JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review Paper of Integration of Unstructured Data of Hate Speech on Facebook Using Sentiment and Emotion

Prof. M.R.Rajput 1) Bhavesh Chandrakant Borole 2) Vaishnavi Pradip Patil
3) Rushikesh Sunil Shrikhande 4) Divya Ninaji Choudhari.
Computer Science Engineering Department, Padm. Dr. VBKCOE, Malkapur, Maharashtra, India.

ABSTRACT *- Hate speech is any expression that targets an individual or group of people because of their colour, national origin, religion, sexual orientation, or other characteristics. Though there are several ways to convey it, including online*

Both online and off, social media's growing popularity has greatly expanded its use and severity. Thus, finding and analyzing the unstructured data of particular social media posts with the intention of spreading hate in the comment sections is the goal of this study. We suggest FADOHS, a unique framework that combines data analysis and natural language processing techniques, as a solution to this problem. Its goal is to make all social media providers aware of how common hate speech is on the platform.

In particular, we examine recent postings and comments on these pages using sentiment and emotion analysis algorithms. Posts that may contain terms that dehumanise people will be examined before being sent to the clustering algorithm for additional analysis. The experimental results show that the suggested FADOHS framework can achieve roughly 10% better precision, recall, and F1 scores than the state-of-the-art method.

Keywords *- Emotion recognition, clustering algorithm, sentiment analysis, data mining.*

INTRODUCTION *- Mark Zuckerberg, the Chief Executive Officer of Facebook, has unequivocally asserted that hate speech and racism find no place on the social media giant's platform [1]. Despite Facebook's implementation of diverse artificial intelligence (AI) techniques to combat hate speech, persistent challenges have surfaced. Notably, the company acknowledges the limitations of its AI technology, revealing in published statistics on hate speech crackdowns that a considerable portion still requires human review, with only 38% being accurately flagged by their technology in a specific quarter [2]. The ongoing dilemma revolves around the intricate question: What constitutes hate speech? This query has sparked continuous debate, leading to diverse definitions such as the expression that spreads, incites, promotes, or justifies hatred, discrimination, or hostility toward a specific group [3], or a direct attack on individuals based on protected characteristics [4].*

Facebook concedes that the inadequacy of AI stems from its inability to discern between hate promotion and the mere description of an experience [5]. Scholars like Sara Chinnasamy and Norain Abdul Manaf note that hate speech can manifest subtly, such as through discussions on controversial topics aimed at eliciting hateful comments [6]. Anat Ben-David and Ariadna Matamoros-Fernandez argue that, despite Facebook's efforts, hate comments remain pervasive, often expressed through subliminal hatred in vicious messages or commentaries [6].

In response to these challenges, this study makes notable contributions. Firstly, it introduces a semi-automatic method for identifying pages discussing sensitive topics. Secondly, an automatic method is proposed for clustering posts from pages focusing on specific topics. Finally, a new framework for hate speech detection is designed and implemented. Experimental results indicate that this framework surpasses state-of-the-art approaches by approximately 10% in terms of precision, recall, and F1 scores. The subsequent sections of the paper are organized as follows: Section II reviews the literature on hate speech detection, Section III delves into the system architecture, Section IV discusses the experiments and their outcomes, Section V presents the results, and Section VI concludes the study while providing a potential framework for future work.

PROBLEM FORMULATION *-*

The problem formulation at the core of this study revolves around the persistent challenges faced by social media platforms, particularly Facebook, in effectively identifying and mitigating hate speech. Despite the implementation of various artificial intelligence (AI) techniques, the platform acknowledges ongoing difficulties, as evidenced by a substantial

portion of hate speech removal requiring human review. The central dilemma lies in defining and accurately identifying hate speech, a task complicated by the diverse interpretations provided by scholars and platforms. This ambiguity is underscored by Facebook's admission that its AI technology is not yet sophisticated enough to distinguish between hate promotion and the mere expression of personal experiences.

Subtle manifestations of hate speech, such as discussions on controversial topics designed to elicit hateful comments, further complicate the issue. This study addresses these challenges by introducing a semi- automatic method for discovering pages discussing sensitive topics, an automatic method for clustering posts from pages focusing on specific topics, and a novel hate speech detection framework. The contributions of this research extend beyond addressing the limitations of existing approaches, with experimental results demonstrating a noteworthy improvement of approximately 10% in precision, recall, and F1 scores compared to state-of-the-art methods. By delving into these complexities, the study aims to enhance our understanding of hate speech detection challenges and proposes innovative solutions to mitigate this pervasive issue on social media platforms.

PROPOSE SYSTEM METHODOLOGY -

The proposed system methodology involves a multi- faceted approach to enhance online payment fraud detection using behavior-based techniques. Firstly, data enhancement strategies will be implemented to address the limitations associated with low-quality user behavioral data. This may include the use of synthetic data generation techniques and feature engineering to augment the dataset.

Secondly, a model augmentation process will be employed to construct behavioral models from diverse perspectives and integrate them effectively. Individual- level models and population-level models, categorized based on the granularity of agents, will be explored and integrated for a comprehensive fraud detection system.

Furthermore, the system will leverage advanced machine learning algorithms and anomaly detection techniques to identify patterns indicative of fraudulent behavior. Continuous validation of transactions will be ensured, transforming fraud detection from a one-time event to a continuous monitoring process.

To maintain user experience, a non-intrusive detection system will be implemented, minimizing the need for user intervention during installation. The system will also consider the evolving nature of cyber threats by incorporating adaptive mechanisms to detect new patterns of fraudulent behavior.

In summary, the proposed system methodology encompasses data enhancement, model augmentation, diverse model perspectives, continuous validation, non-intrusive detection, and adaptability to effectively combat online payment fraud while preserving user experience.

WORKING ON LANGUAGES -

It appears you've provided information about the operating system, coding language, development tool, and database for a software development project. Here's a brief summary and elaboration on each component:

Windows 10: The chosen operating system for your development environment. It provides a user-friendly interface and supports various software applications, making it a popular choice for developers.
Coding Language:

Java: The programming language selected for your project. Java is known for its platform independence, making it suitable for cross-platform applications. It is widely used for web development, mobile applications, and enterprise-level systems.
Development Tool:

Netbeans 8.2: The Integrated Development Environment (IDE) chosen for coding in Java. Netbeans provides a comprehensive set of tools for Java development, including code editors, debugging features, and project management capabilities.
Database Management System (DBMS):

MySQL: The selected database management system for storing and managing your project's data. MySQL is a popular open-source relational database that integrates well with Java applications.
This combination of Windows 10 as the operating system, Java as the programming language, Netbeans
8.2 as the development tool, and MySQL as the database management system forms a coherent and widely used technology stack for software development. It allows for efficient coding, testing, and database management throughout the development life cycle. If you have specific questions or need further assistance related to this technology stack, feel free to ask!

RELATED WORKING -

Certainly, the rapid evolution of online payment services has given rise to a continuous stream of fraudulent activities in online transactions. Addressing this challenge, the use of behavioral models for fraud detection has become a focal point of extensive research, capturing the attention of numerous researchers.

As online transactions proliferate, so do the tactics employed by fraudsters, necessitating innovative approaches for detection. Behavioral models offer a promising avenue, leveraging patterns in user behavior

to identify anomalies indicative of fraudulent activity. The dynamic nature of online fraud requires continuous exploration and refinement of these models, prompting researchers to delve into various methodologies to enhance the effectiveness of fraud detection systems.

This vibrant and evolving field reflects the commitment of researchers to stay ahead of emerging threats in the online payment landscape, aiming to provide secure and reliable services for users. The exploration of behavioral models signifies a proactive response to the challenges posed by the dynamic and sophisticated nature of online transaction fraud.

The proposed system methodology aims to address the pervasive issue of hate speech on Facebook by integrating unstructured data using sentiment and emotion analysis. The system is developed using Java programming language within the Netbeans 8.2 environment and employs MySQL as the database for effective data management.

The working of the system involves several key components. Firstly, data collection is performed, encompassing textual information from Facebook posts and comments. This data undergoes preprocessing, including normalization and feature extraction using sentiment and emotion analysis techniques. The Java programming language facilitates efficient implementation of these functionalities, with Netbeans 8.2 serving as the development platform.

FRONT-END AND BACKEND TECHNOLOGY -

Operating System: Windows 10.
Coding Language: Java.
Development Tool: Netbeans 8.2.
Database: MySQL.

This configuration aligns well for Java development on a Windows environment. Windows 10 serves as the operating system, providing a user-friendly interface and compatibility with a variety of software. Java is chosen as the programming language, known for its portability and versatility. Netbeans 8.2 is selected as the integrated development environment (IDE), offering a comprehensive set of tools for Java development, including code editing, debugging, and project management. MySQL is employed as the database management system, a popular choice for its reliability and open-source nature.

This technology stack is suitable for developing a wide range of applications, from desktop to web applications, and it allows seamless integration between the Java code and the MySQL database. If you have any specific questions or need assistance with this technology stack, feel free to ask!

REFERENCES –

[1] Zuckerberg Refugee Crisis: Hate Speech Has, Place Facebook, Street Guardian, Honolulu, HI, USA, 2010.

[2] Fortune. (2018). Facebook Removed 2.5 Million Pieces Hate Speech 1st Quarter. Accessed: Jul. 16, 2018. [Online]. Available:https://fortune.com/2018/05/15/facebook- hate-speech-removals/.

[3] ILGA. (2018). Hate Crime & Hate Speech. Accessed: May 6, 2018. [Online]. Available: https://www.ilga-europe.org/what-we-do/our advocacy-work/hate-crime-hate-speech

[4] Facebook. (2020). Community Standards Home. Accessed: May 11, 2018. [Online]. Available: https://www.facebook.com/communitystandards/.

[5] CNBC. (2020). Facebook's Artificial Intelligence Still Has Trouble Find ing Hate Speech—But it Finds a Lot of Nudity. Accessed: May 11, 2018. [Online]. Available: https://www.cnbc.com/2018/05/15/facebook artificial-intelligence-still-finds-it-hard-to-identify-hate- speech.html

[6] S. Chinnasamy and N. A. Manaf, ''Social media as political hatred mode in Ts 2018 general election,'' in SHS Web Conf., vol. 53, 2018, p. 2005.

[7] A. Matamoros-Fernández and J. Farkas, ''Racism, hate speech, and social media: A systematic review and critique,'' Telev. New Media, vol. 22, no. 2, pp. 205– 224, Feb. 2021.

[8] F. Del Vigna, A. Cimino, F. Dell-TOrletta, M. Petrocchi, and M. Tesconi,''Hate me, hate me not: Hate speech detection on Facebook,'' in Proc. 1st Italian Conf. Cybersecur. (ITASEC), Venice, Italy, 2017, pp. 86–95.

[9] M. Ahmed, R. Seraj, and S. M. S. Islam, ''The K-means algorithm: A comprehensive survey and performance evaluation,'' Electronics, vol. 9, no. 8, p. 1295, Aug. 2020.

[10] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, ''Student engage□ ment level in an e-Learning environment: Clustering using K-means,'' Amer. J. Distance Educ., vol. 34, no. 2, pp. 137–156, Apr. 2020.

[11] Z. Lv, T. Liu, J. A. Benediktsson, and H. Du, ''Novel land cover change detection method based on K-means clustering and adaptive majority voting using bitemporal remote sensing images,'' IEEE Access, vol. 7, pp. 34425–34437, 2019.

[12] D. Kucukusta, M. Perelygina, and W. S. Lam, ''CSR communication strate□gies and stakeholder engagement of upscale hotels in social media,'' Int. J. Contemp. Hospitality Manage., vol. 31, no. 5, pp. 2129–2148, May 2019.

[13] A. Rodriguez, C. Argueta, and Y.-L. Chen, ''Automatic detection of hate speech on Facebook using sentiment and emotion analysis,'' in Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC), Feb. 2019.

[14] G. C. Santia and J. R. Williams, ''BuzzFace: A news veracity dataset with Facebook user commentary and egos,'' in Proc. 12th Int. AAAI Conf. Web Social Media. Palo Alto, CA, USA, 2018, pp. 531–540.

[15] A. Chopra, A. Dimri, and S. Rawat, ''Comparative analysis of statistical classifiers for predicting news popularity on social web,'' in Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), Jan. 2019, pp. 1–8.

[16] B. Lin, F. Zampetti, G. Bavota, M. Di Penta, M. Lanza, and R. Oliveto, ''Sentiment analysis for software engineering: How far can we go?'' in Proc. IEEE/ACM 40th Int. Conf. Softw. Eng. (ICSE). Gothenburg, Sweden: IEEE, 2018, pp. 94–104.

[17] V. Franzoni, Y. Li, and P. Mengoni, ''A path-based model for emotion abstraction on Facebook using sentiment analysis and taxonomy knowl□edge,'' in Proc. Int. Conf. Web Intell., Aug. 2017, pp. 947–952.

[18] S. Al Mansoori, A. Almansoori, M. Alshamsi, S. A. Salloum, and K. Shaalan, ''Suspicious activity detection of Twitter and Facebook using sentimental analysis,'' TEM J., vol. 9, no. 4, p. 1313, 2020.

[19] S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, ''Discrepancy detection between actual user reviews and numeric ratings of Google app store using deep learning,'' Expert Syst. Appl., vol. 181, p. 115111, 2021.

[20] M. Shad Akhtar, D. Ghosal, A. Ekbal, P. Bhattacharyya, and S. Kurohashi, ''A multi-task ensemble framework for emotion, sentiment and intensity prediction,'' 2018, arXiv:1808.01216.

[21] A. Hussain, A. Tahir, Z. Hussain, Z. Sheikh, M. Gogate, K. Dashtipour, A. Ali, and A. Sheikh, ''Artificial Intelligence–Enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the united kingdom and the united states: Observational study,'' J. Med. Internet Res.,vol. 23, no. 4, Apr. 2021, Art. no. e26627.

[22] U. Bhaumik and D. K. Yadav, ''Sentiment analysis using Twitter,'' in Computational Intelligence and Machine Learning. Singapore: Springer, 2021, pp. 59– 66.