ENTERTAINMENT CONTENT RECOMMENDATION SYSTEM USING MACHINE LEARNING

Assistant Professor: Ms. Karuna Middha[1], Student: Munzir[2], Sarthak Goja[3], Sourabh Choudhary[4]

[2,3,4]UG Student, Dept. of CSE Engineering, Maharaja Agrasen Institute of Technology, Delhi, India.

***

Abstract – Entertainment Content recommendation systems are important for several reasons.

First, they can help users discover new movies that they may not have otherwise found. With so many movies available to watch, it can be overwhelming for users to sift through all of the options and find something that they will enjoy. A recommendation system can narrow down the choices and present the user with a curated selection of movies that are tailored to their personal preferences.

Finally, recommendation systems can improve the efficiency of finding movies to watch. Instead of spending alot of time searching through various movie platforms or scrolling through long lists of movies, a recommendation system can present the user with a short list of options thatare likely to be relevant and interesting to them.

Keywords: content-based approach; sentimental analysis; recommendation system; movie ratings; inductive learning

1. INTRODUCTION

In recent years, there has been a proliferation of movie streaming platforms and other online sources of video content. As a result, consumers have access to a vast and ever-growing selection of movies and TV shows to watch. However, with so much choice, it can be challenging for users to find movies that they will enjoy and that fit their personal preferences.

There has been significant research on the development and evaluation of movie recommendation systems in recentyears. Researchers have explored various methods for collecting and analyzing user data, as well as different approaches for making recommendations. In this paper, we will review the state of the art in movie recommendation systems and discuss the challenges and opportunities that exist in this field. Despite the progress that has been made in this field, there are still many challenges and opportunities for future research. For example, one key challenge is to effectively handle the "cold start" problem, where the system must make recommendations for a new user with little or no previous data. Another challenge is to improve the diversity and novelty of recommendations, while still ensuring that they are relevant and personalized. Finally, researchers are also exploring ways to incorporate additional sources of data, such as social media, to enhance the performance of movie recommendation systems.

2. RELATED WORK

There are two main types of recommender systems: collaborative filtering and content-based filtering. Collaborative filtering relies on user-related information, preferences, and interactions to identify similarities between users and recommend movies that similar users have enjoyed. There are two subtypes of collaborative filtering: model-based and memory-based algorithms. Memory-based methods do not have a training phase and use measures like Pearson correlation coefficient and Cosine similarity to identify similar users. Model-based methods, on the other hand, try to predict user ratings of a movie using estimated models. Collaborative filtering methods can be computationally intensive and may not perform well with sparse data. They also assume that users with similar tastes will rate movies similarly, which may not always be the case. Content-based methods, on the other hand, use information about the content of movies, such as audio and visual features or textual metadata, to find similarities among movies and recommend those that are similar to ones that the user has accessed or searched for. These methods do not incorporate user behavior in their recommendations. In this paper, we will be exploring the latter approach.

3. LITERATURE REVIEW

Nessel stated in the movie oracle that working with examples is an essential part of human interaction and triedto provide a movie recommendation engine based on this behavior. Which of course requires considerably more computing power, as the compared bodies of text are much larger, but the algorithms are essentially the same [3].In a content-based movie recommendation system, the proposed algorithm uses textual metadata of the movies like plot, cast, genre, release year and other production information to analyze them and recommend the most similar ones [2].The paper also analyzes application similarity measure for recommendations forecasting in recommendations systems. It is shown that used method for computing similarity measure in recommendations systems are cosine similarity measure and Pearson correlation coefficient [1]. As the characteristics of movie recommendation go by, the user watching history is very important, so we add content-based recommendation approach. Typically, people have a tendency to think that positive reviews have a positive effect and negative reviews have negative impact. Sentiment analysis will assist us to improve the accuracy of recommendation results. Also, as we explained in our experimental results, it is necessary to make use of distributed system to solve the scalability and timeliness of recommender system [5].

4. TECHNIQUES USED IN METHODOLOGY

The proposed solution aims to enhance the scalability and effectiveness of the movie recommendation system. To efficiently and quickly compute the similarity between movies in the dataset and reduce the computation time of the movie recommendation engine, we employed the cosinesimilarity measure. To determine whether a review is positive or negative, we utilized the Naive Bayes classifier for sentiment analysis.

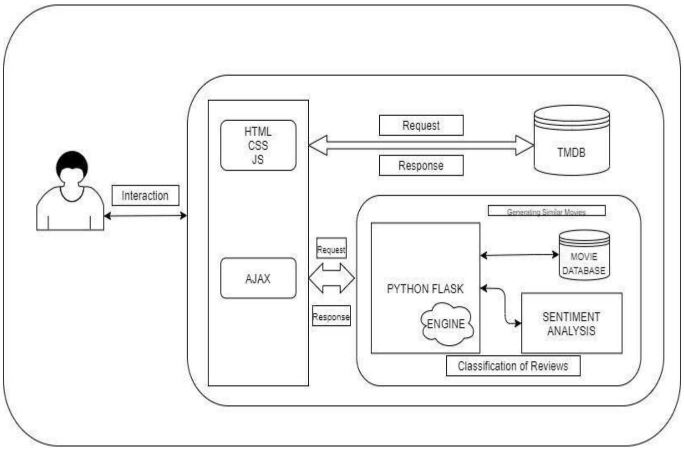


Fig -1: Architecture of the Movie Recommendation System

A. Content-based Filtering

A content-based movie recommendation system utilizes user-provided data, such as ratings, feedback, and reviews, to generate a user profile, which is then used to make recommendations. As the user interacts more with the recommendation system, the engine becomes more precise and reliable. The Term Frequency (TF) and Inverse Document Frequency (IDF) techniques are utilized to retrieve and analyze information, such as movie and article titles. These techniques are used to assess the relative importance of different terms.

To implement a content- based filtering system, the following steps are typically followed:

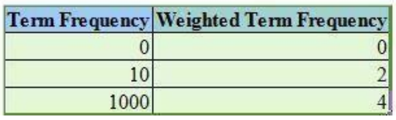
- Terms Representation
- Terms Allocation
- Learning Algorithm Selection
- Provide Recommendations

B. Term Frequency (TF) and Inverse Document Frequency (IDF)

TF, or term frequency, refers to the number of times a specific word appears in a document. IDF, or inverse document frequency, is the reciprocal of the document frequency in a collection of documents. Together, TF-IDF, or term frequency-inverse document frequency, is a statistical measure that assesses the relevance of a word to a document within a larger collection of documents. It is often used in natural language processing (NLP) and machine learning algorithms for text analysis and to score words. In other words, the weight of a word in a document cannot be accurately determined by simply counting its raw frequency, and thus the following equation is used:

Equation:

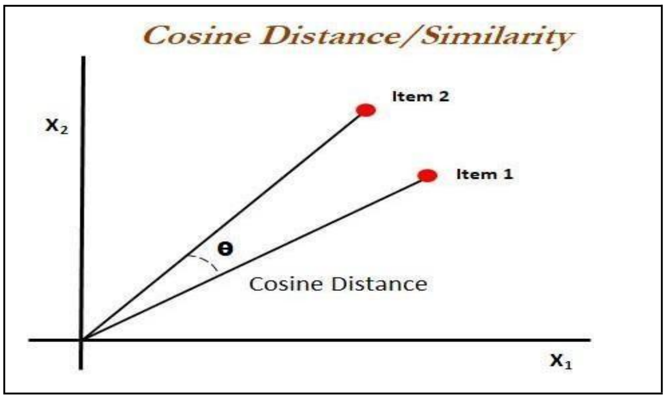
$$w_{t,d} = \begin{cases} 1 + \log_{10} \mathrm{tf}_{t,d}, & \text{if } \mathrm{tf}_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases}$$



Term Frequency	Weighted Term Frequency
0	0
10	2
1000	4

C. Cosine Similarity

The similarity score is a numeric value that ranges from zero to one and is used to determine how similar two items are to each other on a scale from zero to one. This score is obtained by comparing the texts of the two documents and measuring their similarity. The similarity score can therefore be defined as a measure of the similarity between the textual details of two given items. This can be calculated using the cosine similarity measure, which determines the similarity of texts regardless of their size. Cosine similarity is a measure used to calculate the cosine of the angle between two vectors projected in a multi-dimensional space. It is commonly used to determine the similarity between texts.



$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

D. Sentimental analysis

Sentiment analysis is a field within natural language processing that involves the evaluation of subjective opinions, views, or feelings about a particular subject that have been collected from various sources. In more practical business terms, sentiment analysis can be described as a set of tools used to identify and extract opinions and utilize them for the benefit of business operations. Such algorithms delve into the text to identify the underlying attitude towards a subject or its specific elements. An example of a commonly used algorithm in sentiment analysis is the multinomial naive Bayes classifier. This algorithm assumes that the features being analyzed are produced from a simple multinomial distribution. The multinomial distribution defines the probability of observing counts within a number of categories, making it well-suited for features that represent counts or count rates. The basic idea is the same as before, except that instead of modeling the data distribution with a best-fit Gaussian curve, we model it with a best-fit multinomial distribution.

P (positive | overall liked the movie) = P (overall liked the movie | positive) * P (positive) / P (overall liked the movie)

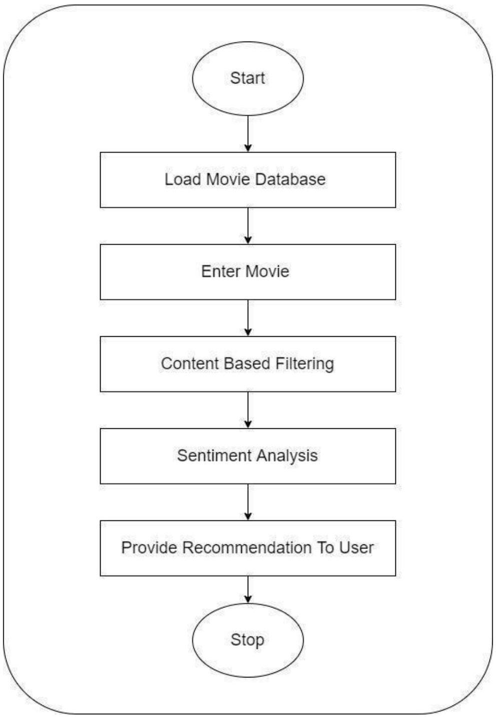


Fig -2: Flowchart of the Movie recommendation system

5. PROPOSED SYSTEM

A. Dataset

For our research, we utilized three different datasets from MovieLens, a collection of datasets generated by the GroupLens Research team for the purpose of evaluating recommender systems. These datasets are commonly used by developers to test their recommendation systems. These are:

1. IMDB 5000 Movie Dataset
2. The Movies Dataset
3. List of movies in 2018 4. List of movies in 2019
5. List of movies in 2020

The Movies dataset includes metadata for all 45,000 films listed in the Full MovieLens dataset, which includes movies released up until July 2017. The data includes information about the cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts and vote averages. This dataset also includes files with 26 million ratings from 270,000 users for 45,000 movies, with ratings ranging from 1 to 5, obtained from the official GroupLens website. The datasets for movies released from 2018 to 2020 were obtained by web scraping their respective Wikipedia pages.

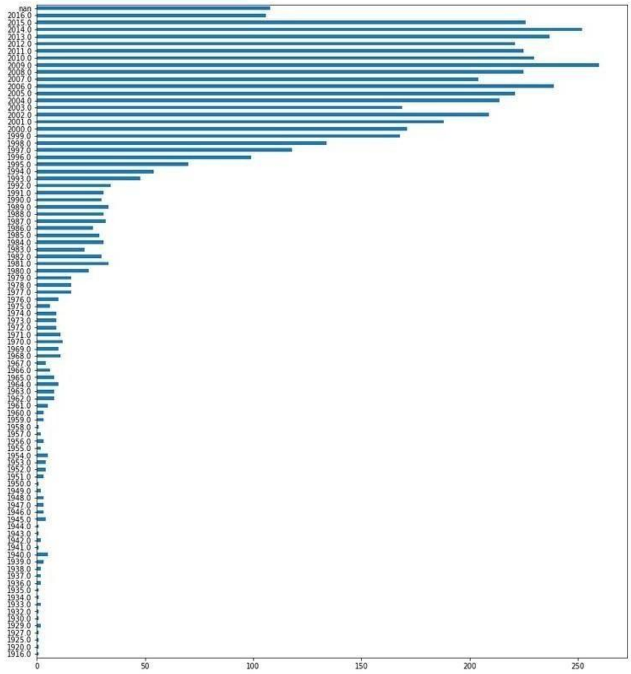


Fig -3: Plotted graph of The Movies Dataset

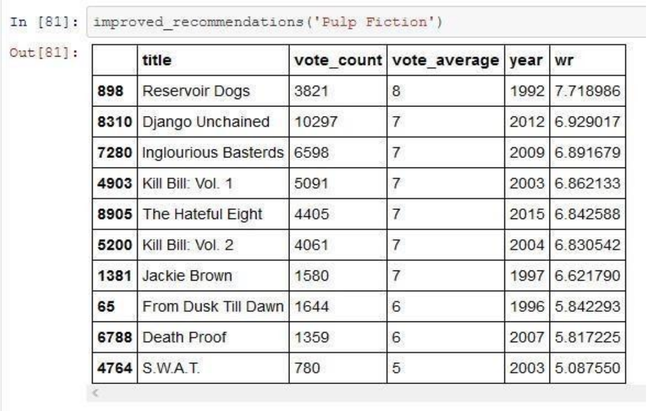
B. Recommendation system quality measures

We have used the TMDB Ratings to come up with our Top Movies Chart. And also IMDB's weighted rating formula to construct the chart.
Mathematically, it is represented as follows:

$$Weighted\ Rating(WR) = (\frac{v}{v+m}.R) + (\frac{m}{v+m}.C)$$

Where,

v represents the number of votes for the movie
m represents the minimum votes required to be listed in the chart
R represents the average rating of the movie
C represents the mean vote across the whole repot



```
In [81]: improved_recommendations('Pulp Fiction')
Out[81]:
```

	title	vote_count	vote_average	year	wr
898	Reservoir Dogs	3821	8	1992	7.718986
8310	Django Unchained	10297	7	2012	6.929017
7280	Inglourious Basterds	6598	7	2009	6.891679
4903	Kill Bill: Vol. 1	5091	7	2003	6.862133
8905	The Hateful Eight	4405	7	2015	6.842588
5200	Kill Bill: Vol. 2	4061	7	2004	6.830542
1381	Jackie Brown	1580	7	1997	6.621790
65	From Dusk Till Dawn	1644	6	1996	5.842293
6788	Death Proof	1359	6	2007	5.817225
4764	S.W.A.T.	780	5	2003	5.087550

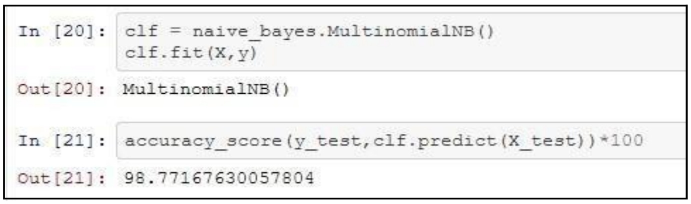
Fig -4: Calculated weighted rating for the dataset.

6. RESULT ANALYSIS

A. Accuracy of Sentimental Analysis Model

The multinomial Naive Bayes algorithm is well-suited for classifying items with discrete features (e.g. word frequencies for text classification).

Accuracy of 98.77% is observed for the dataset provided.



```
In [20]: clf = naive_bayes.MultinomialNB()
         clf.fit(X,y)

Out[20]: MultinomialNB()

In [21]: accuracy_score(y_test,clf.predict(X_test))*100

Out[21]: 98.77167630057804
```

Fig -5: Observed accuracy of sentimental analysis.

B. Results of content-based movie recommendation system

To determine the weighted rating of each film, we will select the 25 most similar movies according to their similarity scores. We will then calculate the vote of the movie that falls at the 60th

percentile of this group. Finally, we will apply IMDB's formula to compute the weighted rating of each movie using this value.

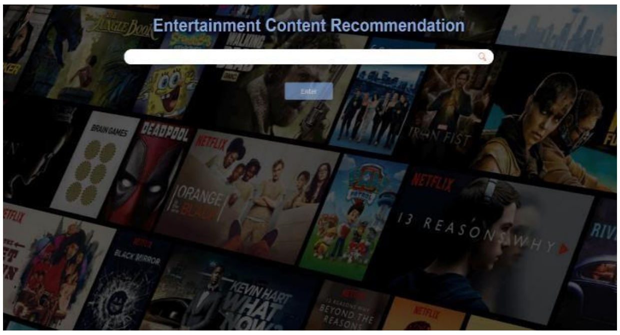


Fig -6: Home page of the Entertainment recommendation system

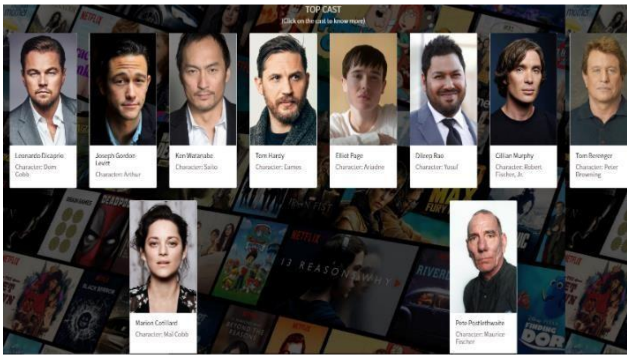
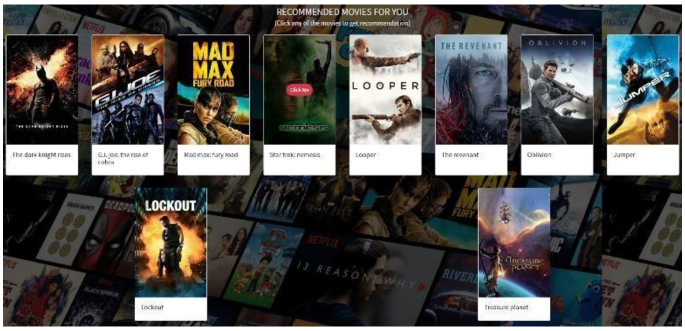
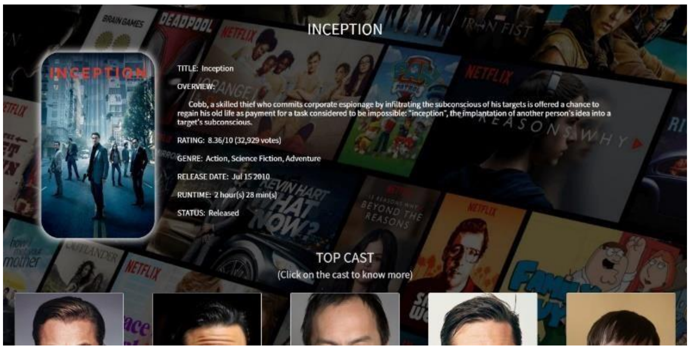


Fig -7: Cast of the Movie



RECOMMENDED MOVIES FOR YOU
(Click any of the movies to get recommendation)

Fig.8 : Recommendation Entertainment Content



INCEPTION

TITLE: Inception

OVERVIEW:

Cobb, a skilled thief who commits corporate espionage by infiltrating the subconscious of his targets is offered a chance to regain his old life as payment for a task considered to be impossible: "inception", the implantation of another person's idea into a target's subconscious.

RATING: 8.36/10 (32,929 votes)

GENRE: Action, Science Fiction, Adventure

RELEASE DATE: Jul 15 2010

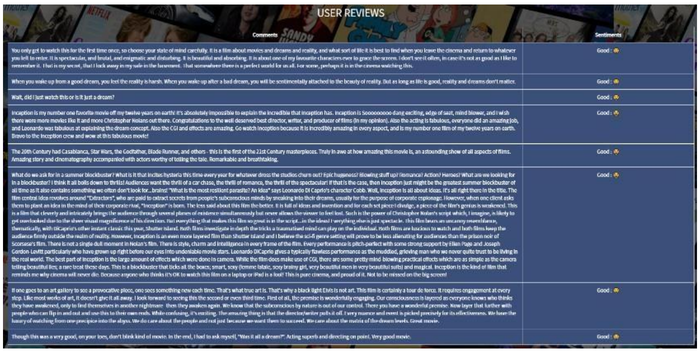
RUNTIME: 2 hour(s) 28 min(s)

STATUS: Released

TOP CAST

(Click on the cast to know more)

Fig -9: Poster and info page of the Movie



USER REVIEWS

Comments	Sentiments
You only get to watch this for the first time once, so choose your state of mind carefully. It is a film about movies and dreams and reality, and what sort of life it is best to find when you leave the cinema and return to whatever you left to enter. It is spectacular, and brutal, and enigmatic and disturbing. It is beautiful and absorbing. It is about one of my favourite characters ever to grace the screen. I don't see it often, in case it's not as good as I like to remember it. That is my secret, that I lock away in my safe in the basement. That somewhere there is a perfect world for us all. For some, perhaps it is in the cinema watching this.	Good : 😀
When you wake up from a good dream, you feel the reality is harsh. When you wake up after a bad dream, you will be sentimentally attached to the beauty of reality. But as long as life is good, reality and dreams don't matter.	Good : 😀
Wait, did I just watch this or is it just a dream?	Good : 😀
Inception is my number one favorite movie off my twelve years on earth! It's absolutely impossible to explain the incredible that inception has. Inception is Sooooooooo dang exciting, edge of seat, mind blower, and I wish there were more movies like it and more Christopher Nolans out there. Congratulations to the well deserved best director, writer, and producer of films (in my opinion). Also the acting is fabulous, everyone did an amazing job, and Leonardo was fabulous at explaining the dream concept. Also the CGI and effects are amazing. Go watch inception because it is incredibly amazing in every aspect, and is my number one film of my twelve years on earth. Bravo to the inception crew and wow at this fabulous movie!	Good : 😀
The 20th Century had Casablanca, Star Wars, the Godfather, Blade Runner, and others - this is the first of the 21st Century masterpieces. Truly in awe at how amazing this movie is, an astounding show of all aspects of films. Amazing story and cinematography accompanied with actors worthy of telling the tale. Remarkable and breathtaking.	Good : 😀
What do we ask for in a summer blockbuster? What is it that incites hysteria this time every year for whatever dross the studios churn out? Epic hugeness? Blowing stuff up? Romance? Action? Heroes? What are we looking for in a blockbuster? I think it all boils down to thrills! Audiences want the thrill of a car chase, the thrill of romance, the thrill of the spectacular! If that is the case, then Inception just might be the greatest summer blockbuster of all time as it also contains something we often don't look for...brains! "What is the most resilient parasite? An idea" says Leonardo Di Caprio's character Cobb. Well, Inception is all about ideas. It's all right there in the title. The film central idea revolves around "Extractors", who are paid to extract secrets from people's subconscious minds by sneaking into their dreams, usually for the purpose of corporate espionage. However, when one client asks them to plant an idea in the mind of their corporate rival, "Inception" is born. The less said about this film the better. It is full of ideas and invention and for each set piece I divulge, a piece of the film's genius is weakened. This is a film that cleverly and intricately brings the audience through several planes of existence simultaneously but never allows the viewer to feel lost. Such is the power of Christopher Nolan's script which, I imagine, is likely to get overlooked due to the sheer visual magnificence of his direction. But everything that makes this film so great is in the script...in the ideas! Everything else is just spectacle. This film bears an uncanny resemblance, thematically, with DiCaprio's other instant classic this year, Shutter Island. Both films investigate in depth the tricks a traumatised mind can play on the individual. Both films are luscious to watch and both films keep the audience firmly outside the realm of reality. However, Inception is an even more layered film than Shutter Island and I believe the sci-fi genre setting will prove to be less alienating for audiences than the prison noir of Scorsese's film. There is not a single dull moment in Nolan's film. There is style, charm and intelligence in every frame of the film. Every performance is pitch-perfect with some strong support by Ellen Page and Joseph Gordon-Levitt particularly who have grown up right before our eyes into undeniable movie stars. Leonardo DiCaprio gives a typically flawless performance as the muddled, grieving man who we never quite trust to be living in the real world. The best part of Inception is the large amount of effects which were done in camera. While the film does make use of CGI, there are some pretty mind-blowing practical effects which are as simple as the camera telling beautiful lies; a rare treat these days. This is a blockbuster that ticks all the boxes; smart, sexy (femme fatale, sexy brainy girl, very beautiful men in very beautiful suits) and magical. Inception is the kind of film that reminds me why cinema will never die. Because anyone who thinks it's OK to watch this film on a laptop or iPad is a fool! This is pure cinema, and proud of it. Not to be missed on the big screen!	Good : 😀
If one goes to an art gallery to see a provocative piece, one sees something new each time. That's what true art is. That's why a black light Elvis is not art. This film is certainly a tour de force. It requires engagement at every step. Like most works of art, it doesn't give it all away. I look forward to seeing this the second or even third time. First of all, the premise is wonderfully engaging. Our consciousness is layered as everyone knows who thinks they have awakened, only to find themselves in another nightmare then they awaken again. We know that the subconscious by nature is out of our control. There you have a wonderful premise. Now layer that further with people who can flip in and out and use this to their own ends. While confusing, it's exciting. The amazing thing is that the director/writer pulls it off. Every nuance and event is picked precisely for its effectiveness. We have the luxury of watching from one precipice into the abyss. We do care about the people and not just because we want them to succeed. We care about the matrix of the dream levels. Great movie.	Good : 😀
Though this was a very good, on your toes, don't blink kind of movie. In the end, I had to ask myself, "Was it all a dream?". Acting superb and directing on point. Very good movie.	Good : 😀

Fig.10: Sentimental analysis of the Entertainement Content recommendation

7. CONCLUSION

Our proposed algorithm utilizes textual metadata, such as the plot, cast, genre, release year, and other production details of movies, to analyze and recommend the most similar films. The user only needs to input a movie of interest, and the system will generate appropriate recommendations. We tested our algorithm on a subset of the movies available on IMDb and found that the cosine similarity measure was effective for forecasting recommendations in recommendation systems. Additionally, we implemented a feature that allows for retraining of the system by rating results as "good" or "bad," resulting in more accurate predictions than simply selecting one movie or providing a single piece of text.

In the future, we plan to track movies searched by users in nearby locations to recommend popular films. We can also consider incorporating the watch history of geographically contextual users (those living in close proximity) with the watch history of the user to provide more "location- relevant" recommendations. Additionally, by using user ratings of movies from websites like Rotten Tomatoes, Metacritic, and IMDb, we can explore the possibility of combining collaborative filtering techniques with our method to create a hybrid model that combines the advantages of both approaches .

REFERENCES

[1] Mykhaylo Schwarz, Mykhaylo Lobur, Yuriy Stekh, Analysis of the Effectiveness of Similarity Measures forRecommender Systems, 978-1-5090-5045- 1/17/$31.00 ©2017 IEEE M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Rujhan Singla, Saamarth Gupta, Anirudh Gupta, Dinesh Kumar Vishwakarma, FLEX: A Content Based Movie Recommender, 978-1-7281-6221-8/20/$31.00 ©2020 IEEE.

[3] Jochen Nessel, Barbara Cimpa, The MovieOracle - Content Based Movie Recommendations, 978-0-76954513-4/11 $26.00 © 2011 IEEE.

[4] Shreya Agrawal, Pooja Jain, An Improved Approach for Movie Recommendation System, 978-1-5090-32433/17/$31.00 ©2017 IEEE

[5] Yibo Wang, Mingming Wang, and Wei Xu, A SentimentEnhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework, Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 8263704

[6] F. Furtado, A, Singh, Movie Recommendation System Using Machine Learning, Int. J. Res. Ind. Eng. Vol. 9, No. 1 (2020) 84–98

[7] Robin Burke. Hybrid recommender systems: Survey and experiments. Usermodeling and user- adapted interaction, 12(4):331–370, 2002.

[8] Erik Cambria. Affective computing and sentiment analysis. IEEE Intelligent Systems, 31(2):102–107, 2016.

[9] Ivan Cantador, Alejandro Bellog ´ ´ın, and David Vallet. Content-based recommendation in social tagging systems. In Proceedings of the Fourth Conference on Recommender systems, pages 237– 240. ACM, 2010.

[10] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In Proceedings of the Fourth Conference on Recommender Systems, pages 39–46. ACM, 2010. ISBN 978-1-60558-906-0