STM JOURNALS

International Journal of
Algorithms Design and
Analysis Review

ISSN: 2584-1866
Volume 2, Issue 2, 2024
July—December
DOI (Journal): 10.37591/IJADAR

https://journals.stmjournals.com/ijadar

Review IJADAR

Machine Learning Innovations for Effective Spam Comment Filtering in Social Networks

Sujata Patil[1,*], Pradudny Phase[2], Anand Walukar[3], Pushpak Pandey[4]

Abstract

The increasing prevalence of social media platforms has revolutionized communication, fostering unparalleled levels of connectivity and data exchange. However, the widespread increase in spam comments presents a serious threat to the integrity of online discussions, potentially undermining the quality of interactions. To confront this issue, our proposed model utilizes machine learning techniques to bolster spam comment detection across various social media platforms. This endeavor involves a thorough investigation encompassing data aggregation and collection, feature engineering, model selection, and meticulous performance evaluation. The ultimate goal is to develop a resilient and precise system capable of effectively discerning genuine user comments from spam content. By leveraging advanced methodologies and algorithms using machine learning, our model aims to mitigate the disruptive impact of spam, preserving the authenticity of online conversations, and sustaining meaningful interactions. Through this comprehensive approach, we aspire to enhance user experience, foster trust in social media platforms, and contribute to the cultivation of a healthier digital environment by classifying real comments and spam comments. Our proposed system not only addresses the technical aspects of spam detection but also emphasizes the broader implications for user engagement and platform integrity as a strong system. By advancing spam detection capabilities with the help of advanced algorithms, we aim to protect the quality of social media interactions and ensure a more trustworthy and enjoyable online community. Ultimately, our work seeks to provide a scalable and effective solution that can be adapted to various social media environments, promoting a safer and more positive user experience.

Keywords: Machine learning, online conversation, real comments, spam comments, strong system

INTRODUCTION

Social media platforms have become an integral part of modern communication, providing avenues for individuals, businesses, and organizations to connect, share information, and engage with diverse communities. However, the rise of social media has also led to the proliferation of spam comments, posing significant challenges to the user experience, content moderation, and platform integrity. Communication media is an essential tool for society and a considerable vector for fraudulent content, such as fake rewards, identity fraud, extortion, phishing, and malware transmission. Comments not only clutter discussions but also undermine trust and authenticity within online communities. Addressing this issue requires robust detection and prevention mechanisms that are capable of swiftly identifying and mitigating spam content. Traditional rule-based approaches to spam

***Author for Correspondence**
Sujata Patil
E-mail: sujata.patil_skncoe@sinhgad.edu

[1]Assistant Professor, Department of Electronics and Telecommunication Engineering, Smt Kashibai Navale College of Engineering, SPPU, Pune, Maharashtra, India
[2-4]Student, Department of Electronics and Telecommunication Engineering, Smt Kashibai Navale College of Engineering, SPPU, Pune, Maharashtra, India

Received Date: June 28, 2024
Accepted Date: August 23, 2024
Published Date: November 06, 2024

Citation: Sujata Patil, Pradudny Phase, Anand Walukar, Pushpak Pandey. Machine Learning Innovations for Effective Spam Comment Filtering in Social Networks. International Journal of Algorithms Design and Analysis Review. 2024; 2(2): 19–24p.

detection often fail to adapt to the evolving nature of spam tactics. As spammers continuously devise new techniques to bypass detection filters, there is an increasing need for more sophisticated and adaptive solutions. Machine learning (ML) offers a promising avenue for enhancing spam comment detection and prevention on social media platforms. By leveraging the power of algorithms to analyze vast amounts of data, ML models can discern patterns and characteristics indicative of spam, thereby enabling a more accurate and efficient detection.

This study explored the application of machine learning techniques for enhancing spam comment detection and prevention on social media platforms. It delves into the challenges posed by spam comments, the limitations of existing detection methods, and the potential of machine learning to effectively address these challenges. By examining recent advancements in ML algorithms, feature engineering, and model training strategies, this study seeks to identify the best practices and novel approaches for combating spam on social media.

Through this comprehensive approach, we aspire to enhance user experience, foster trust on social media platforms, and contribute to the cultivation of a healthier digital environment. Our research not only addresses the technical aspects of spam detection but also emphasizes the broader implications for user engagement and platform integrity. By advancing spam detection capabilities, we aim to protect the quality of social media interactions and ensure a more trustworthy and enjoyable online community. Ultimately, our work seeks to provide a scalable and effective solution that can be adapted to various social media environments, thereby promoting a safer and more positive user experience.

Moreover, this study investigates the role of contextual information, user behavior analysis, and real-time monitoring in improving the efficacy of spam detection systems. By incorporating contextual cues, such as user interactions, content relevance, and temporal dynamics, ML models can better distinguish between genuine contributions and spammy content. Additionally, this study explores the ethical considerations and trade-offs involved in deploying automated moderation systems, emphasizing the importance of transparency, fairness, and user privacy. In summary, this research contributes to the ongoing discourse on combating spam on social media platforms by highlighting the potential of machine-learning approaches. By enhancing our understanding of spam detection mechanisms and leveraging the capabilities of ML algorithms, this study aimed to facilitate the development of more resilient and adaptive systems to safeguard the integrity of online conversations.

LITERATURE SURVEY

Some studies on spam content detection have been conducted. Spam detection is mainly performed in text messages [1], such as Short Message Services (SMS) [2, 3], which employ the UCI SMS dataset with the CNN method using auxiliary hand-engineered features [4]. Spam SMS was also detected using RNN-LSTM and LSTM only, which were also compared with machine-learning methods [5]. In addition to messages, there is considerable spam content on social media platforms. Spam content can be found on social media such as IG, FB, and TW [6].

The purpose of the work of Niranjani et al. [7] was to compare the results between the classification model used in the findings by Chaudhry et al. [8] and an artificial neural network (ANN). They used the same dataset and labeled comments on five popular videos. After data preprocessing, they applied ANN and five measures (i.e., accuracy rate, spam caught rate, blocked ham rate, F1 measure, and MCC). According to the results from these measures, this research using ANN presented a higher accuracy for the F1 measure and MCC than with a similar blocked ham rate [9].

Some researchers have proposed detecting spam comments by analyzing their content. Rdulescu et al. [10] detected spam comments using post-comment theme coherence. They calculated the word co-occurrence frequency between microblogs and comments and then obtained post-comment similarity. They believed that low similarity meant that comments might be spam. However, from our observations, many normal comments are short and inconsistent with the theme [11].

Their results have much room for improvement because they ignore other text-based features. Romero et al. [12] conducted a comparative study using four classifiers (naive Bayes, k-nearest neighbors, neural networks, and support vector machines) for spam comment detection [13]. Support Vector Machines achieved the highest performance (84.6%). By extracting several useful features, these machine learning methods obtain decent results. Kharwal's [14] research focused on detecting spam comments using machine-learning techniques. The research was published on The Clever Programmer website, providing insights into combating spam on online platforms.

METHODOLOGY

Block Diagram of Enhancing Spam Comment Detection and Prevention on Social Media Using a Machine Learning Method Figure 1 illustrates a cohesive system that operates cyclically. As users post comments, the filter component processes the text, and the spam detection component evaluates the comments for spam content. Genuine comments are directed to users' inboxes, whereas spam comments are managed separately. The user feedback loop continually refines system accuracy and adapts. The block diagram illustrates a cohesive system that operates cyclically and is designed to enhance spam comment detection and mitigation on social media platforms. When users post comments, the filter component first processes the text to ensure that it meets the basic standards for content. This initial filtering helps to remove comments that are blatantly inappropriate or irrelevant before they reach the more sophisticated spam detection phase.

The spam detection component then evaluates the comments using advanced machine-learning algorithms and advanced algorithm techniques. This evaluation considers various features of the comments, such as word patterns, the frequency of certain terms, and contextual relevance, to determine the likelihood of a comment being spam. Genuine comments are identified and directed to users' inboxes, ensuring that authentic interactions are promptly preserved and delivered. Spam comments, however, were managed separately. They are flagged and isolated to prevent them from disrupting the user experience. These comments may be stored in a temporary quarantine where further analysis can be conducted or they may be automatically deleted, depending on the severity and confidence level of the spam detection system.

A critical component of this system is the user-feedback loop. Users can provide feedback on the accuracy of spam detection and flagging comments incorrectly identified as spam or missed by the filter. This feedback is instrumental in continually refining the system's accuracy. By incorporating user feedback, the system adapts to evolving spamming techniques, learns from new patterns, and improves its detection capabilities over time.

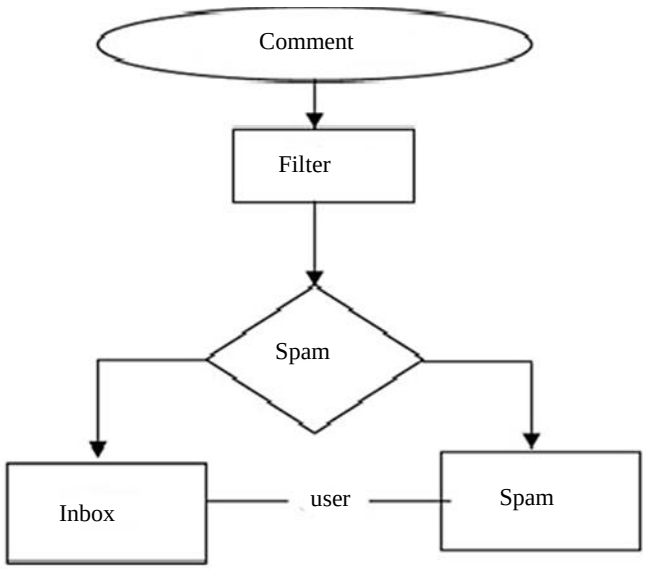

Figure 1. Block diagram of spam comment detection.

This integrated approach aimed to create a safer and more engaging social media experience by significantly reducing the influence of spam content. The system helps build trust and reliability on social media platforms by fostering authentic interactions and maintaining the integrity of online discourse. Ultimately, this cyclical and adaptive spam detection system contributes to the cultivation of a healthier digital environment, enhances user experience, and supports meaningful communication.

RESULT

The implementation of machine learning for advanced spam comment detection and mitigation on social media platforms involves several crucial steps. First, a comprehensive dataset comprising a diverse range of comments spanning both legitimate and spam content must be collected. Each comment should be accurately labeled to facilitate supervised learning. Next, feature extraction techniques were employed to obtain meaningful information from comments. These features may encompass textual content, user metadata such as account age and activity level, linguistic patterns, and sentiment analysis.

Subsequently, a machine learning model is trained using the labeled dataset by employing algorithms such as support vector machines (SVM) and random forests, or deep learning approaches such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs). During training, the model learns to discern patterns and characteristics indicative of spam comments, enabling it to make predictions based on new, unseen data. Once trained, the model is integrated into the social media platform infrastructure to automatically detect and flag potential spam comments in real time.

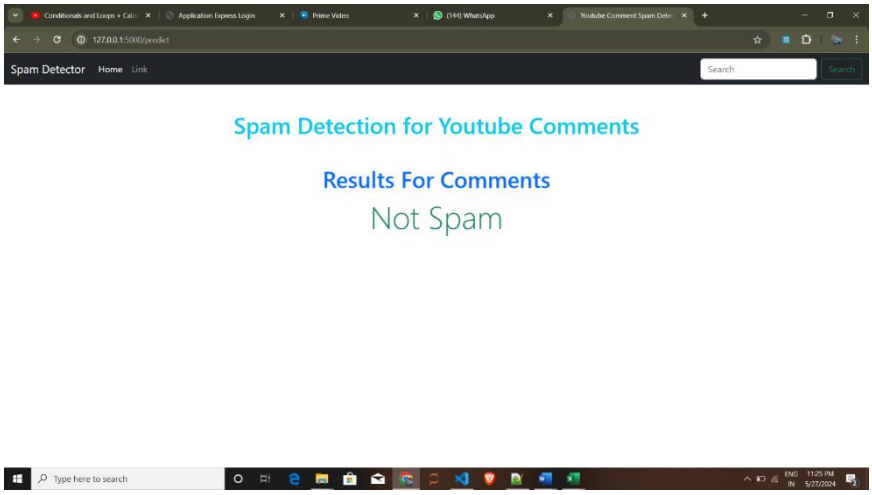

Figure 2. Sample output for not spam.

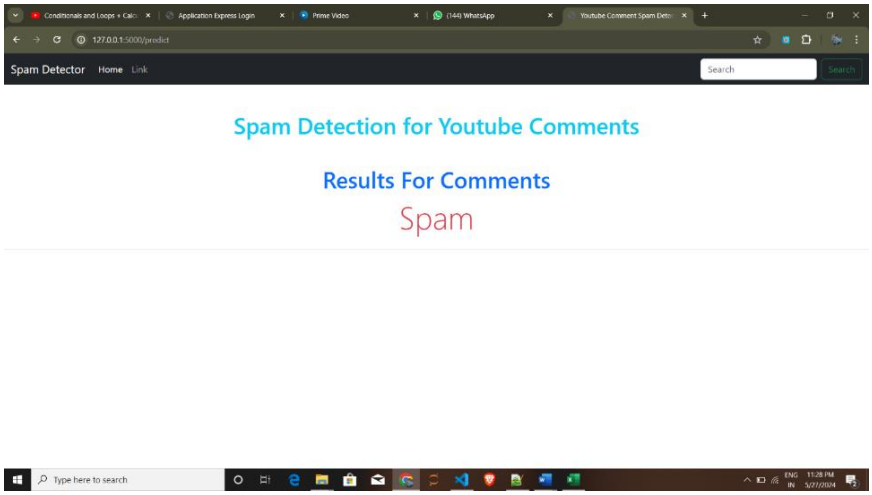

Figure 3. Sample output for spam.

Additionally, continuous monitoring and periodic retraining of the model are essential for adapting to evolving spam tactics and maintaining effectiveness. This ensures that the model remains current with the latest spam techniques and improves its detection capabilities over time. Finally, in conjunction with automated detection, human moderators may review flagged comments to further refine the accuracy of the model and ensure that appropriate mitigation measures are in place.

Through this iterative process of data collection, feature extraction, model training, and ongoing refinement, machine learning can significantly enhance spam detection and mitigation efforts on social media platforms. This comprehensive approach fosters a safer and more engaging online environment for users, preserving the quality of interactions and upholding the integrity of digital communication spaces. By leveraging the power of machine learning, social media platforms can better protect their communities from the disruptive impact of spam, thereby ensuring a more enjoyable and authentic user experience, as shown in Figure 2,3.

CONCLUSION

In this era of digital connectivity, the fight against spam comments on social media platforms has gained significant importance. The constant influx of spam not only disrupts user experiences but also undermines the credibility of online platforms. The proposed model aims to enhance spam comment detection, unveiling a pathway for cleaner and more meaningful online interactions. By integrating advanced machine-learning techniques and adaptive algorithms, the system ensures that social media environments remain conducive to genuine discourse. This approach not only elevates the quality of communication but also reinforces user trust and platform integrity, ultimately contributing to a more positive and productive digital ecosystem.

REFERENCES

1. Tang B, Yang C, Xiang L, Zeng J. Deriving real-time city crowd flows by heterogeneous big urban data. 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3485–94. DOI: 10.1109/BigData.2018.8622370.
2. Aparna VS, Rajan A, Jairaj I, Nandita B, Madhusoodanan P, Remya AA. Implementation of AES algorithm on text and image using MATLAB. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019. pp. 1279–83. DOI: 10.1109/ICOEI.2019.8862703.
3. Rao S, Verma AK, Bhatia T. A review on social spam detection: Challenges, open issues, and future directions. Expert Syst Appl. 2021;186:115742. DOI: 10.1016/j.eswa.2021.115742.
4. Metsis V, Androutsopoulos I, Paliouras G. Spam filtering with Naive Bayes- which Naive Bayes? The Third Conference on Email and Anti-Spam (CEAS 2006), July 27–28, 2006, Mountain View, California, USA. 2006.
5. Alsaleh M, Alarifi A, Al-Quayed F, Al-Salman A. Combating comment spam with machine learning approaches. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2015. pp. 295–300. DOI: 10.1109/ICMLA.2015.192.
6. Garg P, Singh SN. Analysis of ensemble learning models for identifying spam over social networks using recursive feature elimination. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021. p. 713–8. DOI: 10.1109/Confluence51648.2021.9377161.
7. Niranjani V, Agalya Y, Charunandhini K, Gayathri K, Gayathri R. Spam detection for social media networks using machine learning. 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022. pp. 2082–8. DOI: 10.1109/ICACCS54159.2022.9785149.
8. Chaudhry S, Dhawan S, Tanwar R. Spam detection in social network using machine learning approach. In: Batra U, Roy N, Panda B, editors. Data Science and Analytics. REDSET 2019. Communications in Computer and Information Science. Vol 1230. Singapore: Springer; 2020. p. 236–45. DOI: 10.1007/978-981-15-5830-6_20.

9. Sharmin S, Zaman Z. Spam detection in social media employing machine learning tool for text mining. 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, India, 2017. pp. 137–42. DOI: 10.1109/SITIS.2017.32.
10. Rdulescu R, O'Neil B, Klein J. Spam detection in microblogs using theme coherence. J Artif Intell Res. 2014;50:365–95.
11. Shahariar GM, Biswas S, Omar F, Shah FM, Binte Hassan SB. Spam review detection using deep learning. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2019. pp. 27–33. DOI: 10.1109/IEMCON.2019.8936148.
12. Romero DM, Meeder B, Kleinberg J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In: Proceedings of the 20th International Conference on World Wide Web. 2011 Mar 28; p. 695–704. DOI: 10.1145/1963405.1963503.
13. Antony A, Rajendran A, Deepa G. YouTube spam comment detection. In: International Conference on Signal & Data Processing. Springer Nature Singapore; 2022. p. 387–94.
14. Kharwal A. Spam comments detection with machine learning. Thecleverprogrammer [online]. 2022 Aug 2. Available from: https://thecleverprogrammer.com/2022/08/02/spam-comments-detection-with-machine-learning/.

