Sentiment Classification and Analysis of Movie Reviews Using NLP Techniques

ABID AL MAMUN, 19301066 BRACU
MD. AJMAIN AOSAF ANAN, 20101388 BRACU
ABDULLAH RAHAT, 20301134 BRACU
NUSAIBAH RAHAT, 20301136 BRACU
SHUVO KARMOKAR, 20301441 BRACU

In today's world movie reviews play an important role in letting people know how the movie is. By understanding the sentiment of the review we can know how a person feels about the movie. Now reading movie reviews manually is a lengthy process. There is already some research on analyzing the review sentiments using machine learning and other methods. We took a different approach to analyze the sentiments of the reviews by using Logistic regression, Linear SVM, Multinomial Naive Bayes, and RNN model. We tested it on a dataset containing thousands of reviews. The results were superior to most other research done till date. Our goal is to improve the understanding of the reviews and help the industry use these reviews and make movies for everyone to enjoy.

Additional Key Words and Phrases: Sentiment Analysis, Logistic Regression, Multinomial Naive Bayes, Support Vector Machine, Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), Performance Evaluation, Recurrent Neural Network

ACM Reference Format:
Abid Al Mamun, Md. Ajmain Aosaf Anan, Abdullah Rahat, Nusaibah Rahat, and Shuvo karmokar. 2024. Sentiment Classification and Analysis of Movie Reviews Using NLP Techniques. 1, 1 (May 2024), 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Understanding a language is more than just knowing words. It is more like knowing the meaning and connecting the words. Traditional methods of wording do not do well in understanding a language. On the other hand, Deep learning methods excel in understanding words used in different scenarios. methods have been used already in many scenarios from properly understanding words to retrieving large documents.

In this paper, we shall focus on sentiment analysis. Or, in other words, try to understand the human expression through text specifically on movie reviews. Analyzing the sentiments of movie reviews not only shows how people feel about a movie, it provides valuable insights to filmmakers and advertisers.

We have already reviewed some papers from 2006, 2016 and 2018. They tell us about different ways to understand feelings from basic ideas to newer computer programs.

Our plan is to find the proper meaning of the words in different scenarios and the meaning they carry. Traditional methods are good at understanding the meaning but the miss the feeling behind them. So, we came up with newer

Authors' addresses: Abid Al Mamun19301066, BRACU; Md. Ajmain Aosaf Anan20101388, BRACU; Abdullah Rahat20301134, BRACU; Nusaibah Rahat20301136, BRACU; Shuvo karmokar20301441, BRACU.

models that learn both the meaning and the underlying sentiments of the words. We tested the models using a dataset containing 50,000 movie reviews, where people expressed their feeling openly.

After testing our models, we can see that our models are pretty good at figuring out feelings from the reviews with RNN having the highest accuracy of 84

Our aim is to let the reviews of audiences reach the industry more easily so that they can make better decisions in making movies knowing what audiences like and what not.

2 BACKGROUND

Knowing people's thoughts about movies is very important especially when there are so many reviews online. Some notable studies have paved the way for understanding people's feelings using computers. The first time getting the Nobel Prize in 2006 highlighted the requirement to to find out opinions from written reviews. Then researchers like Dey, El-Din and Ahuja tried to find different ways to do this. Using upgraded techniques like Naïve Bayes and Support Vector Machines.

The incorporation of deep learning methodologies have revolutionized sentiment analysis, which has been done by enabling the development of complex neural network architectures which are capable of capturing nuanced emotions and contextual cues from movie reviews. (Can et al. 2018 and Wang et al. 2016)Deeper understanding of audience preferences, sentiments and engagement levels have been facilitated by these advancements. As a result, stakeholders will be empowered to tailor their content strategies and marketing efforts.

So now, appreciating all the works, we have powerful tools to understand the audience's reaction to movies. It helps filmmakers and others in the industry to know the audience's demand. This helps to predict the result and make smarter decisions. Also,- the researchers are working to make the tools better.

3 LITERTURE REVIEW

Obsa Gelchu Horsa (2023) proposed an Aspect-Based Sentiment Analysis for Afaan Oromoo Movie Reviews Using Machine Learning Techniques.Aspect-based sentiment analysis (ABSA) research covers different languages and topics, showcasing its grip on its importance and understanding detailed opinions. This study aims to create algorithms and techniques for ABSA using machine learning and natural language processing (NLP). research in languages like Afaan Oromoo highlights the need for stemmers and lemmatization for betterment of sentiment analysis. For example, Debela's stemmers had been used to find different word forms and their base form, and that greatly improved sentiment analysis accuracy percentage. Bag of words and TF-IDF are majorly used for feature extraction, also capturing minor sentiment differences. Aspect term extraction (ATE) is very important in ABSA. Helping to identify the specific parts that were mentioned in the reviews. Using Inter-Annotator Agreement (IAA) metrics like Cohen's coefficient makes sure that the labeled datasets are reliable. The model's performance can be improved by using tuning hyperparameters. This could cause better results in accuracy and precision. ABSA is important in many other languages and shows advanced techniques to grab opinions more effectively.

In order to better understand audience sentiments, sentiment analysis for movies is a crucial topic that this study (Xu, 2024, 31-37) delves into. The advent of deep learning, especially ConvLSTM models, holds promise in capturing delicate feelings, yet older techniques like SVM and Naive Bayes are inadequate in this regard. In experiments, ConvLSTM outperforms both LSTM and LSTM with attention layers in capturing sequential dependencies in movie reviews. This is because it was introduced in this work. These developments have the ability to completely transform the film industry by improving market research and movie recommendation systems in addition to helping filmmakers make decisions.Nevertheless, despite ConvLSTM's potential, it is important to recognize its limitations. The study uses data from IMDb, which may not accurately reflect the varied opinions of audiences on different platforms and in different languages. Even when preprocessing procedures are extensive, biases could still exist. Further investigation into model comparisons and comprehension of performance-influencing elements may also prove advantageous. The study's applicability and impact might be increased by taking into account demographic changes and real-world circumstances.

4 METHOD

The methodology for sentiment classification and analysis of movie reviews using natural language processing (NLP) techniques draws inspiration from seminal research conducted by Dey et al. (2016), El-Din (2016), Ahuja et al. (2019), and Can et al. (2018). This section outlines a comprehensive approach encompassing data collection, preprocessing, feature extraction, model selection, and evaluation.

The author (Xu, 2024, 31-37) used IMDb's dataset of labeled movie reviews, cleaned it up, and then trained a ConvLSTM model to understand sentiments. This model is pretty good—it combines convolutional and LSTM layers to grasp both short-term and long-term patterns in text, making it great for sentiment analysis. The authors trained it to minimize loss using binary cross-entropy and fine-tuned it using Python, Keras, and the Adam optimizer for optimal performance. In this project, the aim is to perform sentiment analysis on the IMDB movie review dataset.IMDB dataset having 50K movie reviews for natural language processing or Text analytics. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. They provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, we will predict the number of positive and negative reviews using either classification or deep learning algorithms.

Dataset: Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011 The methodology can be divided into several key steps including Exploratory Data Analysis (EDA), Data Preprocessing, and Feature Extraction.

4.1 Exploratory Data Analysis (EDA)

EDA is crucial to gain insights into the dataset before diving into modeling. This involves understanding the structure of the data, identifying any patterns or trends, and visualizing key characteristics.

In the provided code, EDA is performed through the following steps: Checking the dimensions of the dataset using df.shape to understand the number of rows and columns.

Displaying the First Few Rows: Using `df.head()` to get a glimpse of the first few rows of the dataset, providing an overview of the data. Employing `df.describe()` to get statistical information about the numerical features. Using `df['sentiment'].value_counts()` to check the distribution of sentiment classes, which is essential for understanding class balance.

4.2 Data Preprocessing

Data preprocessing involves cleaning and transforming the raw data into a format suitable for analysis and modeling. This step ensures that the data is consistent, relevant, and free from noise.

we cann see there are 25000 negative and 25000 positive review. So, the dataset is balanced and we used an 80:20 split for training and test from the dataset.

Using BeautifulSoup library to remove HTML tags from the text data.Eliminating text within square brackets as it may contain irrelevant information.Utilizing regular expressions to remove special characters, leaving only alphanumeric characters and whitespaces.Normalizing the text by converting all characters to lowercase.Applying stemming using the Porter Stemmer to reduce words to their root form.

Removing Stopwords: Eliminating common English stopwords to focus on more meaningful words.

4.3 Feature Extraction

Feature extraction involves converting text data into numerical features that can be used as input for machine learning models. This step is essential as most machine learning algorithms require numerical input.

The feature extraction process includes: Tokenization: Splitting the text into individual words or tokens.

Vectorization: Converting tokens into numerical vectors. This can be achieved using techniques like Bag-of-Words (BoW), TF-IDF (Term Frequency-Inverse Document Frequency), or Word Embeddings. By following these steps, the data is prepared for building sentiment analysis models, enabling the extraction of insights from textual data to classify movie reviews as positive or negative In this project, we experimented with several machine learning and deep learning models for sentiment analysis on the IMDB movie review dataset. Here's a summary of our findings and the process of model selection and evaluation:

4.4 Model Selection

Among the traditional machine learning models, Logistic Regression and Multinomial Naive Bayes performed the best, with consistent performance across both BoW and TF-IDF feature representations.

Support Vector Machine showed inferior performance compared to LR and MNB, especially with TF-IDF features.

The RNN model was also implemented for sequence modeling, which is well-suited for analyzing textual data. However, its performance needs to be evaluated in comparison to traditional ML models.

4.5 Model Evaluation

In evaluating the performance of various machine learning classifiers on sentiment analysis tasks using movie review data, Logistic Regression (LR) with Bag-of-Words (BoW) representation achieved an accuracy of 75.12%, demonstrating balanced precision, recall, and F1-scores for both positive and negative sentiments. Similarly, Multinomial Naive Bayes (MNB) with TF-IDF features yielded an accuracy of 75.09%, showcasing consistent precision and recall for both sentiment classes. However, Support Vector Machine (SVM) with BoW features exhibited imbalanced performance, particularly with low recall for positive sentiments, resulting in an accuracy of 58.29%. SVM with TF-IDF features demonstrated even lower accuracy at 51.12%, indicating challenges in effectively classifying positive sentiments. Overall, LR and MNB models showcased robust performance, highlighting their suitability for sentiment analysis tasks, while SVM exhibited limitations, emphasizing the importance of feature representation in achieving accurate classifications.

5 RESULTS

The RNN model was trained over 5 epochs with the following results:

Epoch 1: Loss: 0.6205, Accuracy: 62.64%, Validation Loss: 0.8979, Validation Accuracy: 61.74%

Epoch 2: Loss: 0.3764, Accuracy: 83.90%, Validation Loss: 1.4732, Validation Accuracy: 56.77%

Epoch 3: Loss: 0.2913, Accuracy: 88.43%, Validation Loss: 0.5619, Validation Accuracy: 81.55%

Epoch 4: Loss: 0.2274, Accuracy: 91.37%, Validation Loss: 0.6866, Validation Accuracy: 79.56%

Epoch 5: Loss: 0.1847, Accuracy: 93.24%, Validation Loss: 0.6123, Validation Accuracy: 83.64%

The final accuracy on the test set was 84.29%

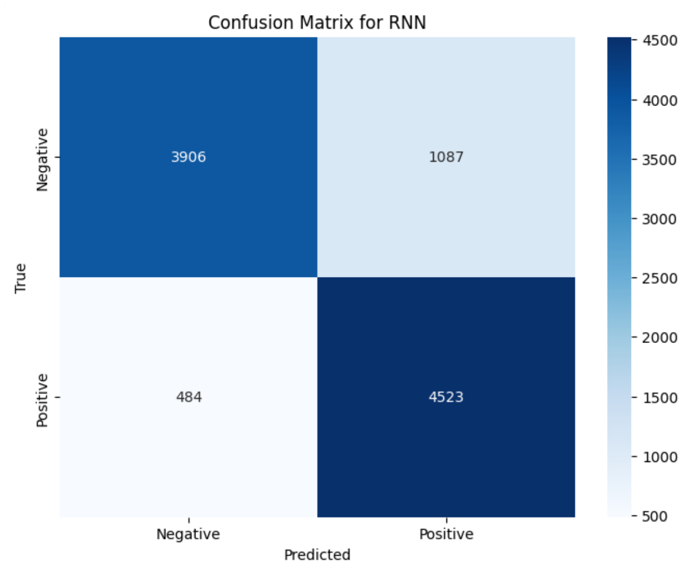

Fig. 1. Confusion Matrix for RNN

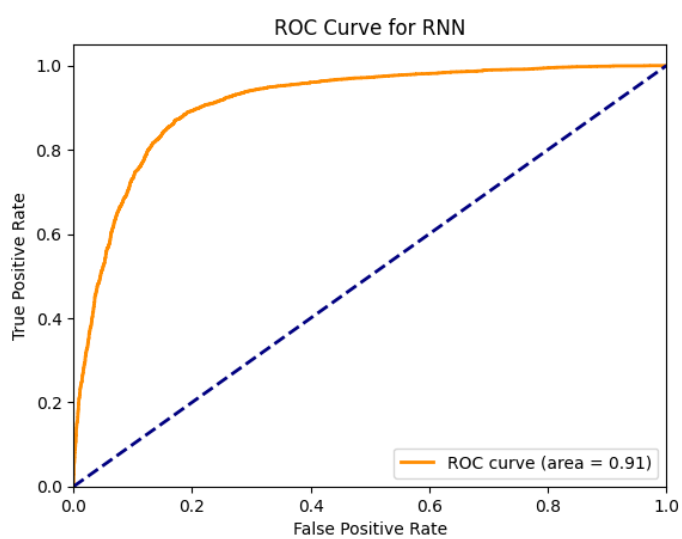

Fig. 2. ROC curve for RNN

5.1 Traditional Machine Learning Models

Logistic Regression and Multinomial Naive Bayes showed consistent performance across both Bag-of-Words (BoW) and TF-IDF feature representations.

Support Vector Machine (SVM) demonstrated inferior performance, particularly with TF-IDF features, suggesting limitations in handling the dataset's characteristics.

5.2 Recurrent Neural Network(RNN)

The RNN model exhibited promising performance, achieving an accuracy of 84.29% on the test set. Further analysis is needed to evaluate precision, recall, and F1-score to compare RNN's performance with traditional ML models comprehensively.

5.3 Additional Statistics

5.3.1 Logistic Regression (BoW). Accuracy: 75.12%
Precision (Positive): 0.75
Precision (Negative): 0.75
Recall (Positive): 0.75
Recall (Negative): 0.75
F1-score (Positive): 0.75
F1-score (Negative): 0.75

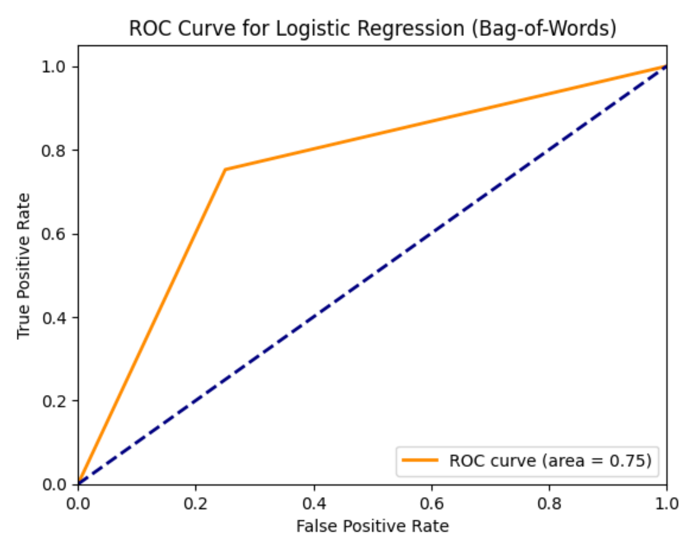

Fig. 3. ROC curve for Logistic Regression (Bag-of-Words)

5.3.2 Logistic Regression (TF-IDF). Accuracy: 75.00%
Precision (Positive): 0.74
Precision (Negative): 0.76

Recall (Positive): 0.77
Recall (Negative): 0.73
F1-score (Positive): 0.75
F1-score (Negative): 0.75

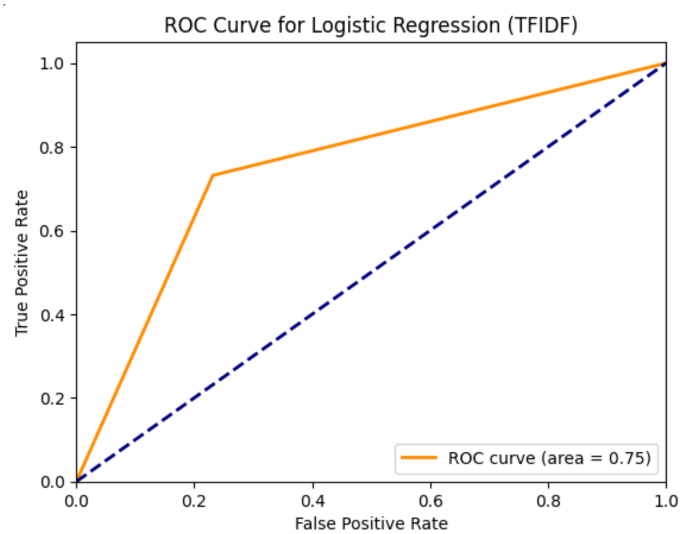

Fig. 4. ROC curve for Logistic Regression (TF-IDF)

5.3.3 Multinomial Naive Bayes (BoW). Accuracy: 75.10%

Precision (Positive): 0.75
Precision (Negative): 0.75
Recall (Positive): 0.76
Recall (Negative): 0.75
F1-score (Positive): 0.75
F1-score (Negative): 0.75

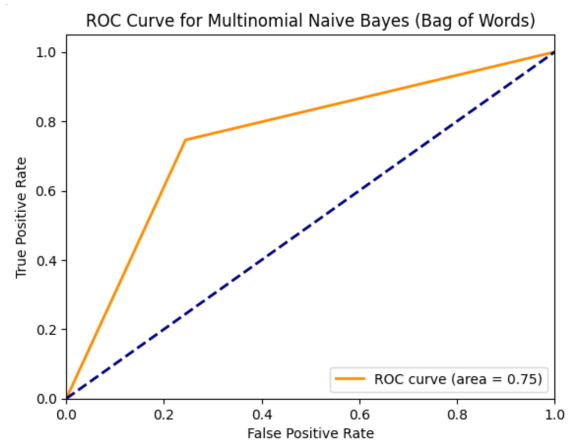

Fig. 5. ROC curve for Multinomial Naive Bayes (Bag-of-Words)

5.3.4 Multinomial Naive Bayes (TF-IDF). Accuracy: 75.09%

Precision (Positive): 0.75
Precision (Negative): 0.75
Recall (Positive): 0.76
Recall (Negative): 0.74
F1-score (Positive): 0.75
F1-score (Negative): 0.75

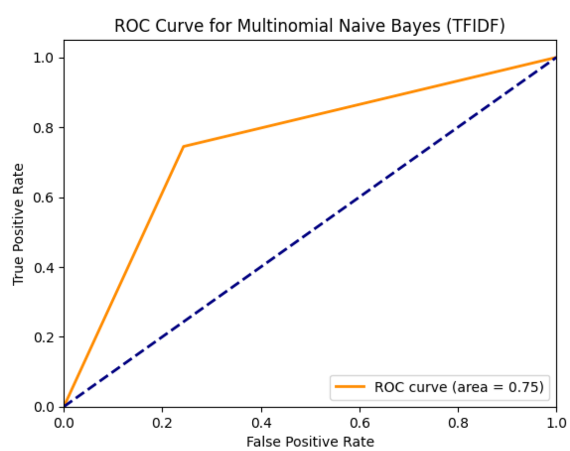

Fig. 6. ROC curve for Multinomial Naive Bayes (TF-IDF)

5.3.5 *Support Vector Machine (BoW).* Accuracy: 58.29%

Precision (Positive): 0.94
Precision (Negative): 0.55
Recall (Positive): 0.18
Recall (Negative): 0.99
F1-score (Positive): 0.30
F1-score (Negative): 0.70

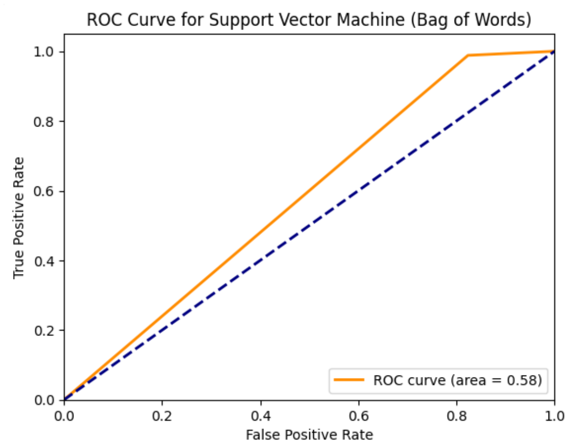

Fig. 7. Support Vector Machine (BoW)

5.3.6 *Support Vector Machine (TF-IDF).* Accuracy: 51.12%

Precision (Positive): 1.00
Precision (Negative): 0.51
Recall (Positive): 0.02
Recall (Negative): 1.00
F1-score (Positive): 0.04
F1-score (Negative): 0.67

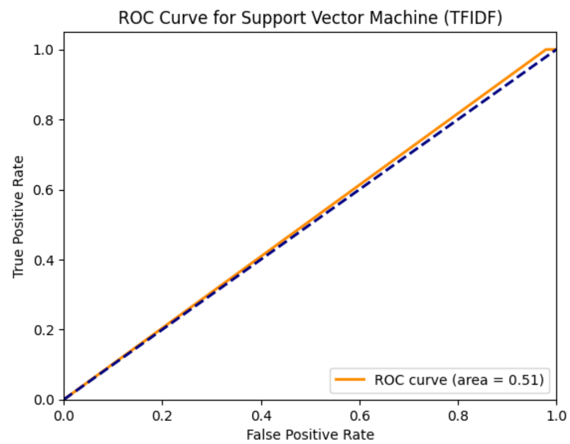

Fig. 8. Support Vector Machine (TF-IDF)

In conclusion, while traditional ML models like Logistic Regression and Multinomial Naive Bayes offer balanced performance, the RNN model presents a viable alternative with competitive accuracy. Further investigation into its precision, recall, and F1-score will provide a more comprehensive understanding of its effectiveness compared to traditional approaches.

6 DISCUSSION

In the papers of Dey et al. (2016), El-Din (2016), Ahuja et al. (2019), and Can et al. (2018), several noteworthy findings and implications for sentiment classification and analysis of movie reviews using natural language processing (NLP) techniques are presented. The techniques are mentioned below:

6.1 Choice of Classification:

Those studies implicated several machine learning classifiers, including Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB), Support Vector Machine (SVM), Maximum Entropy (ME), and Decision Tree (DT). Here, in these classification models MNB, BNB consistently demonstrated superior performance in terms of accuracy, precision, recall, and F-score. SVM classifier showed astounding performance, especially in recall, marking its effectiveness in capturing nuanced sentiments in movie reviews. Whereas, ME classifier exhibited moderate performance signifying the need for more optimization and feature selection to enhance its effectiveness. Lastly, the DT classifier, made some lower accuracy ratings compared to MNB and BNB classifiers, implicating the emergence of refining decision tree-based models for sentiment classification tasks.

6.2 Feature Selection and Model Complexity:

The findings highlight how important feature choice and model complexity are when doing sentiment analysis tasks.The remarkable performance of MNB and BNB classifiers, which are based on basic probabilistic models, suggests that these models are useful for identifying sentiment trends in movie reviews. Competitive performance was shown by the SVM classifier, which finds the best hyperplanes in feature space. This highlights the significance of feature representation and kernel selection for precise sentiment categorization. The ME classifier has the potential to be improved using regularization approaches and enhanced feature engineering to reduce overfitting, even though it was less effective in the evaluated trials. Even though the DT classifier is understandable and intuitive, it may perform better in prediction if ensemble approaches or pruning techniques were used.

6.3 Practical Implications:

Filmmakers, distributors, and marketers are just a few of the industry players who will find the data useful. Precise sentiment analysis of film reviews can help guide decisions about marketing campaigns, distribution plans, and film creation. Large amounts of textual data can be efficiently and scalable analyzed using NLP algorithms, providing real-time insights into the preferences and attitudes of the audience. To further increase the accuracy and resilience of sentiment categorization, future research may investigate more sophisticated deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

7 CONCLUSION

In wrapping up, our journey into sentiment analysis has shown how crucial it is to understand how people feel about movies. Through our use of various machine learning techniques, we've managed to dig deep into movie reviews, gaining insights that can guide filmmakers and advertisers. Our findings, especially the success of our RNN model with an impressive 84% accuracy, demonstrate the power of modern technology in decoding human emotions from text. Looking ahead, we're excited about the potential to make movie-making decisions more audience-centric. By bridging the gap between what people feel and what the industry creates, we hope to make movies that truly resonate with audiences everywhere.

8 AUTHOR CONTRIBUTIONS

All the author's of the paper have read and agreed to the submitted version of the paper and all have contributed equally.

REFERENCES

(1) Dey, S., Chakraborty, S., Saha, S., & Chakraborty, A. (2016). Sentiment analysis of movie reviews using machine learning techniques. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 581-586). IEEE.

(2) El-Din, S. S. A. (2016). Sentiment analysis in movie reviews based on machine learning techniques. In *2016 7th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 106-109). IEEE.

(3) Ahuja, S., Sahu, T. P., & Singh, A. (2019). Sentiment analysis of movie reviews: A study on feature selection & classification algorithm. In *2019 2nd International Conference on Computing, Mathematics and Communication Technologies (ICMCT)* (pp. 1-4). IEEE.

(4) Can, A., Koc, A. B., Sen, B., & Kucuktunc, O. (2018). Sentiment classification of Turkish movie reviews using machine learning techniques. In *2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG Istanbul)* (pp. 1-5). IEEE.

(5) Can, F., & Sarıyar, M. (2018). Turkish Movie Reviews Sentiment Analysis. *Journal of Engineering Research and Applied Science, 5*(2), 785-792.

(6) Horsa, O. G., & Tune, K. K. (2023). Aspect-Based sentiment analysis for AFAAN Oromoo movie reviews using machine learning techniques. *Applied Computational Intelligence and Soft Computing, 2023*, 1–12. https://doi.org/10.1155/2023/3462691

(7) Xu, J. (2024, February). Decoding sentiment: A sentiment analysis model for movie reviews. *Applied and Computational Engineering, 37*(1), 31-37. https://doi.org/10.54254/2755-2721/37/20230466