ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Analysis of Sentiment on IMDB Movie Reviews: A Comparative Study of Algorithms

[1] Mr.Naushad Alam, [2] Khaja Ehtesham Uddin, [3] Hablullah Burhan Uddin, [4] Abdullah Ibrahim Hussaini Syed,
[1]Assistant Professor,Department of CSE-AIML, Lords Institute of Engineering & Technology.
[234] Student Department of CSE-AIML, Lords Institute of Engineering & Technology.

Abstract—

Finding the polarity (sentiment) and intention of a piece of textual data is the core emphasis of Sentiment Analysis, one of the fundamental challenges in natural language processing. Depending on the context, the data might be at the phrase or paragraph level. When we talk to one other like people, our emotions are much simpler to read. However, it gets more difficult to discern the emotions when dealing with robots. Methods from Natural Language Processing (NLP) and other areas are used in Sentiment Analysis. Reviews are a major indicator of a film's success or failure. The reviews will greatly impact a wider audience. Therefore, it is critical to construct a reliable model that correctly categorizes reviews of movies. We use six different ML models to categorize movie reviews in this research, and then we compare and contrast them to find the top classifier according to a number of criteria. Keywords—reviews of films, methods for detecting polarity, feelings, natural language processing, and models for machine learning

INTRODUCTION

People are more likely to express their opinions and thoughts via email, social media, corporate survey forms, and other online mediums due to the prevalence of these services. Typically, they cover topics that elicit strong opinions, such as current events, evaluations of movies and products, etc. the third The purpose of sentiment analysis is to determine the general public's opinion on a certain issue, product, or concept. Sentiment analysis has its roots in the 1950s. Documents on paper were its principal usage. Natural language processing (NLP), statistics, and machine learning are some of the methods used for the study. There are two main categories of sentiment analysis: a) identifying subjectivity and objectivity b) Feature/Aspect-Driven Recognition
Sorting text into categories according to its polarity (emotional value) is the essence of objectivity and subjectivity. egardless of the specific setting, it remains. It is possible to ascertain various feelings or views in regard to various features of an entity or context using feature/aspect-based identification. This category allows for the wide classification of text into different emotions, such as joyful, sad, angry, etc. [7] Textual movie reviews often include a comprehensive summary of the film's good and bad points. A viewer's mindset and their preferred aspects may be better understood using this method. Moviemakers may also benefit from being able to read and comprehend audience emotions. Before deciding whether or not to see a movie, most people look at the reviews written by those who have already seen it. Because of this, the evaluations will be quite consequential and related to its success or failure. In order to boost user engagement and implement new business strategies, sentiment analysis is now employed mostly in areas such as social media monitoring, product analysis, and customer feedback analysis via online forms, among others. [5]

LITERATURE SURVEY

Title: Evaluating the Naive-Bayes Classifier and a Rule-based Approach for Sentiment Analysis in Movie Reviews Vinay Hegde, Dr. B. Satish Babu, and Vihaan Nama are the authors. The primary objective is to use a Naïve Bayes Classifier and an AFINN-111 Rule-based method to do sentiment analysis on movie reviews. The NLTK library, which is used by Naïve Bayes Classifier, has several functions that are extensively employed in sentiment analysis. We use the Movie Reviews dataset included in the NLTK toolkit. In addition, the AFINN-111 dataset was used to develop a rule-based approach. Words in this dataset are rated from -5 (very negative) to +5 (very positive). To determine the emotion, we first deconstruct the whole statement into its component parts and then add the values of each word. At last, the results of the comparison show that the Naïve Bayes strategy is more successful than the Rule-based approach. Restrictions: When reviewing reviews using complicated terms, the Naïve Bayes algorithm failed to provide accurate results. Additionally, it was inaccurate in its predictions when the review included mixed

feelings. [1] B. The Title: Deep Learning-Based Sentiment Analysis Vinod P., Shilpa P. C., Rissa Shereen, and Susmi Jacob are the authors. Classifying user communications posted on Twitter is the primary objective. Using "word2vec," we were able to create semantic word vectors for every word in the input tweets. Then, in order to train and categorize the emotions, LSTM and RNN are fed the extracted features. The suggested model starts by sorting emotions into two broad groups: positive and negative. Within each of these groups, it further subdivides emotions. One dataset includes both positive and negative tweets, another has classes relating to positive emotions, and the third contains classes pertaining to negative tweets; these three datasets are used. The tools used for feature selection include TF-IDF and Doc2Vec. In order to feed the LSTM algorithm, we first need to get a list of words. The LSTM model outperforms the RNN and CNN models, according to the experimental data. Constraints: A system that attempts to determine users' personalities is required for the system to be more tailored to each user. [2] C. Title: An Evaluation of Methods for Twitter Sentiment Analysis S M Salim Reza, Abdullah Al Mueed, Jasiya Fairiz Raisa, and Maliha Ulfat Examining different approaches to sentiment analysis using data from Twitter is the goal. The study is built upon a combination of lexicon-based, Machine Learning, and Deep Learning models. Various datasets are used to test and deploy the models. Out of all the models and datasets tested, those based on machine learning had the best accuracy rate of 98%. Researchers found that self-constructed datasets performed better in terms of accuracy. Disadvantages: Additional study is needed to determine the exactness of opinions in a dynamic setting utilizing new computational methodologies. the third Title: A Sentiment Analysis Survey (D) Members of the author team include RRaja Subramanian, Nukala Akshith, Gogula Narasimha Murthy, Manchala Vikas, Srikar Amara, and Karnam Balaji. If you want to know how to do sentiment analysis on textual data, this article will show you the ropes. Several pre-processing processes are described in the text. Many techniques are used, including those based on lexicons and those based on machine learning. Both BOW and TF-IDF are feature engineering techniques that are used. Neural Networks (RNNs), Long Short-Term Memories (LSTMs), Decision Trees, and Support Vector Machines (SVMs) are some of the models. At last, we use the IMDB Dataset to compare the outcomes of each model. Limitations: Sequence based models provide high accuracy but are difficult to use. Utilizing novel preprocessing and feature extraction techniques to construct simple models should be the focus of future research. [4] E. Title: Utilizing BERT and Convolutional Neural Networks for Sentiment Analysis Ko-Lin Rui and Rui Man The purpose of this research is to use the BERT algorithm and other deep learning techniques to do sentiment analysis on the "ChnSentiCorp" dataset of Chinese-textual restaurant reviews. There are three main components to the BERT model's Encoder-Decoder architecture: the Encoder, the Self-attention layer, and the Decoder. The encoder module is made up of a feedforward neural network and a self-attention layer. The query, key, and value vectors will be computed by the self-attention layer. It is possible to find out where a word is in a phrase by using an extra vector in the encoding and decoding layers. The use of any feature extraction technique leads to the conclusion that BERT-CNN is superior. In the future, we want to construct a 2-way LSTM model and a CNN that can be integrated to improve the efficacy of sentiment analysis algorithms. [5] F. Title: A Bang-lish Sentiment Analysis Based on Deep Learning The disclosure process Rabeya Basri, M.F. Mridha, Md. Abdul Hamidf, and Muhammad Mostafa Monowarf are the authors. In this case, we use attention-based CNN to analyze the tone of Bengali reviews published in English. We compare the outcome to RNN. We make use of a proprietary dataset that includes five thousand brief paragraphs that have been annotated based on the emotions they convey. An input layer, a convolution layer, and a max-p-pooling layer make up the suggested model, with the latter two serving to feed data into the attention layer. Compared to multi-class classification, the models' performance in binary classification was satisfactory. Limitations: For better sentiment analysis results while dealing with Bengali text written in English alphabets, Transformer models that use Encoder-Decoder architecture might be utilized. [6] Mr. G. The Use of Language Models for Sentiment Analysis: A Research Study The authors are Deepti Goel, Nisha Pahal, Spraha Kumawat, and Inna Yadav. The goal of this research is to learn about several models that use deep neural networks to determine if a piece of text is expressing neutral, positive, or negative emotions. Human interference in dataset labeling is removed by the suggested technique. Experimental models include Electra, Roberta, the BERT model, and Bidirectional LSTM. A key statistic that has been used is the Matthews correlation coefficient (mcc). With 81% accuracy, the BERT algorithm has achieved the best result. [7] There are no limitations. Title: Deep Learning and Machine Learning for Sentiment Analysis Yogesh Chandra and Antoreep Jana are the authors. Classifying tweets as good, negative, or neutral is the primary goal of sentiment analysis, which may be achieved using a variety of machine learning classifiers, deep learning models based on polarity, etc. In addition, several performance indicators are used to compare the models' output, and a voting-based categorization approach is also considered. Models trained using LSTM-CNN performed the highest in terms of classification accuracy, according to the findings. In the future, it may be possible to use deep learning algorithms to create architectural plans with prediction

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

accuracy on par with human experts. [8] I. Atiqur Rahman and Md. Sharif Hossen's Machine Learning-Based Sentiment Analysis of Movie Review Data 2. The overarching goal is to apply sentiment analysis to a dataset consisting of movie reviews using five distinct ML methods. Text preprocessing, vector feature development, model training and testing, and result comparison using various performance measure metrics (e.g., accuracy, precision, F1-score, and recall) are the stages involved. Naïve Bayes, Support Vector Machines, Multinomial Naïve Bayes, Maximum Entropy, and Decision Trees are the classifiers that are used. With an accuracy of 88.5%, Multinomial NB has shown to be the most effective method. Conversely, SVM has shown a superior recall value in comparison to others. [9] Looking forward, deep learning algorithms may also be used for polarity classification. Title: Sentiment Analysis and Classification of Movie Reviews (J.) Writers: Sara Tedmori and Mais Yasen Using eight well-known classifiers, the primary goal is to polarity classify the IMDB Movie Reviews dataset. To prepare the dataset for model training, basic text preprocessing is carried out and features are extracted. To evaluate the models, we employed the following metrics: accuracy, recall, precision, and area under the curve (AUC). When compared to the other models, Random Forest fared better, but Ripper Ruler Learning (RRL) did the poorest. Plans for the future include using data from other languages to do sentiment analysis. The accuracy of model outputs may also be enhanced by further study. [10]

III. Block Diagram

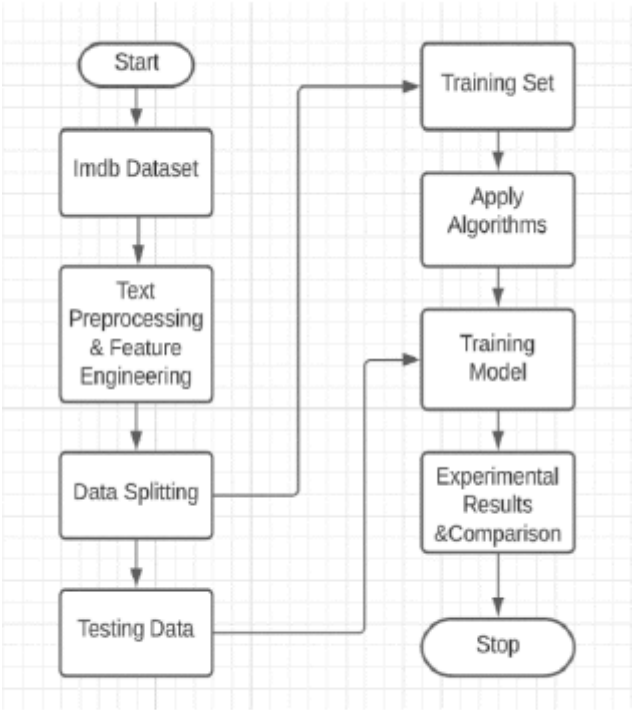

Fig. 1. Block Diagram for proposed system

You can see the whole process laid out in Fig. 1. The first step in feature extraction is to apply text preprocessing to the dataset. Subsequently, many machine learning algorithms are fed data that is either trained or tested. At last, we compare the outcomes of each model.

IMPLEMENTATION

A. Python

Python 3 and subsequent versions, which provide a wealth of libraries and modules for analysis, were employed for our experiments. Part B: Library Resources Pandas is a popular Python package for data analysis. An essential component of data analysis, its features allow users to read data in a variety of forms (e.g., CSV, JSON, etc.) and manipulate the data. 2) The built-in Matplotlib library can generate high-quality visualisations including bar plots, scatter plots, pie charts, and more. Thirdly, NLTK is a popular Python toolkit for preparing human text using natural

language processing techniques. 4) Sklearn: A collection of machine learning models including regression, classification, and more may be found in this library.

EXPERIMENTAL SETUP AND RESULTS

A. Dataset

The "IMDB Dataset" has been culled from the archives of Kaggle. With 50,000 reviews, 25% are favorable and the rest are negative, making it a well balanced sample.

review	sentiment
One of the other reviewers has mentioned that ...	positive
A wonderful little production The filming tech...	positive
I thought this was a wonderful way to spend ti...	positive
Basically theres a family where a little boy J...	negative
Petter Matteis Love in the Time of Money is a ...	positive

Fig. 2. IMDB Movie Reviews Dataset

The categorization and the first five rows of the dataset are shown in Figure 2. Part B: Method The first step is to load the dataset into the Jupyter Notebook Environment and import all the necessary libraries and modules. First, there is text preprocessing, which includes a number of preprocessing operations including tokenization, stemming, stop word removal, etc. 2-Engineering Features: Two feature extraction methods, the Bag of Words (BOW) model and the TF-IDF model, were applied in this case. The first approach involves keeping a dictionary that records every time a word appears. The following model further incorporates frequency, it shows how important a certain text is in a document collection. The third stage is to create training and testing datasets, which consist of 70% and 30% of the whole dataset, respectively. Fourth, Put Algorithms to Work: After we extract features using BOW and TF-IDF, we'll use training data to apply machine learning models. Logistic Regression, Support Vector Machines, Decision Trees, K-Nearest Neighbors, and Xgboost are the techniques that we used. Fifth, test the models using testing data (30%) to determine their correctness and performance. The results are compared after step 4. Common metrics used for this comparison include F1-score, recall, support, and accuracy.

VI. RESULTS

	review	sentiment
0	One reviewers mentioned watching 1 Oz episode ...	1
1	wonderful little production filming technique ...	1
2	thought wonderful way spend time hot summer we...	1
3	Basically theres family little boy Jake thinks...	0
4	Petter Matteis Love Time Money visually stunni...	1

Fig. 3. Output after Text-preprocessing

Following text preparation operations such as stemming, tokenization, stop word removal, etc., the dataset's state is shown in Fig. 3.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

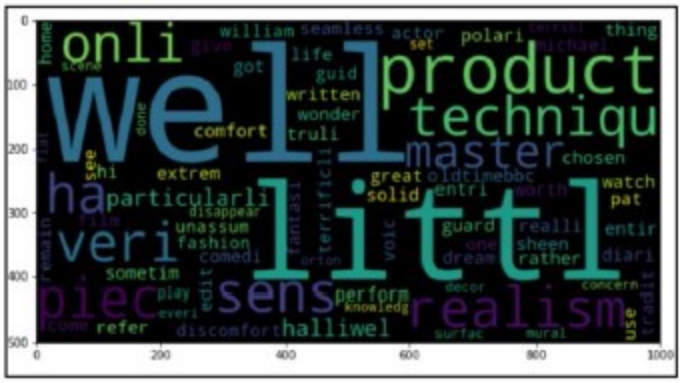

Fig. 4. Word Cloud for Positive Words

Fig. 4 is the visualisation for positive words in reviews.

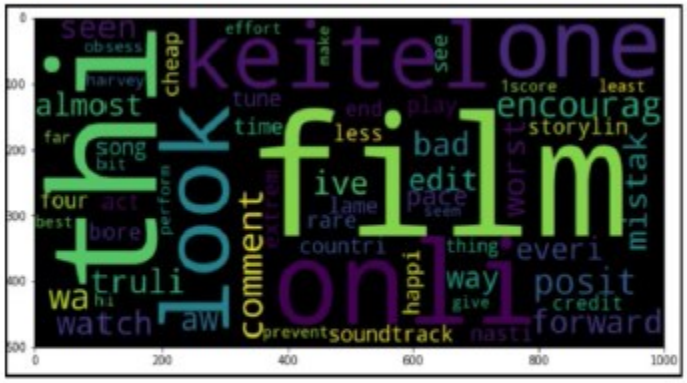

Fig. 5. Word Cloud for Negative words

Fig. 5 is the visualisation for negative words in reviews.

Algorithm	Accuracy
Logistic Regression	90 %
SVM	90 %
Multinomial Naïve Bayes	88 %
Decision Trees	74 %
KNN	53 %
Xgboost	85 %

Fig. 6. Accuracy of different algorithms using BOW technique

Figure 6 shows the results of several methods for feature extraction using the Bag of Words (BOW) methodology.

Algorithm	Accuracy
Logistic Regression	89 %
SVM	89 %
Multinomial Naïve Bayes	89 %
Decision Trees	71 %
KNN	79 %
Xgboost	84 %

Fig. 7. Accuracy of different algorithms using TF-IDF technique

The results of various algorithms' usage of the TF-IDF feature extraction approach are detailed in Figure 7.

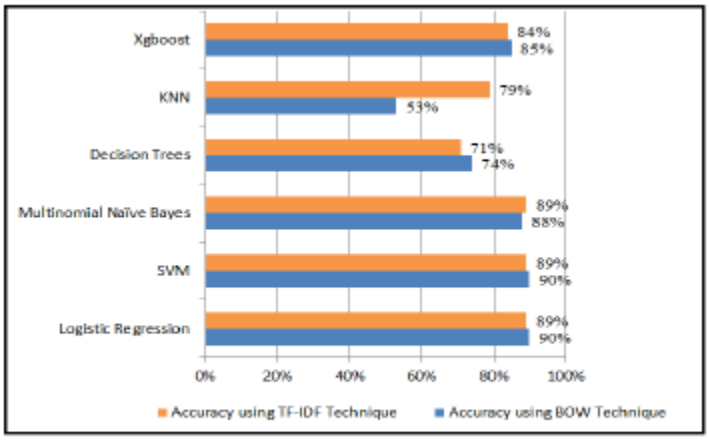

Fig. 8. Comparison between algorithms

The six algorithms are compared in Fig. 8 utilizing BOW and TF-IDF methods.

VII.CONCLUSION

In this research, we used sentiment analysis to a dataset consisting of movies. Logistic Regression, Support Vector Machines, and Multinomial Naive Bayes all achieve almost identical 85% accuracy rates in the experiments. The accuracy was lowest when the same model was trained using the KNN technique. As compared to the BOW model, the accuracy of the TF-IDF, Logistic Regression, SVM, and Naïve Bayes models was 89%. When it came to the TF-IDF model, Decision Trees were the least accurate. Xgboost almost achieved the same level of accuracy with both methods.

REFERENCES

[1] Nama, V., Hegde, V., & Satish Babu, B. (2021). Sentiment analysis of movie reviews: A comparative study between the Naive-Bayes classifier and a rule-based approach. *2021 International Conference on Innovative Trends in Information Technology (ICITIIT)*.
[2] Shilpa, P. C., Shereen, R., Jacob, S., & Vinod, P. (2021). Sentiment analysis using deep learning. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*.
[3] Raisa, J. F., Ulfat, M., Al Mueed, A., & Reza, S. (2021). A review on Twitter sentiment analysis approaches. 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD).
[4] Subramanian,R.R., Akshith ,N., Murthy,G. N., Vikas,M., Amara, S., & Balaji,K. (2021)A Survey on Sentiment Analysis. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

[5] Man, R., & Lin, K. (2021). Sentiment analysis algorithm based on BERT and Convolutional neural network. *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*.
[6] Basri, R., Mridha, M., Hamid, M. A., & Monowar, M. M. (2021). A deep learning based sentiment analysis on Bang-Lish disclosure. *2021 National Computing Colleges Conference (NCCC)*.
[7] Kumawat, S., Yadav, I., Pahal, N., & Goel, D. (2021). Sentiment analysis using language models: A study. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*.
[8] Chandra, Y., & Jana, A. (2020). Sentiment analysis using machine learning and deep learning. *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*.
[9] Rahman, A., & Hossen, M. S. (2019). Sentiment analysis on movie review data using machine learning approach. *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*.
[10] Yasen, M., & Tedmori, S. (2019). Movies reviews sentiment analysis and classification. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. 828