International Journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/875-880

ISSN: 2319-4847

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

[1] Pushpa Soumya, [2] S. Jyothsna,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Article Info

Received: 29-04-2025 ***Revised: 06 -06-2025*** ***Accepted: 17-06-2025*** ***Published:28/06/2025***

Abstract

Finding the polarity (sentiment) and intention of a piece of textual data is the core emphasis of sentiment analysis, one of the key challenges in natural language processing. Data may be presented at various levels, from sentences to whole documents. It is much simpler to convey our feelings when we speak to one another as humans. However, it gets more difficult to discern the emotions when dealing with robots. Text analytics and other natural language processing approaches are used in sentiment analysis. Critical reception is a major factor in determining a film's box office success or failure. A wider audience will be greatly impacted by these evaluations. As a result, creating a reliable model that can correctly categorize movie reviews is crucial. This research uses six different machine learning models to categorize movie reviews, and then compares and contrasts them to find the top classifier according to a number of criteria.

Keywords

topics such as movie reviews, sentiment analysis, polarity identification, natural language processing, and ML models

I.INTRODUCTION

People nowadays are more likely than ever to voice their opinions and thoughts via various online mediums, such as social networking applications, corporate survey forms, email, etc. Elements like extremely polarizing material, evaluations of movies and products, and big societal events are commonplace in their discussions. the third A person's emotional reaction to a product, concept, or issue may be uncovered via sentiment analysis. Sentiment analysis has its roots in the 1950s. Written paper documents were its main usage. Natural language processing (NLP), statistics, and machine learning are some of the methods used for the study. Sentiment analysis may be categorized into two main types: a)Identifying subjectivity and objectivity b. Identification based on features or aspects Sorting text into categories according to its polarity (emotional value) is all there is to subjectivity/objectivity. It doesn't matter in what setting it occurs. Finding out how people feel about various parts of an item or context is possible using feature/aspect-based identification. Text may be grouped into different emotional states using this category, such joyful, sad, angry, etc. [7] In most cases, written evaluations of films highlight both the good and bad parts of the film. Finding out how a viewer felt and what they thought about different parts of the show is helpful. Filmmakers may also

International Journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/875-880

ISSN: 2319-4847

benefit from this skill while trying to decipher and comprehend audience emotions. In order to determine whether a movie is worth seeing, people often read reviews written by those who have already seen it. Therefore, the success or failure of the product is closely related to these evaluations and will have a significant influence. These days, sentiment analysis is most often used to improve user engagement and implement new company strategies in areas such as social media monitoring, product analysis, and consumer feedback analysis via online forms, among others. [5]

II.LITERATURE SURVEY

A. Title: A Rule-based Approach vs. the Naive-Bayes Classifier for Sentiment Analysis of Movie Reviews Vihaan Nama, Vinay Hegde, and B. Satish Babu 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. A dataset called Movie Reviews was taken from the NLTK library. Utilizing the AFINN-111 dataset, a rule-based approach was also developed. Words in this dataset are rated from -5 (very negative) to +5 (very positive). The emotion is determined by adding the values of each word in the phrase. After comparing both approaches, it was determined that Naïve Bayes performed better than the rule-based method. The usage of complicated terms in reviews caused the Naïve Bayes algorithm to fail to give accurate results. When the evaluation included mixed feelings, it also made inaccurate predictions. [1] Section B: Deep Learning for Sentiment Analysis Vinod P., Shilpa P. C., Rissa Shereen, and Susmi Jacob are the authors. The primary objective is to categorize tweets sent by users. We used "word2vec" to build semantic word vectors for each word in the input tweets. In order to train and categorize the emotions, the retrieved features are then fed into LSTM and RNN. The suggested paradigm starts by dividing emotions into positive and negative categories, and then further subdivides each group. All sorts of tweets, both good and bad, make up the first dataset; classes pertaining to positive emotions make up the second; and classes pertaining to negative tweets make up the third. The methods used for feature selection include TF-IDF and Doc2Vec. Long short-term memory (LSTM) is fed the words retrieved in the previous stage. The LSTM model outperforms the RNN and CNN models, according to the experimental data. Cons: In order to make the system more tailored to each individual user, it is necessary to create a system that attempts to determine their personality. [2] in Section C: A Survey of Methods for Analyzing Twitter Sentiment S M Salim Reza, Abdullah Al Mueed, Maliha Ulfat, and Jasiya Fairiz Raisa are the authors. Examining different approaches to sentiment analysis using data from Twitter is the goal. Machine learning and deep learning models, in conjunction with lexicon-based and other hybrid techniques, provide the basis of the investigation. On many datasets, the models are evaluated and put into action. With an accuracy rate of 98%, machine learning based models outperformed all other models and datasets. Results demonstrate that self-constructed datasets provide superior accuracy. Drawbacks: To address these issues, more research using novel computational approaches is needed to determine the exactness of opinions in a dynamic setting. the third A Survey of Sentiment Analysis (D. Title) Srikar Amara, Karnam Balaji, Manchala Vikas, Gogula Narasimha Murthy, Nukala Akshith, and RRaja Subramanian are the authors. An outline of the steps to do sentiment analysis on any textual data type is provided in this work. The article describes in detail the several stages of pre-processing. Various techniques, including those based on lexicons and machine learning, are used. We employ BOW and TF-IDF as our feature engineering approaches. Deep learning techniques like CNN, RNN, and LSTM are among the models, along with Naïve Bayes, Support Vector Machines, Decision Trees, and others. The last step is to use the IMDB Dataset to compare the outcomes of each model. Limitations: Sequence based models are accurate, but they are difficult to put into practice. Utilizing novel preprocessing and feature extraction techniques to construct simple models should be the focus of future research. [4] Subject: E. Title: BERT and CNN-Based Sentiment Analysis Method Writing by Rui Man and Ke Lin In this study, we use the BERT algorithm and other deep learning techniques to conduct sentiment analysis on the "ChnSentiCorp" dataset of Chinese-language hotel reviews. There are three main components to the BERT model's Encoder-Decoder architecture: the Encoder, the Self-attention layer, and the Decoder. A feedforward neural network and a self-attention layer make up the encoder module. The query, key, and

International Journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/875-880

ISSN: 2319-4847

value vectors will be computed by the self-attention layer. To find out where a word is in relation to other words in a phrase, the encoding and decoding layers utilize an extra vector. The results show that BERT-CNN outperforms other methods when it comes to feature extraction. In the future, we want to improve sentiment analysis models by constructing a 2-way LSTM model and a CNN that can be coupled. [5] A Sentiment Analysis on Bang-lish Based on Deep Learning (F. Title) The disclosure Rabeya Basri, M.F. Mridha, Md. Abdul Hamidf, and Muhammad Mostafa Monowarf are the authors. Here, Bengali Reviews published in English are subjected to sentiment analysis using attention-based CNN. An RNN comparison is made with the outcome. We employ a proprietary dataset with 5,000 brief paragraphs labeled with their emotions. There are a total of three layers in the suggested model: the input layer, the convolution layer, and the max-p pooling layer, the latter of which supplies data to the attention layer. Compared to multi-class classification, the models' performance in binary classification was satisfactory. Limitations: Improving the efficacy of sentiment analysis on Bengali text written in English alphabets is possible using transformer models that use encoder-decoder architecture. [6] G. Research on Sentiment Analysis by Means of Language Models Spraha Kumawat, Inna Yadav, Nisha Pahal, and Deepti Goel are the authors. The goal of this research is to learn about several models that use deep neural networks to evaluate and label texts as having neutral, positive, or negative emotions. Human interference in dataset labeling is eradicated by the suggested technique. Roberta, Electra, Bidirectional LSTM, and the BERT model are some of the models used in the experiments. As an important statistic, Matthews' correlation coefficient (mcc) has been used. With 81% accuracy, the BERT algorithm has achieved the best result. [7] Restrictions: 0 H. Evaluating Emotions with the Use of Deep Learning and Machine Learning Yogesh Chandra and Antoreep Jana are the authors. The primary goal is to categorize tweets as good, negative, or neutral using sentiment analysis with the use of several machine learning classifiers, deep learning models based on polarity, etc. Additionally, a voting-based classification method is considered, and several performance indicators are used to compare the models' outputs. The findings showed that LSTM-CNN models performed the best in terms of classification accuracy. Looking forward, it's feasible to build an architecture using deep learning models that can achieve prediction accuracy on par with humans. [8] Part I: Atiqur Rahman and Md. Sharif Hossen's Machine Learning-Based Sentiment Analysis of Movie Reviews The overarching goal is to apply sentiment analysis to a collection of movie reviews using five distinct ML methods. Preprocessing the text, creating features using vectors, training and testing the models, and finally, evaluating the outcomes using various performance measures including recall, accuracy, precision, and F1-score are all part of the process. Decision Trees, Maximum Entropy, Multinomial Naïve Bayes, SVM, and Naïve Bayes are the classifiers that are used. With an accuracy of 88.5%, Multinomial NB has shown to be the most effective method. Contrarily, SVM has shown superior recall value in comparison to others. [9] Looking forward, deep learning algorithms may also be used for polarity classification. J. Title: Analyzing and Classifying the Sentiment of Movie Reviews Writers: Sara Tedmori and Mais Yasen Our primary goal is to use eight well-known classifiers to do polarity classification on the IMDB Movie Reviews dataset. To prepare the dataset for model training, basic text preprocessing is carried out and features are extracted. There was a comparison of the models using evaluation criteria such as recall, accuracy, f1-score, precision, and AUC. When compared to the other models, Random Forest did the best, while Ripper Ruler Learning (RRL) performed the poorest. Looking forward, it will be able to do sentiment analysis using data from a variety of languages. In addition, there is room for improvement in the accuracy of model outputs via more study. [10]

III.BLOCK DIAGRAM

International Journal of application or innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/875-880

ISSN: 2319-4847

Fig. 1. Block Diagram for proposed system

You can see the whole process laid out in Fig. 1. The dataset is subjected to text preprocessing before feature extraction can begin. Afterwards, several machine learning algorithms are fed data that has been split into training and test sets. Lastly, we compare the outcomes of all the models.

IV.IMPLEMENTATION

Section A. Python Python 3 and subsequent versions, which provide a wealth of libraries and modules for analysis, have been used for our experimental goals. B. Bookstores 1) Pandas is a popular Python package for data analysis. It can read data in a variety of formats (CSV, JSON, etc.) and manipulate data, two essential components of every data analysis project. 2) Matplotlib is an in-built library that 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.

V.EXPERIMENTAL SETUP AND RESULTS

A. Database Our team has compiled the "IMDB Dataset" from several kaggle warehouses. With 25,000 favorable evaluations and 25,000 negative reviews, this dataset is quite balanced.

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 first five rows of the dataset, together with their categorization, are shown in Figure 2. B. Methodology 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 preparation, which entails a number of operations including tokenization, stemming, stop word removal, and so on. 2-Engineering Features: The Bag of Words (BOW) model and the Tree-Frame-IDF (TF-IDF) model are the feature extraction methods employed in this case. A dictionary that records every occurrence of a term is kept up-to-date using the first approach. While both models employ frequency, the latter places more focus on the relative value of individual texts within a corpus of documents. Thirdly, we create training and testing datasets by dividing the whole dataset into two parts: training (70%) and testing (30%). 4) Algorithm Application: Following the feature extraction using BOW and TF-IDF, we will train machine learning models on the training data. Logistic Regression, Support Vector Machine, Decision Trees, K-Nearest Neighbors, and Xgboost are the methods that were used. 5) We compare the results and test the models with testing data (30%) after step 4 to see how well they work and how accurate they are. Metrics such as F1-score, recall, support, and accuracy are used for the comparison.

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

Fig. 3 shows the status of dataset after performing text preprocessing like stemming, tokenisation, stop words removal etc.

International Journal of application or Innovation in engineering and management

Ijaiem.com/June 2025/ Volume 14/Issue 1/Article No-1/875-880

ISSN: 2319-4847

Fig. 4. Word Cloud for Positive Words

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

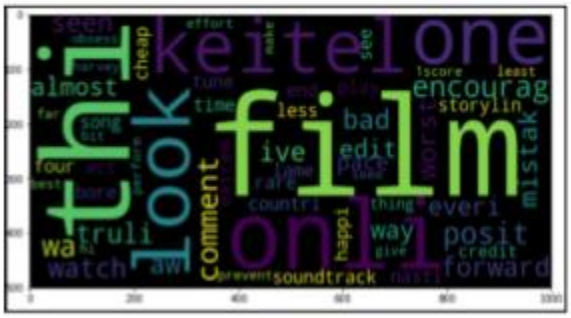

Fig. 5. Word Cloud for Negative words

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

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

Fig. 6. Accuracy of different algorithms using BOW technique

Fig. 6 gives details about performance of different algorithms when Bag Of Words (BOW) feature extraction technique is used.

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

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

Fig. 7 gives details about performance of different algorithms when TF-IDF feature extraction technique is used.

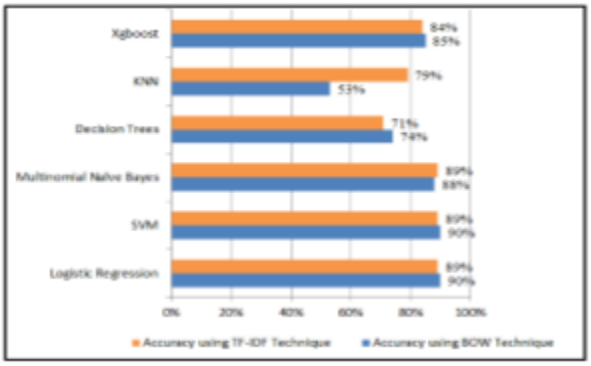

Fig. 8. Comparison between algorithms

Fig. 8 shows comparison between all 6 algorithms using both BOW and TF-IDF techniques.

VII.CONCLUSION

In this study, we used sentiment analysis to a dataset consisting of movies. The experimental findings show that Multinomial Naive Bayes, Logistic Regression, SVM, and the BOW model all achieve an accuracy of 85% or above. However, the same model had the lowest accuracy when trained using the KNN approach. While the BOW model achieved

an accuracy of 89%, the TF-IDF model, Logistic Regression, SVM, and Naïve Bayes also achieved 89%. When it came to the TF-IDF model, decision trees were the least accurate. Xgboost achieved almost identical results for the two 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).