Movie recommendation using Sentiment Analysis and Ajax

[1] Aarushi Sharma, [2] Kandyala Lekhana, [3] Keerthana M, [4] Keshav Anil Rathi,
[5] Dr. Sumithra Devi K A
[1][2][3][4]BE Students, Department of Information Science and Engineering
[5] Dean Academics & Head, Department of Information Science and Engineering
[1][2][3][4][5] Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India

Abstract

Movie proposal is an important aspect in Web for the users. The whole idea revolves around the accumulation of client's inclinations, surveys, and feelings to help them find appropriate motion pictures advantageously. These recommendation systems help users with the recommended movies based on their choice, thus increase their option for selection of a movie. This can be achieved using Sentiment Analysis which helps to analyze the sentiments of the user based on his/her choice. In these recent years, Sentiment Analysis has become the art of models for many of the recommendation systems, the conclusion from such models is more promising and represent more precise prediction with its continuous advancement.

Keywords**:** Recommendation System, Collaborative filtering, Sentiment Analysis, Content based filtering

INTRODUCTION

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 large list of applications for recommendation programs is available provided to the user. Recommendations for recommendations
programs have increased slightly and more recently installed on almost every online platform used by people. The content of this program is different from movies, podcasts, books and videos, colleagues and news on social media, in goods on commerce websites, to people in commercial and dating websites. Usually, these programs exist is able to retrieve and filter data about user preferences, and can use this Intel to improve their suggestions in the next season. For example, Twitter can analyze your
in conjunction with several issues on your wall to understand what kinds of stories interest you. Many times,these programs can be improved on the basis of a large number of people. For example, if Flipkart notifications that a large number of users are buying a modern laptop too buy a laptop bag. They can recommend a new laptop bag a customer who has just put a laptop in his cart. Due to-advances in recommendation programs, users continuously expect positive results. They have a low service level that
they cannot make appropriate recommendations. If the music the streaming app cannot preview and play the song if the user likes, the user will just stop using it[1] This has led to great value by technology companies in the redesign of their recommendations. However, the problem is far more complex than it seems.Every user has different likes and dislikes. Moreover, I say the taste of a single customer can vary depending on the size a number of factors, such as emotions, season, or type of activity
the user does. For example, the type of music one would like to listen to during a workout varies greatly from person to person the kind of music he could listen to while preparing dinner. They should find new places to decide more about customer, while determining almost everything already known about the customer. Two in scrutiny important methods are widely used to recommend programs[2]. Another content-based filter, where we try shape user preferences using retrieved data, too suggest items based on that profile. One is collaborative filtering, where we try to integrate users equally together and use data about the group to make customer recommendations.

PROBLEM STATEMENT

The current Movie Recommendation model is not fully capable of delivering the estimated number of movies as expected. Another emerging problem is the factors considered in the predictive model. Few things can be described as actor names, genres, movie title, directors. These factors lead to successful movie predictions and provide more accurate results from the model.

One of the biggest problems facing recommendation programs is that they need a lot of data to successfully make recommendations[3]. A good recommendation system first requires object data (from catalog or other method), in which case it should take and analyze user data (behavioral events), and then the magic algorithm using Sentiment Analysis does its job. Most user data is a recommendation system to work with, which strengthens the chances of getting good recommendations. But it can be a problem - to get good recommendations, you need more users, to get more data for recommendations.

Another major reason for failure is not enough to change user preferences[4]. The problem here is that while today the user may have some purpose when browsing. Additionally, if you are given a set of users with their previous movie rating ratings, can we predict the rating they will respond to a movie they have not rated before. To resolve this database found in Kaggle it needs to be updated regularly.

PROPOSED METHODOLOGY

Over the past decade, many site integration programs have been created and are used for example Netflix, Amazon, and Google. These suggestion systems use a combination of strategies such as a content-based approach, a collaborative approach, a knowledge-based approach, a user-based approach, a hybrid approach, etc.

This paper uses content-based filters that 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, but users like we have, and we like. To determine whether two users are comparing or not, this program looks at the movies both viewers and how they evaluated them. To be effective, this type of recommendation requires standards, not all users calculating things continuously. Some of them are rare or even worthless! This leads to a cold launch problem: If there was another movie, no one else would like it or watch it, so you won't have this on your recommended list, even if you like it. Another feature of this process is the variation in recommendations, which may be lucky or unfortunate, depending on the case.

On the other hand, content-based filters have proven to be effective, this type of filter does not involve different users but ourselves: It has no problem with cold startup because it uses attributes, such as characters, director name, genre, movie length, so our favorite movies are recommended immediately. By looking at what we like, the algorithm will simply select items similar to what we already like to recommend. In this case, there will be little variation in the recommendations, but this will apply whether user prices or not.[5]

This paper uses content-based filters that 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, but users like we have, and we like.

METHODOLOGY

1. Content based RS using Sentiment Analysis

A content-based recommender works with data that the user provides, either explicitly (rating) or implicitly (clicking on a link). Based on that data, a user profile is generated, which is then used to make suggestions to the user by analyzing the sentiments on the reviews given by the user for that movie. Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative or neutral. Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc), urgency (urgent, not urgent) and even intentions (interested v. not interested). With the help of Sentiment analysis, the movie with most positive feedback will be suggested first.[6]

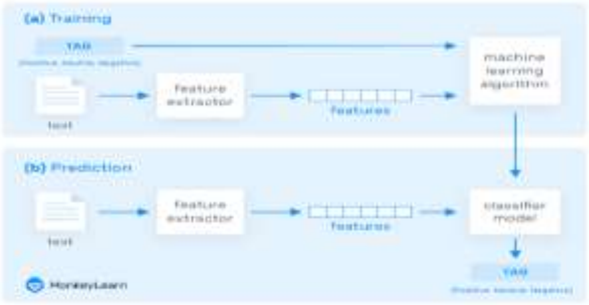


Fig1: Workflow in Sentiment Analysis

The project can be composed of the follows:

1.1 Data Collection

Movie tricks (title, category, duration, rating, banner, etc.) are extracted using the API via TMDB. Kaggle is the world's largest data science community with powerful tools that allow users to discover and publish data sets, test and create models in a web-based data-science credit Credit.csv from Kaggle reveals details about the response to various movies and character reviews.[2]

1.2 Data Pre-processing

Collected data must be labeled or tested with a feature extractor. Certain data features are selected in the Data Frame. Movie metadata contains information such as Movie title, director name, actor name etc on line 5043 and 28 columns. Credit.csv file downloaded from Movie Database, Kaggle contains 45476 rows and 8 columns.[7]

1.3 Web scraping from Wikipedia using Python

Web Removal is a method of copying where certain data is collected and copied on the web, usually uploaded to a local site or spreadsheet, for later retrieval or analysis. Removing a web from a web page involves downloading and uninstalling it. The unavailability of data on Kaggle for 2018 and 2019 creates a reason to remove the features of movies and their characters of this age on Wikipedia.

1.4 Data Processing

After generating the datasets, we prepare the datasets using the “pandas” library in python to create different data frames. The datasets will be divided into different groups and frames that is required for the next step, that is, transferring the data to the frontend using Flask(A python web framework).[5]

The dataset will be combined as a multiple list as a dictionary which can be passed to the html file so that it can be processed easily and the order of information will be preserved.

KEY:VALUE pair where KEY=[Movie, Cast, Cast_Details]

1.5 Sentiment Analysis

Indigenous language processing is used to process and process text data and text voting. We use Tfidfvectorizer to convert text into price representations used by separator and fit_transform to convert all available information into vector.[3]

To use this Multinominal NB algorithm is used which is usually the first solution for the emotional analysis task. The basic idea of the Naive Bayes process is to find opportunities for literary classes through opportunities for word and paragraph interaction.

A similar score is used to determine which object is most similar to the user's preference. Basically, it is a set of values between zero to one that helps determine how much of two things are similar to each other on a scale of zero to another. And this is found using the similarity of Cosine.

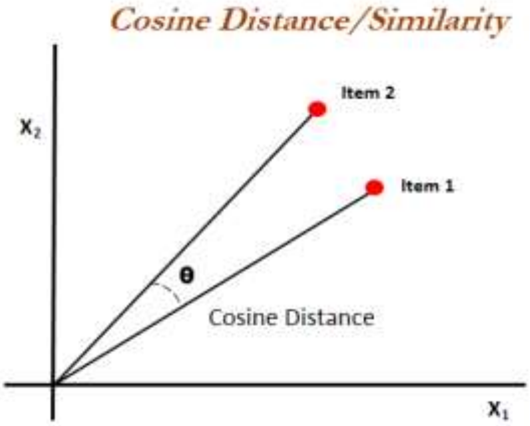


Fig 2: Cosine Similarity

1.6 Showing the Data on the Client Side using AJAX
Ajax is a collection of web development strategies using many web technologies on the customer side to create smaller web applications. With Ajax, web applications can send and retrieve data from a server in sync. Using Ajax we can display python Data on the client side in the form of Request and Feedback.[7]

- / The request method will send the client a search in the form of a query

RESULT

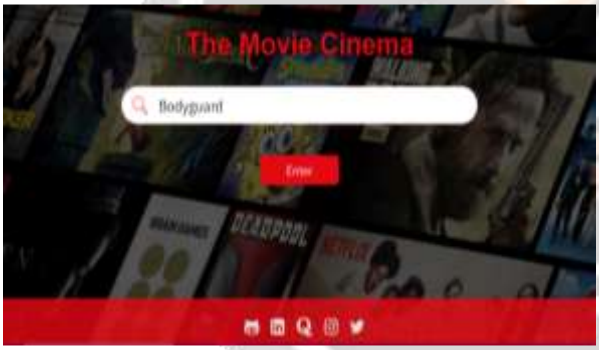
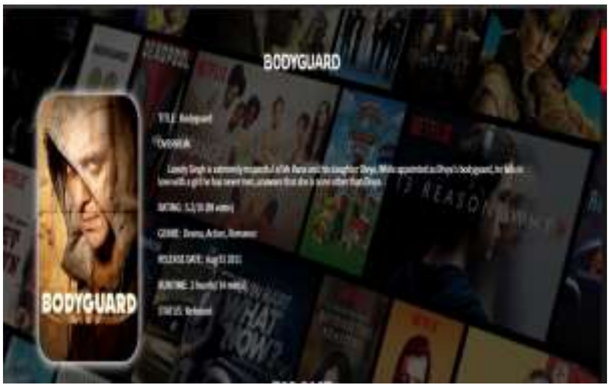


Fig 3: Landing Page

This is the landing page of our project, the input field is a form that takes a string of movie and /post to the back-end for the pre-processing from the database and the results are reverted back to the client side using AJAX.



BODYGUARD

TITLE: Bodyguard

Overview:

[illegible]

RATING: [illegible]

GENRE: Drama, Action, Romance

RELEASE DATE: [illegible]

RUNTIME: [illegible]

STATUS: Released

Fig 4: Search Result

The output generated by the Search engine will be the movie name entered by the user. This is the /post result of our project, where in the string gets processed at the back-end and the details of movie along with genre, top cast, name, year of realized etc. are displayed at the client side.

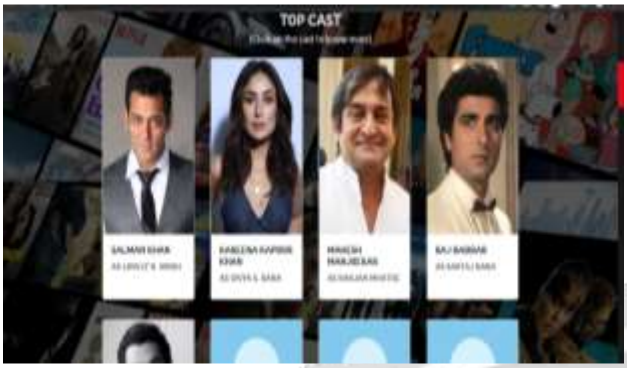


Fig 5: Cast Details

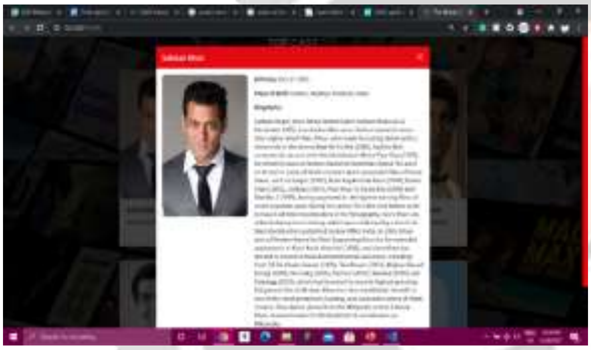


Fig 6: Scrapped details from Wikipedia

Using Web scrapping the details of the cast is displayed in the cast section. This is the end result of Web scraping part where in the data of each actors worked in the movie will be fetched from Wikipedia using web scraping

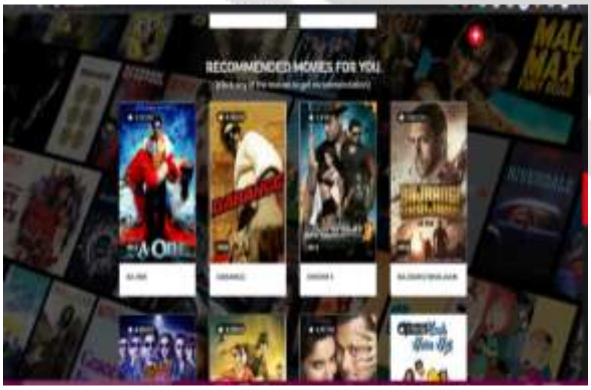
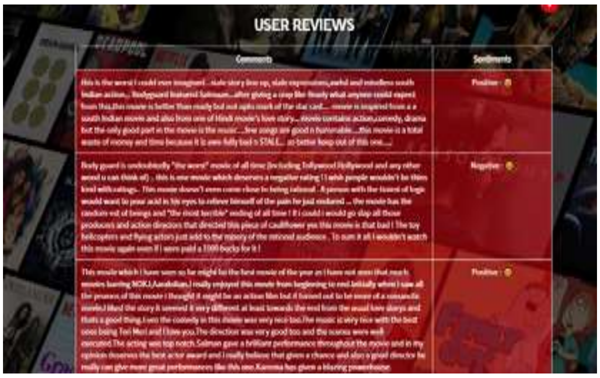


Fig 7: Recommended Movies

Using Sentiment Analysis, the movies are recommended either on the type of genre, same cast, directors etc. The movies with the highest rating achieved using Content-based filtering are displayed. This is the end result of content-based filtering



USER REVIEWS

Comments	Sentiments
this is the worst i could ever imagined ...stale story line up, stale expressions,awful and mindless south Indian action... Bodyguard featured Salmaan...after giving a crap like Ready what anyone could expect from this,this movie is better than ready but not upto mark of the star card.... movie is inspired from a a south Indian movie and also from one of Hindi movie's love story... movie contains action,comedy, drama but the only good part in the movie is the music....few songs are good n hummable....this movie is a total waste of money and time because it is awe fully bad n STALE... so better keep out of this one....!	Positive :
Body guard is undoubtedly "the worst" movie of all time (including Tollywood Hollywood and any other wood u can think of) .. this is one movie which deserves a negative rating ! I wish people wouldn't be this kind with ratings.. This movie doesn't even come close to being rational . A person with the least of logic would want to pour acid in his eyes to relieve himself of the pain he just endured ... the movie has the [illegible] of beings and "the most terrible" ending of all time ! If i could i would go slap all those producers and action directors that directed this piece of cauliflower yes this movie is that bad ! The toy helicopters and flying actors just add to the misery of the retarded audience . To sum it all i wouldn't watch this movie again even if i were paid a 1000 bucks for it !	Negative :
This movie which I have seen so far might be the best movie of the year as I have not seen that much movies barring MOKJ,Aandolan,I really enjoyed this movie from beginning to end.Initially when I saw all the promos of this movie i thought it might be an action film but it turned out to be more of a romantic comedy.I liked the story it seemed it very different at least towards the end from the usual love storys and thats a good thing.I see the comedy in this movie was very nice too.The music is very nice with the best ones being Teri Meri and I love you.The direction was very good too and the scenes were well executed.The acting was top notch.Salman gave a brilliant performance throughout the movie and in my opinion deserves the best actor award and I really believe that given a chance and also a good director he really can give more great performances like this one.Kareena has given a blazing powerhouse	Positive :

Fig 8: Sentiment Analysis Part

This is the sentiment analysis part, where the users reviews are displayed and those reviews are categorized into positive or negative using sentiment analysis.

CONCLUSION

In this paper, we provide a recommendation program based on emotional analysis and TMDB data set out in this study. Metadata and social networking are used as important elements in making movie recommendations. Emotional analysis has the advantage of extracting information about how the audience reacts to a particular film, which can be used to make future suggestions. In the future, we would like to improve the suggestion process by extracting relevant data from the user's response, preferences, and so on. With the new technology, there will be enough resources to achieve this and all the information will be integrated into the most powerful RS. In our experiment, we used a static database that only looks at movies released until 2017. The building has the potential to be investigated in a dynamic environment, with new films being added regularly.

The accuracy of the project is 98% approx. 2% accuracy is affected because data is not available from the TMDB update database as the project depends on the updated database, it is difficult to get all the details of all updated movies from the database.

REFERENCES

1. A. Rashid, I. Albert, D. Cosley, S. Lam, S. McNee, J. Konstan, and J. Riedl (2002) Learning new user preferences in recommender systems: Getting to know you. The proceedings of the international conference on intelligent user interfaces are published in this book (21) Yu, K., Tresp, V., and Yu, S. Information filtering using a nonparametric hierarchical Bayesian framework. Pages 353–360 in SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and Development in Information Retrieval. The ACM Press published this book in 2004. (25)

2. P. Zigoris and Y. Zhang are the authors of this paper. User profiling using Bayesian adaptive methods with explicit and implicit input. Presented at the 2006 Conference on Information and Knowledge Management (27)

3. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? : sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—Volume 10 (EMNLP '02), pp. 79–86, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA, July 2002.

4. “Thumbs up or thumbs down?” by P. D. Turney, in Proceedings of the Association for Computational Linguistics' 40th Annual Meeting, pp. 417–424, Philadelphia, Pennsylvania, July 2002.

5. J. R. Priester and R. E. Petty, “The Gradual Threshold Model of Ambivalence: Relating the Positive and Negative Bases of Attitudes to Subjective Ambivalence,” Journal of Personality and Social Psychology, vol. 71, no. 3, pp. 431–449, 1996.

6. “An intelligent movie recommendation system through group-level sentiment analysis in microblogs,” Neurocomputing, vol. 210, pp. 164–173, 2016. H. Li, J. Cui, B. Shen, and J. Ma, “An intelligent movie

recommendation system through grouplevel sentiment analysis in microblogs," Neurocomputing, vol. 210, pp. 164–173, 2016

7. Clayton J Hutto and Eric Gilbert. Vader: A simple rule-based model for analyzing sentiment in social media text. The Eighth International Conference on Weblogs and Social Media, held in 2014, is a collection of papers presented at the conference.