

Original Article

CineInsight: NLP-Driven Movie Recommendation Enhancement for Over-The-Top Platforms

K. Nandhini¹, S. Muthukumaran², N. Gnanasankaran³, G. Rakesh⁴, K. Muthuchamy⁵

^{1,2}Department of Computer Applications, School of Computing Sciences, VELS Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India.

^{3,4}Department of Computer Science, Thiagarajar College, Madurai, Tamil Nadu, India.

⁵Department of Information Technology, School of Computing Sciences, VELS Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India.

²Corresponding Author : muthumphil11@gmail.com

Received: 15 October 2024

Revised: 19 November 2024

Accepted: 09 December 2024

Published: 30 December 2024

Abstract - Computers, through language, understand, interpret, and interact with humans, enabling Natural Language Processing (NLP), a sector of Artificial Intelligence (AI). Selecting a movie to watch may be very difficult because so many options are accessible on different streaming services; people have different likes and preferences, and others are unaware of fantastic films. Objective: In today's digital landscape, creating and executing a movie recommendation system is crucial to addressing the issues of information overload, improving user satisfaction, and maintaining competitiveness in the entertainment sector. Method: This study proposed the CineInsight NaiveFlix Algorithm for a movie recommendation, which leverages a movie review dataset gathered from the websites of YIFY and IMDB. After that, the preprocessed data was pipelined, and the essential stop words in the English language were extracted to improve the Naïve Bayes (NB) model. Subsequently, the audience reviews of the film were categorized as either positive or negative. Results: After comparing the suggested method's performance to the conventional NB model and the linear support vector classification algorithm, it was discovered that the suggested CineInsight NaiveFlix method performs better in categorizing audience movie reviews.

Keywords - Natural language processing, Movie recommendation system, Naïve Bayes algorithm.

1. Introduction

Based on a user's taste, interests, and online browsing activities, a recommendation system is a software application that suggests products, movies, and other media to them. A movie recommendation system assists users in making movie suggestions based on their preferences by utilizing collaborative and content-based filtering [1]. These days, movie libraries are digitally converted to websites that stream content, such as Netflix, HBO, YouTube, and Over the Top (OTT). App switching has allowed users to watch specialized content provided by several OTT services. In contrast to set-top boxes, OTT providers only provide users with videos upon request. OTT offers two services. Movies are viewed via Video on Demand (VOD), while sporting events are streamed live. These days, internet access is provided via fibre, Wi-Fi, LTE, DSL, and other technologies. Delivering video material over OTT may be significantly impacted by the speed and performance of the network. With a bit rate of 6 Mb/Sec for HD screens and 250 Kb/Sec for mobile and tablets, the OTT provider sent the video content [2]. Owing to the scarcity of theatres to screen low-budget films and the need to spend more money on publicity and advertising,

small and medium-sized filmmakers are shifting the distribution of their films to OTT platforms. The movie is purchased by the OTT platforms, who then directly give the content to the viewer. By providing quality content, OTT platforms can generate more money if they can retain customers over the long term. One of the most critical factors in keeping consumers with OTT service providers is the calibre of the video content they offer.

1.1. How Sentiment Analysis Impacts Audience in Choosing a Film

Sentiment analysis, sometimes known as opinion mining, is a branch of NLP that converts user opinions about movies, products, and other media into valuable insights. Consumer's subconscious mind plays a significant role in their decision to see a movie [3]. Customers communicate their profound feelings and thoughts about a movie through metaphors they employ to describe it rather than utilizing a rating scale to quantify their thoughts. Social media is a readily accessible venue for reaching a target audience and can generate significant buzz about a film in preparation for its release. A single tweet or post from a well-known



audience member or celebrity can spark a wave of attention in their followers and increase the popularity of a movie. Inviting fans to make their own trailers, posters, and Twitter hashtags can result in free film marketing and aid in reaching a large audience [4]. Public opinion of a film might impact an audience's decision to watch it, as provided by reviews of the film written by celebrities and well-known content creators.

1.2. Challenges Present in Sentiment Analysis by NLP Algorithms

Sometimes, people use sarcasm to convey their negative ideas, for example, "I spent two hours sleeping nicely in the theatre". The NLP technologies may identify the input for this kind of review as favourable even though it is truly negative. Both positive and negative reviews of the film are difficult to categorize, as evidenced by comments like "It was a very Terrible Experience" that are used to describe it. When the audience uses emotionally charged language with various meanings, sentiment analysis utilizing NLP methods may become extremely difficult to discern the response [5].

The audience's reaction, such as "The movie was not unpleasant," cannot be classified as negative because the present NLP methods lack a robust enough negation detection mechanism. People may use multiple languages to describe their feelings, yet most NLP methods only support English for sentiment analysis. To recognize emotions in multilingual information, NLP methods must be developed. Currently, Emojis are also used by individuals to convey their feelings; however, as most NLP methods treat Emojis as special characters, it is impossible to detect the emotions accurately.

1.3. Research Contributions

Given the enormous number of movies saved on streaming platforms, a movie recommendation system must be created to help consumers choose relevant and engaging films from a massive digital library.

- As part of this research, the CineInsight NaiveFlix Algorithm was created. This approach leverages NLP and sentiment analysis to assist users in overcoming the difficulties involved in selecting a film to watch.
- The Term Frequency-Inverse Document Frequency (TF-IDF) method is implemented to extract essential attributes from the dataset after preprocessing it using the pipeline technique.
- Using the most recent English stop terms, the CineInsight NaiveFlix approach was used to categorize the audience review into good and negative categories.
- The suggested approach outperforms the two techniques in classifying the audience review, as demonstrated by a comparison with the NB and Support Vector approach (Linear SVC).
- The suggested approach correctly classifies reviews as favourable or bad and accepts fresh reviews that the audience dynamically enters.

2. Literature Review

N. Pavithara et al. [6] created a recommendation system for movies users should watch based on audience reviews using NB and Support Vector Machines (SVMs). The authors employed Cosine Similarity to recommend movies based on two datasets gathered from the IMDB website. NLP was used to analyze customer review sentiment to determine if the review was favourable or unfavourable. The writers' module allows users to enter the name of a movie and provides a list of recommended films to view. The authors demonstrated that SVM performs better at determining whether a movie review is good or negative than the NB model. Yillin Zhang and Lingling Zhang [7] employed Linear Discriminant Analysis to examine audience comments and determine the viewers' feelings toward the films. The author thoroughly analyses the body of research on NLP in movie recommendations. The IMDB dataset the authors gathered for this study includes 2,85,000 reviews. The movie titles, adjectives like "good," "bad," "classic," and thematic terms like "love," "life," and "war," among others, were filtered out during preprocessing. The dataset was subjected to BERT and NB models, categorising customer reviews into 25 themes. Jitendra Agrawal and Sonu Airen [8] employed the Partitional Weighted Co-Clustering Algorithm to filter Netflix and Amazon movies according to user presence. The 4000 movie ratings, ranging from 1 to 5, were gathered from 6000 individuals for the MovieLens dataset and used by the author. The movie titles are placed in a column, and the users are arranged in a row to construct the Co-Clustering matrix. The movie's rating can be found in the relevant row and column value. 16 clusters were created from the dataset, with an RMSE value of 1.2852 and an MAE of 0.7814 for neutral evaluations.

Using the Cosine Similarity Principle, Pooja Bagane et al. [9] created a movie recommendation for the audience based on the star, genre, and other factors. The writers keep the movie's genre in a dataframe and then utilize the TF-IDF vectorizer function to transform user reviews from text to vector format. Using a Lambda function, movie similarity is arranged in descending order. The authors' module used the Cosine Similarity Function to recommend the top 10 movies relevant to the submitted movie. The module took the movie name as input and saved it as a similarity matrix. A Collaborative Filtering approach was created by Gopal Behera and Neeta Nain [10] to propose movies to users based on their tastes. The writers keep a matrix with the customer ratings for each product, ranging from 1 to 5. The performance of the suggested method was compared using Single Value Decomposition (SVD), Non-Negative Feature (NNF), and RecSVD++ methods on the MovieLens dataset by the author. The results demonstrated that the Collaborative Filtering Method performs better than average when recommending a movie. Raj et al. [11] present a methodology that uses Large Language Models (LLMs) for movie genre prediction and recommendations. Zhao et al. [12] introduce

M5, a model for OTT recommendations that utilizes multi-modal embeddings, extraction, and scenario differentiation to capture user preferences and generate hybrid scores for recommendations. Arvind et al. [13] use textual review analysis and Machine Learning (ML) methods, namely K-Means, Linear Regression, and SVM Regression, to address issues like churn and biased content recommendations on OTT platforms.

Yang, Xu, and Tu [14] utilize a co-attention networking-based sentiment analysis method for aspect-level sentiment classification. At the same time, a Softmax Discriminant Classifier predicts movie success, employing review sentiments and movie information as input variables. Cakar, Aytekin, and Ozcan [15] introduce a novel solution that employs textual content—overviews, reviews, plots, and subtitles—to generate improved content descriptions. Anand, Azam, and Sagar [16] designed a movie recommendation system utilizing classical ML methodologies, namely Content-based and Collaborative Filtering, to suggest movies based on user ratings, watch history, and movie data. Wijati et al. [17] aim to analyze user sentiment using the K-Nearest

Neighbors (KNN) approach to classify reviews into negative, neutral, and positive categories. Oh et al. [18] introduce a Music Recommender System (MUSE) utilizing self-supervised learning and transition-based augmentation. It also uses item- and similarity-assisted matching strategies. Mathebula, Modupe, and Marivate [19] propose a Language Feature Extraction and Adaptation for Reviews (LFEAR) model by integrating Retrieval-Augmented Generation (RAG) and Auto-Regressive Fine-Tuning (ARFT) to enhance sentiment analysis across various sectors. Leem et al. [20] present a Dimension REduction, Clustering, and Classification (DRECE) model.

Despite the advancements in movie recommendation systems, limitations comprise challenges in handling diverse user preferences, addressing biases in the data, and achieving transparency in the decision-making process. Current models mostly need help accurately predicting preferences for new or less popular movies. Moreover, incorporating contextual factors, namely user mood or external influences, remains a research gap, needing additional exploration to improve personalization and system flexibility.

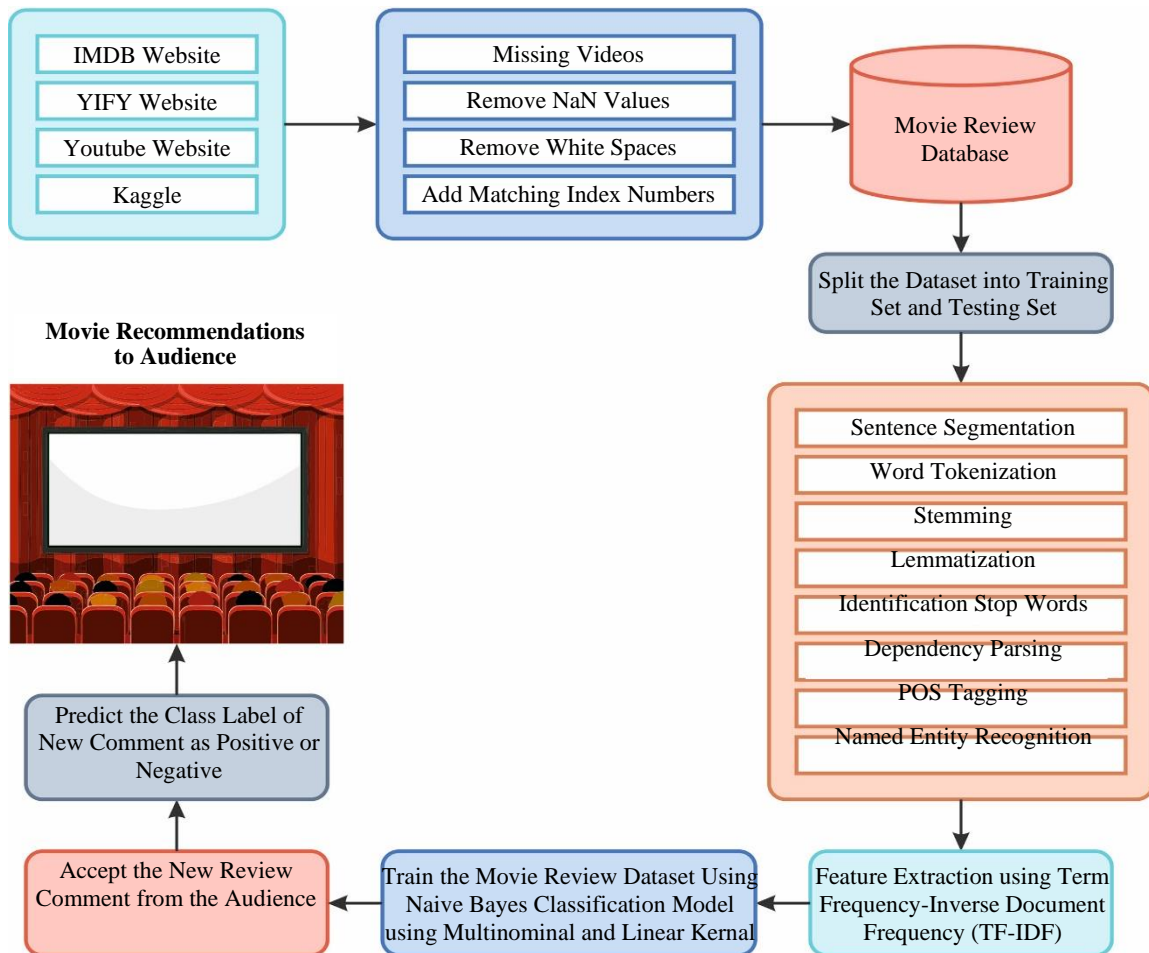


Fig. 1 Architecture of cineInsight NaiveFlix algorithm

3. Proposed Cineinsight Naiveflix Algorithm For Movie Recommendation

The CineInsight NaiveFlix Algorithm's architecture is depicted in Fig. 1 for movie suggestion. The information required for this research was gathered from reviewer comments on the YIFY, YouTube, and IMDB websites. Every audience member's response was recorded in a text file and categorized as favourable or negative. Next, whitespaces, NaN values, and missing values were eliminated from the dataset as part of preprocessing. After that, the dataset is

Algorithm 1: CineInsight NaiveFlix Algorithm

Input: (Movie Review Dataset- D_{movie} , Y_i -Label, X_i -Reviews)

Output: Classification of Movie Reviews as Positive, Negative

1. Start the process
2. Perform Data Preprocessing in the movie review Dataset D_{movie}
 - a. Foreach X_i in D_{movie} do
 - b. Begin
 - c. Perform sentence segmentation using $P\left(\frac{X}{Y}\right) = \frac{1}{Z(X)} \exp\left(\sum_i \sum_j \lambda_j f_j(y_{i-1}, y_i, X, i)\right)$, where $Z(X)$ -normalization factor, λ_j -model parameters, f_j -feature function to capture relationships in variables. Equation (1)
 - d. Perform word tokenization using $Tokenize(Text) = \{Token_1, Token_2, \dots, Token_n\}$, where $Tokenize(Text) = \{word_1, word_2, \dots, word_n\}$. Equation (2)
 - e. Perform porter stemming operations using $PorterStem(word) = Step1(word) \rightarrow Step2(word) \rightarrow \dots \rightarrow StepN(word)$, where $Step1(word) = RemoveSuffix(word, \{"ing", "ly", \dots\})$. Equation (3)
 - f. Perform Lemmatization using $Lemmatize(word) = lemma$ Equation (4)
 - g. Identifying stop words [\text{StopWordIdentification}(D_{movie}) = \{\text{IsStopWord}(w_1), \text{IsStopWord}(w_2), \dots, \text{IsStopWord}(w_n)\} \setminus \}. Equation (5)
 - h. Perform dependency parsing using $DependencyParse(sentence) = DependencyTree$. Equation (6)
 - i. Perform Parts of Speech Tagging using $POS_Tagging(W) = arg\ max_T P(T | W)$, here W is the word sequence, T is the corresponding POS tag. Equation (7).
 - j. Perform Named Entity Recognition using $NER(W) = \{(w_i, e_i) | w_i \in W, e_i \in E\}$, where E is the set of possible named entity categories.

segmented into testing and training sets (70:30). The movie review dataset underwent pipelining, and the TF-IDF method was utilized to identify the dataset's significant features. The performance of the CineInsight NaiveFlix classifier was evaluated after it was applied to the pipelined dataset. The most frequently used stop words in the English language were filtered to enhance the classifier's performance, and the movie review dataset was once more employed with hyperparameter tuning. Algorithm 1 illustrates the pseudocode for the CineInsight NaiveFlix Algorithm.

Equation (8).

k. End

3. Perform Sentiment Analysis to classify the audience review in D_{movie} .

l. For each review X_i in D_{movie} do

m. Begin

n. Perform TF-IDF Vectorization using $TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$, where

$TF(t, d) =$

$\frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}$, and

$IDF(t, D) =$

$\log\left(\frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}\right)$,

Equation (9).

o. Apply the NB approach to classify the movie review using

$P\left(\frac{Class}{Features}\right) = \frac{P\left(\frac{Features}{Class}\right) \times P(Class)}{P(Features)}$, where individual features

in text data are extracted using $P(Features | Class) = P(Feature1 | Class) \times P(Feature2 | Class) \times \dots \times P(Feature_n | Class)$. Equation (10).

p. Give the selected stop words to the algorithm using

$P\left(\frac{Feature_{e_i}}{Class}\right) = \frac{\text{Count of documents containing } Feature_{e_i} + 1}{\text{Total number of records in the class} + \text{Vocabulary size}}$

Equation (11).

q. Classify the class label as positive or negative for each review in the D_{movie} row.

r. End

s. Classify the class label for each audience review

4. Accept the new movie review from a new user and give input to the trained model.

5. Predict the class of the movie review as positive or negative

6. Stop the process.

4. Experimental Results

The suggested CineInsight NaiveFlix Algorithm was implemented in Python 3.10 using Spyder as the code editor. This recommendation system was created using the "en_core_web_sm" English vocabulary library and the NLP library spacy.

Following preprocessing, the movie review dataset depicted in Figure 2 was pared down from its original 2000 records to 1938 records categorized as favourable or negative. The TfidfVectorizer method from the sklearn.feature_extraction.text library was used for vectorization, and the Pipeline method from the sklearn library was used for the process. A multinomialNB kernel was used to build the NB algorithm.

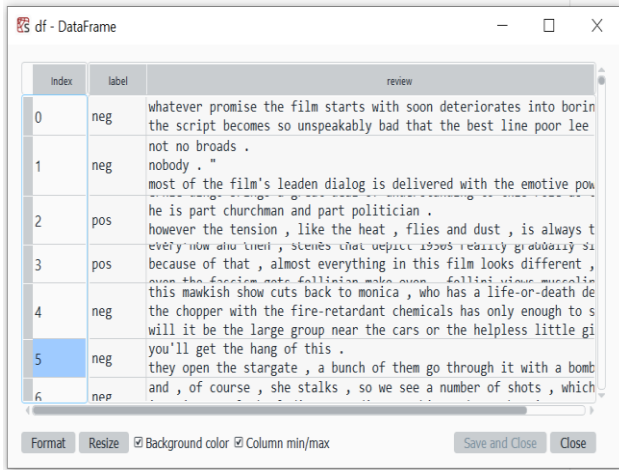


Fig. 2 Movie review dataset collected for this research

The confusion matrix provided in Figures 3 to 5 was utilized to gauge the effectiveness of the several models employed in this study, and the findings were tallied in Table 1. It has been demonstrated that the suggested CineInsight NaiveFlix works best at predicting the user's positive and negative remarks.

Table 1. Performance of the Training Set

Measures	NB	Linear SVC	CineInsight NaiveFlix
Accuracy	0.764	0.846	0.852
Error Rate	0.236	0.154	0.148
Recall	0.769	0.846	0.852
Specificity	0.769	0.852	0.852
Precision	0.796	0.852	0.852
1-Precision	0.203	0.148	0.148
F-Measure	0.76	0.84	0.852
False Positive Rate	0.283	0.153	0.149
False Negative Rate	0.229	0.012	0.147
False Discovery Rate	0.202	0.153	0.147
Negative Predicted value	0.796	0.846	0.852

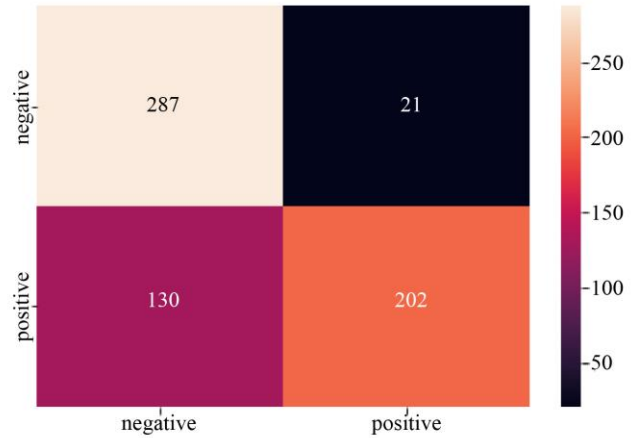


Fig. 3 Confusion matrix for NB approach

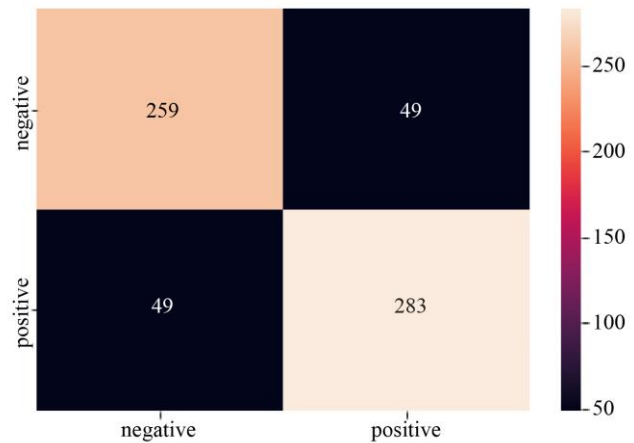


Fig. 4 Confusion matrix for linear SVC

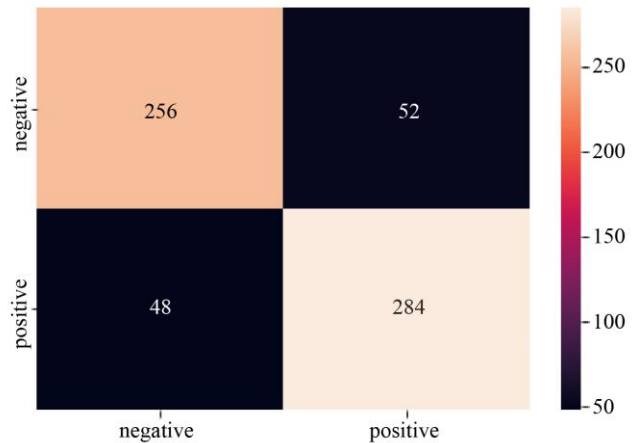
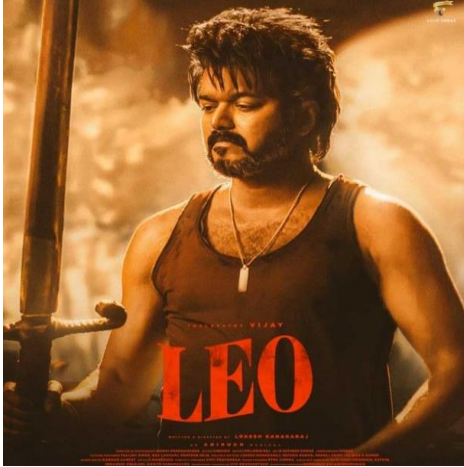


Fig. 5 Confusion matrix for CineInsight NaiveFlix approach

One of the CineInsight NaiveFlix Algorithm modules suggested in this research takes the user-prompted new movie comment and accepts it. Following the entry, the comment is pipelined and fed into the CineInsight NaiveFlix Algorithm, which forecasts that viewers will not have any positive or negative things to say about the movie. Figures 6 and 7 demonstrate some user-initiated reviews and the label the suggested algorithm anticipated.



```
In [36]: myreview="Leo- The hero actioing was good, \but the concept for the movie is lacking,\ The graphics used in this film was very bad,\ The logic in the fight scenes was worst. The subramani character for the wolf is also have no use in this film, Sandy master character brings the real character of Leo"
```

```
In [37]: print(text_clf_lsvc2.predict([myreview]))
['neg']
```

Fig. 6 Sample one of Movie Review Predicted by the CineInsight NaiveFlix Algorithm



```
In [32]: myreview="Oppenheimer-A non spoiler detail review: It's actually the best scientific Biopic made after The theory of everything \
...: A Cult classic cinema \
...: Christopher Nolan outshines in his technicality experimentation of direction with subjectives and objectives clearly shown in Black And white that too on an IMAX reel! \
...: Ita an clever, intelligence and Bold move to stay with your opinion on what you want rather then being a fearful about your work"
```

```
In [33]: print(text_clf_lsvc2.predict([myreview]))
['pos']
```

Fig. 7 Sample two of Movie Review Predicted by the CineInsight NaiveFlix Algorithm

5. Conclusion

This research proposed a CineInsight NaiveFlix Algorithm to design a recommendation system for watching movies on an online streaming platform. The preprocessing and pipelining done in the movie review dataset help classify the positive and negative reviews present in the dataset. It has been proved that giving selected stop words to the algorithm improves its performance and predicts the prompted review with higher accuracy. The proposed algorithm was also compared with NB and Linear SVC, proving that the CineInsight NaiveFlix algorithm outperforms well in classifying the dataset. The recommendation system encounters limitations, including difficulty with new or unpopular movies, reliance on user review quality, potential biases, and challenges adapting to changing user preferences. Future enhancements include incorporating contextual data for personalization, improving sentiment analysis, using multi-modal data, addressing ethical concerns, and developing real-time recommendation algorithms based on user behaviour. Also, the proposed algorithm will be tested with a vast volume of datasets that help choose films related to user interest.

References

- [1] P. Arokiaraj et al., "Movie Recommendation System Using Machine Learning," *Advances in Science and Technology*, vol. 124, pp. 398-406, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [2] Vipin Kumar et al., "The Ultimate Recommendation System: Proposed Pranik System," *Multimed Tools and Applications*, vol. 83, pp. 43177-43198, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [3] Amir Saleh et al., "Improving the Major Recommendation Systems: Analysis of Hybrid Naive Bayes-based Collaborative Filtering and Fuzzy Logic," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 8, no. 4, pp. 789-798, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [4] Pretty Paul, and Rimjhim Padam Singh, "A Weighted Hybrid Recommendation Approach for User's Contentment using Natural Language Processing," *AIP Conference Proceedings*, vol. 2705, no. 1, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [5] Yeresime Suresh et al., "Naive Bayes Classifier based Movie Recommendation System," *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, pp.753-758, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [6] N. Pavitha et al., "Movie Recommendation and Sentiment Analysis using Machine Learning," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 279-284, 2022. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [7] Yilin Zhang, and Lingling Zhang, "Movie Recommendation Algorithm Based on Sentiment Analysis and LDA," *Procedia Computer Science*, vol. 199, pp. 871-878, 2022. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [8] Sonu Airen, and Jitendra Agrawal, "Movie Recommender System Using Parameter Tuning of User and Movie Neighbourhood via Co-Clustering," *Procedia Computer Science*, vol. 218, pp. 1176-1183, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [9] Pooja Bagane et al., "Automotive Movie Recommendation System based on Natural Language Processing," *International Journal on Research and Innovation Trends in Computing and Communication*, vol. 11, no. 5, pp. 300-303, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [10] Gopal Behera, and Neeta Nain, "Collaborative Filtering with Temporal Features for Movie Recommendation System," *Procedia Computer*

- Science, vol. 218, pp. 1366-1373, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [11] Subham Raj et al., "Transformative Movie Discovery: Large Language Models for Recommendation and Genre Prediction," *IEEE Access*, vol. 12, pp. 186626-186638, 2024. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [12] Pengyu Zhao et al., "M5: Multi-Modal Multi-Interest Multi-Scenario Matching for Over-the-Top Recommendation," *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5650-5659, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [13] Vidisha Arvind et al., "Decoding OTT: Uncovering Global Insights for Film Industry Advancement and Cultural Exchange," *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp. 1300-1309, 2024. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [14] Gelan Yang, Yiyi Xu, and Li Tu, "An Intelligent Box Office Predictor Based on Aspect-level Sentiment Analysis of Movie Review," *Wireless Networks*, vol. 29, pp. 3039-3049, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [15] Prakhar Anand, Sajid Azam, and Shreddha Sagar, "A Hybrid Approach Towards a Movie Recommendation System," *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 1696-1708, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [16] Mahmut Cakar, Tevfik Aytakin, and Hunkar Ozcan, "Movie Tag Prediction Using Multi-label Classification with BERT," *International Conference on Computer and Communication Engineering*, pp. 31-40, 2024. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [17] Damar Wijati et al., "Sentiment Analysis of Application Reviews using the K-Nearest Neighbors (KNN) Algorithm," *PIXEL Computer Science Research Embedded and Logic Systems*, vol. 12, no. 1, pp. 209-218, 2024. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [18] Yunhak Oh et al., "Muse: Music Recommender System with Shuffle Play Recommendation Enhancement," *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 1928-1938, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [19] Miehleketo Mathebula, Abiodun Modupe, and Vukosi Marivate, "Fine-Tuning Retrieval-Augmented Generation with an Auto-Regressive Language Model for Sentiment Analysis in Financial Reviews," *Applied Sciences*, vol. 14, no. 23, pp. 1-43, 2024. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]
- [20] Subeen Leem et al., "Towards Data-Driven Decision-Making in the Korean Film Industry: An XAI Model for Box Office Analysis Using Dimension Reduction, Clustering, and Classification," *Entropy*, vol. 25, no. 4, pp. 1-29, 2023. [[CrossRef](#)] [[Googel Scholar](#)] [[Publisher Link](#)]