# International Journal of Research Publication and Reviews

# A Review on Sentiment Classification Analysis of Twitter Data Based on Machine Learning Techniques

*Md Wasim Reza[1], Prof. Manvendra Singh Divakar[2], Dr. Devendra Kumar Vajpai[3]*

[1]MTech Scholar, CSE, Department, NIRT Bhopal Wasimreza94@gmail.com
[2]Assistant Professor, CSE, Department, NIRT, Bhopal divakarsingh9613@gmail.com
[3]Professor, HoD, CSE, Department, NIRT, Bhopal, devendrabajpai13@gmail.com

**ABSTRACT**

In the era of social media dominance, Twitter stands as a rich source of real-time public opinion and sentiment. This paper realm of sentiment classification analysis of Twitter data, employing advanced machine learning techniques. By harnessing the vast amount of text data generated on Twitter, our research aims to develop robust models for accurately classifying tweets into sentiment categories such as positive, negative, and neutral. we explore various machine learning algorithms, including but not limited to support vector machines (SVM), random forests, and recurrent neural networks (RNN). We leverage state-of-the-art natural language processing (NLP) techniques for feature extraction and preprocessing, incorporating methods such as word embeddings and sentiment lexicons. The evaluation of our models involves rigorous experimentation on large-scale Twitter datasets, encompassing diverse topics and domains. Performance metrics such as accuracy, precision, recall, and F1 score are utilized to assess the effectiveness and generalization capabilities of the developed classifiers.

*Keywords*— *Sentiment Analysis, Text Mining, Natural language Processing, Deep Learning*

## Introduction

In today's digital era, an enormous amount of textual data is generated daily through social media, online reviews, and other online platforms. Extracting valuable insights from this vast amount of data has become a crucial task, particularly in understanding public opinion, customer sentiment, and market trends. Sentiment analysis, also known as opinion mining, has emerged as a powerful technique to automatically analyze and classify sentiments expressed in text. By leveraging the advancements in machine learning (ML) and natural language processing (NLP), sentiment analysis has gained significant attention and witnessed remarkable progress in recent years.

This review paper aims to provide a comprehensive overview and analysis of sentiment analysis techniques utilizing ML and NLP approaches. Sentiment analysis involves determining the sentiment polarity, such as positive, negative, or neutral, expressed in a piece of text. It has broad applications across diverse domains, including market research, brand reputation management, customer feedback analysis, political sentiment tracking, and social media monitoring. The integration of ML and NLP techniques has proven to be effective in handling the inherent challenges of sentiment analysis. ML algorithms provide the ability to learn from data and make predictions or classifications, while NLP techniques enable the understanding and interpretation of human language. Together, they form a powerful combination for sentiment analysis tasks. This review begins by outlining the fundamental steps involved in sentiment analysis. Data preprocessing techniques, such as tokenization, stemming, and removing stopwords, are discussed, as they are essential for transforming raw text into a suitable format for analysis. Feature extraction methods, including bag-of-words, n-grams, and word embeddings, are explored, as they capture the meaningful representation of text that ML algorithms can process.
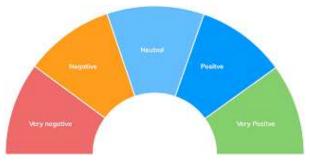


**Figure 1. Sentiment Analysis**

The review then delves into various ML algorithms commonly used for sentiment classification. Traditional algorithms such as Naive Bayes, Support Vector Machines (SVM), and decision trees are examined, along with their strengths and limitations. Additionally, the review explores more advanced techniques, including ensemble methods and deep learning models like recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which have demonstrated state-of-the-art performance in sentiment analysis tasks. Furthermore, the role of NLP techniques in sentiment analysis is thoroughly explored. Part-of-speech tagging, named entity recognition and sentiment lexicon creation are discussed as vital components for enhancing sentiment analysis accuracy and capturing linguistic nuances. The review also addresses challenges faced in sentiment analysis, including sarcasm detection, sentiment inconsistency, and domain adaptation, and explores potential solutions and strategies.

The practical applications of sentiment analysis across various domains are discussed, emphasizing the value it brings in understanding customer opinions, market trends, and social media sentiments. The review also highlights emerging trends in sentiment analysis research, such as the integration of multimodal data, transfer learning, and the ethical considerations surrounding bias and fairness in sentiment analysis. By providing a comprehensive analysis of sentiment analysis techniques using ML and NLP, this review paper serves as a valuable resource for researchers, practitioners, and enthusiasts in the field. It aims to inspire further advancements and innovations in sentiment analysis, enabling more accurate and insightful sentiment interpretation in the ever-evolving landscape of textual data.

## BACKGROUND AND RELATED WORK

Numerous studies have investigated the application of ml and nlp techniques in sentiment analysis, aiming to improve accuracy and overcome the limitations of traditional approaches. researchers have explored various ml algorithms and nlp techniques to tackle sentiment analysis challenges effectively.

One commonly used ml algorithm for sentiment analysis is naive bayes, which assumes independence between features and calculates the probability of each sentiment class given the text features. Support vector machines (svm) have also been widely employed, using a hyper plane to separate positive and negative sentiment classes. decision trees and random forests have been explored due to their interpretability and ability to capture nonlinear relationships in the data.

In recent years, deep learning models have gained prominence in sentiment analysis. Recurrent neural networks (rnns), especially long short-term memory (lstm) networks, have been utilized to capture sequential dependencies in text, allowing for context-aware sentiment classification. convolutional neural networks (cnns) have demonstrated success in capturing local patterns and hierarchical structures in text, making them suitable for sentiment analysis tasks.

Researchers have also focused on leveraging nlp techniques to enhance sentiment analysis performance. techniques such as part-of-speech tagging, named entity recognition, and syntactic parsing have been used to extract linguistic features that contribute to sentiment classification. Sentiment lexicons and word embeddings have been employed to capture semantic information and sentiment associations.

Transfer learning has emerged as another area of interest, where pre-trained models, such as bert (bidirectional encoder representations from transformers), have been fine-tuned on sentiment analysis tasks. this approach has shown promise in improving performance by leveraging large-scale pre-training on general language understanding.

Additionally, studies have addressed specific challenges in sentiment analysis, such as sarcasm detection, sentiment inconsistency, and domain adaptation. methods involving irony and sarcasm detection through linguistic cues, sentiment inconsistency modeling, and domain adaptation techniques have been proposed to handle these challenges effectively.

Overall, the combination of ml algorithms and nlp techniques has advanced sentiment analysis capabilities significantly. by leveraging the power of machine learning to learn from data and the linguistic analysis of natural language processing, sentiment analysis models have become more accurate and robust in capturing and interpreting sentiment in textual data.

Opinion mining and sentiment analysis are related in a sense that opining mining deals with analyzing and summarizing expressed opinions whereas sentiment analysis classifies opinionated text into positive and negative. Aspect extraction is a crucial problem in sentiment analysis. Model proposed in [20] utilizes topic model for aspect extraction and support vector machine learning technique for sentiment classification of textual reviews. The goal is to automate the process of mining attitudes, opinions and hidden emotions from text.

In [22], the users learn people's opinions, attitudes, and emotions towards the given posts. Lexicon based language is characterized as a polarity decision to check the words that can help the system to categorize as positive, negative and neutral tweets. Sentiment analysis is used by various parties for the marketing of products, used by public figures to analyze their activities in order to gain followers and views respectively.

In [23] author review of the existing techniques for both Emotion and sentiment detection is presented. As per the paper's review, it has been analyzed that the lexicon-based technique performs well in both sentiment and emotion analysis. However, the dictionary-based approach is quite adaptable and straightforward to apply, whereas the corpus based method is built on rules that function effectively in a certain domain. As a result, corpus-based approaches are more accurate but lack generalization. The performance of machine learning algorithms and deep learning algorithms depends on the pre-processing and size of the dataset. Nonetheless, in some cases, machine learning models fail to extract some implicit features or aspects of the text. In situations where the dataset is vast, the deep learning approach performs better than machine learning. Recurrent neural networks, especially the LSTM

model, are prevalent in sentiment and emotion analysis, as they can cover long-term dependencies and extract features very well. But RNN with attention networks performs very well.

In [24] authors discussed based on the sentiment analysis taxonomy, it has opinion mining to have the opinion polarity classification, subjectivity detection, opinion spam detection, opinion summarization and argument expression detection. On the other hand, emotion mining has the emotion polarity classification, emotion detection, emotion cause detection and emotion classification. If it is based on granularity level, it has sentence level, document level and aspect/entity level of sentiment analysis. As for the machine learning approaches, it has semi-supervised learning, unsupervised learning and supervised learning of sentiment analysis.

## FINDINGS OF THE SURVEY

The survey finds the Importance of Preprocessing. Preprocessing techniques such as tokenization, stop-word removal, and stemming/lemmatization play a crucial role in sentiment analysis. They help to improve the quality of textual data by reducing noise and extracting relevant features. Various feature extraction methods, including bag-of-words, n-grams, and word embeddings, have been explored. Word embeddings, such as Word2Vec and GloVe, have shown promising results by capturing semantic relationships between words. Traditional machine learning algorithms, such as Naive Bayes, SVM, and Decision Trees, have been widely used for sentiment analysis. However, their performance is limited by the reliance on handcrafted features and their inability to capture complex contextual information. Advancements with Deep Learning: Deep learning models, particularly CNN, RNN (LSTM, GRU), and their variants, have demonstrated significant improvements in sentiment analysis. These models excel at capturing contextual information and have achieved state-of-the-art results. Transfer learning, specifically leveraging pretrained language models like BERT, has emerged as a powerful technique in sentiment analysis. Fine-tuning pretrained models with domain-specific data has shown promising results and reduced the need for large labeled datasets.

Challenges and Open Problems: Challenges in sentiment analysis include detecting sarcasm, irony, and negation, handling multilingual data, and performing domain-specific sentiment analysis. The lack of standardized evaluation metrics and annotated datasets poses further challenges.

Applications and Future Directions: Sentiment analysis has various applications, including social media analytics, brand reputation management, and political sentiment tracking. The future direction of research includes addressing the challenges mentioned above, exploring novel techniques for sentiment analysis, and focusing on real-time and dynamic sentiment analysis in evolving online environments.

Overall, the survey highlights the advancements made in sentiment analysis using NLP and ML techniques and provides valuable insights into the current state of the field. It emphasizes the importance of deep learning models and transfer learning approaches while addressing the challenges that still need to be tackled for further advancements in sentiment analysis.

## CONCLUSION

The survey on Sentiment Analysis Using Natural Language Processing and Machine Learning highlights the significant progress made in the field. The advancements in preprocessing techniques, feature extraction, and the adoption of deep learning models have greatly improved sentiment analysis performance. Transfer learning approaches, particularly leveraging pretrained models, have shown remarkable results in capturing semantic information and domain-specific sentiment analysis. However, challenges such as detecting sarcasm, handling multilingual data, and developing standardized evaluation metrics remain. The survey emphasizes the need for further research in these areas to overcome the existing limitations and drive the future development of sentiment analysis in real-world applications.

### References

[1] Samaneh Moghaddam., Martin Ester.: On the Design of LDA Models for Aspect-basedOpinion Mining. In: ACM, 2012.

[2] Samaneh Moghaddam and Martin Ester : The FLDA Model for Aspect based Opinion Mining:Addressing the Cold Start Problem. ACM conference, pp. 909-918, 2011

[3] Lei Zhang and Bing Liu : Aspect and Entity Extraction for Opinion Mining. Springer-Verlag Berlin Heidelberg 2014., Studies in Big Data book series, Vol 1, pp. 1-40, Jul. 2014.

[4] Zhen Hai, Kuiyu Chang, and Jung-jae Kim : Implicit Feature Identification via Co-occurrence Association Rule Mining. Springer-Verlag Berlin Heidelberg 2011., LNCS 6608, pp. 393-404, Jul. 2011.

[5] Haiping Zhang, Zhengang Yu, Ming Xu, Yueling Shi : Feature-level Sentiment Analysis for Chinese Product Reviews. IEEE, pp. 135-139, 2011

[6] Zhen Hai, Kuiyu Chang, Gao Cong : One Seed to Find Them All:Mining Opinion Features via Association. ACM CIKM'12., LNCS 6608, pp. 255-264, Nov. 2012

[7] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang : Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Volume 26, No. 3 pp. 623-634, 2014.

[8] Wei Jin., Hung Hay Ho., Rohini K. Srihari : OpinionMiner: A Novel Machine Learning System for Web Opinion Mining and Extraction. In: ACM, 2009.

[9] Hui Song, Yan Yan, Xiaoqiang Liu : A Grammatical Dependency Improved CRF Learning Approach for Integrated Product Extraction. IEEE International Conference on Computer Science and Network Technology, pp. 1787-139, 2012.

[10] Luole Qi and Li Chen : Comparison of Model-Based Learning Methods for Feature-Level Opinion Mining. IEEE International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 265-273, 2011.

[11] Arjun Mukherjee and Bing Liu: Aspect Extraction through Semi-Supervised Modeling. In: Association for Computational Linguistics.,vol. 26, no. 3, pp. 339-348, Jul. 2012.

[12] Liviu, P.Dinu and Iulia Iuga.: The Naive Bayes Classifier in Opinion Mining:In Search of the Best Feature Set. Springer-Verlag Berlin Heidelberg, 2012.

[13] Xiuzhen Zhang., Yun Zhou.: Holistic Approaches to Identifying the Sentiment of Blogs Using Opinion Words. In: Springer-Verlag Berlin Heidelberg, 5–28, 2011.

[14] M Taysir Hassan A. Soliman., Mostafa A. Elmasry., Abdel Rahman Hedar, M. M. Doss.: Utilizing Support Vector Machines in Mining Online Customer Reviews. ICCTA (2012).

[15] Ye Jin Kwon., Young Bom Park.: A Study on Automatic Analysis of Social NetworkServices Using Opinion Mining. In: Springer-Verlag Berlin Heidelberg, 240–248, 2011.

[16] Anuj Sharma., Shubhamoy Dey: An Artificial Neural Network Based approach for Sentiment Analysis of Opinionated Text. In: ACM, 2012.

[17] Yulan He. : A Bayesian Modeling Approach to Multi-Dimensional Sentiment Distributions Prediction. In: ACM, Aug. 2012.

[18] Danushka Bollegala, David Weir and John Carroll : Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, pp. 1-14, 2012.

[19] Andrius Mudinas., Dell Zhang., Mark Levene. : Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis. In: ACM, Aug. 2012.

[20] Vamshi Krishna. B, Dr. Ajeet Kumar Pandey, Dr. Siva Kumar A. P "Topic Model Based Opinion Mining and Sentiment Analysis" 2018 International Conference on Computer Communication and Informatics (ICCCI -2018), Jan. 04 – 06, 2018, Coimbatore, INDIA

[21] Rita Sleiman, Kim-Phuc Tran "Natural Language Processing for Fashion Trends Detection" Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET 2022) 20-22 June 2022, Prague-Czech Republic

[22] 1d.sai tvaritha, 2nithya shree j, 3saakshi ns 4surya prakash s, 5siyona ratheesh, 6shimil shijo "a review on sentiment analysis applications and approaches" 2022 JETIR June 2022, Volume 9, Issue 6 www.jetir.org (ISSN-2349-5162)

[23] Pansy Nandwani1 · Rupali Verma1 "A review on sentiment analysis and emotion detection from text" https://doi.org/10.1007/s13278-021-00776-6

[24] Hoong-Cheng Soong, Norazira Binti A Jalil, Ramesh Kumar Ayyasamy, Rehan Akbar "The Essential of Sentiment Analysis and Opinion Mining in Social Media" 978-1-5386-8546-4/19/$31.00 ©2019 IEEE