazzani, ''A framework for collaborative, content-based and demo- graphic filtering,'' *Artif. Intell. Rev.*, vol. 13, no. 5, pp. 393–408, 1999.

[9] R. Van Meteren and M. Van Someren, ''Using content-based filtering for recommendation,'' in *Proc. Mach. Learn. Inf. Age, MLnet/ECML2000 Workshop*, vol. 30, 2000, pp. 47–56.

[10] P. B. Thorat, R. M. Goudar, and S. Barve, ''Survey on collaborative filtering, content-based filtering and hybrid recommendation system,'' *Int. J. Comput. Appl.*, vol. 110, no. 4, pp. 31–36, Jan. 2015.