Plagiarism - Report

Originality Assessment

16%

Overall Similarity

Date: May 24, 2024
Matches: 387 / 2439 words
Sources: 12

Remarks: Low similarity detected, check with your supervisor if changes are required.

Verify Report:

WeCare: A User-focused Sentiment Analysis Framework for Social Media Interactions

Mohammad Airad Khana, Mohd. Irshadb, Pankaj Mauryac, Piyush Kumard, Ms. Garima Gakhare

a, b, c, d Department 1 of Information Technology, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, India

e Professor of Inderprastha Engineering College, Ghaziabad, Uttar Pradesh,

IndiaaAIRADKHAN102000@gmail.com, cPankajmaurya4041@gmail.com, bmohdirshadmalik007@gmail.com, drajputpiyush234@gmail.com

Abstract — This paper addresses the sentiment analysis of WhatsApp chats, focusing on the challenges and opportunities presented by conversational data. The introduction outlines the prevalence of sentiment analysis in social media platforms and the significance of accurate sentiment categorization. The literature review highlights previous work in sentiment analysis, emphasizing methods for categorizing sentiment polarity and feature selection. The proposed system introduces a comprehensive approach to sentiment analysis tailored to WhatsApp chats, considering the unique characteristics of conversational data. Key features include user-centric inclusion, nuanced extraction of sentiment, context-aware analysis, and effective handling of conversational nuances. The method section details the proposed system's emphasis on real-time sentiment tracking, user-specific sentiment scores, and integration with WhatsApp's API for broader applications. The results section summarizes the positive outcomes of the proposed method, including user-centric inclusion, nuanced extraction, context-aware illustration, and effective content handling. The future scope section discusses potential enhancements such as real-time sentiment alerts, multilingual support, detailed sentiment analysis, improved user feedback systems, personalized sentiment metrics, and collaboration capabilities. The conclusion highlights the significance of the "WeCare" sentiment analysis project in improving our understanding of digital conversations, particularly within the context of WhatsApp chats. The project's commitment to user-centric design, adaptability, and ethical considerations positions it as a dynamic tool for enhancing digital interactions and well-being. Finally, acknowledgments recognize the contributions of the development team and academic mentor.

Keywords — Nuanced Extraction of Sentimen, Context-Aware Analysis, Conversational Nuances, Chats Analysis, Sentiment Analysis.

I. Introduction

In today's digital era, social media platforms have become integral channels for individuals to express their thoughts, share experiences, and engage in conversations. From forums and micro-blogs to online communities, social networking sites offer a diverse range of avenues for communication and interaction [1]. With the proliferation of these platforms, researchers and developers have gained access to valuable data through application programming interfaces (APIs), enabling the collection and analysis of user-generated content [2].

For instance, WhatsApp, one of the leading social media platforms, provides various APIs, including the REST API, Search API, and Streaming API, facilitating data retrieval and analysis for researchers and developers. These APIs empower users to gather real-time information, search specific content, and create customized projects, laying a robust foundation for sentiment analysis and data-driven insights. However, despite the vast potential of online data for sentiment analysis, several challenges persist, hindering the accurate interpretation of sentiments. One significant challenge stems from the open nature of social media platforms, where users can freely post content without stringent quality control measures. This lack of control introduces the risk of encountering spam or irrelevant content, which can distort sentiment analysis results [6]. Moreover, distinguishing genuine opinions from fake sentiments or spam presents a formidable task for researchers and analysts, highlighting the need for robust methodologies and techniques to filter and analyze data effectively [3]. Another challenge lies in the absence of a ground truth for sentiment analysis, where the veracity of sentiment labels— whether positive, negative, or neutral—is not always ascertainable. While 2 sentiment analysis algorithms rely on labeled datasets for training and validation, the availability of accurate ground truth labels remains inconsistent across online platforms. This limitation poses a significant hurdle for sentiment analysis models, impacting their ability to deliver reliable insights and predictions [6]. In light of these challenges, this 1 paper focuses on addressing the complexities of sentiment analysis within the context of WhatsApp chats—a ubiquitous mode of communication for millions of users worldwide. WhatsApp, with its conversational format and intimate user interactions, presents a unique set of 2 challenges and opportunities for sentiment analysis. By developing a user-focused 1 sentiment analysis framework tailored specifically for WhatsApp interactions, this

research aims to enhance our understanding of digital conversations and improve sentiment analysis accuracy [4].

Through a comprehensive exploration 2 of sentiment analysis methodologies, feature selection techniques, and real-time sentiment tracking, this paper proposes the "WeCare" framework—a dynamic and adaptable solution for analyzing sentiments in WhatsApp chats. By integrating user-centric design principles, contextual awareness, 1 and ethical considerations, "WeCare" endeavors to provide users with actionable insights and enhance their digital well-being. In the following sections, we delve into the 2 challenges and opportunities of sentiment analysis, review relevant literature, present the proposed framework, discuss its methodology, highlight the results, explore future scope, and conclude with reflections on 1 the significance of the "WeCare" project in advancing sentiment analysis research.

II. Literature Survey

Sentiment analysis, a crucial aspect of opinion investigation, revolves around the classification of sentiment extremity within textual data. This classification aims to assign each 10 piece of text into distinct sentiment categories, typically positive, negative, or neutral. However, sentiment analysis encompasses multiple levels of granularity, including document-level, sentence-level, and even content and perspective-level categorization.

Fang 1 et al. (2015) [5] present a thorough investigation into sentiment analysis using product review data sourced from Amazon.com, addressing fundamental challenges in sentiment polarity categorization at both sentence and review levels. Through meticulous exploration, they outline a robust process 2 for sentiment analysis, emphasizing the extraction of subjective content and the identification of sentiment tokens. Their methodology involves extracting sentiment sentences, performing part-of-speech tagging, and identifying negation phrases to enhance sentiment analysis accuracy. Furthermore, they propose 2 algorithms for sentiment score computation and feature vector generation, laying the groundwork for effective sentiment categorization. The study's experiments, conducted 3 on a large dataset of product reviews, yield promising outcomes, demonstrating the efficacy of their approach in both sentence-level and review-level categorization. Additionally, they acknowledge limitations in implicit 1 sentiment analysis and provide insights into future research

directions, underscoring the importance of further exploration to enhance sentiment analysis methodologies [5]. Sentiment categorization presents a fundamental classification challenge, requiring the identification of features containing sentiment or opinion data prior to classification. Throb and Lee [5] proposed a methodology for feature selection in sentiment analysis, advocating for the extraction of subjective sentences while excluding objective ones. Their text-categorization approach, leveraging techniques such as minimum cuts, focuses on identifying subjective content, thus aiding sentiment analysis tasks [7].

Fig 1. Public Opinion on Cyberbullying: A Visual Representation of Diverse Perspective.

Building upon this foundation, Gann et al. [8] delved into sentiment tokenization based on Twitter data, selecting 6,799 tokens and assigning each a sentiment score, termed as TSI (Total Sentiment Index). This approach facilitates the identification of tokens expressing either positive or negative sentiment, contributing to a deeper understanding of sentiment within textual data. While these studies offer valuable insights into sentiment analysis methodologies and feature selection techniques, contemporary challenges in digital communication extend beyond mere sentiment classification. With the proliferation of online platforms, instances of online bullying and cyberbullying have surged, posing serious threats to individuals' well-being and safety [9]. Recent research indicates a concerning rise in online bullying, attributed to various factors such as anonymity, accessibility, and the asynchronous nature of online communication [9]. The shift towards digital platforms as primary modes of interaction has amplified opportunities for individuals to engage in harmful behaviors, including threatening or harassing others [9]. The pervasive nature of social media and messaging apps like WhatsApp has facilitated the spread of misinformation and fake news, further exacerbating online bullying incidents. The anonymity afforded by online platforms often emboldens perpetrators, leading to increased instances of threatening behavior towards others, particularly vulnerable populations such as children and adolescents [10].

In light of these contemporary challenges, sentiment analysis frameworks tailored to online communication platforms like WhatsApp assume heightened significance. By incorporating nuanced approaches to sentiment analysis and context-aware analysis, such frameworks not only enable the

detection of sentiment extremity but also offer insights into the emotional dynamics of digital conversations, thereby contributing to the mitigation 4 of online bullying and the promotion of digital well-being.

Fig 2. Dual Bar Graph illustrating the prevalence of Online and Physical Bullying Across Age Groups (14-19).

Fig 3. Frequency of Bullying Incidents Across Grade Levels: 1 A Comparative Analysis.

III. Methodology

To effectively conduct 2 sentiment analysis on WhatsApp chats, the "WeCare" framework employs a multi-faceted methodology encompassing data collection, preprocessing, sentiment analysis algorithms, and integration with WhatsApp's API for real-time analysis and broader applications.

☐ 11 Data Collection: The first step involves gathering WhatsApp chat data from users. This process is facilitated by a user-friendly interface and an export tool integrated into the "WeCare" framework. Users can conveniently export their WhatsApp chats while ensuring compliance with privacy regulations. 1 The collected data is then subjected to preprocessing to handle informal language, emojis, and other nuances typical of conversational data.

☐ Preprocessing: The preprocessing phase is crucial for cleaning and preparing the data for sentiment analysis. It involves text normalization, tokenization, and handling special characters and emojis. Techniques such as word stemming or lemmatization may also be applied to standardize the text further. This ensures that the subsequent 2 sentiment analysis algorithms operate on clean and consistent data, improving accuracy.

☐ Sentiment Analysis Algorithms: The core of the methodology lies in the sentiment analysis algorithms employed by the "WeCare" framework. These algorithms are designed to capture nuanced sentiments expressed in WhatsApp chats. The framework utilizes techniques such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to understand the sequential and contextual nature of chat conversations. By analyzing user-specific data, including message length and response time, the framework provides granular insights into individual participants' sentiments.

Integration with WhatsApp API: To enable real-time sentiment tracking and broader applications, the "WeCare" framework integrates with WhatsApp's API. This integration allows for seamless data retrieval and analysis directly from the WhatsApp platform. Users can receive alerts for sentiment fluctuations during conversations, empowering them to respond promptly. Furthermore, the integration facilitates cross-platform analysis, extending the framework's utility beyond WhatsApp to other messaging platforms.

 Ethical Considerations: Throughout the methodology, ethical considerations such as privacy, security, and compliance with legal regulations are paramount. The framework incorporates data encryption and anonymization techniques to protect user privacy. Additionally, adherence to ethical guidelines ensures that 2 the sentiment analysis process is conducted responsibly, without infringing on users' rights or compromising their confidentiality.

 Continuous Improvement: An essential aspect of the methodology involves continuous improvement based on user feedback and technological advancements. The "WeCare" framework incorporates a feedback loop to gather insights on 1 the accuracy of sentiment predictions. This feedback informs iterative refinements to the sentiment analysis algorithms, enhancing their performance over time.

IV. Results

The outcomes of implementing the "WeCare" sentiment analysis framework for WhatsApp interactions.

Fig 4. "WeCare" Dashboard User Interface.

Our results showcase the efficacy of the framework in providing nuanced insights into user sentiments, enhancing digital well-being, and facilitating real-time analysis. Additionally, we provide visual representations through screenshots to illustrate key features and functionalities of the framework.

Fig 5. Sentiment Fluctuations in Group Chat Over Time.

Fig 6. Activity Maps of Group Chat Over Time.

Fig 7. Active user and Most Uses sentimental words in Group Chat.

Fig 8. Words and Emoji frequency i.e. have Directly Affects Sentiment Analysis.

V. Discussion

The "WeCare" sentiment analysis framework is recognized for its user-centric design, which prioritizes users' needs and promotes digital well-being by offering tailored insights into WhatsApp chats. The framework's ability to capture nuanced sentiments allows for a deeper understanding of emotional dynamics, while context-aware analysis and real-time insights empower users to navigate conversations effectively. Ethical considerations, including privacy protection, are paramount, urging ongoing research into privacy-preserving techniques. Future directions include multilingual support, enhanced user feedback mechanisms, personalized metrics, expansion to other platforms, and collaboration with mental health experts. Ultimately, "WeCare" represents a significant advancement in sentiment analysis, poised to positively impact digital interactions and well-being.

VI. Conclusion

The "WeCare" sentiment analysis framework presents a comprehensive and innovative solution for understanding and enhancing digital interactions within WhatsApp chats. By addressing the challenges of sentiment analysis through user-centric design, nuanced sentiment extraction, and context-aware analysis, the framework empowers users with actionable insights tailored to their

specific conversations. [1] Through the integration of real-time analysis capabilities and ethical considerations such as privacy protection, "WeCare" not only facilitates meaningful interactions but also promotes digital well-being in an increasingly connected world. As digital communication continues to evolve, the framework's adaptability and potential for future enhancements position it as a dynamic tool for researchers, developers, and end-users alike. By fostering collaboration and ongoing refinement, "WeCare" stands poised to make a significant impact on sentiment analysis research and contribute to the creation of more positive and constructive online environments.

VII. Acknowledgement

VIII. References

[1] Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. [5] A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168.

[2] Gordon, W. (2024, April 2). Understanding Customer Behavior and Preferences. Nutshell. Retrieved from https://www.nutshell.com/blog/understanding-customer-behavior-and-preferences

[3] Muhammed T, S., & Mathew, S.K. (2022). The disaster of misinformation: [1] a review of research in social media. [3] International Journal of Data Science and Analysis, 13, 271–285. https://doi.org/10.1007/s41060-022-00311-6

[4] Wankhade, M., Rao, A.C.S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

[5] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 5. https://doi.org/10.1186/s40537-015-0015-2

[6] Wankhade, M., Rao, A.C.S., & Kulkarni, C. [1] (2022). A survey on sentiment analysis methods,

applications, and challenges. Artificial Intelligence Review, 55(10), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

[7] Walaa Medhat, Ahmed Hassan, Hoda Korashy. 8 (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093-1113. doi: https://doi.org/10.1016/j.asej.2014.04.011.

[8] Chawla, P., Hazarika, S., Shen, H-W. (2020). 9 Token-wise sentiment decomposition for ConvNet: Visualizing a sentiment classifier. *Visual Informatics, 4*(2), 132-141. doi: https://doi.org/10.1016/j.visinf.2020.04.006.

[9] Nahar, V., Unankard, S., Li, X., Pang, C. 4 (2012). Sentiment Analysis for Effective Detection of Cyber Bullying. In: Sheng, Q.Z., Wang, G., Jensen, C.S., Xu, G. (eds) Web Technologies and Applications. APWeb 2012. *Lecture Notes in Computer Science*, vol 7235. Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-642-29253-8_75.

[10] Hook, K., & Verdeja, E. (2022). Social Media Misinformation and the Prevention of Political Instability and Mass Atrocities. *Human Security & Governance*, *Protecting Civilians & Human Security Program*, *Atrocities Prevention Study Group Project*. Retrieved from [https://docs.wixstatic.com/ugd/e64469_ea7b4527e03e42b88c46fc35eeaf35e4.pdf].

Sources

1
https://link.springer.com/article/10.1007/s10462-023-10651-9
INTERNET
6%

2
https://dl.acm.org/doi/10.1007/s10462-022-10144-1
INTERNET
3%

3
https://link.springer.com/article/10.1007/s13735-023-00308-2
INTERNET
2%

4
https://link.springer.com/chapter/10.1007/978-3-642-29253-8_75
INTERNET
2%

5
https://jyx.jyu.fi/handle/123456789/76144
INTERNET
1%

6
https://rcciit.org/students_projects/projects/cse/2018/GR20.pdf
INTERNET
<1%

7
https://www.who.int/europe/news/item/27-03-2024-one-in-six-school-aged-children-experiences-cyberbullying--finds-new-who-europe-study
INTERNET
<1%

8
https://sci-hub.se/10.1016/j.asej.2014.04.011
INTERNET
<1%

9
https://www.researchgate.net/publication/340433170_Token-wise_sentiment_decomposition_for_ConvNet_Visualizing_a_sentiment_classifier
INTERNET
<1%

10
https://towardsdatascience.com/a-guide-to-text-classification-and-sentiment-analysis-2ab021796317
INTERNET
<1%

11
https://ijrti.org/papers/IJRTI2404089.pdf
INTERNET
<1%

12
https://medium.com/@francis.everett/cyberbullying-the-silent-scourge-of-the-digital-world-4ff255cf991c
INTERNET
<1%

EXCLUDE CUSTOM MATCHES   OFF

EXCLUDE QUOTES   ON

EXCLUDE BIBLIOGRAPHY   ON