International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 6 Issue 1, January 2025

# A Review of Sentiment Analysis and Emotion Detection from Text using Different Models

Aastha Chourasiya[1], Aasma Khan[2], Kashish Bajaj[3], Muskan Tomar[4], Tarun Kohli[5], Dipti Chauhan[6]
Student, Prestige Institute of Engineering Management & Research, Indore[1,2,3,4,5]
Professor, Prestige Institute of Engineering Management & Research, Indore[6]

***Abstract:*** *Sentiment analysis and emotion detection form an important branch of Natural Language Processing that deals with identifying and classifying emotions, attitudes, or opinions within a text dataset. The concepts of sentiment analysis and emotion detection, their application, and their challenges are studied. In this paper, we have discussed different approaches related to sentiment analysis, like the machine learning models of VADER, Roberta, Naive Bayes, SVM, etc. We also discussed the difficulties that arise due to ambiguity, context, sarcasm, irony, and cultural and linguistic variability. Our study indicates how NLP, machine learning, and cognitive science should proceed with continuous research and development to develop more accurate analysis tools.*

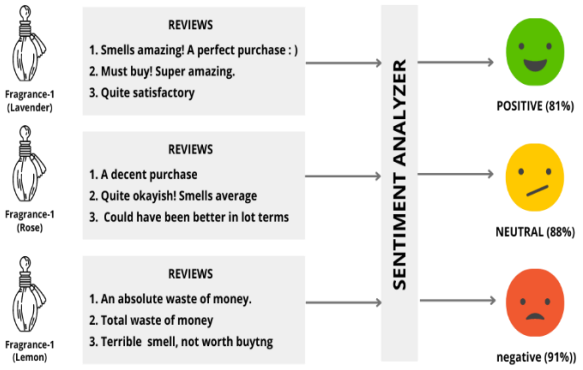
***Keywords: Sentiment Analysis, Natural language processing, VADER model, Roberta Model, Naïve Bayes model***

## 1. Introduction

Sentiment analysis refers to the emotions, attitudes, or opinions with regard to an object or subject. In this context, analysis falls into the classification of myriad kinds of subjects based on their effect and an opinion containing its polarity. Using a complex detection model, one would be justified in arguing that the class of types of emotions represented in a textual document can be clearly distinguished and distinguished. Concepts of control lead us into the domain of machine learning as well as the framework of classification [1]. It has become possible to observe a recent surge in various distinct domains, such as sentiment analysis and detection of antisocial and toxic behavior on social media, where automatic recognition of emotions can be beneficial.

Natural Language Processing has two main parts: Human Language Understanding (HLU) and Human Language Generation (HLG). But due to ambiguity of language, HLU is more difficult than HLG. But it is the former that is considered more challenging, especially with the existing ambiguities in natural language. Speech recognition, document summarization, question answering, speech synthesis, and machine translation are some applications of Natural Language Processing. Affect and sentiment analysis are essentially two specific areas of research in natural language processing, with emotion recognition as yet another central aspect of the field [2]. Although sometimes used interchangeably, they are quite different from one another sentiment analysis basically asks whether the data or information in question has a positive, negative, or neutral attitude.

.



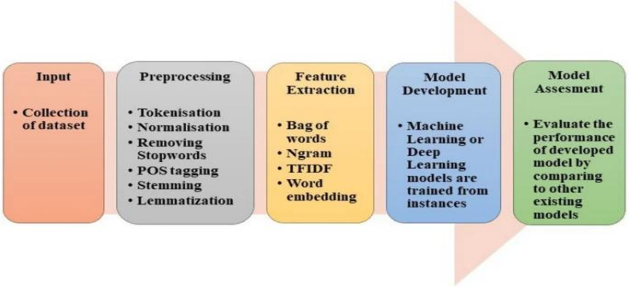

[1] Fig.1 Decoding Emotions Using Text Data: Natural Language Processing for Sentiment Analysis

Conversely, emotion detection helps identify human emotions such as angry, happy, or sad. The terms 'emotion detection', 'affective computing', 'emotion analysis', and 'emotion identification' are also sometimes used to reference these same constructs. this should be rephrased, shortened, and written in speech format [3]. Social networking has become more common as a tool of health promotion in which professionals and most citizens use within the health sector. For example, people have expressed their opinions regarding Covid-19. Considering clinical guidelines that advise patients to keep themselves away from their loved ones and close friends, an improvement in their emotional well-being was not going to be observed [4]. Healthcare workers should integrate automatic emotion recognition to decrease the misery of depression in patients. People always voice their feelings or opinions on the Internet via posts, and whenever an individual is in a state of self-pity, others will extend help to that individual, therefore, solving the problem and stopping the individual from forming worse mental conditions. There are many approaches for emotion recognition and classification based on the processing of different kinds of data such as Brain signal processing (e.g., EEG), Voice/speech processing, eye Processing of facial movement and Text processing.

## 2. Background Work

It is based on old computer processing, which was more uncomplicated than today's but not more acceptable. Among the most widespread techniques were lexicon-based methods, which refer to dictionaries of predefined words used in specific sentiments. One can distinguish between the words based on predefined sentiment dictionaries. The technique proceeds by giving each word or phrase a particular value that expresses the degree to which the given text is positive or negative. However, these approaches suffered from dependency on the coverage and quality of the lexicon and often failed to take into account nuanced language use, such as negations ("not good" was negative) or intensifiers ("very good" is highly positive). Rule-based systems improved on the lexicon-based approach by adding hand-crafted linguistic rules. The goal was to support richer situations such as changes in the context that affect sentiment, negations, and modifier effects as in the comparison of sentences like "barely good" and "extremely good". Though the systems were interpretable for less complex tasks, they had to step down in the face of the complexity of the tasks involving detecting sarcasm, irony, idioms, slang words, and informal speech. It also marks a period when machine learning will be introduced to the sentiment analysis domain. Naive Bayes, Support Vector Machines, and decision trees early models relied on feature engineering [5]. Feature engineering represented the text using handcrafted features, such as a bag-of-words, n-grams, or part-of-speech tagging. They do much better than the lexicon-based and rule-based approach but cannot yet understand the deeper context and semantic relationship in the text. Such was the absence of contextual embedding techniques, due to which it often faltered in very complex scenarios of sentiment-like opinion words with nuanced thoughts or in conflicting sentiments. These form the basis on which these advanced neural techniques are nowadays used in sentiments.

## 3. Materials And Methods




[5] Fig.2 Basics steps to perform sentiment analysis and emotion detection

**Different models were used for sentiment analysis:**
Sentiment analysis is a natural language processing technique used to determine the emotional tone or attitude conveyed by a piece of text. The various models developed for sentiment analysis, each with different strengths and weaknesses such as Vader, Bert, Random Forest, Gradient Boosting, Roberta, and Naïve Bayes, Support Vector Machine(SVM) , Artificial Neural Network etc. Some of these models are summarized below:

1. VADER (Valence Aware Dictionary and sentiment Reasoner)

VADER is a rule-based sentiment analysis tool that relies on a pre-constructed lexicon of words associated with sentiment scores [6]. It is particularly suitable for analyzing social media texts and informal language. Unlike machine learning models, VADER doesn’t require training data but instead uses heuristics to account for contextual nuances such as capitalization, punctuation, and emoticons (e.g., "GREAT!!!" is more positive than "great"). Its simplicity makes it possible to run it in real-time, with high speed, and that's why it is used where interpretability and speed are essential. However, it lacks more complex linguistic constructs, such as sarcasm and complicated sentence

structures, which makes it less accurate in specific scenarios.

2. Roberta (Robustly Optimized BERT Approach)

Roberta is a transformer-based model designed as an improvement over BERT to capture the contextual meaning of words in text [7]. It further improves BERT's capabilities by removing the next sentence prediction task and using larger mini-batches and more training data, which leads to a more robust understanding of language. In sentiment analysis, Roberta is particularly effective at identifying complex patterns in text, such as subtle emotions and contextual dependencies, which makes it highly effective for nuanced datasets. However, this accuracy comes at a computational cost, requiring significant resources for training and inference, which may not be ideal for real-time or resource-constrained applications

3. Naive Bayes

Naive Bayes is a probabilistic machine learning model based on Bayes' Theorem, assuming independence between features. It is widely used in text classification tasks, including sentiment analysis, due to its simplicity and efficiency [8]. The model computes a probability of a positive, negative, or neutral sentiment label over the input text given based on the frequency of words in labeled training data. Its independence assumption though makes computations simpler, leads sometimes to suboptimal performance in capturing the relation between words, especially more in length or complexity when used in sentences. However, Naive Bayes is valued for its speed and interpretability, making it a strong baseline for sentiment analysis tasks

SVM (Support Vector Machine):

A Support Vector Machine is a supervised classification algorithm, broadly used for sentiment analysis tasks. SVM works by finding the best hyperplane that separates the data points of different classes, say positive, negative, and neutral sentiments. For text classification, it uses features extracted from the text, such as TF-IDF scores or word embeddings, to represent the data. SVM is particularly good at handling high-dimensional datasets when the data is linearly separable [9]. It does not suffer from overfitting, especially in small datasets, because it focuses on maximizing the margin between classes. However, its performance heavily depends on the choice of kernel functions and feature engineering, and it might struggle with highly imbalanced or non-linear datasets. Despite these limitations, SVM is still a sensible choice for sentiment analysis on restricted computational resources.

**Comparison of models :**

VADER is a lexicon and rule-based model that has been developed to perform sentiment analysis. It means that VADER applies heuristics and maintains a sentiment lexicon that captures the intensity of sentiment in cases of negations and intensifiers. Its powerversion is fairly low and it seems to be particularly suited for short or informal language such as posting on social networking sites. However, owing to lack of contextual understanding of meanings, VADER does not perform well on such tasks. On the other hand, RoBERTa is stated to be an improved, optimized variant of BERT that's better suited for deep comprehension of context and accurately performing tasks such as sentiment analysis and text classification. However, on the other hand, RoBERTa or similar models require much computational power and vast datasets. Whereas onNaive Bayes is a probabilistic model that considers all features to be independent and is effective for text classification if the features are simple such as word counts. This is effective from a computational point of view but its accuracy is taken a hit for complex datasets. Meanwhile, SVM is a type of ML algorithm whose objective is to use a hyperplane to create a model which classifies data within it and is suitable for small or medium datasets. In addition, it's quite robust for text classification problems especially with TF-IDF features. However, its effectiveness is highly sensitive to hyperparameter settings, plus it's also computationally expensive to train on large datasets.

**Challenges in sentiment analysis and emotion analysis**:

Sentiment analysis and emotion analysis are two related but distinct tasks in the NLP field that pose unique challenges. One of the major challenges sentiment analysis faces is dealing with ambiguity and context, where words and phrases can have different meanings depending on the situation. Moreover, sarcasm and irony are not easy to identify because the literal meaning of the text can be the opposite of the intended meaning. Idioms, colloquialisms, and regional expressions also make it difficult for a machine learning model, which is not trained on various datasets [10]. Moreover, negation words and modal verbs can alter the sentiment of a sentence, but identifying these nuances is hard. Emotion analysis also poses some problems, such as determining emotion intensity and gradations where the emotions may vary from slight to intense and cannot distinguish one emotion from another. Further, defining and classifying emotions can be hard due to the presence of many theories and models of emotions. Contextual understanding is vital for the correct analysis of emotion, and the implicit emotions, which are not overtly stated but may be inferred from the text, are difficult to identify. Emotions can change or move over time, and these changes are hard to track down, especially in lengthy texts or conversations. Another challenge is cultural and linguistic variations. Sentiment and emotion can be expressed

differently among cultures and languages. Machine learning models might suffer from overfitting in a particular dataset or domain, which limits their ability to generalize to unseen data [11]. Finally, it is hard to understand the reasons behind the decisions of sentiment and emotion analysis models, which is a necessary step toward trustworthiness and accountability. These challenges will be overcome through continuous research and development in NLP, machine learning, and cognitive science to develop more accurate and reliable analysis tools.

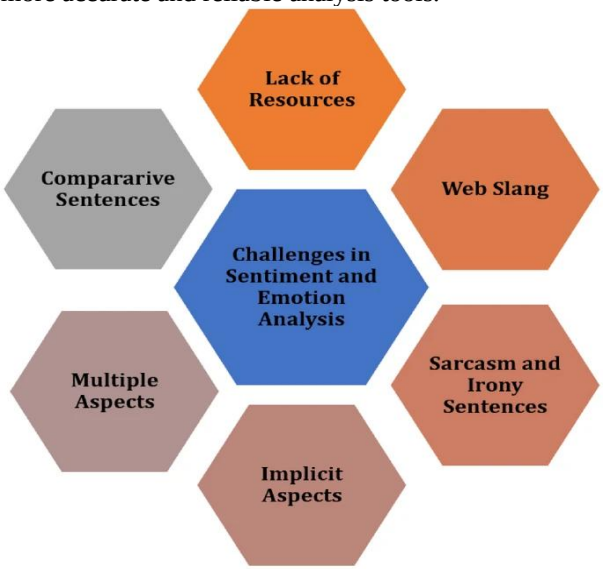



Fig. 3 Challenges in sentiment analysis and emotion detection

## 4. Conclusion

Sentiment analysis and emotion detection are highly complicated tasks requiring advanced NLP techniques to correctly identify and classify emotions, attitudes, or opinions portrayed in text data. Though promising results have come from machine learning models like VADER, Roberta, Naive Bayes, and SVM, there are certain shortcomings associated with them. Such ambiguity, context, sarcasm, irony, and cultural and linguistic variations should be addressed through continuous research and development in NLP, machine learning, and cognitive science. In that way, it is possible to unlock the full potential of sentiment analysis and emotion detection, enabling applications in fields such as healthcare, social media, and customer service. Ultimately, this study underlines the importance of interdisciplinary research in advancing our understanding of human emotions and developing more effective tools for sentiment analysis and emotion detection.

## References

[1] Chauhan, D., & Jain, J. K. (2023, January). Profiling Network Traffic by Using Classification Techniques in Machine Learning. In International Conference on Smart Trends in Computing and Communications (pp. 113-123). Singapore: Springer Nature Singapore.

[2] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.

[3] Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., & Prendinger, H. (2018). Deep learning for affective computing: Text-based emotion recognition in decision support. Decision support systems, 115, 24-35.

[4] Chauhan, D., & Jain, J. K. (2025). Classifying Sleep Health and Lifestyle Patterns: A Machine Learning Approach Using IoT and Cloud. In Revolutionizing Healthcare Systems Through Cloud Computing and IoT (pp. 151-178). IGI Global.

[5] Khule, M., Motwani, D., & Chauhan, D. (2024, September). Enhancing Network Intrusion Detection with Support Vector Machines: A Comparative Study of Feature Selection Techniques. In 2024 5th International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1281-1286). IEEE..

[6] Mändli, F. Automated Sentiment Analysis vs. Manual Coding. THREE ESSAYS ON DYNAMIC MEDIA EXPOSURE AND THE RELATED METHODOLOGICAL ISSUES OF CONTROL VARIABLE SELECTION AND AUTOMATED SENTIMENT ANALYSIS, 103.

[7] Abdal, M. N., Oshie, M. H. K., Haue, M. A., & Islam, K. (2023, December). A Transformer Based Model for Twitter Sentiment Analysis using RoBERTa. In 2023 26th International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.

[8] Jiang, L., Li, C., Wang, S., & Zhang, L. (2016). Deep feature weighting for naive Bayes and its application to text classification. Engineering Applications of Artificial Intelligence, 52, 26-39.

[9] Parmar, M., & Tiwari, A. (2024, January). Enhancing Text Classification Performance using Stacking Ensemble Method with TF-IDF Feature Extraction. In 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 166-174). IEEE.

[10] Caroppo, A., Leone, A., & Siciliano, P. (2020). Comparison between deep learning models and traditional machine learning approaches for facial

expression recognition in ageing adults. Journal of Computer Science and Technology, 35, 1127-1146.

[11] Jain, J. K., Waoo, A. A., & Chauhan, D. (2022). A literature review on machine learning for cyber security issues. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 374-385.

[12] M.joshi, S.Birla, H.Pal, K. Khatri, M.Kadwal, D. Salitra, "AI-Driven Intrusion Detection System: Leveraging Deep Learning for Network Security" Nanotechnology Perceptions ISSN 1660-6795 Q4 Journal with SJR 0.11 Volume Vol.20, S10 Sept 2024.

[13] M.kadwal et al "Automatic Facial Emotions Recognition Based on Deep Learning" Samdarshi ISSN: 2581-3986 Vol 16, Issue 4, September 2023.

[14] K. Paul Joshua, D. Srinivasa Rao, Govinda Patil, Mohit Kadwal, Jitendra Choudhary, "Cascaded ANN Based Clustering and Optimized Routing Path Selection in Mobile Adhoc Networks" SSRG International Journal of Electrical and Electronics Engineering Volume 10 Issue 6, 81-93, June 2023.