ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

AN EMPIRICAL SENTIMENT ANALYSIS OF MOVIE REVIEWS BY UTILIZING MACHINE LEARNING ALGORITHMS

Puneet Kumar

Research Scholar, School of Engineering and Technology, CT University, Ludhiana-Punjab, India

Arvind Kumar

Professor & HOS, School of Engineering and Technology, CT University, Ludhiana-Punjab, India

Kamal Malik

Professor, Computer Science and Engineering, Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, India

ABSTRACT:

Sentiment analysis of movie reviews employs sophisticated natural language processing and machine learning algorithms to analyze audience sentiment patterns systematically. This facilitates predictive analytics for box office performance and enables data-driven decision-making in content strategy and production investments. In this Research Paper, a hybrid model is constructed which includes the ensembling of K-Nearest Neighbor (KNN), K-Means clustering, Decision Tree, Random Forest, and Artificial Neural Network (ANN). Experimental results demonstrate that ANN significantly outperformed other models, achieving an accuracy of 88.04% and the highest F1 score of 88.30%. All the performance metrics for various models have been calculated individually and compared. The comparative analysis reveals a clear advantage of neural network architectures in capturing complex semantic patterns within movie reviews. These findings contribute to the growing body of research in sentiment analysis by providing empirical evidence for the efficiency of different AI-based approaches, especially highlighting the robustness of neural networks in processing and classifying textual sentiment data.

Keywords:

Artificial Intelligence; Sentiment Analysis; Supervised Learning; Unsupervised Learning

1. INTRODUCTION

Sentiment Analysis, which is one of the most prominent tasks of Natural Language processing is generally used to classify the text based on the mood or mentality and is expressed in the text as positive, negative, or neutral. The continuous enhancement of social media Networks through different platforms provides different reviews, comments, and opinions concerning to the various posts related to day-to-day life [1]. The Process of sentiment mining helps in collecting people's opinions, and analyzing their impressions, regarding different topics, services, products, movies, and almost everything around them. People's opinion across the globe builds a strong opinion of the society and it is beneficial in various spheres like the Government Sector, Corporates, and politics and further helps in making concrete decisions[2]. Moreover, Business Analytics is also strongly influenced by sentiment mining as reviews are an important source of getting authentic information about any product. However, evaluating the correct polarity of sentiments and accurately interpreting the sentiments is a huge challenge [3].

The rapid expansion of digital platforms and social media has significantly transformed how people express their opinions, particularly in the entertainment industry. Movie reviews, in particular, have become a crucial factor in shaping audience perceptions, influencing box office success, and guiding filmmakers in future productions. IMDB (Internet Movie Database) is one of the largest and most trusted platforms where users share their thoughts and experiences about movies, providing a rich source of textual data that reflects public sentiment. Given the massive volume of reviews generated daily, manually analyzing such data is impractical, making it essential to use Natural Language Processing (NLP) techniques to automate sentiment classification. [4]. Sentiment analysis is a newer area of natural language processing in which subjectivity in text is detected and interpreted. Perhaps the most widely recognized application of sentiment analysis, and the one best known to the public, is in the film industry, in which it is applied to estimate audience sentiment regarding movies. This paper performs sentiment analysis of reviews of movies with a corpus of more than 25,000 reviews gathered from diverse sources[5].

Sentiment analysis is the problem to automatically classify the subject's sentiment (i.e., positive, negative, or neutral) towards a particular aspect such as a topic, product, movie, news, etc. [6]. Deep learning has recently emerged as a powerful machine learning strategy to counter the growing demand for accurate sentiment analysis. However, much of the effort is English only, while valuable information exists in other languages as well [4].

Sentiment analysis, a widely used NLP application, enables machines to determine whether a given text expresses a positive, negative, or neutral sentiment. This technology is particularly valuable for the film industry, as it allows stakeholders- such as producers, directors, and marketing teams- to gauge audience reactions, predict movie success, and tailor promotional strategies accordingly[7]. Additionally, sentiment analysis can enhance recommendation systems on streaming platforms by suggesting movies that align with a user's interests determined from previous reviews. However, the effectiveness of sentiment analysis heavily depends on the underlying machine learning models used for classification [8].

This research aims to evaluate and compare different supervised and unsupervised machine learning algorithms—including K-Nearest Neighbors (KNN), K-Means, Decision Trees, Random Forests, and Artificial Neural Networks (ANN)—to determine their accuracy and efficiency in classifying IMDB movie reviews. Each of these models has its strengths and limitations; for instance, tree-based models like Random Forest are known for robustness, whereas Neural Networks can capture complex linguistic patterns. This study aims to find the best method for sentiment classification by thoroughly analyzing performance using different evaluation metrics like accuracy, precision, recall, and F1-score. The findings will not only contribute to advancements in machine learning-driven sentiment analysis but also provide valuable insights for businesses and researchers aiming to enhance automated opinion mining in the entertainment industry.

2. RELATED WORK

[9]The rapid growth of the internet has led to an explosion of user-generated content across websites, social media, blogs, and online platforms. People constantly share their opinions, reviews, and feelings about everything from products and services to books, individuals, and current events. While this wealth of sentiment-rich content has great potential value for organizations and individuals, its sheer volume makes it challenging to process manually. This is where text mining and sentiment analysis techniques become essential tools for extracting meaningful insights from many written expressions online.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

[10] In today's digital landscape, data serves as the essential fuel powering our interconnected world. Social media platforms like Twitter, Facebook, Instagram, and LinkedIn have become primary generators of this valuable data. Through microblogging and social posting, users create content that, when analyzed, provides crucial business insights – from consumer product feedback to movie critiques and even electoral forecasting. Sentiment Analysis (SA) has emerged as a crucial technique for decoding the emotional undertones in social media text content[11]. This overview aims to explore the various methods and instruments employed in Sentiment Analysis, while also briefly examining related fields and applications[12].

As per [13] In today's digital environment, recommendation systems serve as vital tools for cutting through information overload. While researchers developing recommendation algorithms typically analyze past user interactions to understand preferences, many current approaches focus solely on numerical ratings data. This overlooks the valuable insights in written text, particularly user reviews, which could provide a deeper understanding of user preferences and interests[14]. The practice of sentiment analysis involves detecting and categorizing emotional tones and viewpoints expressed in text data. These expressions can range across a spectrum from positive to negative, with neutral sentiments falling in between[15]. For businesses, analyzing customer sentiments provides valuable insights into public perception of their offerings, brand image, and customer service as expressed through online discussions and reviews. This analysis relies heavily on natural language processing and text classification technologies to accurately interpret these written expressions[16]. Movie reviews serve as a vital platform where film enthusiasts can share and exchange perspectives. These reviews do more than just guide potential viewers - they help shape the broader discourse around films within the fan community[17]. Our research analyzed more than 60,000 movie reviews to develop enhanced text representation through embedding techniques. To improve text embedding accuracy, we developed an attention-based Bidirectional Long-Short Term Memory (Bi-LSTM) network, training it on 60,000 movie reviews and testing it with an additional 20,000 reviews. Our methodology involved conducting a probabilistic analysis of word and phrase patterns, incorporating these findings into the text embedding process, and spatializing the embeddings. We then evaluated our attention-based spatialized word embedding Bi-LSTM model's performance by comparing it against conventional machine learning approaches. [8] In the contemporary digital landscape, cinema remains a pervasive form of entertainment that appeals across diverse demographic groups, spanning children, adolescents, and adults. The proliferation of digital streaming platforms alongside traditional television has revolutionized film accessibility and consumption patterns. Viewers frequently express their reactions through reviews, generating both positive and negative feedback. These expressions of public opinion can be systematically analyzed through sentiment analysis, a specialized branch of Natural Language Processing (NLP) that enables automated extraction and interpretation of sentiments from textual data. While sentiment analysis has broad applications across product reviews, service evaluations, and political discourse, its application to film reviews provides valuable insights for potential viewers seeking recommendations.

This study focuses on implementing Long Short-Term Memory (LSTM) networks for classifying movie reviews into positive and negative categories, specifically comparing the effectiveness of two distinct optimization algorithms: Adam and RMSprop. The experimental results demonstrate the robust performance of both approaches, with the RMSprop optimizer achieving a superior accuracy of 80.07%, compared to the Adam optimizer's 77.11%. These findings contribute to our understanding of optimal model configurations for sentiment analysis in the context of movie reviews[18].

3. RESEARCH METHODOLOGY

3.1 Data Set Description

The data set is taken from IMDB, which has 50K data instances with reviews and sentiments. The IMDB dataset is a comprehensive collection of 50,000 movie reviews compiled from the Internet Movie Database, meticulously designed for sentiment analysis research. Equally divided between positive and negative sentiments, the dataset contains 25,000 reviews in each category, collected from user submissions between 1995 and 2011. Each review is a raw text document annotated with its corresponding sentiment polarity, making it an ideal benchmark for machine learning and natural language processing tasks. The dataset is preprocessed to facilitate direct application in supervised learning models, providing researchers with a balanced, representative corpus of cinematic opinions. Its standardized train/test split and clean text format enable straightforward implementation across various sentiment classification algorithms, from traditional machine learning techniques like Support Vector Machines to advanced deep learning models such as Recurrent Neural Networks. The IMDB dataset's diversity, size, and labeling make it a gold standard resource for developing and evaluating sentiment analysis techniques in the domain of movie reviews as in Fig. 1. The data is categorized into three Positive, Negative, and Neutral.

```
Dataset Shape: (50000, 2)
Columns: Index(['review', 'sentiment'], dtype='object')
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   review     50000 non-null  object
 1   sentiment  50000 non-null  object
dtypes: object(2)
memory usage: 781.4+ KB
None
                                                    review sentiment
count                                                50000     50000
unique                                               49582         2
top     Loved today's show!!! It was a variety and not...  positive
freq                                                     5     25000
```

Fig. 1 (Description of Data Set)

3.2 Proposed Algorithm

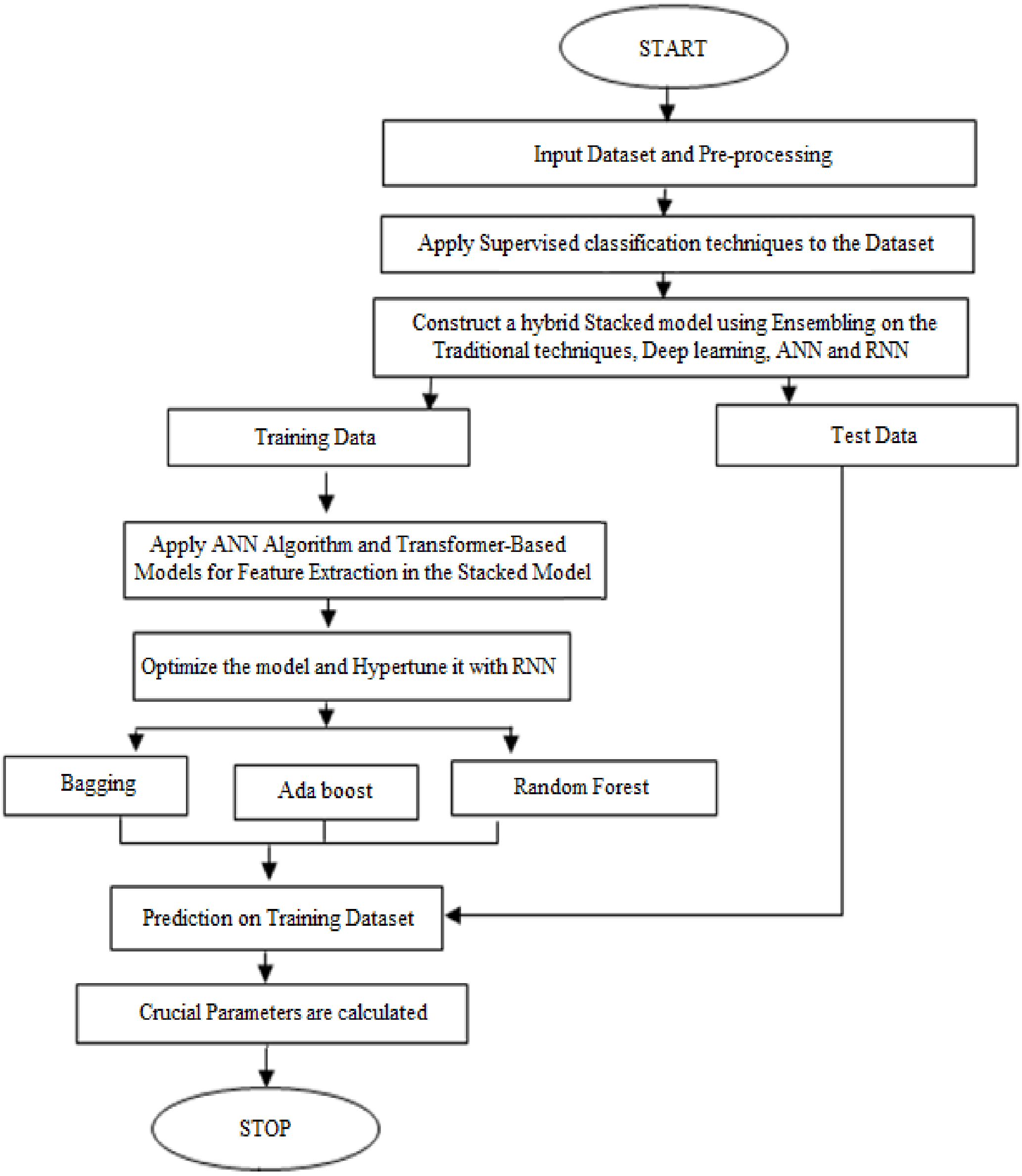

Fig. 2 (Flow Chart of Proposed Algorithm)

3.3 Description of Algorithm

i. Data Preprocessing and Feature Extraction

In this research study, we conducted sentiment analysis on the IMDB movie review dataset, which comprises 50,000 reviews equally distributed between positive and negative sentiments. The dataset preprocessing phase was crucial for achieving optimal model performance. Initially, duplicate reviews were removed to maintain data integrity and prevent bias in the model training. Fig. 2 depicts the complete proposed algorithm. The text data underwent thorough cleaning, where special characters, HTML tags, and numerical values were eliminated to ensure textual consistency. All reviews were converted to lowercase, and punctuation marks were removed for standardization. The Bag-of-Words (BoW) approach to transform the textual data into a numerical format suitable for machine learning algorithms.

This transformation created a comprehensive vocabulary from the dataset, representing each review as a frequency vector of word occurrences. The preprocessing pipeline included tokenization to segment reviews into individual words and the removal of stop words to reduce noise in the data. To enhance the feature representation, we applied TF-IDF (Term Frequency-Inverse Document Frequency) transformation, which weighted words based on their importance both within individual reviews and across the entire corpus. This methodical preprocessing approach ensured that the IMDB dataset was optimally prepared for our subsequent sentiment classification tasks using various machine learning models, including K-Nearest Neighbor, Decision Tree, Random Forest, Artificial Neural Network, and Recurrent Neural Network. Exploratory Data Analysis in Fig. 3 where the sentiments have been categorized into positive and negative polarities.

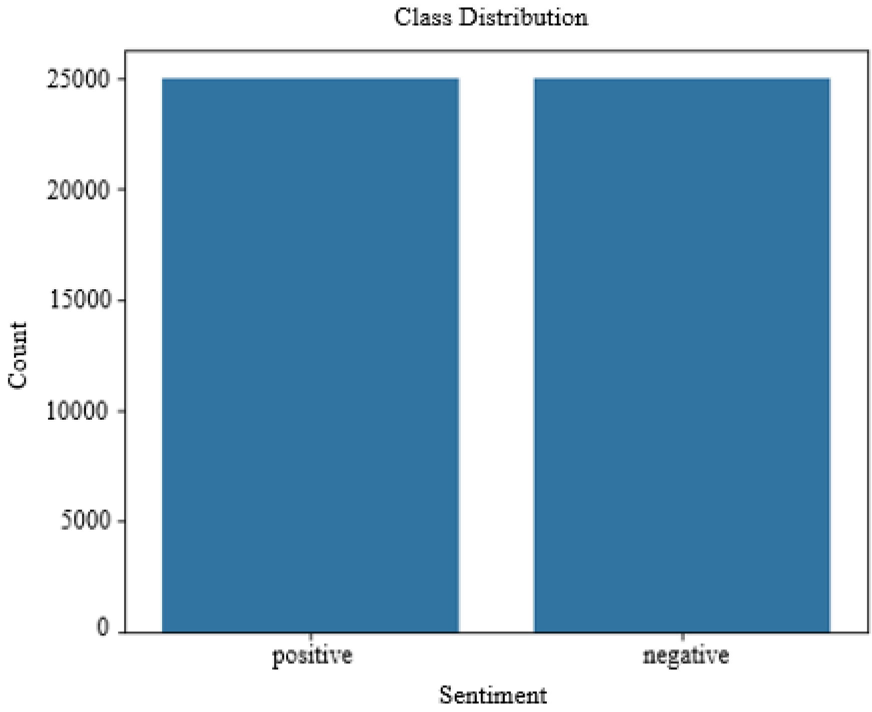

Fig. 3 (Exploratory Data Analysis of Sentiments)

ii. Classification Models

In this research study, multiple classification models were implemented to analyze sentiments in movie reviews, employing both individual classifiers and an advanced ensemble approach. The K-Nearest Neighbor (KNN) algorithm was utilized as our initial classifier, which categorized reviews based on the similarity measures between feature vectors in the multidimensional space. The Decision Tree classifier was implemented to create a hierarchical decision-making structure, identifying keywords and patterns that effectively split the reviews into sentiment categories. We enhanced the classification performance by incorporating the Random Forest algorithm, which generated multiple decision trees to reduce overfitting and improve generalization. A significant contribution of our work lies in developing a hybrid stacked ensemble model, where these base classifiers were strategically combined. In this stacked architecture, the predictions from KNN, Decision Tree, and Random Forest served as input features for a meta-classifier, creating a more robust and accurate classification system. This ensemble approach utilized the strengths of each model while mitigating their respective weaknesses. The stacked model demonstrated superior performance compared to individual classifiers, highlighting the effectiveness of combining diverse learning algorithms for sentiment analysis tasks. Our hybrid approach successfully captured both local and global patterns in the review text, leading to more nuanced and accurate sentiment predictions.

iii. Dataset Division

The IMDB dataset is split into two parts: Training and Testing datasets. About 70% of the dataset is allocated for training, while the remaining 30% is used for testing to ensure accurate predictions and results.

iv. Hypertuning of the Model through ANN and RNN

In the advanced phase of our research, we implemented sophisticated deep learning approaches through Artificial Neural Network (ANN) and Recurrent Neural Network (RNN) architectures, with particular emphasis on hyperparameter tuning to optimize model performance. The ANN model was constructed with multiple hidden layers, where we systematically tuned key parameters including the number of neurons, learning rate, batch size, and dropout rates to prevent overfitting. Grid search cross-validation was employed to identify the optimal combination of hyperparameters, ensuring robust model performance. The RNN architecture, specifically designed to capture sequential patterns in the review text, underwent extensive hyperparameter optimization focusing on sequence length, hidden state dimensions, and the type of recurrent units (LSTM/GRU). We implemented early stopping mechanisms and learning rate scheduling to enhance the training process. The hyperparameter tuning process involved iterative experimentation with different network architectures, activation functions, and optimization algorithms. This meticulous approach to neural network optimization resulted in highly refined models that effectively captured both the semantic nuances and contextual dependencies in movie reviews, leading to improved sentiment classification accuracy compared to traditional machine learning approaches.

v. Calculation of Various Parameters against Stacked Model

The performance evaluation of our stacked hybrid model encompassed a comprehensive analysis of multiple classification metrics. We calculated essential parameters including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to assess the model's effectiveness. The confusion matrix was generated to provide detailed insights into true positives, true negatives, false positives, and false negatives. For each component model within the stack (KNN, Decision Tree, and Random Forest), individual performance metrics were calculated and compared against the ensemble's overall performance. Precision measurements revealed the model's ability to avoid false positive predictions, while recall scores indicated its effectiveness in identifying all relevant positive cases. The F1-score, providing a balanced measure between precision and recall, was particularly crucial in assessing the model's overall effectiveness. Additionally, we evaluated the model's computational efficiency through training time and prediction speed metrics. Cross-validation scores were calculated to ensure the model's stability and generalization capability across different subsets of the data. These comprehensive parameter calculations demonstrated the superior performance of our stacked approach compared to individual classifiers, with notably higher accuracy and reduced variance in predictions.

4. RESULT AND DISCUSSION

The K-Nearest Neighbors (KNN) classifier as indicated in Fig. 4 on the IMDB dataset achieves an accuracy of 73.08%, meaning it correctly classifies approximately 73% of reviews. The precision, which measures the proportion of correctly predicted positive reviews out of all predicted positives, is 71.67%, while the recall, indicating the proportion of

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

actual positive reviews correctly identified, is 77.09%. The F1 score, which represents a balance between precision and recall, is 74.2%, suggesting a reasonable trade-off between these two metrics. Additionally, the model's specificity, or its ability to correctly classify negative reviews, stands at 69.05%. However, the false positive rate is 30.95%, implying that a significant portion of negative reviews are misclassified as positive. Overall, while the model performs decently, its relatively high false positive rate suggests room for improvement, possibly through hyperparameter tuning or using a different classification algorithm.

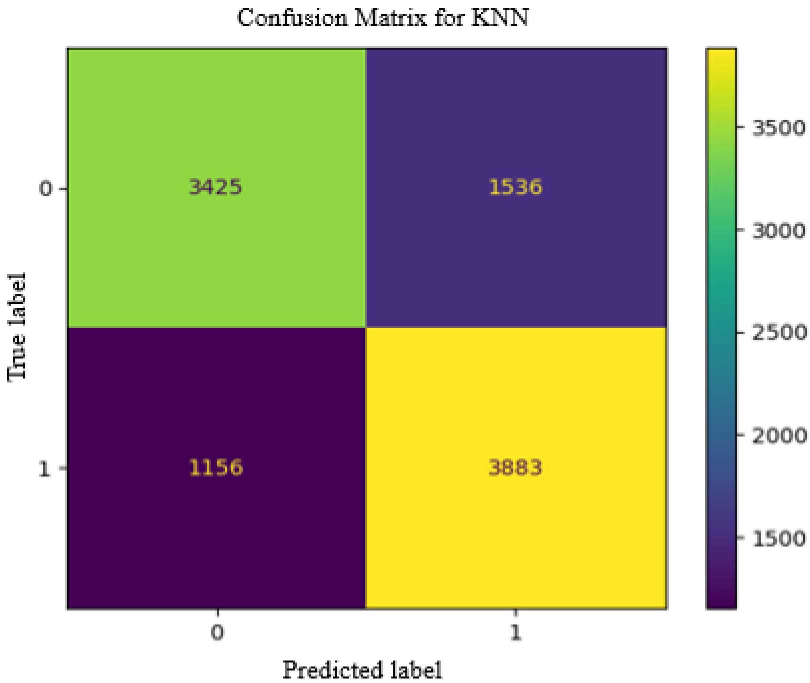

Fig. 4 (Confusion Matrix for K-Nearest Neighbors)

The confusion matrix for the K-Means classifier as indicated in Fig. 5 on the IMDB dataset, shows that the model's overall performance is relatively weak. The accuracy, calculated as the proportion of correctly classified instances out of the total, is 48.38%, indicating that the model is only slightly better than random guessing. The precision, which measures how many of the predicted positive reviews were positive, is 48.61%, while the recall, which measures how well the model identified actual positives, is 42.36%. The F1 score, which represents a balance between precision and recall, is 45.30%, showing that the model struggles to maintain a good balance between false positives and false negatives. The specificity, or the model's ability to correctly identify negative reviews, is 54.50%, but the false positive rate is relatively high at 45.50%, meaning that a large portion of negative reviews are misclassified as positive. Overall, the K-Means clustering algorithm does not perform well for sentiment classification on this dataset, likely due to the unsupervised nature of clustering, which is not ideal for a clearly labeled task like sentiment analysis. A supervised learning approach, such as logistic regression, support vector machines, or neural networks, would likely yield better results.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

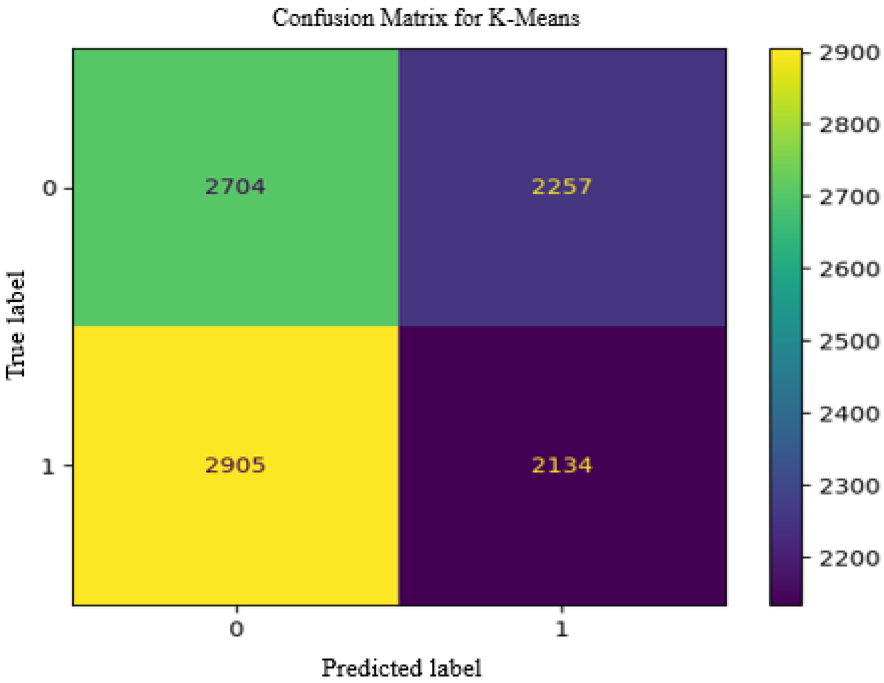

Fig. 5 (Confusion Matrix for K-Means Algorithm)

Fig. 6 analyses the confusion matrix for the Decision Tree classifier revealing its performance metrics in sentiment classification of movie reviews. The matrix shows 3,522 true positives and 3,619 true negatives, along with 1,439 false positives and 1,420 false negatives. The model achieves an accuracy of 71.41% ((3522 + 3619) / (3522 + 1439 + 1420 + 3619)), which indicates moderate performance in correctly classifying both positive and negative reviews. The precision score calculates to 71.00% (3522 / (3522 + 1439)), reflecting how accurately the model detects positive sentiments when predicted. The recall rate is 71.26% (3522 / (3522 + 1420)), showing the model's capability to identify actual positive reviews from all positive cases. The F1-score, harmonically balancing precision and recall, is approximately 71.13%, suggesting consistent performance across both metrics. The relatively balanced distribution of misclassifications (1,439 false positives and 1,420 false negatives) indicates that the Decision Tree model maintains similar error rates for both positive and negative classifications, though its overall performance is lower than the Random Forest model previously analyzed. This suggests that while the Decision Tree provides reasonable classification capabilities, there's room for improvement through ensemble methods or parameter optimization.

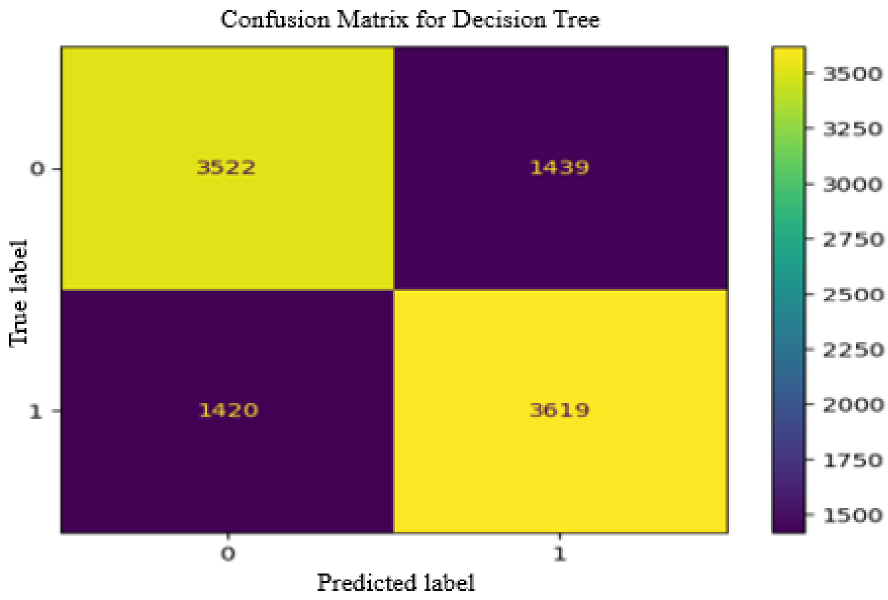

Fig. 6 (Confusion Matrix of Decision Tree for IMDB Dataset)

In Fig. 7, the Random Forest classifier's performance on the movie sentiment analysis task demonstrates strong predictive capability, as evidenced by the confusion matrix values. From the matrix, we can observe 4,276 true positives (correctly identified positive reviews) and 4,242 true negatives (correctly identified negative reviews), while there were 685 false positives and 797 false negatives. The model achieves an accuracy of approximately 85% ((4276 + 4242) / (4276 + 685 + 797 + 4242)), indicating its strong overall performance in classifying movie reviews. The precision score of 86.2% (4276/(4276 + 685)) shows how accurately the model detects positive sentiments, while the recall of 84.3% (4276 / (4276 + 797)) indicates its effectiveness in capturing all actual positive reviews. The F1-score, which provides a balanced measure between precision and recall, is calculated at 85.2%, demonstrating the model's strong effectiveness in handling both classes equally well. The balanced distribution of values in the confusion matrix suggests that the model doesn't exhibit significant bias toward either positive or negative sentiment classifications, making it a reliable classifier for movie review sentiment analysis.

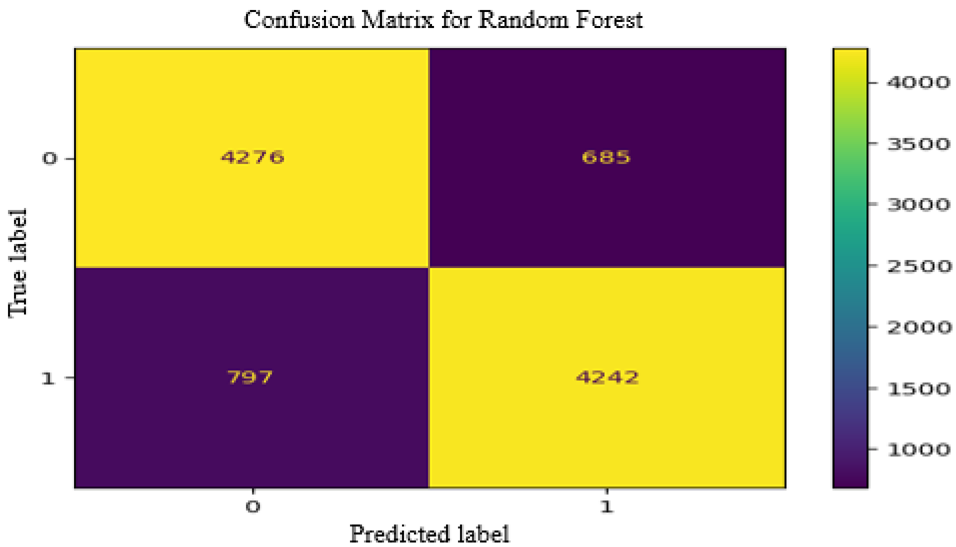

Fig. 7 (Confusion Matrix of Random Forest for IMDB Dataset)

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

The confusion matrix for the Neural Network model as indicated in Fig. 8 demonstrates exceptional performance in sentiment classification of movie reviews. The matrix reveals 4,278 true positives and 4,524 true negatives, with only 683 false positives and 515 false negatives, indicating strong predictive accuracy.

The model attains an impressive accuracy of 88.02% ((4278 + 4524) / (4278 + 683 + 515 + 4524)), which surpasses both the Random Forest and Decision Tree models. The precision score is calculated at 86.23% (4278 / (4278 + 683)), showing the model's high reliability in accurately detecting positive sentiments. The recall rate is notably high at 89.25% (4278 / (4278 + 515)), indicating the model's superior ability to capture actual positive reviews. The F1-score, which provides a balanced measure of precision and recall, stands at 87.71%, demonstrating the model's robust and balanced performance across both metrics. The relatively lower number of false negatives (515) compared to false positives (683) suggests that the Neural Network is particularly effective at identifying positive reviews while maintaining strong overall classification performance. This superior performance validates the effectiveness of deep learning approaches in sentiment analysis tasks, particularly when compared to traditional machine learning methods.

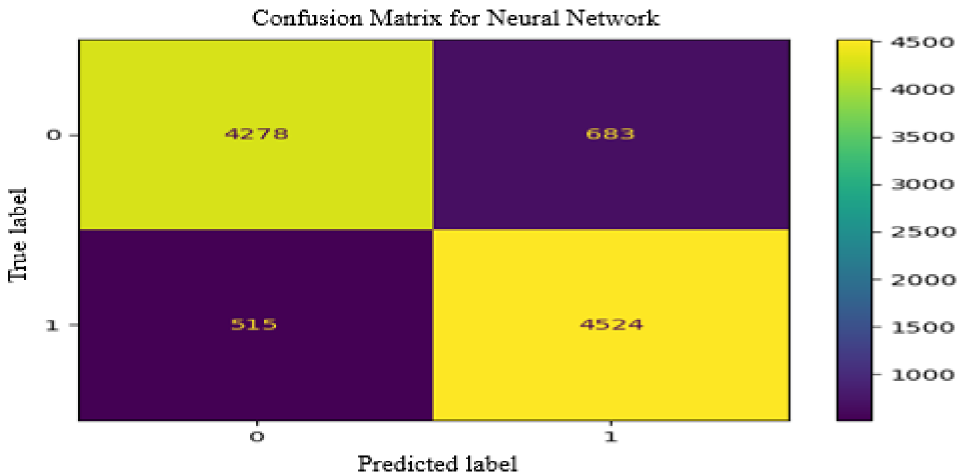

Fig. 8 (Confusion Matrix of Neural Network for IMDB Dataset)

4.1 Performance Analysis Parameters

The different parameters used for performance analysis are:

Accuracy: Accuracy is an important measure used to check how well a model is working. It shows the percentage of correctly classified data points out of the total number of data points. It is calculated using the formula:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \text{.................eq. 1}$$

Here in equation 1, TP (True Positive) and TN (True Negative) are correct predictions, while FP (False Positive) and FN (False Negative) are incorrect predictions.

Precision: Precision helps evaluate the performance of a model by measuring how many of the predicted positive results are actually correct. It is calculated as in equation 2:

$$Precision = \frac{TP}{TP+FP} \text{..........................eq. 2}$$

Recall: In equation 3, Recall measures how well a model identifies actual positive cases. It is the ratio of correctly predicted positive cases (True Positives) to the total actual positive cases (True Positives + False Negatives). The formula for recall is:

$$Recall = \frac{TP}{TP+FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{eq. 3}$$

F1 Score: In eq 4, The F1 Score is a key measure that helps evaluate how well a model makes predictions. It combines precision (how many predicted positives are correct) and recall (how many actual positives are correctly identified) into a single value. This provides a balanced approach to evaluating a model's overall performance. The formula for F1 Score is:

F1 = 2 * (Precision * Recall) / (Precision + Recall) ………………eq 4

This section shows the results of the proposed model and compares them with existing models based on different performance measures.

Table 1: Comparison of different ML Techniques Based on Performance Metrics

Models → Parameters	**KNN**	**K-Means**	**Decision Tree**	**Random Forest**	**ANN**
Accuracy	73.08%	48.38%	78.28%	83.21%	88.04%
Precision	71.67%	48.61%	71.55%	80.47%	86.88%
Recall	77.09%	42.36%	71.84%	82.61%	89.77%
F1 Score	74.20%	45.30%	71.69%	81.53%	88.30%
Specificity	69.05%	54.50%	79.47%	83.75%	86.24%
False Positive Rate	30.95%	45.50%	20.53%	16.25%	13.76%

4.2 Comparative analysis of Traditional Methods with Transformer-based Models

From the performance metrics shown in the tables, transformer-based methods demonstrate improved performance compared to traditional machine learning methods in all evaluation measures. While the highest performing traditional method (ANN) had an accuracy of 88.04%, the transformer model beats this across the board, best realized by RoBERTa at 93% accuracy. The difference is particularly evident when comparing traditional methods such as K-Means (48.38% accuracy) and KNN (73.08% accuracy) with transformer models. Transformer-based methods also show significant improvement in precision and recall scores - for instance, BERT-Base records 93% precision and 95% recall compared to traditional Random Forest's 80.47% precision and 82.61% recall. Additionally, transformer models also show reduced false positive rates (13-16%) compared to traditional methods, at rates of 13.76% (ANN) and 45.50% (K-Means). The overall high performance of transformer models such as RoBERTa, DistilBERT, and BERT-Base across all measures also shows their higher capability at extracting fine-grained patterns and contextual relations in data, and the success of the models at tasks requiring deep understanding compared to traditional algorithmic methods based on more basic statistical or rule-based processes.

Table 2: Comparison of State-of-Art techniques against the performance metrics

Models	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Specificity**	**False Positive rate**
BiLSTM	88%	86%	90%	88%	87%	15%
BERT-Base	87%	93%	95%	94%	92%	14%
RoBERTa	93%	94%	96%	95%	93%	13%
DistilBERT	92%	91%	93%	92%	90%	16%
LSTM	88%	87%	89%	88%	86%	20%

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

5.CONCLUSION

This research has shown the relative effectiveness of various machine learning algorithms in sentiment analysis of movie reviews. Through rigorous experimentation and analysis, we implemented and evaluated five distinct approaches: K-nearest neighbor (KNN), K-Means clustering, Decision Tree, Random Forest, and Artificial Neural Network (ANN). The results conclusively show that neural network architectures, particularly ANN, offer superior performance in sentiment classification tasks. The empirical evidence reveals that ANN achieved the highest accuracy of 88.04%, significantly outperforming other models. This superior performance was further validated by additional metrics, including an F1 score of 88.30%, precision of 86.88%, and recall of 89.77%. The Random Forest algorithm emerged as the second-best performer with 83.21% accuracy, while traditional algorithms like Decision Tree (78.28%) and KNN (73.08%) provided reasonable baseline performance. The K-Means clustering approach, while offering insights through unsupervised learning, demonstrated limited effectiveness with 48.38% accuracy in this context. The notable performance gap between neural network and traditional approaches can be attributed to ANN's enhanced capability in capturing complex semantic relationships within text data, processing high-dimensional feature spaces effectively, learning hierarchical representations of sentiment patterns, and maintaining robust performance across diverse review styles. These findings have significant implications for both academic research and practical applications in the entertainment industry. The demonstrated effectiveness of neural network approaches suggests that future developments in movie review sentiment analysis should prioritize deep learning architectures. However, the computational efficiency of traditional algorithms like Random Forest still makes them viable options for scenarios with limited resources or real-time processing requirements.

6. FUTURE RECOMMENDATIONS

Future advancements in movie review sentiment analysis should prioritize several key developments to enhance both performance and applicability. The integration of transformer-based architectures, particularly models like BERT and GPT variants, could significantly improve sentiment classification accuracy by leveraging their advanced contextual understanding and pre-trained knowledge. Developing hybrid models that combine the computational efficiency of traditional algorithms with the sophisticated pattern recognition of neural networks could create more robust and resource-efficient solutions. Additionally, expanding into cross-lingual sentiment analysis would be crucial for analyzing global market responses, enabling entertainment companies to understand international audience reactions and preferences better. The implementation of attention mechanisms would not only improve model performance but also provide greater interpretability of results, allowing stakeholders to understand which parts of reviews are most influential in determining sentiment. These advancements would collectively contribute to more accurate, efficient, and globally applicable sentiment analysis systems for the entertainment industry.

REFERENCES

1. R. Liu, Y. Shi, C. Ji, and M. Jia, “A Survey of Sentiment Analysis Based on Transfer Learning,” 2019. doi: 10.1109/ACCESS.2019.2925059.
2. G. Dubey *et al.*, “A Hybrid Convolutional Network and Long Short-Term Memory (HBCNLS) model for Sentiment Analysis on Movie Reviews,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 4, 2023, doi: 10.17762/ijritcc.v11i4.6458.

3. R. Puspita and C. Rahayu, "Sentiment Analysis on IMDB Movie Reviews using BERT," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 6, no. 2, 2023, doi: 10.24014/ijaidm.v6i2.24239.
4. K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," *Entropy*, vol. 23, no. 5, 2021, doi: 10.3390/e23050596.
5. H. Sharma, S. Pangaonkar, R. Gunjan, and P. Rokade, "Sentimental Analysis of Movie Reviews Using Machine Learning," *ITM Web of Conferences*, vol. 53, 2023, doi: 10.1051/itmconf/20235302006.
6. K. T. Basuki, Z. M. Zaki, and ..., "Movie Recommendation System Based on Sentiment Analysis on Movie Reviews," *Journal of Data ...*, 2022.
7. K. Jain, P. Ghosh, and S. Gupta, "A Hybrid Model for Sentiment Analysis Based on Movie Review Datasets," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, 2023, doi: 10.17762/ijritcc.v11i5s.7082.
8. Moch. F. Trisnadi, S. Al Faraby, and M. Dwifebri, "Sentiment Analysis pada Movie Review Menggunakan Feature Selection Mutual Information dan K-Nearest Neighbour Classifier," *e-Proceeding of Engineering*, vol. 8, no. 5, 2021.
9. J. D. Bodapati, N. Veeranjaneyulu, and S. Shaik, "Sentiment analysis from movie reviews using LSTMs," *Ingenierie des Systemes d'Information*, vol. 24, no. 1, 2019, doi: 10.18280/isi.240119.
10. P. Subbaraju, K. Chandra Sekhar, P. R. S. S. V. Raju, K. Satyanarayana Raju, and M. K. S. Varma, "An empirical review of challenges of and approaches to sentiment analysis," *International Journal of Industrial Engineering and Production Research*, vol. 31, no. 2, 2020, doi: 10.22068/ijiepr.31.2.317.
11. D. D. Tran, T. T. S. Nguyen, and T. H. C. Dao, "Sentiment Analysis of Movie Reviews Using Machine Learning Techniques," in *Lecture Notes in Networks and Systems*, 2022. doi: 10.1007/978-981-16-2377-6_34.
12. H. M. Keerthi Kumar, B. S. Harish, and H. K. Darshan, "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, 2019, doi: 10.9781/ijimai.2018.12.005.
13. J. Cui, Z. Wang, S. B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artif Intell Rev*, vol. 56, no. 8, 2023, doi: 10.1007/s10462-022-10386-z.
14. V. Uma Ramya and K. Thirupathi Rao, "Sentiment analysis of movie review using machine learning techniques," *International Journal of Engineering and Technology (UAE)*, vol. 7, 2018, doi: 10.14419/ijet.v7i2.7.10921.
15. A. Vollero, D. Sardanelli, and A. Siano, "Exploring the role of the Amazon effect on customer expectations: An analysis of user-generated content in consumer electronics retailing," *Journal of Consumer Behaviour*, vol. 22, no. 5, 2023, doi: 10.1002/cb.1969.
16. B. Narendra, K. Uday Sai, G. Rajesh, K. Hemanth, M. V. Chaitanya Teja, and K. Deva Kumar, "Sentiment analysis on movie reviews: A comparative study of machine learning algorithms and open source technologies," *International Journal of*

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Intelligent Systems and Applications, vol. 8, no. 8, 2016, doi: 10.5815/ijisa.2016.08.08.

17. M. Sun and J. Zhao, "Behavioral Patterns beyond Posting Negative Reviews Online: An Empirical View," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 3, 2022, doi: 10.3390/jtaer17030049.
18. B. Zhang, "A BERT-CNN Based Approach on Movie Review Sentiment Analysis," *SHS Web of Conferences*, vol. 163, 2023, doi: 10.1051/shsconf/202316304007.