A Comprehensive Study on the Use of SVM, Random Forest, and Naive Bayes for Sentiment Analysis in the Film Industry: Insights from IMDb Reviews

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Kruthik Selvan
Dept. of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
ch.en.u4aie22081@ch.students.amrita.edu

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Opinion mining, also referred to as sentiment analysis, has become increasingly vital as a method for understanding public sentiment across diverse fields. Our study uses IMDb movie reviews as our main dataset and focuses on sentiment analysis in the film business. Because it has a wealth of annotated data collected from widely available web-based destinations, IMDb is a great tool for creating and assessing sentiment analysis models Our study attempts to evaluate how well three widely used machine learning techniques Random Forest, Naive Bayes, and Support Vector Machines classify feelings as precisely as possible in IMDb movie reviews. Through the application of several machine learning approaches and natural language processing strategies, our goal is to offer a thorough analysis of sentiment classification in the film business Our methodology includes grid search, cross-validation, and hyperparameter tweaking to maximize each model's performance. Our goal is to optimize the sentiment categorization accuracy of IMDb movie reviews through these procedures. Using important performance criteria including F1-score, recall, accuracy, and precision, we evaluate these algorithms' efficacy Our tests' outcomes provide insightful information on the advantages and disadvantages of SVM, Random Forest, and Naive Bayes for sentiment analysis of IMDb reviews. In order to further our understanding of audience preferences in the film industry, we also examine the significance of sentiment trends and attitudes indicated in these evaluations By providing a thorough analysis of sentiment analysis methodologies applied to IMDb movie reviews, this research advances the fields of sentiment analysis and film studies Furthermore, our research has practical implications for distributors, producers, and other film industry players, offering them insightful information to guide strategic choices in this fiercely competitive and quickly changing environment.**

***Index Terms*—IMDb movie reviews,Random Forests,Support Vector Machines,Naive Bayes**

I. Introduction

In the digital age, film industry doesn't just make feature films; it also captures many of the thoughts and feelings of people all over world in one Hovie buffs are able to freely review movies and voice their opinions thanks to websites like IMDb. Sentiment research has become important in this context to understand and analyze audience reactions to films. Cognitive analysis under the general heading of Natural Language Processing focuses on identifying and analyzing psychological phenomena in text This process reveals the beliefs, attitudes and emotions expressed by people
[1].

Our research seeks to explore sentiment analysis in film industry, with a particular emphasis on IMDb ratings. One of largest online movie and television guide websites, IMDb offers a wealth of essays reflecting a wide range of opinions and feelings from people all over world

We want to investigate the effectiveness of reliable machine learning methods in predicting behaviors in IMDb movie reviews. So we explore their community for knowledge. [2].

Sensitivity analysis is important in today's environment for several reasons. First, it reveals values of viewers' tastes, attitudes, and emotions. This allows studios, producers and filmmakers to better determine the reception of their work. Filmmakers can adjust to audience preferences by understanding the emotions expressed by their audiences and altering their strategies, choices and creative processes accordingly [3]

Sensitivity analysis is important in today's environment for several reasons. First, it reveals the values of viewers' tastes, attitudes, and emotions. This allows studios, producers and filmmakers to better determine the reception of their work. Filmmakers can better adjust to audience preferences by understanding the emotions expressed by their audiences and altering their strategies, choices, and creative processes accordingly.In any case, thought search is a key device used in many organizations, apart from the entertainment world It is essential for activities such as social media management, reputation management, customer feedback analysis and marketing. Sentiment analysis is a tool that organizations use to analyze customer perceptions of their products, identify new trends, and predict problems before they arise By using sentiment analysis, companies can provide increased customer satisfaction, increased brand loyalty, and gained competitive advantage. [4]

Machine learning techniques for opinion examination errands consolidate incorporate Random Forest, Support Vector Machines, and Naive Bayes classifiers. Contingent upon what kind of textual material is being examined, each algorithm has advantages and disadvantages that must be considered before applying.

Support Vector Machines are gaining popularity for their potential to process high-dimensional data and correctly categorize sophisticated datasets. SVM accomplishes by separating good and bad sentiments as many hyperplanes as feasible in distinct feature areas. The crucial reason SVM is such a frequent choice for sentiment evaluation chores is that it is a very robust and generalizative algorithm, particularly when evaluating huge datasets such as the IMDb evaluations.Random Forest functions similarly to a group of decision trees working together. During training, it generates a large number of decision trees, and to determine the final classification, it votes to integrate each tree's predictions. In sentiment analysis, Random Forest excels in managing overlapping and noisy features, which enables it to successfully extract minute details from textual input. High accuracy in sentiment categorization tasks is the outcome of this.The Bayes theorem and feature independence serve as the foundation for naive Bayes classifiers. They work quite well in sentiment analysis jobs despite their simplicity, especially when dealing with text data Their exceptional computational efficiency, capacity to manage large amounts of text data, and intuitive design make them ideal for real-time sentiment analysis applications.

In this work, we examine and assess how well Random Forest, Naive Bayes, and SVM classifiers can infer sentiment from IMDb movie reviews We hope to provide insightful information on their advantages and disadvantages when it comes to textual data analysis in the film business through this inquiry. Our thorough investigation aims to improve sentiment analysis methods and provide a deeper knowledge of audience attitudes regarding movies in the current digital era.

II. LITERATURE REVIEW

By utilizing a variety of machine learning models on the IMDB Review Dataset, Shubham Kumar Singh and Neetu Singla explored the field of sentiment analysis in 2023 [5]. Their goal was to determine which models could most accurately predict whether movie reviews were favorable or unfavorable. Thus, they put many models to test, including BiLSTM, LSTM, LSVM, Decision Tree, and Naive Bayes. They discovered that BiLSTM model outperformed others in terms of accuracy, precision, and recall. Singh and Singla (2023) emphasised significance of optimising model settings and appropriately prepping data to achieve optimal outcomes in sentiment analysis assignments.

The topics of the 2021 paper "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," by V. Umarani, A. Julian, and J. Deepa, are deep learning and supervised machine learning methods for sentiment analysis. Their study assesses the performance of various classifiers, including Multinomial Naive Bayes, Logistic Regression, Support Vector Machine, Random Forest, K-nearest neighbor, Decision tree, Long Short-Term Memory, and Convolutional Neural Network, using standard datasets. In terms of k-fold cross-validation, accuracy, recall, F1-score, RoC-Curve, and running time, experimental findings demonstrate their efficacy. [Umarani et al., 2021).

Priza Pandunata, Yanuar Nurdiansyah, and Fitri Dwi Alfina started a study in 2023 that delved deeply into IMDb evaluations of Avatar 2 [7]. Rather than concentrating on general viewpoints, they examined particular elements such as the narrative and illustrations. They employed a Support Vector Machine model to balance the data with SMOTE and analyze sentiments after sorting through 3198 reviews. Their findings were a treasure for marketers and producers in addition to being fascinating for movie buffs. These revelations may be used to further comprehend viewer preferences and direct the creation of upcoming motion pictures.

Anuj Sharma and Shubhamoy Dey unveiled a brand-new method for sorting through internet evaluations in 2013 [8]. combination of boosting methods and weak SVM classifiers, they dubbed it Boosted SVM. Their method's approach to addressing the drawbacks of traditional techniques like Naive Bayes and regular SVMs is quite fascinating. Better still, it outperforms individual SVM classifiers in terms of accuracy in addition to improving SVM performance as a whole.They demonstrated the idea's effectiveness in interpreting feelings by testing it on datasets of hotel and movie reviews (Sharma & Dey, 2013).

"Sentiment Analysis on Product Reviews Data Using Supervised Learning: A Comprehensive Review of Recent Techniques" was written by Nahili, Rezeg, and Kazar in 2020 [9]. This study delves deeply into contemporary text mining and sentiment analysis methods The study examines a number of sentiment analysis-related topics, including handling characteristics, information gathering, data preparation, and application of machine learning and natural language processing methods. The study also explores sentiment analysis's practical applications, including its usage in business intelligence, big data analytics, and consumer review mining. The study's ultimate goal is to help researchers choose the methodologies that will best meet their unique requirements (Nahili, Rezeg, & Kazar, 2020)..

Topal and Ozsoyoglu (2016) [10] looked at the emotion analysis of IMDb movie reviews to help people choose movies that they should know about. In order to suggest movies with desirable emotional patterns, study suggests using emotion maps created from the sentiments of all reviews combined. Emotion maps were produced after an analysis of IMDb movie reviews and ratings. As part of the research technique, these maps were then used to cluster movies according to the opinions of reviewers, and genre-specific emotional patterns were investigated.

III. Methodology

A. *Model Architecture*

In fig 1 we explain our proposal in step-by-step process .

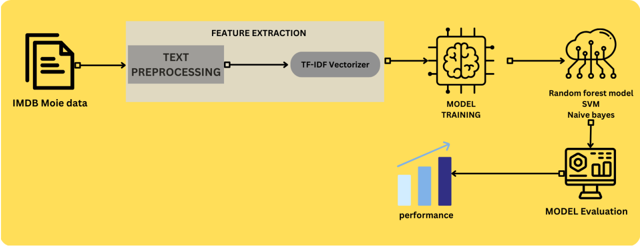

Fig. 1. Example of a figure caption.

B. *SVM - An Overview*

Support Vector Machines (SVM) [11] is a broadly involved algorithm for text classification undertakings because of its ability capacity to deal high-dimensional data effectively. With regards to opinion investigation, SVM tries to gain proficiency with a choice limit that isolates positive and negative feeling classes by expanding the margin between classes.

For our IMDb review dataset W, SVM can be applied to classify the sentiment of movie reviews x_i into positive $(y = +1)$or negative $(y = 1)$ classes. The classification decision can be represented by:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \tag{1}$$

where:

- x_i and y_i represent the training samples and their corresponding class labels, respectively.
- α_i are the learned Lagrange multipliers.
- $K(x_i, x)$ denotes the kernel function, which computes the similarity between x_i and x.
- b is the bias term.

C. *Naive Bayes*

Text classification applications [13] often use the Naive Bayes classifier due to its efficiency and effectiveness in handling high-dimensional data. This probabilistic classifier is straightforward but effective. Because Nev Bayes assumes that each word in a document is independent of the presence of other words, it is particularly suited for text classification tasks in sentiment analysis.

For our IMDb review dataset W, Naive Bayes can be applied to classify movie reviews xi into positive $(y = +1)$ or negative $(y = 1)$ sentiment classes. The classification decision can be formulated using Bayes' theorem as:

$$\hat{y} = \underset{y \in \{-1,+1\}}{\text{argmax}} \; P(y) \prod_{i=1}^{n} P(x_i | y) \tag{2}$$

where:

- $\hat{y}$ represents the predicted sentiment label for the input review x.
- $P(y)$ is the prior probability of sentiment class y.
- $P(x_i|y)$ is the conditional probability of word x_i given sentiment class y.
- n is the total number of words in the vocabulary.

D. *Random Forest*

Random Forest [12]is a dependable solution for text classification issues since it is an effective ensemble learning technique for handling noisy features and high-dimensional data. In sentiment analysis, Random Forest constructs many decision trees during training, then votes to aggregate their predictions to establish the final classification.

For our IMDb review dataset W, Random Forest can be employed to classify movie reviews xi into positive (y=+1) or negative (y=1) sentiment classes. The classification decision can be represented as:

$$\hat{y} = \text{mode}\left(f_1(x), f_2(x), .., f_T(x)\right) \tag{3}$$

where:

- $\hat{y}$ represents the predicted sentiment label for the input review x.
- $f_t(x)$ denotes the prediction of the t-th decision tree.
- T is the total number of decision trees in the Random Forest.

Before applying Random Forest to text classification, we tokenize the text in the IMDb review dataset W, eliminate stop words, and use word embeddings or TF-IDF to convert words to numerical feature vectors. We utilize the retrieved features and matching sentiment labels to train the Random Forest model on the training set after splitting the dataset into training and testing sets.

E. *Data Collection*

In the data collection phase of our study, we retrieved the IMDb movie reviews dataset from the official IMDb website, which hosts a vast repository of user-generated reviews and ratings spanning various genres and release years. To ensure the dataset's diversity and relevance, we applied stringent inclusion and exclusion criteria. We included reviews from a broad spectrum of genres and release years to capture a comprehensive range of opinions and sentiments expressed by IMDb users. Conversely, reviews with insufficient textual content or those in non-English languages were excluded to maintain data consistency and readability. Employing a systematic sampling approach, we randomly selected a subset of movie reviews across different genres and ratings to mitigate bias and ensure dataset representation. Before sentiment analysis, we preprocessed the collected data by removing irrelevant characters, punctuation marks, and stopwords, and conducted tokenization to structure the textual data suitably for analysis. Through these meticulous procedures, we aimed to curate a robust and unbiased dataset that accurately reflects IMDb users' sentiments and opinions in their movie reviews.

F. Data visualization

*1) **Feature Importance Plot**:* The feature importance plot, where the x-axis most likely represents importance scores, shows the relevance of words or phrases in the sentiment classification. It's noteworthy that phrases like ”bad,” ”worst,” and ”wasted” frequently indicate negative sentiment. Although there isn't a set scale, larger bars indicate more significance. When writing about this in a research paper, it's critical to clarify the significance of feature importance, reference the data source (such as IMDb reviews), and draw attention to relevant features like ”bad” and ”worst.” But it's important to recognize its limitations, like its reliance on model parameters and dataset properties.

G. Data Pre-Processing

In order to prepare the dataset for IMDb movie reviews, we first clean it up by eliminating duplicates, dealing with missing values, and correcting formatting mistakes. After that, the text is tokenized to make it more readable. Lowercasing maintains consistency by treating word variations equally. Then, stopwords—common but unimportant words—are eliminated, emphasizing terms that convey emotion. Lemmatization and stemming help to further standardize words. For extra context, feature engineering methods like sentiment lexicons or n-grams may be used. In order to train and assess models independently, we finally divided the data into training, validation, and test sets. By standardizing and improving the dataset, these procedures guarantee that significant insights can be gleaned from it for sentiment analysis.

H. Model Training

In training machine learning models for sentiment analysis on IMDb movie reviews, we start by preprocessing dataset to clean text data and normalize it using techniques like stemming or lemmatization. Following preprocessing, we convert text into numerical features using methods like Bag-of-Words, TF-IDF, or word embeddings.For sentiment analysis, we employ Support Vector Machines , Random Forest, and Naive Bayes classifiers due to their effectiveness with text data. Models are trained on preprocessed data, with training/validation splits for performance evaluation and hyperparameter tuning.After training and tuning, models are assessed on a separate test dataset. Performance metrics like accuracy, precision, recall, and F1-score gauge their effectiveness. This methodology outlines the steps for training and evaluating machine learning models for sentiment analysis on IMDb movie reviews.

I. Data Processing

In the data processing stage, our focus lies in readying the IMDb movie review dataset for machine learning model training. Initially, we meticulously clean the dataset, rectifying inconsistencies, errors, and noise. This entails removing duplicate entries, handling missing values, and rectifying formatting issues. Subsequently, we embark on text preprocessing to render the raw text data amenable to analysis. We proceed by eliminating stopwords—common, yet insignificant words like articles, prepositions, and conjunctions. Post-stopwords removal, we employ stemming or lemmatization to standardize words by reducing them to their base or root form. This enhances analysis consistency and effectiveness by consolidating word variations. Furthermore, we may conduct feature engineering, employing techniques such as n-grams to capture word sequences or sentiment lexicons for added contextual insights. Finally, the dataset undergoes division into training, validation, and test sets. This segregation facilitates model training, validation, and subsequent assessment of generalization performance on unseen data subsets.

J. Model Evalution

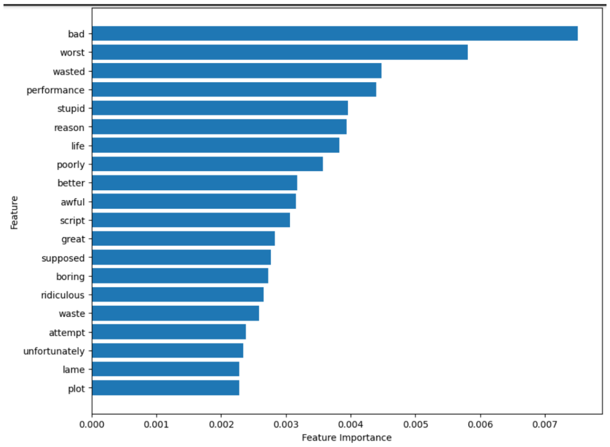

Fig. 2. Feature Importance Plot

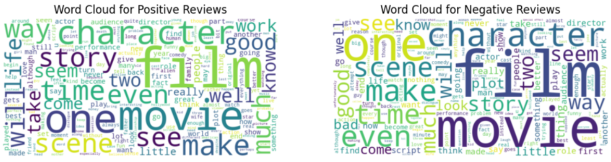

Fig. 3. word cloud for positive and negative review.

IV. RESULT

We used IMDb reviews for this extensive study to examine the efficacy of Random Forest, Naive Bayes, and Support

TABLE I
PERFORMANCE ANALYSIS OF MACHINE LEARNING CLASSIFIERS FOR SENTIMENT ANALYSIS

Classifier	***Accuracy***	***Recall score***	***F1 score***
Random forest	0.84	0.84	0.84
SVM	0.83	0.83	0.83
Naive bayes	0.79	0.79	0.79

[a] Sample of a Table footnote.

Vector Machine classifiers for sentiment analysis in the film business. Promising insights regarding these classifiers' ability to extract sentiments from movie reviews were obtained from our investigation.

A. Accuracy

Accuracy measures proportion of correctly classified instances out of the total instances in dataset. It is calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

B. F1-Score

F1-Score is harmonic mean of precision and recall, providing a balance between two metrics. It is calculated as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

C. Recall

Recall measures proportion of correctly predicted positive instances out of all actual positive instances. It is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

D. ROC Curve

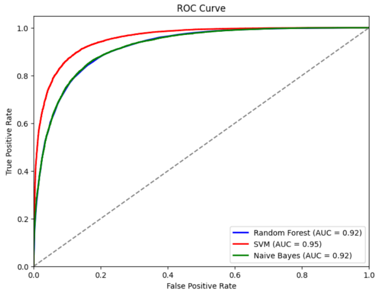

Fig. 4. ROC Curve .

Random Forest: The ROC curve for Random Forest classifier appears to have an Area Under the Curve of 0.92. This is a good score, indicating that the model is good at distinguishing between positive and negative classes.
SVM: ROC curve for SVM classifier has highest AUC among the three at 0.95. This suggests that the SVM model performs the best at distinguishing between the positive and negative classes in this dataset.
Naive Bayes: The ROC curve for the Naive Bayes classifier has an AUC of 0.92, which is the same as Random Forest classifier.
In general, an ROC curve closer to the top-left corner indicates better performance. So, based on this ROC curve, the SVM classifier performs best, followed by the Random Forest and Naive Bayes classifiers.

It's important to note that the AUC is just one metric for evaluating a machine learning model. Other factors, such as the specific task and the cost of misclassification, may also be important to consider when choosing a model.

*E. **Confusion Matrix***

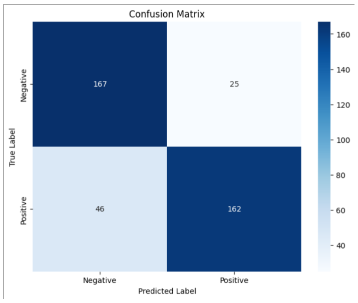

Fig. 5. confusion matrix for Random forest model.

confusion matrix provides a comprehensive overview of a classification model's performance. In this particular case, the confusion matrix appears to be evaluating a random forest model. Rows represent actual labels, while columns represent predicted labels. diagonal indicates correct predictions, while off-diagonal elements signify incorrect predictions. Here's a breakdown of confusion matrix: Positive: Refers to positive class. Negative: Refers to negative class. Predicted labels: Labels predicted by the model. True labels: Actual labels of the data.
True Negative: Number of negative instances correctly predicted. False Positive: Number of negative instances incorrectly predicted as positive. False Negative: Number of positive instances incorrectly predicted as negative. True Positive: Number of positive instances correctly predicted.

Predicted Classes: The values in the columns represent the classes that the SVM model predicted.
Actual Classes: The values in the rows represent the true classes of the data.

True Positives: Located in the top left corner, this value indicates the number of instances correctly classified as positive by the model. In this case, it appears to be 5.

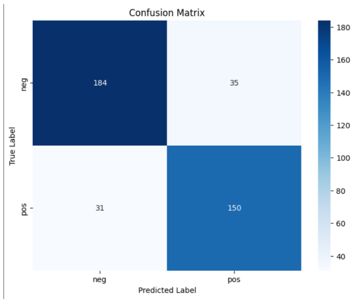

Fig. 6. confusion matrix for svm model.

False Negatives: The value in the top right corner represents the number of instances that were actually positive but were incorrectly predicted as negative by the model. This value is 1.

False Positives: Positioned in the bottom left corner, this value indicates the number of instances that were actually negative but were incorrectly classified as positive by the model.

True Negatives: While not explicitly mentioned in the provided excerpt, it is the remaining value in the matrix, typically located in the bottom right corner. It represents the instances that were correctly classified as negative by the model.

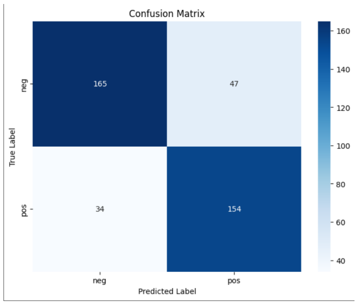

Fig. 7. confusion matrix for Naive bayes model.

Predicted Class Labels: These are typically listed in the columns and represent the classes the model predicted the data points belong to. In a binary classification, these might be ”Positive” and ”Negative.”

Actual Class Labels: Listed in the rows, these represent the true classes of the data points used for evaluation.

F. *Comparative Analysis*

Comparative analysis involves comparing the performance metrics (such as accuracy, precision, recall, etc.) of different models to determine which one performs better on a given task.

With an accuracy score of 0.84, the SVM classifier showed impressive performance, demonstrating its ability to successfully categorize movie reviews into positive and negative attitudes. Its margin-based categorization strategy worked very well for identifying intricate patterns in the text data.

With an accuracy score of 0.82, the Random Forest classifier also performed competitively. By combining the predictions of several decision trees and utilizing the ensemble learning technique, Random Forest was able to attain resilience, reducing overfitting and improving.

(Fig.4.) Positive Reviews: Words: good, story, acting, performance, characters, see, movie, even, really, well Interpretation: These words suggest that viewers enjoyed the film’s plot, acting, and characters. They found the movie to be good overall, even exceeding expectations. Words like ”see” and ”movie” indicate that viewers recommend watching the film.
Negative Reviews: Words: bad, time, boring, movie, plot, first, not, make, sense, way Interpretation: These words suggest that viewers found the movie to be bad, boring, and not worth their time. Words like ”plot” and ”make sense” imply that viewers disliked the movie’s storyline. Overall, the word cloud provides a quick summary of the sentiment expressed in positive and negative reviews of the movie. It shows that positive reviews focus on the movie’s good qualities, while negative reviews focus on its flaws.

V. CONCLUSION

This study comprehensively evaluated the performance of three popular machine learning algorithms, namely Support Vector Machine , Random Forest, and Naive Bayes, for sentiment analysis in film industry based on IMDb reviews.

Random Forest classifier exhibited the highest performance with an accuracy, F1 score, and recall score of approximately 0.84. This indicates its effectiveness in accurately classifying sentiment in movie reviews.

SVM classifier also performed well, achieving an accuracy, F1 score, and recall score of approximately 0.83. SVM showed robust performance, making it a viable option for sentiment analysis tasks in the film industry.

On other hand, Naive Bayes classifier demonstrated slightly lower performance compared to SVM and Random Forest, with an accuracy, F1 score, and recall score of approximately 0.79. While Naive Bayes is known for its simplicity and efficiency, its performance may be limited in more complex sentiment analysis tasks.

Overall, this study provides valuable insights into the effectiveness of SVM, Random Forest, and Naive Bayes algorithms for sentiment analysis in the film industry, offering guidance for researchers and practitioners seeking to leverage machine learning techniques for analyzing IMDb reviews and understanding audience sentiment towards movies.

VI. Future Research

Examine more intricate feature engineering techniques, such as word embeddings or contextualized representations, to enhance the models' capacity to identify sentiment in movie surveys and focus more subtle semantic data.

To work on Random Forest, Naive Bayes, and SVM models' performance in sentiment analysis tasks, investigate enhancement methodologies such hyperparameter tuning and model ensembling. Multimodal Analysis: To enhance the input features and offer a more thorough grasp of audience emotion toward movies, combine textual evaluations with multimodal data sources, such as user ratings or movie trailers. Fine-grained Sentiment study: To get more in-depth information on how viewers view movies, expand the study to include fine-grained sentiment analysis. In this system, sentiments are partitioned into a couple of classes, similar to positive, neutral, and negative.

By addressing these research trajectories, future studies can advance the precision and efficacy of sentiment analysis systems for IMDb movie reviews, ultimately contributing to a more informed understanding of audience preferences and feedback within film industry.

References

[1] Yenter, A., & Verma, A. (2017, October). Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis. In 2017 IEEE 8th annual ubiquitous computing, electronics and mobile communication conference (UEMCON) (pp. 540-546). IEEE.

[2] Dahir, Ubaid Mohamed, and Faisal Kevin Alkindy. ”Utilizing machine learning for sentiment analysis of IMDB movie review data.” International Journal of Engineering Trends and Technology 71.5 (2023): 18-26.

[3] Danyal, Mian Muhammad, et al. ”Sentiment Analysis Based on Performance of Linear Support Vector Machine and Multinomial Naïve Bayes Using Movie Reviews with Baseline Techniques.” Journal on Big Data 5 (2023).

[4] Jassim, M. A., Abd, D. H., Omri, M. N. Machine Learning-based New Approach to Films.

[5] Singh, S. K., & Singla, N. (2023). Sentiment Analysis on IMDB Review Dataset. Journal of Computers, Mechanical and Management.

[6] Umarani, V., Julian, A., Deepa, J. (2021). Sentiment analysis using various machine learning and deep learning Techniques. Journal of the Nigerian Society of Physical Sciences, 385-394.

[7] Pandunata, P., Nurdiansyah, Y., & Alfina, F. D. (2023). Aspect-Based Sentiment Analysis of Avatar 2 Movie Reviews on IMDb Using Support Vector Machine. In E3S Web of Conferences (Vol. 448, p. 02041). EDP Sciences.

[8] Sharma, A., & Dey, S. (2013). A boosted svm based ensemble classifier for sentiment analysis of online reviews. ACM SIGAPP Applied Computing Review, 13(4), 43-52.

[9] Nahili, W., Rezeg, K., & Kazar, O. (2020, June). Sentiment analysis on product reviews data using supervised learning: A comprehensive review of recent techniques. In Proceedings of the 10th International Conference on Information Systems and Technologies (pp. 1-6).

[10] Topal, K., & Ozsoyoglu, G. (2016, August). Movie review analysis: Emotion analysis of IMDb movie reviews. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1170-1176). IEEE.

[11] Naeem, M. Z., Rustam, F., Mehmood, A., Ashraf, I., Choi, G. S. (2022). Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. PeerJ Computer Science,8,e914.

[12] Tripathi, S., Mehrotra, R., Bansal, V., Upadhyay, S. (2020, September). Analyzing sentiment using IMDb dataset. In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp.30-33).IEEE.

[13] Baid, P., Gupta, A., Chaplot, N. (2017). Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications,179(7),45-49

[14] Sudhir, P., Suresh, V. D. (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. Global Transitions Proceedings,2(2),205-211.

[15] Hourrane, O., Idrissi, N. (2019, October). Sentiment classification on movie reviews and twitter: an experimental study of supervised learning models. In 2019 1st International Conference on Smart Systems and Data Science (ICSSD) (pp.1-6).IEEE.

[16] Venkateswara, S. Sentimental Analysis of Movie Reviews using NLP Techniques.

[17] Ghosh, M., Sanyal, G. (2018). An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. Journal of BigData,5(1),44.

[18] Gupta, K., Jiwani, N., Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. Revue d'Intelligence Artificielle,37(1),1.