IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14920

The Sentiment Spectrum: A Comparative Study Using NLP, Machine Learning and Deep Learning.

Faman Bushra[1], Seema Khanum[2], Hema Prabha[3]

Dept of MCA, Surana College Autonomous, Bangalore[1-3]

Abstract: A Comparative Study of Sentimental analysis using the NLP and different machine learning and deep learning techniques focuses on analyzing the subjective information conveyed with the expression. This encompasses appraisals, opinions, attitudes or emotions towards a particular subject, individual, or entity. Conventional sentiment analysis solely consider text modalities and the derives sentiment by identifying the semantic relationships between the words in the sentences. Despite this, some expressions, such as exaggeration, sarcasm and humor pose a challenge for the automated detections when conveyed only through the texts. The similar approach can precisely determine the implied sentiment polarities which contain all positive, neutral, Negative sentiment. This research communities can shown significant interest with the topic because of its potential for both the practical application and education related research. With this fact, the paper aims to present all analysis of recent ground breaking research studies which helps the deep learning models in many modalities and works.

Keywords: Sentiment Analysis, Deep Learning, Emotion Recognition, Opinion Mining, Sarcasm and Humor Detection.

I. INTRODUCTION

Natural Language Processing has made remark- able progress in understanding and generating human language. Traditional sentiment analysis or classifica- tion models have been effective at handling simple, di- rect expressions of opinion or emotion. However, real- world communication is rarely that straightforward. People often use sarcasm, irony, humor, mixed emo- tions, and implicit meanings, which are difficult for ma- chines to interpret accurately. By leveraging the future advancement methods and techniques that is contextual embed(e.g., BERT, GPT), multimodal learning, emotion classification, and commonsense reasoning, researchers aim to make ma- chines more adept at interpreting the richness of human languages. The application of the sentimental analysis tools on the free text data which is taken from medical and research fileds in health domains. This is maily attributed to the state of art applications which is designed for the different area of domain. Such as social media and ecommerce website and there is some knowledge which is related to performance in the survey detailed data. In NLP the sentimental analysis is means extracting the sentimental and detailed opinion from the text similar topics and entities. Which can be events, places, people and organizarion. This is similar to the task of text classification like negative, positive and neutral data. The research done in this Area is taken place in the different areas of range and granularities in the sentimental analysis which occur in different stages.

Figure 1: Sentiment analysis

DOI: 10.17148/IJARCCE.2025.14920

Same inside exploring various aspects similar to the entities, which is connected to different levels of sentiments[9]. Affective computings are the field in integration that can traverses psychology, computer science,psychology and cognitive sciences. Based on the nmber of years, sentimental analysis is vigorous domain in the explorationl affective type of computing. [10]In the initial days, NLP systems are mostly rule-based. The development of rule-based sys- tems is quite difficult as it requires significant human in- tervention in the form of domain expertise to frame the rules. It is required to reframe the rules with even with a small change in the input data which makes it expen- sive and laborious. Machine learning systems to some extent brought flexibility in developing NLP systems. Machine learning systems learn the rules during train- ing and thereby avoids the laborious process of man- ual rule framing. However, the main drawback in machines learning the models is the requirement of feature engineering which again requires domain expertise. In straightforward terms, and multimodal sentiment analysis, MSA is a abbreviation which is defined as the type of the sentiment analysis where the multiple forms of data that is videos, audios and images are taken into considerations. More than the types of data formats are passed to the models which is helpful to predict the sentiments. Now, MSA is burgeoning the research field. And which is crucial for everyone to work more furthermore since everyone have many internet community platforms, that is like twitter, facebook, etc

,.Where the people can easily express their thoughts and opinions in the pictures, sounds, and video recordings and not only by the text. So, as a result of this enormous volume of the rich opinion multimodal Content and an efficient analysis system is very necessary. In this case, MSA may either be bimodal, which combines any 2 modalities (text+ images or text + audio or images + audio), or tri- modal, which combines all three modalities (text + au- dio + images).The multimedia content will be helpful to provide the more descriptive detailed informations about the topics particularly; hence, MSA is an indispensable area of study.[10] To evalu- ate and complement the literature, we implemented two practical models for sentiment classification using the IMDB movie review dataset: a TF-IDF + Logistic Re- gression model and a Convolutional Neural Network (CNN) based deep learning model. The logistic regres- sion classifier achieved an accuracy of 89%, effectively handling sparse vector space representations, while the CNN model, trained on padded sequences of tokenized text, achieved a test accuracy of 87%, highlighting the trade- offs between feature-based and deep learning ap- proaches. These experiments also included visualiza- tions such as confusion matrices and ROC curves for interpretability and performance comparison.[5]

In the recent times, there is a lot of trending application fields of the sentimental analysis which is affecting the computing, aspect extraction, knowledge extraction, text summarization, movie review, product recommendation, opinion mining and language understanding. Two of the covid-19 survey have datasets which were used in this study, both the teams collected from the national institutes of health (NIH) and standard university. The data collected were used to assess the general topics experiences by the participants during the pandemics lockdown. In the sentimental analysis, an in depth of analysis is used to the determination of the strength of feelings which is known as the sentimental scoring. Sentimentak analysis is conducted in mainly at the three level sentences and aspects levels. However the document levels can only discover general polarities and not particular emotions for the each entity (Sachdeva and Kumar 2021). Its gol is to classify all the done opinion texting as a only one single type that is, negative and positive vision for a product.[2]

1.1 Introduction to the Naive Bayes

Naive Bayes is excellent in the terms of data processing speed, as will be described in the next section. Herein, we analyze the speed of pattern analysis by machine learning using naïve Bayes and NLP by keyword. The aim of all similar experiment was to measure the processing speed when the pattern of analysis was to performing using the proposed machine learning or NLP methods. [2]

1.2 Logistic regression

Logistic regression is the algorithm used in the classifica-tion tasks and predictive analysis. It uses linear regres- sion equation to make discrete binary outputs but un- like linear regression, its cost function is Sigmoid func- tion. This function is an S-shaped curve and can also be called the logistic function. The hypothesis of all this algorithms tends to limit many of the logistic functions between the 0 and 1.[1]

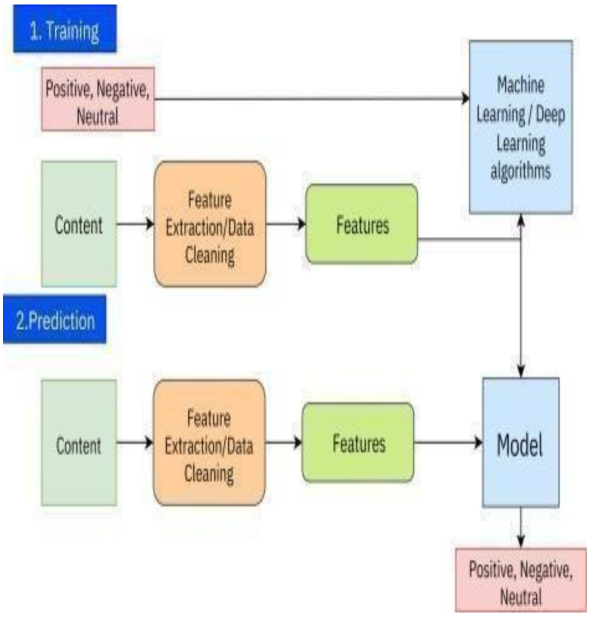

Figure 2: This figure illustrates the working process of sentiment analysis.[8]

1.3 BERT

In sentiment analysis, BERT has demonstrated state- of-the-art performance across various datasets and do- mains, including customer feedback, product review and social media posts. Its ability to fine-tune on domain-specific data while retaining its pretrained lin- guistic knowledge makes it a powerful tool for extract- ing and classifying sentiments from complex and in- formal texts. Furthermore, BERT has proven effective in handling sentiment expressions involving sarcasm, negation, and implicit emotions, which were tradition- ally challenging for earlier models.
This research aims to leverage the strengths of BERT for sentiment classification and explore its capa- bilities in capturing nuanced and complex expressions in the textual datas. The main goal is to achieve not only high ac-curacy but also improve the interpretability of sentiment predictions in real-world contexts. [4]

II. LITERATURE SURVEY

M. Wankhade et al. (2022) presented a comprehensive survey of sentiment analysis techniques, applications, and challenges. Their work highlighted issues such as sarcasm detection, context-based sentiment analysis, high computation costs, and difficulty in detecting neu- tral or mixed sentiments.

Ananya Pandey and D. K. Vishwakarma (2024) explored multimodal sentimental analysis using the deep learning for many videos, audios, images and text. They identified limitations like data scarcity, weak fusion strategies, heavy computation requirements, and limited model in- terpretability.

J. R. Jim et al. (2024) reviewed advancements in NLP-based sentiment analysis, covering applications, datasets, and ML/DL models. The authors pointed out challenges in explainability, domain adaptation, nu- anced sentiment detection, and data availability.

Md. Shofiqul Islam et al. (2024) discussed deep learning models and proposed a hybrid CRDC model. Their study emphasized problems such as data imbal- ance, domain adaptability, linguistic complexity, model interpretability, and high computational demand.

M. S. U. Miah et al. (2024) developed a cross- lingual sentiment analysis framework using translation and an ensemble of RoBERTa, mBERT, and GPT-3. They noted issues including limited language coverage, translation errors, small datasets, and high computa- tional cost.

J. A. Lossio-Ventura et al. (2024) compared Chat- GPT and fine-tuned OPT against other widely used sen- timent analysis tools for COVID-19 text. They found that more datasets and domains are needed, alongside a reliable gold standard and cost efficiency.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14920

K. Bu et al. (2024) reviewed the role of prompt learning in sentiment analysis. They identified diffi- culties in prompt generation, interpretability, handling mixed sentiments, and the need for benchmark datasets.

III. METHODOLOGY

3.1 Dataset

The IMDB Movie Review Dataset is employed for all experiments. It comprises 50,000 preprocessed reviews labeled as either positive or negative. The dataset is equally divided into 25,000 training samples and 25,000 test samples to ensure a balanced binary classification task

3.2 Data preprocessing

Preprocessing is an essential step for improving Lower- casing: All the text is converted to lowercase. Tokenization: sentences are split into the individual tokens(words). Stopword Removal (for traditional models). Padding and Sequence Handling: For deep learning models, token sequences are padded to a fixed length to maintain uniform input shape.

3.3 Naïve Byaes

The main goal of the research is to analyses the data from the surveys and to decide whether it is suitable to be analyzed with the use of the discussed data mining methods. A graphical description of the processes involve in sentiment analysis is detailed in Figure 1 below.

3.4 Logistic regression

\ This study helps to employ two different approaches which helps in the sentiment analysis on text to sentiment analysis on textual data: the traditional machine learning model and the deep learning model. The objective is to evaluate and compare their performance in classifying sentiments of movie reviews from the IMDB data. learning model and a Deep Learning Model. The ob- jective is to evaluate and compare their performance in classifying sentiments of movie reviews from the IMDB dataset.

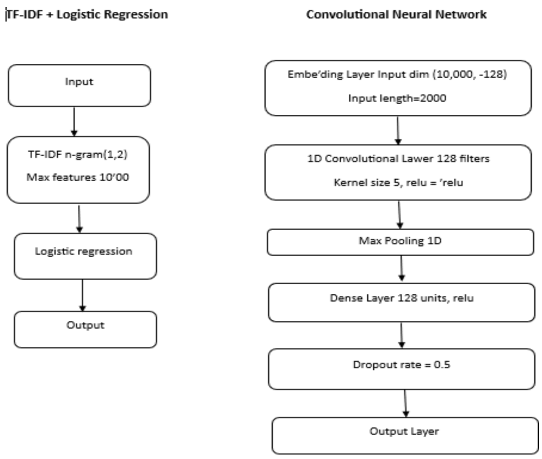

Figure 3: model architecture

3.4.1 Model Architectures

TF-IDF + Logistic Regression A classical pipeline combining TF-IDF vectorization with Logistic Regression is implemented as a baseline. The model trans- forms text into sparse numeric vectors and applies a logistic function to classify sentiment. Vectorizer: TF- IDF with n-gram range.

3.5 CNN

CNN-Based Model Architecture The core classification model consists of an embeddings layer followed by the multiple one dimentional convolutional layers with max- pooling, ReLU activation, and dropout for regularization. These layers extract local semantic features such as phrase patterns and n-gram sentiments. A flattening layer is then used to convert the feature maps into a one- dimensional feature vector, the folloed dense layer and the sigmoid

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14920

activated output layer for binary classification. Model Training and Evaluation Comparative Analysis with Transformer Models To benchmark the CNN model, transformer-based models such as BERT and RoBERTa are fine-tuned on the same dataset using HuggingFace's Transformers library. These models are evaluated under the similar metrics. Additionally, for transformer-based models, attention visualizations is used in the understanding how different the tokens can influence sentiment decisions. These methods allow insights into model behavior and alignment with ethical AI principles. Model Deployment Consideration (Optional) For potential real-time applications, lightweight models are further compressed using pruning and quantizations, making the sutable deployment in edge the environments like mobile devices. The CNN model is compiled with a binary cross- entropy loss function and optimized using the Adam optimizer. The training is con- ducted over 5–10 epochs with a batch size of 64, using 20p of the training set as validation data. The model's performance is evaluated using accuracy, precision,recall,F1-score,and-AUC

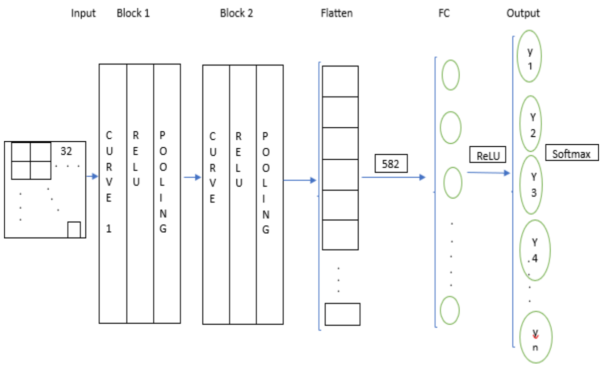

Figure 4: CNN architecture

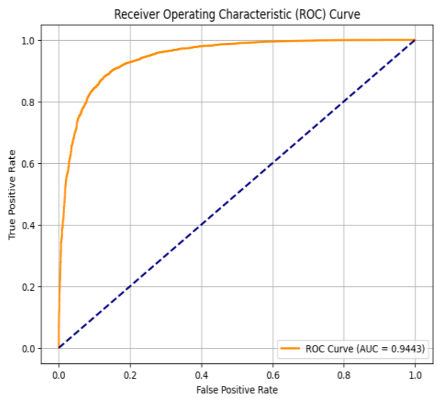

Figure 5: ROC Curve

The Roc curve will show how well the model will separates the positive and negative sentiments. With an AUC of 0.9443, the Logistic Regression model demonstrates excellent discriminative power, meaning it can be reliable distinguish positive from the negative reivews. The curve's step rise the forward top-left corner indicates the more high positive rate with very few of the false positives, proving all the model performs far better than the random guessing.

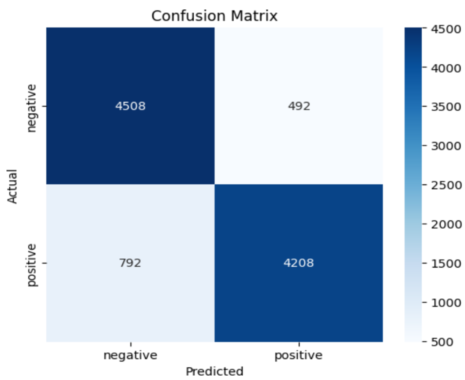

Figure 6: Confusion Matrix

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14920

True Negatives (TN = 4508)
These are reviews that were actually negative and correctly predicted as negative. This shows the model's strong ability to identify negative sentiments accurately.

False Positives (FP = 492)
Negative reviews that the model wrongly classified as positive. These misclassifications indicate that the model occasionally gets "fooled" by negative sentences containing superficially positive words (e.g., sarcasm, irony).

False Negatives (FN = 792)
Positive reviews predicted it was negative. This highlights that the model sometimes fails to capture subtle positivity (e.g., mild praise or implicit appreciation).

True Positives (TP = 4208)
Reviews that were truly positive and correctly predicted as positive. A high TP count reflects the model's strength in recognizing positive sentiment patterns in the dataset. Overall Interpretation

Accuracy =(Tp+tn)/total= (4508 + 4208) / 10,000
= 88.98%, matching your reported result.
The confusion matrix confirms that Logistic Regression handles negative reviews slightly better than positive ones, but overall, both classes are balanced with precision and recall around 0.89.

Misclassifications (FP + FN) mainly arise due to contextual complexity (sarcasm, mixed emotions), which aligns with your discussion about the limitations of traditional models vs. CNN/transformers.

IV. RESULTS AND DISCUSSION

4.1 Results
Table 1 shows the observation.

Table 1

Input	Output
I love this new phone, it works perfect.	The predicted sentiment is positive.
The service was slow and the food was cold.	The predicted sentiment is negative.

4.2 Logistic Regression
The model achieved an overall accuracy of **88.98%**on the test set. Table 4 presents the detailed classification report.

Table 2: Classification Report logistic regression

Class	Precision	Recall	F1-score	Support
Negative	0.90	0.88	0.89	5000
Positive	0.88	0.90	0.89	5000
Accuracy	-	-	0.89	10000
Macro Avg	0.89	0.89	0.89	10000
Weighted Avg	0.89	0.89	0.89	10000

Table 3: Best model parameters for the logistic regressions

Parameter	Value
Class	1
Class Weight	Balanced
Penalty	L2
Solver	Liblinear

4.3 CNN Model
The Convolutional Neural Network (CNN) model was trained for 5 epochs. The highest validation accuracy of

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14920

88.48% was achieved in Epoch 2. However, slight overfitting was observed after Epoch 3, as validation ac- curacy plateaued while training accuracy continued to increase. The final test accuracy was **87.16%**. Training and validation metrics are summarized in Table 5.

Table 4: CNN Training and Validation Results per Epoch

Epoch	Accuracy	Loss	Val_Accuracy	Val_Loss
1	0.6702	0.5715	0.8727	0.3103
2	0.8925	0.2721	0.8848	0.2793
3	0.9415	0.1645	0.8842	0.2996
4	0.9723	0.0866	0.8763	0.3638
5	0.9825	0.0521	0.8727	0.4475

[4]

[5]

On the test dataset, the CNN achieved **87.16%** ac- curacy and demonstrated reliable performance for sen- timent classification. For a sample input, the predicted sentiment was **positive**.

V. CONCLUSION

This study compared the logistic regressions and the CNN for the sentiment analysis on the IMDB dataset. Logistic regression with the TF-IDF achieved **88.98% accuracy**, showing strong performance and interpretability, while CNN reached **87.16%**, effectively capturing local text patterns. Logistic Regression proved efficient for feature- based tasks, whereas CNN offered better adaptability for contextual learning. Both struggled with sarcasm and mixed emotions. Overall, Logistic Regression remains a reliable baseline, but deep learning models that is like deep learning and CNN— and more advanced transformers such as **BERT or RoBERTa**—hold greater potential for handling complex sentiments.

REFERENCES

[1]. Aliman, G., Nivera, T. F. S., Olazo, J. C. A.,Ramos, D. J. P., Sanchez, C. D. B., Amado, T. M.,Arago, N. M., Jorda Jr, R. L., Virrey, G. C., and Valenzuela, I. C. Sentiment analysis using logistic regression. *Journal of Computational Innovations and Engineering Applications*, 7(1):35–40, 2022.

[2]. Back, B.-H. and Ha, I.-K. Comparison of senti- ment analysis from large twitter datasets by naïve bayes and the natural language processing methods. Journal of the information and the communications *con- vergence engineering*, 17(4):239–245, 2019.

[3]. Bu, K., Liu, Y., and Ju, X. Efficient utilization of pre-trained models: A review of sentiment anal- ysis via prompt learning. *Knowledge-Based Sys- tems*, 283:111148, 2024.

[4]. Devlin, J., Chang, M.-W., Lee, K., and Toutanova,K. Bert: the pre-training of deep bidirectional of tranformers for the language understanding. In proceedings the 2019 conferences of all the North American chapters of the associations for the computational linguistics: human language technologies, volume 1 the long and the short papers, page of 4171–4186, 2019.

[5]. Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli,K. Z., Zulkifli, N. S. A., Rahman, M. M., and Moni,M. A. Challenges and the future in deep learnings for the sentimental analysis: the comprehensives reviewing and the proposed novel hybrid approach.Artificial IntelligenceReview,57(3):62,20.

[6]. Jim, J. R., Talukder, M. A. R., Malakar, P.,Kabir M. M., Nur, K., and Mridha, M. Recent advancem- ents and challenges of nlp-based senti- ment analy- sis:A state-of-the-art review. Natural Language Pro- cessing Journ, 6:100059, 2024.

[7]. Lossio-Ventura, J. A., Weger, R., Lee, A. Y.,Guinee, E. P., Chung, J., Atlas, L., Linos, E., and Pereira, F. The comparison of GPT and fine- tuning opens thr pre-trained the transformers(that is OPT) against all the widely used the sentimental analysis tools, sentiment analysis of the covid-19 survey of data. Mental Health, 11:e50150, 2024.

[8]. Pandey, A. and Vishwakarma, D. K. Progress the achievemnts, and the multimodal sentimental analysis challenges using the machine learning and deep learnin:. Ap-plied Soft Computing, 152:111206, 2024.