Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 6s (2025)
https://www.jneonatalsurg.com

Sentiment Analysis on Social Media Opinions: A Survey of Machine Learning and Lexicon-Based Approaches

G Radha[1], Dr. K Chandrashekhar[2]

[1]Research Scholar, Department of Computer Application, Aurora University, Hyderabad, Telangana, India
Email ID: aud22sccoa01@aurora.edu.in

[2]Head of the Department of Computer Science and Engineering, Aurora University, Hyderabad, Telangana, India
Email ID: hodcse@aurora.edu.in

Cite this paper as: G Radha, Dr. K Chandrashekhar, (2025) Sentiment Analysis on Social Media Opinions: A Survey of Machine Learning and Lexicon-Based Approaches. *Journal of Neonatal Surgery*, 14 (6s), 24-29

ABSTRACT
Social media platforms, particularly Twitter, have become a rich source of public opinion, influencing various sectors such as politics, business, and social movements. Sentiment analysis plays a crucial role in interpreting and classifying emotions embedded in user-generated content. This paper provides a comprehensive survey of sentiment analysis methodologies, comparing machine learning-based and lexicon-based approaches. The study explores different frameworks, techniques, and challenges associated with sentiment classification, including sarcasm detection, data noise, and domain dependency. The paper reviews key contributions from recent studies, such as Qi et al. (2023), Karim et al., Dong & Lian (2021), Maqsood et al. (2020), Niu et al. (2021), and Nedjah et al. (2022), highlighting advancements and research gaps. Finally, the study identifies future research directions for improving sentiment analysis models by integrating deep learning, multimodal approaches, and real-time sentiment tracking.

Keywords: Sentiment Analysis, Social Media, Machine Learning, Lexicon-Based Analysis, Twitter, Natural Language Processing, Big Data Analytics

1. INTRODUCTION

The rise of social media platforms has revolutionized how individuals express opinions, making sentiment analysis an essential tool for understanding public perception. Platforms like Twitter generate vast amounts of textual data daily, which, when analyzed effectively, can provide valuable insights for businesses, policymakers, and researchers. However, extracting meaningful information from social media posts presents several challenges, including text ambiguity, sarcasm detection, and varying linguistic styles. This paper presents a survey of sentiment analysis techniques applied to social media data, with a focus on machine learning and lexicon-based approaches.

The information revolution has fundamentally reshaped economies by centering them around the flow of data, information, and knowledge. With the advent and subsequent rise of the Internet, the way content is consumed and disseminated has undergone a dramatic transformation. Social platforms have emerged as prime examples of interaction among millions of individuals, where vast amounts of data are generated daily. These platforms facilitate information exchange networks, which thrive on user engagement and interaction, as highlighted by Wang et al. (2010).

Twitter has become a focal point for researchers and organizations due to the sheer volume and variety of data it produces. The platform's ability to generate unstructured data quickly from diverse sources offers significant insights into various trends and dynamic changes within society, markets, and other environments. Businesses and organizations aim to leverage this information to better understand these changes, thereby minimizing risks and maximizing opportunities.

2. LITERATURE REVIEW

Recent studies have explored various sentiment analysis techniques, emphasizing both machine learning (ML) and lexicon-based approaches.

- Qi et al. (2023) compared lexicon-based sentiment analysis with ML-based models using Twitter data, highlighting that ML models generally outperform lexicon-based techniques in handling complex linguistic structures.

- Karim et al. introduced an ensemble model incorporating logistic regression and deep learning to improve sentiment prediction accuracy.
- Cano-Marin et al. conducted a systematic review of Twitter's predictive capabilities, demonstrating its effectiveness in forecasting trends.
- Dong & Lian (2021) reviewed public opinion analysis on social media, discussing challenges such as misinformation and bias in sentiment models.
- Georgiadou et al. (2020) analyzed Brexit-related sentiment, showcasing how big data analytics can be leveraged in international negotiations.
- Hidayatullah et al. (2018) utilized topic modeling for football news sentiment classification, illustrating domain-specific sentiment analysis challenges.
- Hu & Wang (2020) examined the impact of geotagging removal on sentiment analysis, emphasizing location-based sentiment variations.
- Khan et al. (2020) explored stock market predictions using sentiment from social media and news sources, demonstrating real-world applications of sentiment analysis.
- Maqsood et al. (2020) applied deep learning techniques to forecast stock market trends using sentiment analysis on local and global events.
- McKay (2020) analyzed typed laughter expressions in tweets, emphasizing the role of digital communication in sentiment analysis.
- Niu et al. (2021) examined the application of big data analytics in business intelligence for decision-making.
- Nedjah et al. (2022) introduced computational intelligence for sustainable power transformer maintenance, showing the impact of sentiment analysis in engineering applications.
- Moss et al. (2021) discussed new collaboration models in conferences and how sentiment analysis can improve virtual interactions.
- Schultz et al. (2011) analyzed crisis communication through Twitter and traditional media, demonstrating sentiment analysis applications in public relations.

These studies provide a foundation for understanding the evolution of sentiment analysis and its applications in diverse fields.

Understanding Social Networks

According to Streeter and Gillespie (1993), a social network consists of a finite set of connected social units. The analysis of such networks necessitates an understanding of these connections, or the links among individuals and groups. Connectivity, defined as the relationships and interactions between individuals and social units, is a crucial element in the analysis of social networks. This connectivity is essential for determining who is and who is not a member of a network, thereby defining the criteria for membership.

Most of the data within social networks are unstructured, making it challenging to measure and analyze. However, this data can be examined through various characteristic features. Twitter, for instance, produces large volumes of unstructured data rapidly from heterogeneous sources, providing a wealth of information that can be analyzed to glean insights. This unstructured data is characterized by its variety, volume, and velocity, which poses both challenges and opportunities for researchers and organizations.

Leveraging Twitter Data

Researchers and organizations harness Twitter data to gain a better understanding of dynamic changes and trends in various environments, including society and markets. This understanding enables businesses to mitigate risks associated with these changes and capitalize on emerging opportunities. The digital transformation, coupled with advancements in computational intelligence, has paved the way for sophisticated analysis of social networks. Twitter's big data nature, characterized by its variety, volume, and the speed with which it is generated, makes it a valuable resource for such analysis.

The rapid evolution of artificial intelligence (AI)-based technologies, supported by the continuous development of supercomputing power, has opened new frontiers in the analysis of massive datasets. These advancements have enabled the application of AI and machine learning techniques to analyze Twitter data, uncovering patterns and trends that were previously difficult to detect. This has significant implications for various domains, including economics, politics, healthcare, and marketing, among others.

Evolution Of AI And Computational Intelligence

The ongoing development of AI-based technologies and supercomputing power has revolutionized the analysis of massive datasets. AI and machine learning techniques have proven to be particularly effective in analyzing unstructured data, such as that generated by social networks. This has enabled researchers and organizations to uncover patterns and trends within the data, providing valuable insights that can inform decision-making processes.

The continuous advancements in AI and computational intelligence have expanded the applicability of these techniques to various domains. From economics and politics to healthcare and marketing, the ability to analyze large volumes of data quickly and accurately has opened new avenues for research and practical applications. This has led to the development of predictive systems that leverage Twitter data to forecast trends and changes in various environments.

Twitter as a Predictive System

Given these advancements, Twitter is evaluated as a potential predictive system. Systematic literature reviews (SLRs) on Twitter have been conducted since 2012, with Aarts et al. (2012) being the first to our knowledge. They evaluated the state-of-the-art, identifying the characteristics of online social platforms concerning actors, messages, and networks. These reviews have highlighted the potential of Twitter data to serve as a predictive tool, providing valuable insights into various trends and changes.

The application of innovative techniques, such as machine learning and graph analysis, has further enhanced the predictive potential of Twitter data. By conducting massive SLRs and leveraging these advanced techniques, researchers have been able to identify gaps and opportunities for developing predictive applications of user-generated content (UGC) on Twitter. This has significant implications for both scholars and business leaders, paving the way for new research lines and practical applications.

Applications in Healthcare and Education

The potential applications of Twitter data as a predictive system are vast and varied. Samaras et al. (2020) demonstrated how Information and Communication Technologies (ICT) have influenced intelligent healthcare systems. They highlighted the use of Twitter as a real-time data source for raising awareness about epidemics, facilitating knowledge transfer among patients and the medical community. This has underscored the value of Twitter data in improving public health outcomes and informing healthcare strategies.

In the field of education, Guraya (2016) explored the use of Twitter and Facebook to understand their impact on educational practices. The study revealed that these platforms could be powerful tools for enhancing learning experiences and facilitating communication between educators and students. The broader implications of these findings suggest that social media platforms, including Twitter, can play a significant role in shaping educational strategies and outcomes.

3. SENTIMENT ANALYSIS TECHNIQUES

Sentiment analysis approaches are broadly classified into two main categories:

3.1 Lexicon-Based Approaches

Lexicon-based sentiment analysis relies on predefined sentiment dictionaries to classify text.

Examples: AFINN, SentiWordNet

Strengths: Simple and interpretable

Limitations: Struggles with sarcasm, domain dependency

3.2 Machine Learning-Based Approaches

Machine learning models leverage algorithms to classify sentiment based on learned patterns.

Algorithms: SVM, KNN, Logistic Regression, Neural Networks

Strengths: Better handling of complex linguistic structures

Limitations: Requires labeled datasets, computationally expensive

4. CHALLENGES IN SENTIMENT ANALYSIS

Despite advancements, several challenges persist in sentiment analysis:

- **Sarcasm Detection:** Many models struggle to recognize sarcastic statements.
- **Multilingual Sentiment Analysis:** Sentiment varies across languages, requiring robust multilingual models.
- **Context Understanding:** Short social media posts often lack sufficient context, making sentiment classification difficult.
- **Data Noise and Ambiguity:** Social media text includes slang, emojis, and inconsistent grammar, complicating preprocessing.

Comparative Analysis:

Sl.no.	**Author(s)**	**Year**	**Study Focus**	**Techniques Used**	**Findings**	**Challenges Identified**
1	Qi et al.	2023	Twitter sentiment analysis	Lexicon & ML-based	ML models outperform lexicon-based	Need for better sarcasm detection
2	Karim et al.	2022	Sentiment prediction	Ensemble learning, LR & DL	Improved sentiment accuracy	High computational cost
3	Cano-Marin et al.	2021	Twitter as a predictive system	Systematic literature review	Effective in forecasting trends	Data variability
4	Dong & Lian	2021	Public opinion analysis	NLP & ML techniques	Identified misinformation impact	Bias in sentiment models
5	Georgiadou et al.	2020	Brexit sentiment analysis	Big Data Analytics	Used sentiment analysis in negotiations	Context-dependent results
6	Hidayatullah et al.	2018	Football news sentiment	Topic modeling	Classified domain-specific sentiment	Requires domain adaptation
7	Hu & Wang	2020	Geotagging & sentiment	NLP & ML	Geotagging affects sentiment patterns	Privacy concerns
8	Khan et al.	2020	Stock market predictions	ML classifiers & sentiment	Sentiment influences stock trends	Real-time sentiment challenges
9	Maqsood et al.	2020	Stock forecasting	DL-based sentiment analysis	Used local/global event sentiment	High data processing needs

10	McKay	2020	Digital communication sentiment	Twitter laughter patterns	Found new forms of digital sentiment	Informal text processing issues
11	Niu et al.	2021	Business intelligence	Big Data Analytics	Decision-making through sentiment	Requires integration with real-time models
12	Nedjah et al.	2022	Power transformers maintenance	Computational intelligence	Sustainability-based sentiment analysis	Application-specific challenges
13	Moss et al.	2021	Sentiment in virtual collaborations	NLP techniques	Sentiment trends in remote work	Limited generalizability
14	Schultz et al.	2011	Crisis communication	Sentiment via Twitter & blogs	Twitter provides real-time responses	Potential misinformation

5. Future Research Directions

Several areas require further exploration to enhance sentiment analysis:

- Deep Learning Integration: Leveraging transformer models like BERT for improved contextual understanding.
- Multimodal Sentiment Analysis: Combining text, images, and videos for sentiment classification.
- Real-Time Sentiment Tracking: Developing automated systems for continuous sentiment analysis.
- Domain-Specific Sentiment Models: Creating specialized sentiment models for fields such as finance, healthcare, and politics.

6. Conclusion

This survey highlights the key advancements and challenges in sentiment analysis on social media. While machine learning-based models generally outperform lexicon-based methods, issues such as sarcasm detection, multilingual processing, and data ambiguity remain unresolved. Future research should focus on developing hybrid models that integrate deep learning, multimodal analysis, and real-time processing to enhance sentiment analysis accuracy.

This comprehensive analysis underscores Twitter's predictive potential through the intelligent analysis of user-generated content. The systematic literature review reveals gaps and opportunities for developing predictive applications, particularly in social network analysis and public health. These findings pave the way for new research lines with significant implications for both scholars and business leaders. By leveraging the vast amounts of data generated on Twitter, researchers and organizations can gain valuable insights into various trends and changes, informing decision-making processes and enhancing outcomes across multiple domains. 3. Sentiment Analysis Techniques

REFERENCES

[1] Qi et al., "Sentiment Analysis Using Twitter Data: A Comparative Application of Lexicon- and Machine-Learning-Based Approaches," Social Network Analysis and Mining, 2023.

[2] Karim et al., "Anticipating Impression Using Textual Sentiment Based on Ensemble LRD Model," Elsevier, 2022.

[3] Cano-Marin et al., "Twitter as a Predictive System: A Systematic Literature Review," Elsevier, 2021.

[4] Dong & Lian, "A Review of Social Media-Based Public Opinion Analyses: Challenges and Recommendations," Technology in Society, 2021.

[5] Georgiadou et al., "Big Data Analytics and International Negotiations: Sentiment Analysis of Brexit Negotiating Outcomes," International Journal of Information Management, 2020.

[6] Hidayatullah et al., "Twitter Topic Modeling on Football News," IEEE Conference on Computer and Communication Systems, 2018.

[7] Hu & Wang, "Understanding the Removal of Precise Geotagging in Tweets," Nature Human Behavior, 2020.

[8] Khan et al., "Stock Market Prediction Using Machine Learning Classifiers and Social Media, News," Journal of Ambient Intelligence and Humanized Computing, 2020.

[9] Maqsood et al., "A Local and Global Event Sentiment Based Efficient Stock Exchange Forecasting Using Deep Learning," International Journal of Information Management, 2020.

[10] McKay, "Some Distributional Patterns in the Use of Typed Laughter-Derived Expressions on Twitter," Journal of Pragmatics, 2020.

[11] Niu et al., "Organizational Business Intelligence and Decision Making Using Big Data Analytics," Information Processing & Management, 2021.

[12] Nedjah et al., "Sustainable Maintenance of Power Transformers Using Computational Intelligence," Sustainable Technology and Entrepreneurship, 2022.

[13] D. Shanthi, R. K. Mohanty and G. Narsimha, "Application of machine learning reliability data sets", Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS), pp. 1472-1474, 2018.

[14] D Shanthi, "Smart Water Bottle with Smart Technology", Handbook of Artificial Intelligence, Benthem Science Publishers, Pg. no: 204-219, 2023

[15] Moss et al., "Forging a Path to a Better Normal for Conferences and Collaboration," Nature Astronomy, 2021.

[16] Schultz et al., "Is the Medium the Message? Perceptions of and Reactions to Crisis Communication via Twitter, Blogs, and Traditional Media," Public Relations Review, 2011.

[17] SCImago Research Group, "Description of SCImago Journal Rank Indicator," Retrieved February 24, 2022, from https://www.scimagojr.com/journalrank.php.

[18] D Shanthi , Smart Healthcare for Pregnant Women in Rural Areas, Medical Imaging and Health Informatics, Wiley Publishers,ch-17, pg.no:317-334, 2022

[19] "The Blue brain project Hil sean:Markram Henry", *International conference of IEEE*, 2008.