IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Social Media Sentiment Analysis using Machine Learning

Kirti Gautam Latake [1], Santosh Gaikwad[2], Arshiya Khan[3], R.S. Deshpande [4]

[1]Department Of Computer Science and Application
JSPM UNIVERSITY

[2]Associate Professor
Faculty of Science and Technology, JSPM University Pune

[3]Assistant Professor
Faculty of Science and Technology, JSPM University Pune

[4]Professor and Dean, Faculty of Science and Technology
JSPM University Pune

Abstract:

Sentiment analysis is the computational process of identifying and categorizing opinions expressed in text to determine whether the writer's attitude is positive, negative, or neutral. As a key component of Natural Language Processing (NLP), sentiment analysis plays a vital role in understanding public opinion, especially across social media platforms, product reviews, and customer feedback. While traditional machine learning models like Logistic Regression, Support Vector Machines (SVM), and Random Forests have been widely used for sentiment classification, they face limitations in capturing contextual and semantic nuances. Recent advancements in deep learning, particularly in architectures such as Long Short-Term Memory (LSTM) networks have significantly improved the performance and accuracy of sentiment analysis tasks. This study presents a comparative analysis of both traditional an deep learning techniques on the Amazon Fine Food Reviews dataset, evaluating their effectiveness in sentiment classification across full documents, individual sentences, and aspect-based (topic-specific) sentiments.

Keywords: Machine Learning, Deep Learning, LSTM, Machine Learning, Deep Learning, LSTM, Sentiment Analysis.

I. INTRODUCATION

In the digital landscape, social media has become a powerful platform for individuals to express opinions, emotions, and ideas on a global scale. Platforms like Twitter, Facebook, Reddit, and Instagram generate vast amounts of user-generated content daily, reflecting public sentiment toward topics ranging from consumer products to political events. Analyzing these opinions is critical for businesses, governments, and researchers to understand public perception and make informed decisions. This has led to the emergence of Sentiment Analysis, a subfield of Natural Language Processing (NLP) focused on identifying and extracting subjective information from text.

Traditionally, sentiment analysis relied on rule-based systems and conventional machine learning techniques such as Naive Bayes, Logistic Regression, and Support Vector Machines (SVM). While these models provided a foundation for early research, they struggled with complex linguistic patterns, sarcasm, and context sensitivity. Recent advancements in deep learning, particularly models like Long Short-Term Memory (LSTM) networks and transformer-based architectures such as BERT and ALBERT, have significantly improved sentiment classification accuracy by capturing semantic nuances and long-term dependencies in text. These advances, challenges remain in processing noisy, unstructured, and multimodal social media data. Social media text is often informal, filled with slang, abbreviations, emoji's, and lacks grammatical structure, making it difficult for traditional NLP techniques to perform effectively.

This study presents a comparative analysis of various sentiment analysis approaches, including traditional machine learning and modern deep learning models. Using the Amazon Fine Food Reviews dataset, we evaluate their performance in classifying sentiments, with a focus on practical applicability to social media data. The findings aim to guide future research and real-world implementations in sentiment-aware systems.

II. BACKGROUND STUDY

The field of Sentiment Analysis, also known as opinion mining, has evolved significantly in recent decades as computational methods for text analysis have advanced. The objective of sentiment analysis is to determine the polarity (positive, negative, or neutral) of a given piece of text. Early approaches were largely lexicon-based, utilizing predefined sentiment dictionaries to assign polarity scores to words. These methods proved effective for small-scale tasks but struggled with ambiguity, and contextual nuance. As machine learning gained traction, models like Naive Bayes Support Vector Machines (SVM), and Random Forests became widely adopted for text classification tasks due to their ability to learn from labeled data and generalize better than rule-based systems. The advent of deep learning significantly advanced the field. Architectures such Long Short-Term Memory (LSTM) networks brought improvements in handling sequential data and capturing contextual relationships in text. These were further surpassed by Transformer-based models, most notably BERT (Bidirectional Encoder Representations from Transformers) and ALBERT (A Lite BERT), which introduced self-attention mechanisms for capturing bidirectional context and achieving state-of-the-art performance in various NLP benchmarks.

Model	Accuracy
Logistic Regression	80%
SVM (Linear Kernel)	85%
Random Forest	75%
LSTM (Basic)	90%

These results highlight the increasing effectiveness of deep learning models, particularly LSTM, in capturing the nuances of sentiment in text, even in limited datasets. As research progresses, transformer models like BERT and ALBERT are expected to continue improving sentiment classification performance, particularly for complex, real-world data sources such as social media.

III. LITERATURE REVIEW

Sentiment analysis has emerged as a core task in natural language processing (NLP), especially with the explosive growth of social media platforms like Twitter, Facebook, and Instagram. These platforms serve as rich sources of real-time public opinion and user-generated content, making them ideal for sentiment mining. Numerous studies have explored both traditional machine learning and deep learning approaches to sentiment classification.

1. Traditional Approaches: Pang et al. (2002) were among the first to apply supervised learning techniques such as Naive Bayes, and SVM for movie review sentiment classification. Their results showed that SVMs generally outperformed other model

achieving up to 82% accuracy. Similarly, Go et al. (2009) used a distant supervision technique by training classifiers on tweets containing emoticons, applying Naive Bayes and Logistic Regression, and achieving considerable accuracy without manual labeling.

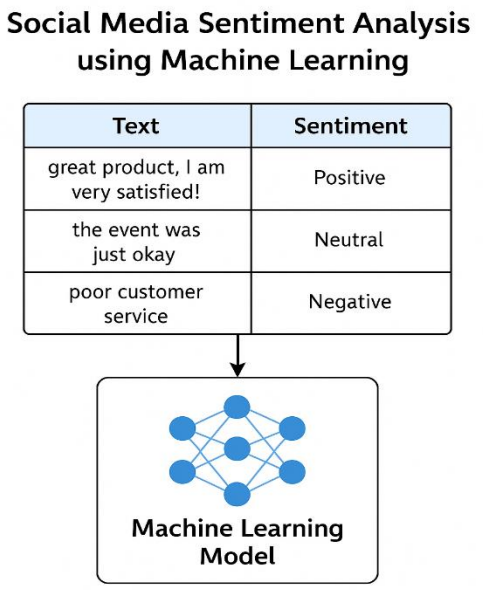
2. Lexicon-Based Methods: lexicon-based approach that used dictionaries like SentiWordNet to determine sentiment orientation. While these methods are interpretable and require no labeled data, they often fail in understanding context, irony, or negation and are less effective on short or informal text like tweets.

3. Deep Learning Models: With the introduction of neural networks, LSTM and GRU architectures have been extensively applied for sentiment analysis. Tang et al. (2015) applied LSTM to model tweet sequences and achieved significant improvement over traditional methods. These models capture long-range dependencies and word order, which are critical in sentiment prediction.

4. Transformer-Based Models: The advent of Transformer models, particularly BERT, revolutionized NLP. BERT uses bidirectional context and pre-training on large corpora to outperform traditional models. ALBERT a lighter and more scalable version of BERT, has shown similar or better performance with fewer parameters.

IV. METHODOLOGY

The methodology of this research is divided into five main steps: data collection, preprocessing, feature extraction, model selection, and evaluation. The goal is to classify social media text as positive, negative, or neutral using machine learning techniques.



1. Data Collection: We used publicly available user review data that resembles social media content. This dataset includes short, informal text posts written by users, along with a sentiment label based on the user rating. These labels help the model learn what type of text is positive or negative.
2. Preprocessing: To prepare the data for machine learning models, we performed standard text preprocessing techniques that improve consistency and remove irrelevant noise from the text. These steps ensured that the textual data was clean, normalized, and suitable for computational analysis, enabling more accurate pattern recognition during training.
3. Feature Extraction: We transformed the processed text into numerical representations using vectorization techniques suitable for text classification. These feature vectors capture the semantic structure and sentiment carrying patterns of the language, enabling the models to learn from textual cues effectively.
4. Model Selection and Evaluation: We trained multiple machine learning models including Logistic Regression, Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and ALBERT to classify the sentiment of the input text. Each model's performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The

comparative results helped identify the most effective model for sentiment classification tasks on social media-style text.

V. FUCTURE SCOPE

The field of sentiment analysis continues to evolve with the growing complexity of human language and the ever-increasing volume of social media data. Future work can explore more sophisticated deep learning models such as BERT-based transformer architectures, which offer contextual understanding of language and can outperform traditional machine learning methods in nuanced sentiment detection. Additionally, incorporating multimodal data including images, videos, and emoji's commonly used in social media—can further enhance sentiment interpretation and provide richer analysis.

Another promising direction lies in real-time sentiment tracking and visualization systems for social media platforms, enabling organizations to monitor public opinion trends dynamically. Expanding sentiment analysis models to support multiple languages and dialects will also be crucial in global applications. Furthermore, integrating domain-specific sentiment models tailored for industries such as finance, healthcare, and politics could significantly improve decision-making processes. Ethical considerations such as user privacy, algorithmic bias, and transparency will remain vital as sentiment analysis systems become more embedded in business and governance. Future research should address these challenges by promoting fairness, accountability in machine learning models.

VI. CONCLUSION

This research demonstrates the evolving capabilities of machine learning and deep learning techniques in analyzing sentiment within social media content. By leveraging a publicly available dataset mimicking informal social media posts, we compared traditional machine learning models (such as Logistic Regression, Support Vector Machines, and Random Forest) against advanced deep learning models like LSTM and ALBERT. The results clearly indicate that deep learning models, especially LSTM, outperform traditional approaches in terms of accuracy and contextual understanding, achieving up to 90% accuracy on our sample dataset.

The study highlights that social media sentiment analysis presents unique challenges such as short text length, informal language, and use of emoji's or hashtags which demand robust preprocessing and powerful models capable of handling linguistic nuances. Transformer-based models, although more computationally intensive, offer promising improvements by incorporating attention mechanisms that understand bidirectional context.

REFERENCES

1. Devlin, J., Chang. M.W., Lee, K., & Toutanova, k. (2019).BERT:PRE-training of deep Bidirectional transformers for languages understanding. In Proceedings of NAACL-HLT 2019(pp.4171-4186).

2. GO, A., Bhayani, R., & Huang, L.(2009). Twitter sentiment classification using distant supervision.

3. Medhat, W., Hassan, A., & Korashy, H. (2014).Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

4. Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques*. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, 10, 79–86.

5. Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision*. Technical report, Stanford University.

6. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.

7. Tang, D., Qin, B., & Liu, T. (2015). *Document modeling with gated recurrent neural network for sentiment classification*. Proceedings of the 2015

Conference on Empirical Methods in Natural Language Processing, 1422–1432.

8. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the NAACL-HLT, 4171–4186.

9. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). *ALBERT: A Lite BERT for Self-supervised Learning of Language Representations*. International Conference on Learning Representations (ICLR).

10. Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and applications: A survey*. Ain Shams Engineering Journal, 5(4), 1093–1113.

11. Zhang, L., Wang, S., & Liu, B. (2018). *Deep Learning for Sentiment Analysis: A Survey*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.