International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Sentiment Classification of Imdb Movie Reviews Using Naturl Language Processing Techniques

P. Anusha[1], E. Naveen Kumar[2], G. Sravanthi[3], E. Rohitha[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India

[2,3,4]Student , Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India

Abstract- Sentiment analysis is a crucial task in natural language processing (NLP) that aims to determine the overall sentiment or opinion expressed by a reviewer towards a movie. This study focuses on the sentiment analysis of IMDB movie reviews using various machine learning and NLP techniques. The findings indicate that feature selection can enhance the accuracy of sentiment-based classification, but the effectiveness depends on the specific method and number of features selected. The paper also presents a comprehensive comparison of traditional machine learning techniques and advanced transformer-based models for sentiment analysis of IMDB movie reviews. The results provide insights into choosing appropriate methods for accurate and timely sentiment analysis on IMDB data. The study employs feature extraction techniques such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and word2vec. Feature selection using methods like chi-square is shown to improve classification performance.

Keywords- Bag-of-words, Natural Language Processing, Sentiment analysis.

I. INTRODUCTION

Sentiment analysis plays a crucial role in Natural Language Processing (NLP) by determining the overall sentiment or opinion expressed in textual data. In this study, we focus on sentiment analysis applied to IMDB movie reviews, aiming to understand and classify the sentiments expressed by reviewers towards movies. The study employs a range of machine learning and NLP techniques to achieve accurate sentiment analysis and investigates the impact of feature selection on classification performance. The findings of this study highlight the importance of feature selection in enhancing the accuracy of sentiment-based classification tasks.

Feature selection methods such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and word2vec are explored to extract relevant features from IMDB movie reviews. Specifically, the study investigates how feature selection techniques, such as chi-square, can improve classification performance by selecting the most informative features for sentiment analysis. Additionally, the study presents a comprehensive comparison between traditional machine learning techniques and advanced transformer-based models for sentiment analysis on IMDB movie reviews. By evaluating the effectiveness of different methods and feature extraction techniques, the research aims to provide insights into choosing appropriate approaches for accurate and timely sentiment analysis on IMDB data.

Overall, this study contributes to the field of sentiment analysis by exploring various techniques and methods to improve classification accuracy and effectiveness, particularly focusing on IMDB movie reviews as a rich source of sentiment-laden textual data.

II. LITERATURE SURVEY

Mohammed S.Albarrak et.al.,In this paper we analyse firms' dissemination decisions using Twitter, developing a comprehensive measure of the amount of financial information that a company makes available to investors (iDisc) from a big data of firms' tweets (1,197,208 tweets). Using a sample of 4131 firm-year observations for 791 non-financial firms listed on the US NASDAQ stock exchange over the period 2009–2015, we find evidence that iDisc significantly reduces the cost of equity. These results are pronounced for less visible firms which are relatively small in size, have a low analyst following and a small number of investors. Highly visible firms are less likely to benefit from iDisc in influencing their cost of equity as other communication channels may have widely disseminated their financial information. Our investigations encourage managers to consider the benefits of directly spreading a firm's financial information to stakeholders and potential investors using social media in order to reduce firm equity premium (COE).

Sundong Kim et. al., In this paper proposes an attention-based multi-layer friend recommendation model to mitigate information overload in social networks. We first constructed the basic user and item matrix via convolutional neural networks (CNN). Then, we obtained user preferences by using the relationships between users and items, which were later inputted into our model to learn the preferences between friends. The error performance of the proposed method was compared with the traditional solutions based on collaborative filtering. A comprehensive performance evaluation was also conducted using large-scale real-world datasets collected from three popular location-based social networks. The experimental results revealed that our proposal outperforms the traditional methods in terms of recommendation performance.

Qian Li et. al., This paper fills the gap by reviewing the state of the art approaches from 1961 to 2020, focusing on models from shallow to deep learning. We create a taxonomy for text classification according to the text involved and the models used for feature extraction and classification. We then discuss each of these categories in detail, dealing with both the technical developments and benchmark datasets that support tests of predictions. A comprehensive comparison between different techniques, as well as identifying the pros and cons of various evaluation metrics are also provided in this survey. Finally, we conclude by summarizing key implications, future research directions, and the challenges facing the research area.

Qusay Al-Maatouk et. al., This research hypothesizes that TTF applied to social media for learning will affect technology, task, and social characteristics that in turn improve students' satisfaction and students' academic performance. It also posits that the behavioral intent to use social media for learning will affect comprehension efficiency, ease of use, and enjoyment, all of which also improve students' satisfaction and students' academic performance. The data collection questionnaire was conducted with 162 students familiar with social media. Quantitative structural equation modeling was employed to analyze the results. A significant relationship was found between technology, task, and social features with TTF for utilizing social media for academic purposes, all of which fostered student enjoyment and improved outcomes. Similarly, a clear relationship was found between comprehension efficiency, ease of use, and enjoyment with behavioral intentions to utilize social media for academic purposes that positively affected satisfaction and achievement. Therefore, the study indicates that TTF and behavioral intentions to use social media improve the active learning of students and enable them to efficiently share knowledge, information, and discussions. We recommend that students utilize social media in pursuit of their educational goals. Educators should also be persuaded to incorporate social media into their classes at higher education institutions.

Raghad Alshalan and Hend Al-Khalifa With the rise of hate speech phenomena in the Twittersphere, significant research efforts have been undertaken in order to provide automatic solutions for detecting hate speech, varying from simple machine learning models to more complex deep neural network models. Despite this, research works investigating hate speech problem in Arabic are still limited. This paper, therefore, aimed to investigate several neural network models based on convolutional neural network (CNN) and recurrent neural network (RNN) to detect hate speech in Arabic tweets. It also evaluated the recent language representation model bidirectional encoder representations from transformers (BERT) on the task of Arabic hate speech detection. To conduct our experiments, we firstly built a new hate speech dataset that contained 9316 annotated tweets. Then, we conducted a set of experiments on two datasets to evaluate four models: CNN, gated recurrent units (GRU), CNN + GRU, and BERT. Our experimental results in our dataset and an out-domain dataset showed that the CNN model gave the best performance, with an F1-score of 0.79 and area under the receiver operating characteristic curve (AUROC) of 0.89.

Existing System

Support Vector Machines (SVM) stand as stalwarts in the realm of machine learning, particularly renowned for their prowess in classification tasks. At the heart of SVMs lies the endeavor to find a hyperplane within the feature space that effectively delineates distinct classes, thereby facilitating accurate classification of data points. This hyperplane, strategically positioned to maximize the margin between classes, serves as the linchpin of SVM's success in handling both linearly separable and non-linearly separable data. One of the key strengths of SVMs is their versatility in handling high-dimensional data, making them well-suited for complex datasets where other algorithms might falter. By leveraging the concept of kernels, SVMs can transform input data into higher-dimensional spaces, potentially unraveling intricate patterns and relationships that might be obscured in lower-dimensional representations.

Furthermore, SVMs exhibit robustness in the face of overfitting, thanks to their ability to maximize the margin between classes, thus promoting generalization to unseen data. This attribute makes SVMs a favored choice in scenarios where model interpretability and predictive performance on new instances are paramount.

Existing System Disadvantages

- Computational Complexity: SVMs can be computationally intensive, especially when dealing with large datasets. The training time and memory requirements can escalate significantly as the size of the dataset increases, posing challenges in scalability.
- Selection of Kernel and Hyperparameters: The performance of SVMs is heavily influenced by the choice of kernel function and hyperparameters. Selecting the appropriate kernel and tuning hyperparameters such as regularization parameters can be a non-trivial task, requiring careful experimentation and optimization.
- Sensitivity to Noise: SVMs can be sensitive to noisy data or outliers, potentially leading to suboptimal performance or overfitting if not appropriately handled. Preprocessing steps such as data cleaning and outlier removal may be necessary to mitigate this issue.
- Limited Interpretability: While SVMs excel in classification accuracy, they often provide limited insight into the underlying decision-making process. The resulting models may lack interpretability compared to more transparent algorithms like decision trees or logistic regression.
- Scalability Issues: As the size of the dataset grows, the computational demands of SVMs can become prohibitive. Training SVMs on very large datasets may require specialized techniques such as stochastic gradient descent or distributed computing frameworks.

Proposed System

Natural Language Processing (NLP) stands at the forefront of cutting-edge technology, revolutionizing how we interact with and extract insights from vast amounts of textual data. At its core, NLP encompasses a myriad of techniques and methodologies aimed at enabling computers to understand, interpret, and generate human language in a manner that mimics human cognitive abilities. From sentiment analysis and text classification to machine translation and question answering systems, NLP empowers us to unlock the wealth of information embedded in text, making it a cornerstone of modern artificial intelligence applications. One of the key challenges in NLP lies in the ambiguity and complexity inherent in natural language. Human language is rich in nuances, context-dependent meanings, and linguistic variations, presenting formidable hurdles for machines to overcome. NLP techniques such as tokenization, part-of-speech tagging, syntactic parsing, and semantic analysis are deployed to deconstruct and analyze text at various levels, enabling machines to derive meaningful insights and infer relationships between words, phrases, and sentences.

The advent of deep learning, particularly models like recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer-based architectures such as BERT and GPT, has revolutionized NLP by enabling more sophisticated language modeling, sentiment analysis, and natural language understanding tasks. These deep learning models, coupled with pre-training on large corpora of text, have significantly advanced the state-of-the-art in tasks such as text generation, language translation, and sentiment classification. Despite the remarkable progress, NLP still grapples with challenges such as handling ambiguity, understanding context, and preserving privacy and ethical considerations in language processing. The quest for more accurate, interpretable, and ethically sound NLP systems continues to drive research and innovation in the field, promising a future where machines seamlessly converse, comprehend, and create in natural language, fostering new frontiers in human-computer interaction and knowledge discovery.

III. METHODOLOGY

Modules

1. Data set

A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

2. Pre-Processing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way.

3. Splitting

Data splitting is the act of partitioning available data into. two portions, usually for cross-validatory purposes. one portion of the data is used to develop a predictive model. and the other to evaluate the model's performance.

- Training Data: Used for train the model or given as input to the to the learning model
- Testing Data: Used for test the model or given as input to the model for prediction.

4. Apply Algorithm

Natural Language Processing (NLP) stands at the forefront of cutting-edge technology, revolutionizing how we interact with and extract insights from vast amounts of textual data. At its core, NLP encompasses a myriad of techniques and methodologies aimed at enabling computers to understand, interpret, and generate human language in a manner that mimics human cognitive abilities. From sentiment analysis and text classification to machine translation and question answering systems, NLP empowers us to unlock the wealth of information embedded in text, making it a cornerstone of modern artificial intelligence applications.

One of the key challenges in NLP lies in the ambiguity and complexity inherent in natural language. Human language is rich in nuances, context-dependent meanings, and linguistic variations, presenting formidable hurdles for machines to overcome. NLP techniques such as tokenization, part-of-speech tagging, syntactic parsing, and semantic analysis are deployed to deconstruct and analyze text at various levels, enabling machines to derive meaningful insights and infer relationships between words, phrases, and sentences.

The advent of deep learning, particularly models like recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer-based architectures such as BERT and GPT, has revolutionized NLP by enabling more sophisticated language modeling, sentiment analysis, and natural language understanding tasks. These deep learning models, coupled with pre-training on large corpora of text, have significantly advanced the state-of-the-art in tasks such as text generation, language translation, and sentiment classification.

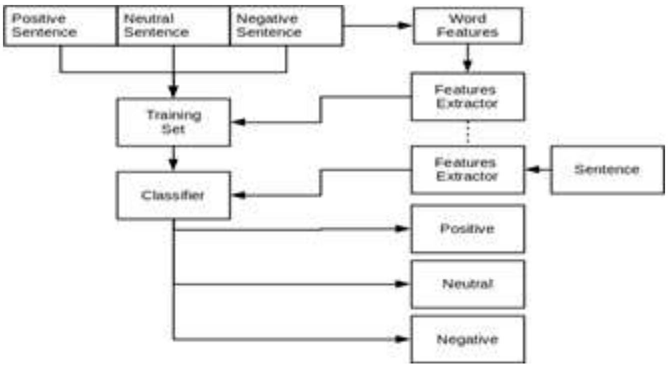
5. Visualization

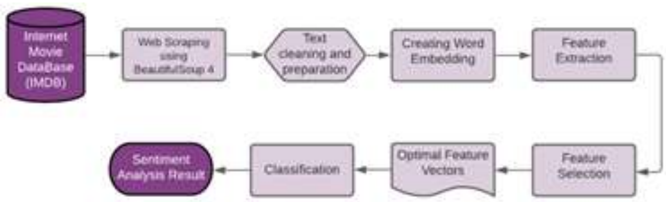
Visualization is a technique that uses an array of static and interactive visuals within a specific context to help people understand and make sense of large amounts of data. The data is often displayed in a story format that visualizes patterns, trends and correlations that may otherwise go unnoticed.

6. Accuracy

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total Predictions.

IV. SYSTEM ARCHITECTURE



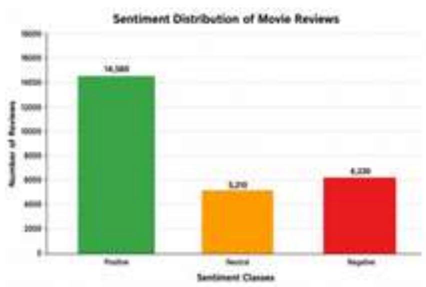


The above architecture illustrate the working process of sentiment analysis using Natural Language Processing (NLP) and machine learning techniques. The first diagram explains the basic sentiment classification architecture where sentences are categorized into positive, neutral, and negative classes. Initially, labeled sentences are collected and converted into a training dataset. Important word features are extracted from the text using a feature extraction process, and these features are used to train a classifier model. When a new sentence is given as input, the same feature extraction process is applied, and the classifier predicts the sentiment category of the sentence.

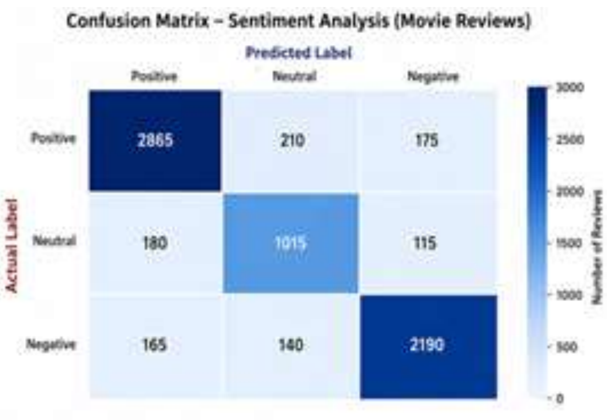
International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

The second diagram represents a complete sentiment analysis workflow using the IMDB movie review database. The process starts with web scraping using BeautifulSoup4 to collect movie reviews from the internet. After collecting the data, text cleaning and preparation are performed to remove unnecessary information and improve data quality. Word embeddings are then created to represent text numerically, followed by feature extraction and feature selection to identify the most relevant information. Optimal feature vectors are generated and passed to a classification model, which produces the final sentiment analysis result. Together, these diagrams demonstrate how machine learning and NLP techniques are used to analyze opinions and emotions from textual data effectively.

V. EXPERIMENTAL RESULTS



The bar graph represents the distribution of movie review sentiments classified into positive, neutral, and negative categories. The graph shows that positive reviews are the highest with 14,560 reviews, indicating that most users expressed favorable opinions about the movies. Neutral reviews are the lowest with 5,210 reviews, while negative reviews account for 6,230 reviews, showing comparatively fewer dissatisfied responses.



This represents a confusion matrix used to evaluate the performance of a sentiment analysis classification model for movie reviews. The rows indicate the actual sentiment labels, while the columns represent the predicted sentiment labels, showing how accurately the model classifies positive, neutral, and negative reviews. Higher values along the diagonal cells indicate correct predictions, demonstrating that the model achieved high accuracy in sentiment classification with fewer misclassification errors.

VI. CONCLUSION

This study has demonstrated the critical role of feature selection in enhancing the accuracy of sentiment analysis for IMDB movie reviews, employing a variety of machine learning and NLP techniques. By exploring methods such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and word2vec for feature extraction, and utilizing chi-square for feature selection, we have shown that selecting the most informative features significantly improves classification performance. Additionally, the comprehensive comparison between traditional machine learning techniques and advanced transformer-based models provides valuable insights into the effectiveness of different approaches for sentiment analysis.

This research not only advances our understanding of sentiment classification but also offers practical guidance for choosing appropriate methods for analyzing sentiment-rich textual data. The findings underscore the importance of leveraging advanced NLP techniques and thoughtful feature selection to achieve more accurate and reliable sentiment analysis, paving the way for future enhancements and applications in various domains.

REFERENCES

1. Baid, P.; Gupta, A.; Chaplot, N. Sentiment analysis of movie reviews using machine learning techniques. Int. J. Comput. Appl. 2017, 179, 45–49.
2. Pang, B.; Lee, L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. arXiv 2004, arXiv:cs/0409058.
3. Elghazaly, T.; Mahmoud, A.; Hefny, H.A. Political sentiment analysis using twitter data. In Proceedings of the International Conference on Internet of things and Cloud

Computing, Cambridge, UK, 23 February—22 March 2016; pp. 1–5.
4. Pratiwi, A.I. On the feature selection and classification based on information gain for document sentiment analysis. Appl. Comput. Intell. Soft Comput. 2018, 2018, 1407817.
5. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. arXiv 2002, arXiv:cs/0205070.
6. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. Expert Syst. Appl. 2016, 57, 117–126.
7. Zou, H.; Tang, X.; Xie, B.; Liu, B. Sentiment classification using machine learning techniques with syntax features. In Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 7–9 December 2015; pp. 175–179.