

A Review on Sentiment Analysis with Machine Learning in Social Media

Mukut Prasad¹, Vaibhav Patel², Anurag Shrivastava³

¹MTech Scholar, Department of Computer Science, NIIST, Bhopal
mukutprasad1992@gmail.com

² Associate professor, Department of Computer Science, NIIST, Bhopal

³ HOD, Department of Computer Science, NIIST, Bhopal

Abstract: Sentiment analysis, a crucial component of natural language processing (NLP), holds significant importance in understanding public opinion, market trends, and social dynamics on various social media platforms. This review comprehensively examines the landscape of sentiment analysis with machine learning techniques in social media, aiming to provide insights into the current state, methodologies, challenges, and future directions in this domain. Through an extensive analysis of existing literature, this review highlights the diverse approaches and methodologies employed in sentiment analysis, ranging from traditional machine learning algorithms to state-of-the-art deep learning architectures. It explores the effectiveness of different feature extraction techniques, sentiment lexicons, and model architectures in capturing the nuanced sentiment expressed in social media texts.

Furthermore, the review discusses the challenges inherent in sentiment analysis, including the handling of sarcasm, irony, and context-dependent sentiments, as well as the impact of data biases and domain adaptation. Ethical considerations regarding privacy, user consent, and algorithmic fairness are also addressed. By synthesizing current research findings and identifying areas for future exploration, this review aims to provide researchers, practitioners, and policymakers with a comprehensive understanding of sentiment analysis in social media and inspire advancements towards more accurate, robust, and ethically sound sentiment analysis systems.

Keywords— *Sentiment Analysis, Social media, Natural language Processing, Deep Learning*

1. INTRODUCTION

In recent years, social media platforms have become ubiquitous, serving as primary sources of information, communication, and expression for billions of users worldwide. Within this vast digital landscape, sentiment analysis has emerged as a pivotal tool for understanding and analyzing the opinions, attitudes, and emotions expressed in user-generated content. Sentiment analysis, also known as opinion mining, encompasses the computational techniques used to extract, quantify, and interpret sentiment from text data [5].

This review focuses on the application of machine learning techniques in sentiment analysis within the context of social media platforms. With the exponential growth of user-generated content on platforms such as Twitter, Facebook, and Instagram, there is a pressing need for effective sentiment analysis methods capable of processing large volumes of data in real-time. Machine learning algorithms, ranging from traditional classifiers to sophisticated deep learning models, have been extensively employed to address the complexities of sentiment analysis in social media. These algorithms leverage various features extracted from textual data, including word embeddings, syntactic structures, and semantic representations, to infer sentiment polarity and intensity. Throughout this review, we aim to provide a comprehensive overview of the methodologies, advancements, challenges, and applications of sentiment analysis with machine learning in social media [6]. By synthesizing insights from existing literature, we seek to elucidate the strengths and limitations of current approaches, identify emerging trends, and highlight opportunities for future research and innovation in this rapidly evolving field [7].



Figure 1. Sentiment Analysis

The pervasive influence of social media on public discourse, consumer behavior, and societal trends underscores the importance of sentiment analysis as a means of extracting actionable insights from the vast reservoir of user-generated content. Sentiment analysis not only enables businesses to gauge customer satisfaction, sentiment towards products or services, and emerging trends but also empowers policymakers and researchers to monitor public opinion, track sentiment shifts, and identify potential crises or opportunities [8]. Despite its widespread adoption and utility, sentiment analysis in social media presents numerous challenges. The inherently noisy, informal, and context-dependent nature of social media text poses difficulties for traditional natural language processing techniques. Moreover, the prevalence of sarcasm, irony, slang, and emoticons further complicates sentiment interpretation, necessitating sophisticated computational approaches capable of capturing nuanced linguistic cues and contextual nuances [9].

Furthermore, ethical considerations surrounding user privacy, data consent, and algorithmic fairness are paramount in the development and deployment of sentiment analysis systems in social media. As such, this review endeavors to explore not only the technical aspects of sentiment analysis with machine learning but also the ethical implications and societal ramifications of its application in the realm of social media. Through a holistic examination, this review aims to contribute to the advancement of sentiment analysis research and practice in social media while promoting ethical and responsible deployment of these technologies [10].

2. LITRETURE REVIEW

Many Researchers work on sentiment analysis with machine learning in social media explores a broad range of methodologies, from traditional algorithms to cutting-edge deep learning architectures. It delves into the challenges of handling context-dependent sentiments and data biases while emphasizing the ethical considerations essential for responsible technology deployment in this domain.

Author's [1] proposes a comprehensive review of the multifaceted reality of sentiment analysis in social networks. In this review authors explore new aspects such as temporal dynamics, causal relationships, and applications in industry. Authors also applied, and discuss the practical applicability of emerging Artificial Intelligence methods. Authors emphasizes the importance of temporal characterization and causal effects in sentiment analysis in social networks, and explores their applications in different contexts such as stock market value, politics, and cyber bullying in educational centers.

Authors [2] gaining a comprehensive understanding of the application of sentiment analysis to user-generated data, the work identifies the challenges and issues in the existing sentiment analysis research. Using the PRISMA framework, the author's report the objectives of sentiment analysis tasks, the general implementation process, the algorithms adopted and how they are used in different domains. Afterward, by comparing aspects of different studies, the paper presents several challenges and issues related to datasets, languages of the review text, analysis methods and evaluation metrics in the existing work.

In [3] 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 [4] 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.

3. FINDINGS OF THE SURVEY

While extensive research has been conducted on sentiment analysis with machine learning in social media, several notable gaps persist, warranting further investigation. One such gap lies in the development of robust models capable of effectively handling context-dependent sentiments, sarcasm, irony, and colloquial expressions prevalent in social media text. Current methodologies often struggle to accurately interpret nuanced linguistic cues, leading to suboptimal performance and limited applicability in real-world scenarios.

Moreover, there is a need for comprehensive studies addressing the ethical implications of sentiment analysis in social media, particularly concerning user privacy, data consent, and algorithmic fairness. Ethical considerations are paramount in ensuring responsible technology deployment and mitigating potential societal harms arising from biased or intrusive sentiment analysis systems. Furthermore, while existing research has explored various feature representation techniques and model architectures, there remains room for innovation in the design of context-aware models capable of adapting to diverse social media platforms and user demographics. Addressing these gaps is crucial for advancing the field of sentiment analysis with machine learning in social media, enabling the development of more accurate, robust, and ethically sound systems with broader applicability and societal impact.

4. 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]. Margarita et al. "A review on sentiment analysis from social media platforms" Science Direct: [www.elsevier.com/locate/eswa https://doi.org/10.1016/j.eswa.2023.119862](https://doi.org/10.1016/j.eswa.2023.119862)
- [2]. Qianwen Ariel Xu "A systematic review of social media-based sentiment analysis: Emerging trends and challenges" <https://doi.org/10.1016/j.dajour.2022.100073>, April 2022; science direct
- [3]. Pansy Nandwani1 · Rupali Verma1 "A review on sentiment analysis and emotion detection from text" <https://doi.org/10.1007/s13278-021-00776-6>
- [4]. 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
- [5]. Samaneh Moghaddam., Martin Ester.: On the Design of LDA Models for Aspect-based Opinion Mining. In: ACM, 2012.
- [6]. Samaneh Moghaddam and Martin Ester : The FLDA Model for Aspect based Opinion Mining: Addressing the Cold Start Problem. ACM conference, pp. 909-918, 2011

- [7]. 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.
- [8]. 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.
- [9]. Haiping Zhang, Zhengang Yu, Ming Xu, Yueling Shi : Feature-level Sentiment Analysis for Chinese Product Reviews. IEEE, pp. 135-139, 2011
- [10]. 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
- [11]. 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.
- [12]. Wei Jin., Hung Hay Ho., Rohini K. Srihari : OpinionMiner: A Novel Machine Learning System for Web Opinion Mining and Extraction. In: ACM, 2009.
- [13]. 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.
- [14]. 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.
- [15]. 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.
- [16]. 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.
- [17]. Xiuzhen Zhang., Yun Zhou.: Holistic Approaches to Identifying the Sentiment of Blogs Using Opinion Words. In: Springer-Verlag Berlin Heidelberg, 5–28, 2011.
- [18]. 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).
- [19]. Ye Jin Kwon., Young Bom Park.: A Study on Automatic Analysis of Social Network Services Using Opinion Mining. In: Springer-Verlag Berlin Heidelberg, 240–248, 2011.
- [20]. Anuj Sharma., Shubhamoy Dey: An Artificial Neural Network Based approach for Sentiment Analysis of Opinionated Text. In: ACM, 2012.
- [21]. Yulan He. : A Bayesian Modeling Approach to Multi-Dimensional Sentiment Distributions Prediction. In: ACM, Aug. 2012.
- [22]. 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.
- [23]. Andrius Mudinas., Dell Zhang., Mark Levene. : Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis. In: ACM, Aug. 2012.
- [24]. 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
- [25]. 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
- [26]. Id.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)