Literature Review: Sentiment Analysis of Social Media Posts

Krish Patel and Charles Woo

February 12, 2025

Abstract—This paper reviews sentiment analysis techniques applied to social media data, emphasizing machine learning and NLP methods. Recent studies are examined, highlighting methodologies, findings, and research gaps that our project aims to address. Social media platforms, like Twitter, are pivotal in reflecting public opinion, and sentiment analysis interprets emotional tones behind social media posts. By analyzing these sentiments, businesses can tailor marketing strategies, policymakers can gauge public reaction, and researchers can study social trends.

I. Introduction

In the digital age, social media platforms like Twitter have become pivotal in reflecting the pulse of public opinion on a wide array of topics, ranging from consumer products to political movements. Sentiment Analysis, a branch of Natural Language Processing (NLP), plays a crucial role in interpreting the emotional tones behind social media posts. By analyzing these sentiments, businesses can tailor marketing strategies, policymakers can gauge public reaction, and researchers can study social trends. This literature review examines recent studies that utilize machine learning and NLP techniques for sentiment analysis of social media posts, highlighting current methodologies, findings, and research gaps that our project aims to address.

II. Sentiment Analysis

Sentiment Analysis involves the use of computational techniques to identify and categorize opinions expressed in textual data. It aims to determine the emotional tone behind a series of words, helping to understand the attitudes, opinions, and emotions expressed within an online mention. In the context of social media, sentiment analysis can uncover insights into public opinion and sentiment trends.

A. Techniques and Approaches

- **Lexicon-Based Approaches:** Utilize predefined lists of words (lexicons) with associated sentiment scores to analyze text.
- **Machine Learning Approaches:** Involve training algorithms on labeled data to predict sentiment. Common algorithms include Naïve Bayes, Support Vector Machines (SVM), and deep learning models like LSTM and BERT.
- **Hybrid Approaches:** Combine lexicon-based and machine learning techniques to improve accuracy and robustness.

III. Literature Review

A. Overview of Selected Studies

Borkar and Kolhe (2019) conducted a comprehensive review of sentiment analysis techniques applied to Twitter data using machine learning. Their paper highlights that Twitter, characterized by its short, informal messages, presents unique challenges for sentiment analysis due to its unstructured and noisy text. The study evaluates various machine learning algorithms, including Naïve Bayes, Support Vector Machines (SVM), and Maximum Entropy, highlighting their effectiveness in classifying sentiments expressed in tweets. Key components such as data collection, preprocessing, feature extraction, and sentiment classification are discussed. The review identifies challenges like handling imbalanced data, sarcasm, and slang, and proposes recommendations to improve sentiment classification performance, including the integration of domain-specific lexicons and context-aware analysis. The study provides a comprehensive review of various machine learning techniques applied to sentiment analysis of Twitter data, addressing the challenges of extracting meaningful insights from the vast amounts of unstructured and noisy textual data inherent in Twitter posts. The primary objective is to evaluate the effectiveness of different machine learning algorithms in classifying sentiments expressed on Twitter. The review identifies that SVM demonstrated higher accuracy in classification tasks due to its ability to handle high-dimensional data and find the optimal separating hyperplane.

Gupta and Reddy (2022) provided an in-depth analysis of current sentiment analysis techniques, focusing on the integration of NLP and machine learning methods. Their survey aims to identify the strengths and weaknesses of various approaches and suggest directions for future research. In the work, they reviewed over 150 research papers published between 2015 and 2021, categorizing sentiment analysis techniques into lexicon-based, machine learning-based, and hybrid approaches. Further, they examined the effectiveness of different NLP techniques like part-of-speech tagging, parsing, and semantic analysis and analyzed machine learning models ranging from traditional algorithms to advanced deep learning frameworks. The researchers identified that hybrid models combining lexicon-based methods with machine learning algorithms offer improved accuracy and robustness.

Lopez, Singh, and Kim (2021) investigated how sentiment analysis can be applied to health-related social media posts to gauge public opinion on health issues, treatments, and policies. The work aimed to understand the challenges of processing

medical terminology and the implications for public health monitoring. In the paper, the researchers gathered a dataset of health-related posts from platforms like Twitter and health forums using keywords such as "vaccination," "mental health," "COVID-19," etc. The researchers found that deep learning models, particularly LSTM networks, outperformed traditional machine learning algorithms in accurately classifying sentiments in health-related posts.

Obulapuram et al. (2023) reviewed sentiment analysis methods on social media platforms, discussing challenges such as colloquialisms, irony, and negation. This review notes that the accuracy performance of lexicon-based and machine learning techniques is similar, and that combining both is shown to improve overall performance.

Wang et al. (2023) explored deep learning-based sentiment analysis for social media, analyzing both text and images to determine sentiment. The study finds that text outperforms images in unimodal sentiment recognition and that multimodal analysis is better overall. Text and image may have no correlation.

Aarts et al. (2020) discussed developing a practical application for sentiment analysis on social media textual data, using an ensemble classifier method to classify a broad range of emotions from text. In the research the researchers found that Ensemble classifier can classify a broad range of emotions but multi-class labeling has drastically lower performance.

Reference	Approach	Algorithm	Dataset	Key Findings
Borkar and Kolhe (2019)	Supervised Learning	Na¨ıve Bayes, SVM, MaxEnt	Twitter API	SVM outperformed NB and MaxEnt; challenges with sarcasm and slang.
Kim, Chen, and Park (2021)	Deep Learning	Bi-LSTM, CNNs	Multi-platform	Bi-LSTM achieved highest accuracy; deep learning models handle variability in language.
Gupta and Reddy (2022)	Hybrid Approaches	Lexicon + ML	Literature Review	Hybrid models improved accuracy; deep learning advancements noted.

TABLE I – Continued from previous page

Reference	Approach	Algorithm	Dataset	Key Findings
Lopez, Singh, and Kim (2021)	Health-Related Analysis	LSTM, SVM, Naïve Bayes	Health Forums, Twitter	LSTM outperformed traditional models; need for real-time and multilingual analysis.
Obulapuram et al. (2023)	Lexicon	N/A	Social Media	Accuracy Performance of Lexicon-Based and Machine Learning techniques are similar.
Wang et al. (2023)	Deep Learning	N/A	Social Media	Text outperforms images in unimodal sentiment recognition
Aarts et al. (2020)	Ensemble Classifier	N/A	Social Media	A drastic drop in performance for multi-class labeling

TABLE I: Literature Review on Sentiment Analysis of Social Media Posts using ML Techniques

IV. SYNTHESIS AND CONCLUSION

A. Overall Trends

- **Advancements in Machine Learning Algorithms:** Continuous improvement and adaptation of machine learning algorithms for sentiment analysis, particularly in handling unstructured and noisy social media data.
- **Deep Learning and Hybrid Models:** Growing adoption of deep learning models and hybrid approaches that combine traditional NLP techniques with advanced algorithms for enhanced performance.
- **Domain-Specific Challenges:** Recognition of the unique challenges posed by different domains, such as health-related content requiring specialized preprocessing and analysis techniques.
- **Multimodal Analysis:** Integration of multiple sources of information, such as text and images, to improve sentiment analysis.

B. Research Gaps

- **Multilingual Sentiment Analysis:** A significant gap exists in analyzing non-English social media data, which is crucial for a comprehensive understanding of global sentiments.
- **Real-Time Processing Capabilities:** Limited research on developing models that can process and analyze data in real-time, which is essential for timely decision-making.
- **Handling Contextual Nuances:** Existing models struggle with interpreting sarcasm, irony, and slang prevalent in social media language.
- **Multimodal Sentiment Analysis:** Need more effective methods for combining information from different modalities (text, images, audio) to enhance sentiment analysis.

C. Project Relevance

Our project intends to address these gaps by:

- **Developing Real-Time, Multilingual Models:** Implementing efficient algorithms capable of processing streaming data across multiple languages to provide immediate sentiment insights.
- **Enhanced Contextual Understanding:** Utilizing advanced deep learning models with attention mechanisms to better interpret contextual nuances like sarcasm and slang.
- **Domain Adaptability:** Designing adaptable frameworks that can be customized for specific domains, such as healthcare, to improve accuracy and relevance.
- **Multimodal Approach:** Implementing methods for integrating and analyzing information from both text and images.

By focusing on these areas, our project aims to contribute meaningful advancements to the field of sentiment analysis of social media posts.

V. REFERENCES

REFERENCES

[1] S. Borkar and K. Kolhe, "A Review on Sentiment Analysis of Twitter Data Using Machine Learning Techniques," *International Journal of Emerging Technologies and Innovative Research*, vol. 6, no. 6, pp. 691-696, June 2019.

[2] B. Kim, L. Chen, and D. Park, "A Comprehensive Study on Sentiment Analysis Techniques for Social Media Data," *IEEE Access*, vol. 9, pp. 47735-47757, 2021.

[3] S. Gupta and M. Reddy, "Sentiment Analysis Using NLP and Machine Learning: A Survey," *ACM Computing Surveys*, vol. 54, no. 4, Article 85, pp. 1-34, 2022.

[4] E. Lopez, K. Singh, and D. Kim, "A Systematic Literature Review on Sentiment Analysis of Health-Related Social Media Posts Using Machine Learning Techniques," *Journal of Biomedical Informatics*, vol. 113, p. 103621, 2021

[5] T.V. Sai Obulapuram, J. A. Sa Reddy Yendreddy, L. S. Kotari, V. T. Tirumalasetty, T. Santhi Se and S. S. Imambi, "A Review: Sentiment Analysis Methods and their use in Social Media Platforms," 2023 *International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*

[6] Zhe Wang Ying Liu, le Fang, and Daxiang Li. 20023 Deep Learning-Ilased Sentiment Analysis for Socia Media. In Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattem Recognition (AIPR 22)

[7] Aarts, Jiang, Chen (2020): A Practical Application for sentiment analysis on social media textual data.