IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

EMOJI-AWARE AND CONTEXTUAL SPAM FILTERING ON SOCIAL MEDIA USING ADVANCED ENSEMBLE MACHINE LEARNING

[1] Dr. P. Veeresh, [2] S Nikhath Fathima Bee,[3] R Krupavathi,[4] M Ragasandeepthi,[5] B Srinithya

[1]M. Tech., Ph. D, Professor,[2345] B.Tech Students

Department of Computer Science & Engineering

St. Johns College Of Engineering & Technology, Yerrakota, Yemmiganur, Kurnool

ABSTRACT

The proliferation of spam comments on social media poses a growing challenge to user engagement, content authenticity, and platform trustworthiness. Traditional spam detection models primarily rely on textual features, often overlooking the expressive and contextual elements embedded in social media interactions. This study proposes an emoji-aware and context-driven spam filtering framework that integrates emotional cues from emojis and contextual information from post-comment pairs to enhance spam detection capabilities.

By analyzing not just the comment text, but also the semantics of the corresponding post and the emojis used, the model captures nuanced patterns commonly exploited by spammers. To effectively model these diverse features, we utilize advanced ensemble machine learning techniques, including Random Forest, Gradient Boosting, and a hybrid Voting Classifier, which collectively improve accuracy and reduce false positives.

Experimental results on curated social media datasets demonstrate the superiority of the proposed method over conventional approaches, showing significant gains in precision, recall, and F1-score. The findings underscore the value of incorporating non-textual and contextual indicators in spam detection and pave the way for more intelligent, adaptive, and user-aware content moderation systems.

I. INTRODUCTION

1.1 Over View

Spam comments on social media disrupt user interactions and spread misinformation. Traditional spam detection models mainly focus on text but often ignore emojis and the context between posts and comments, making them less effective. Emojis play a key role in communication, conveying emotions and intent, but are frequently overlooked in spam detection. Additionally, analyzing comments in isolation without their corresponding posts makes it difficult to determine relevance.

To address these issues, this research introduces a new spam detection approach that incorporates emoji-based features and post-comment pair analysis. By considering both textual and non-textual elements, this method enhances spam classification accuracy. The study uses the SpamID-Pair dataset for Indonesian social media and demonstrates that combining emojis and contextual analysis significantly improves spam detection performance.

Social media enables people to share their ideas and aspirations, collaborate, conduct business, promote products, and participate in politics. Popular social media platforms include Face book (FB) for more formal or semi-formal text and image media, YouTube (YT) for semi-formal videos, Tik-Tok (TT) for non-formal videos, Instagram (IG) for semiformal and non-formal text, images, and videos, and Twitter (TW) for semi-formal and non-formal text and images [1]. These social media have large user bases, are fully and well functioning, and are used by celebrities to increase their popularity.

Public figures who have large numbers of followers on social media include celebrities.

ISSN 2347–3657
Volume 13, Issue 2, 2025

Many celebrities utilize social media for promoting their activities, increasing their popularity, interacting with their followers, and other purposes. The more famous the celebrities are, the greater number of followers they have. With more followers, celebrities can interact with their fans more frequently [2]. As is characteristic of Web 2.0, users can now comment creatively on celebrities' feeds.

TW, YT, and IG are frequently used in spam detection research because these social media contain a lot of spam accounts and spam content. Particularly in Indonesia, spam content is usually found in comments against Indonesian artists, especially on IG [2]. Figure 1 depicts an example of a post and spam comments on social media in Indonesia of the@ abutting account. Spam comments are very annoying and can disrupt the flow of information in the comments on a given post/status.

Another problem is the limited publicly available datasets for identifying spam text on social media. Most datasets on social media are found in English, and obtaining datasets in other languages, including Indonesian, is challenging. Many researchers conducted similar studies using their own collected datasets without sharing them.

SpamID-Pair1 is a dataset provided for spam content detection in the Indonesian language available in Medley Data Repository. Spam ID-Pair provides posts from Indonesian artists and their comments as pairs labeled spam/not spam. This dataset includes many emojis, which are widely used on social media. Users on social media frequently utilize emojis to describe their emotions and intentions. However, in various research in the Natural Language Processing (NLP) field, most emoji features are discarded/not used [3]. Studies of spam content detection have been previously conducted. However, detecting spam content, particularly spam comments, is difficult due to multiple causes, for example: 1) the very unstructured and abnormal form of comment text; 2) the number of symbols and emoticons used by users; 3) the number of typos, intentional abbreviations, non-standard words, and mixed language usage; 4) some content is intentionally camouflaged to avoid being detected as spam, such as using the \/ sign instead of the letter V which becomes unreadable by the system; 5) the comments are spam but contain very subtle ads; and 6) the system fails to recognize the semantic meaning or semantic relationship between posts and comments. These issues are complex, require investigations, and necessitate many mutually supporting solution modules. 1SPAMID-PAIR on Medley Data Repository (https:// data medley com/datasets/fj5pbdf95t).

1.2 Motivation

Spam detection on social media has become an increasingly pressing challenge as digital interactions continue to expand. With the rise of social networking platforms, businesses, influencers, and everyday users rely heavily on these spaces for communication, promotion, and engagement. However, the prevalence of spam comments—ranging from promotional content, phishing attempts, misinformation, and harmful or irrelevant messages—has led to disruptions in user experiences and online discourse. Traditional spam filtering methods have often focused on keyword-based and rule-based techniques, but spammers have adapted by using more sophisticated methods, including the use of emojis and seemingly natural linguistic patterns to evade detection.

Given this evolving threat, it is imperative to develop more advanced spam detection mechanisms that account for these new trends. Emojis, for instance, are extensively used in online communication, often altering the sentiment and meaning of a message. Spammers exploit this feature by embedding emojis in misleading ways, making it harder for traditional models to differentiate between spam and

ISSN 2347–3657

Volume 13, Issue 2, 2025

legitimate content. Similarly, many spam comments are highly context-dependent, meaning that their spam-like nature only becomes apparent when analyzed in conjunction with the post they are responding to. Hence, a robust approach is needed to improve the accuracy of spam detection by incorporating these elements—emoji features and post-comment relationships—into a machine learning framework.

Ensemble methods in machine learning provide an effective way to enhance classification accuracy by combining multiple models to make more reliable predictions. These methods help in reducing bias, improving generalization, and addressing the limitations of individual classifiers. By leveraging ensemble learning, the project aims to create a more resilient and accurate spam detection system that takes into account both linguistic and contextual features of social media comments.

1.3 Problem Definition

Spam comments on social media platforms present significant challenges for users, platform administrators, and content creators. The primary problem lies in distinguishing spam from legitimate comments effectively, given the evolving tactics employed by spammers. Traditional detection methods primarily rely on text-based analysis, often failing when faced with spam that incorporates emojis, abbreviations, or misleading phrases that appear natural. Additionally, current methods frequently treat comments as standalone texts, ignoring the context in which they appear.

This project seeks to address these shortcomings by enhancing spam detection through a two-fold approach: leveraging emoji features to capture additional semantic information and considering post-comment pairs to provide contextual understanding. The hypothesis is that by integrating these two aspects into the detection framework and employing ensemble machine learning models, it will be possible to achieve a more accurate classification of spam versus non-spam comments.

The problem is further compounded by the ever-changing nature of spam tactics. Spammers continuously adapt to detection mechanisms, making it necessary to develop models that can generalize well across various types of spam and remain effective against emerging spam strategies. Therefore, the project aims to build a system that is robust, adaptable, and capable of improving spam detection performance beyond traditional text-based approaches.

Spam comments on social media negatively impact user experience, disrupt meaningful discussions, and spread misleading information. Traditional spam detection methods rely mainly on text analysis but often fail to detect spam that includes emojis or depends on post-comment context.

Ignoring Emojis – Many spam detection models discard emojis, even though they play a crucial role in conveying intent and meaning in social media conversations. This oversight reduces detection accuracy.

Lack of Context Awareness – Most existing approaches analyze comments in isolation without considering the original post. As a result, comments that may appear legitimate on their own could be spam when evaluated in context.

Higher False Positives and Negatives – Without emoji analysis and post-comment relationships, traditional models may incorrectly classify genuine comments as spam or fail to detect actual spam.

To overcome these issues, this study proposes an enhanced spam detection approach that integrates emoji-based features and post-comment pair analysis, leading to improved classification accuracy.

1.4 Objective Of The Project

The primary objective of this project is to enhance the detection of spam comments on social media by incorporating emoji-based features and post-comment contextual analysis within an ensemble machine learning framework. The project aims to improve the accuracy of spam classification by addressing the limitations of traditional text-based spam detection methods.

One of the key objectives is to extract and analyze the role of emojis in spam detection. Emojis serve as an important communication tool, adding nuance and sentiment to messages. However, they can also be used to mask the true intent of a spam comment, making detection more difficult. This project seeks to develop feature extraction techniques that effectively encode emoji usage, ensuring that their impact on spam classification is properly considered.

Another major objective is to establish a method for analyzing post-comment relationships. Many spam comments appear benign when viewed in isolation but reveal their spam-like nature when examined in the context of the post they respond to. The project will develop a model that can process and learn from both the original post and the associated comments, improving the detection of contextually misleading spam.

Additionally, the project aims to leverage ensemble learning techniques to enhance classification performance. By combining multiple machine learning models, the goal is to create a detection system that minimizes false positives and false negatives, ensuring higher reliability and effectiveness in identifying spam. The effectiveness of different ensemble strategies—such as bagging, boosting, and stacking—will be explored to determine the most efficient approach for this task.

The primary objective of this project is to enhance spam comment detection on social media by integrating emoji-based features and post-comment relationships to improve accuracy and reduce false detections.

To analyze the role of emojis in spam comments and incorporate them as key features in spam detection models. To develop a post-comment pair approach that evaluates the relationship between a comment and its associated post for better spam classification. To improve spam detection accuracy by combining textual, emoji, and contextual features. To reduce false positives and false negatives in spam detection models, ensuring that genuine comments are not wrongly classified as spam. To test and validate the proposed approach using the SpamID-Pair dataset, focusing on Indonesian social media spam detection.

1.5 Limitations Of The Project

Despite the advancements and improvements targeted by this project, certain limitations must be acknowledged. One limitation is the dependency on dataset quality and availability. Spam detection models require large, diverse, and well-labeled datasets to perform effectively. However, obtaining high-quality labeled data for spam detection can be challenging due to the dynamic nature of social media and the evolving tactics of spammers.

Another limitation is the computational complexity associated with ensemble methods. While ensemble learning techniques improve model performance, they also increase computational overhead. Training multiple models and combining their predictions requires more processing power and memory, which may not be feasible for real-time spam detection on large-scale social media platforms with millions of active users.

The project also faces challenges in emoji interpretation. Emojis often carry different meanings depending on cultural context, platform rendering, and user intent. While the model will attempt to incorporate emoji-based features, there may still be cases where emojis are used in ways that are difficult to classify accurately.

Additionally, contextual spam detection using post-comment pairs introduces complexity in terms of data representation and model training.

ISSN 2347–3657

Volume 13, Issue 2, 2025

Establishing meaningful relationships between posts and comments requires advanced natural language processing (NLP) techniques, which can be computationally expensive and may require additional fine-tuning to avoid misclassifications.

Finally, the project may struggle with generalization across different social media platforms. The nature of spam and user behavior varies across platforms, meaning that a model trained on one platform (such as Twitter or Instagram) may not generalize well to another (such as Facebook or YouTube). Ensuring cross-platform effectiveness remains a challenge that may require platform-specific fine-tuning.

1.6 Organization Of The Report

This report is structured to provide a comprehensive understanding of the research problem, methodology, and implementation approach. It begins with an introduction to the problem of spam detection on social media, outlining the challenges posed by traditional methods and the need for an improved approach that incorporates emoji features and post-comment context.

The subsequent sections delve into the related work and literature review, discussing previous research on spam detection methods, the role of emojis in NLP tasks, and the effectiveness of ensemble learning techniques in classification problems. This review helps establish the foundation for the proposed approach and highlights the gaps in existing research that this project aims to address.

The results and discussion section presents the findings of the project, analyzing the performance of different ensemble methods and assessing the impact of incorporating emoji and post-comment features. It includes performance metrics such as accuracy, precision, recall, and F1-score, comparing the proposed approach with baseline models to demonstrate its effectiveness.

Finally, the report concludes with a summary of key findings, limitations, and potential future work. It discusses how the proposed spam detection approach can be further improved, including possible enhancements in dataset collection, model optimization, and real-world deployment. The conclusion reiterates the importance of this research in enhancing online security and improving user experiences on social media platforms.

II. LITERATURE SURVEY

Some research on spam content detection has been conducted previously. Spam detection was mainly done in text messages , such as in the Short Message Services (SMS) , which employed the UCI SMS dataset with the CNN method using auxiliary hand-engineered features . Spam SMS was also detected using RNN-LSTM and LSTM only, which were also compared to machine learning methods . Besides messages, there is much spam content on social media. Spam content can be found on social media like IG, FB, and TW .

Article detected spam content based on spammers' accounts on IG in English. This study used Random Forest (RF) to detect the text content datasets totaling 1983 and 953808 media using their proposed method with special hand-engineered addition features. The significant handengineered features are a) the presence/absence of mention tags to another users; b) the hashtags number used, particularly the hashtags used that are not related to the content; c) the presence or absence of repeated words; d) specific keywords which tend to be spam as defined; and e) the presence/absence of watermarks on images. Using hand engineered features and k=10 in k-fold validation, the result reached 96.27%. Utilizing features that necessitated manual extraction was one of the limitations of the research.

The research differed from in that it employed Indonesian rather than English and did not detect spam posts but rather spam comments. The dataset used in came from a publicly available dataset of Indonesian accounts. However, in contrast to what the authors did, the spam comments referenced in the study were Indonesian-language comments with promotional purposes (such as advertising products). The combination of 1) keyword, 2) content text, and 3) hand-engineered features were employed. The handcrafted characteristics included the number of capital letters, the comment length, and the number of emoticons. Methods used in did not use the emoji features. The keyword feature in the study consisted of specific keywords identified as selling/promoting particular products and extracted using an NLP regular expression pattern. Finally, the text features were extracted and weighted through various TF-IDF, Bag of Words, and FastText techniques configurations. Naive Bayes, SVM, and XGBoost were the classification methods used. Based on , it was found that using all of the features (features 1, 2, and 3) resulted in an F1 score of 96%. According to the research presented in , the employed characteristics were highly contingent on the dataset and cannot be applied to all new data, particularly for keywords retrieved using regular expressions.

Research on I ndonesian spam comment detection, particularly on Instagram, was still rare. A study in employed the Naive Bayes (NB) algorithm to detect Indonesian spam comments with a 72% accuracy rate. In contrast, employed the opposite Naive Bayes algorithm, Complementary Naïve Bayes (CNB), because it used an unbalanced dataset between non-spam and spam comments. With more non-spam comments than spam, the CNB algorithm could achieve an accuracy of 92%, while SVM only achieved 87%. Recent research on social media spam detection, including methods, results, datasets, emoji usage, and post context,.

SpamID-Pair is one of the available datasets and is taken from social media. The hallmark of this dataset is that it includes a large number of emojis that are included in the content. This dataset is also distinctive because the data consists of pairs of posts and comments labeled as spam or non-spam. The social media used in this dataset is IG. The reason is that IG is a popular social media with many users, and many public figures use it. Consequently, much spam is detected,especially in the comments of public figures on Instagram.

IG data contains informal language, lots of emoticons/emojis, some of typos and abbreviations, lots of code mixes (mixed languages), comments of varying lengths but relatively short (1-3 sentences with five words each), a post-reply structure with no hierarchical data, and mention tags (using the symbol '@').

The pre-processing phase was nearly identical to numerous studies that employed text data. NLP techniques were required for most pre-processing in detecting spam remarks or posts. Several references, such as and explained the importance of text pre-processing before further processing. Tokenization, case-folding, n-gram features, stemming, post-tagging, and stop-words removal were the methods that were used. Based on these pre-processing techniques, stemming techniques had the least significant effect . Besides pre-processing, most features in many NLP research features were the text. Some research used tokens feature in the form of BoW or weighted tokens in the form of TF-IDF.

III. METHODOLOGY

EXISTING SYSTEM

Some research on spam content detection has been conducted previously. Spam detection was mainly done in text messages, such as in the Short Message Services (SMS), which employed the UCI SMS dataset with the CNN method using auxiliary hand-engineered features. Spam SMS was also detected using RNN-LSTM and LSTM only, which were also compared to machine learning methods. Besides messages, there is much spam content on social media. Spam content can be found on social media like IG, FB, and TW.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

Article detected spam content based on spammers' accounts on IG in English. This study used Random Forest (RF) to detect the text content datasets totaling 1983 and 953808 media using their proposed method with special hand-engineered addition features. The significant hand engineered features are a) the presence/absence of mention tags to another users; b) the hashtags number used, particularly the hashtags used that are not related to the content; c) the presence or absence of repeated words; d) specific keywords which tend to be spam as defined; and e) the presence/absence of watermarks on images. Using hand engineered features and k=10 in k-fold validation, the result reached 96.27%. Utilizing features that necessitated manual extraction was one of the limitations of the research.

The research differed from in that it employed Indonesian rather than English and did not detect spam posts but rather spam comments. The dataset used in came from a publicly available dataset of Indonesian accounts. However, in contrast to what the authors did, the spam comments referenced in the study were Indonesian-language comments with promotional purposes (such as advertising products). The combination of 1) keyword, 2) content text, and 3) hand-engineered features were employed. The handcrafted characteristics included the number of capital letters, the comment length, and the number of emoticons. Methods used in did not use the emoji features. The keyword feature in the study consisted of specific keywords identified as selling/promoting particular products and extracted using an NLP regular expression pattern. Finally, the text features were extracted and weighted through various TF-IDF, Bag of Words, and FastText techniques configurations. Naive Bayes, SVM, and XGBoost were the classification methods used. Based on , it was found that using all of the features (features 1, 2, and 3) resulted in an F1 score of 96%.

According to the research presented in , the employed characteristics were highly contingent on the dataset and cannot be applied to all new data, particularly for keywords retrieved using regular expressions.

DISADVANTAGES OF EXISTING SYSTEM

- Despite advancements in spam detection, current systems have several limitations that reduce their effectiveness:
- Ignoring Emojis in Spam Classification: Many spam messages use emojis strategically to evade detection. Emojis can replace words, alter the meaning of a comment, or hide spam intent. Existing models fail to analyze emoji sentiment and intent, leading to misclassification.
- Lack of Post-Comment Contextual Understanding: Most systems analyze only the comment without considering the original post. A comment may appear normal in isolation but could be spam when viewed in the context of the post. Without contextual awareness, models struggle to differentiate between relevant and spam comments.
- High False Positives and False Negatives: False Positives: Genuine user comments are mistakenly flagged as spam. False Negatives: Actual spam comments are not detected and remain visible. These errors reduce user trust in the spam detection system and require manual moderation.
- Adaptability to Evolving Spam Techniques: Spammers frequently change tactics, using new wording, symbols, and emoji combinations to bypass detection. Traditional models require continuous retraining to adapt to these new spam patterns.
- Computational Complexity: Deep learning-based systems require high processing power and large datasets for

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

training. Real-time spam detection on large social media platforms can be resource-intensive and expensive.

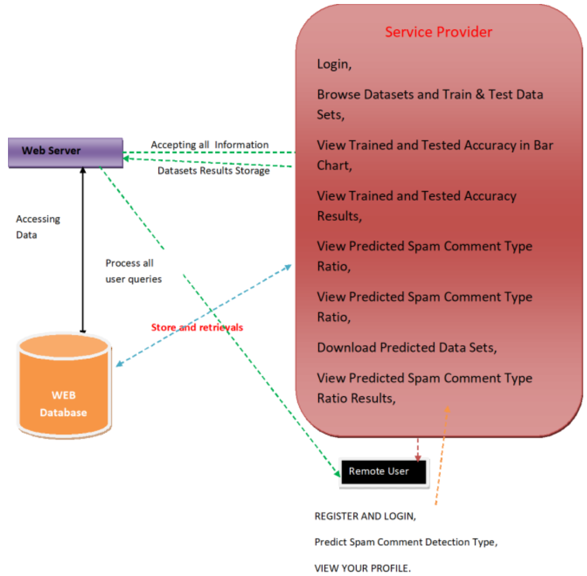
PROPOSED SYSTEM

In this paper, the authors compared and explored the SpamID-Pair dataset collected from 12 celebrities with over 15 million followers with different machine learning techniques according to plus Complement Naïve Bayes (CNB) and Extra Tree (ET). This research made a contribution by providing comprehensive experimental results of spam detection performance (accuracy and F1) between nonemoji and emoji features with various combinations of hyperparameter scenarios (n-grams features, balanced/unbalanced data, the use of comment-only/post-comment pairs approach) using state-of-the-art machine learning and ensemble voting methods as well as their analysis [10]. This research also offers a new approach that uses post and comment text as pair-stacked input in machine learning to identify spam comments based on the posting context. This research uses NLP techniques on the Indonesian SpamID-Pair dataset.

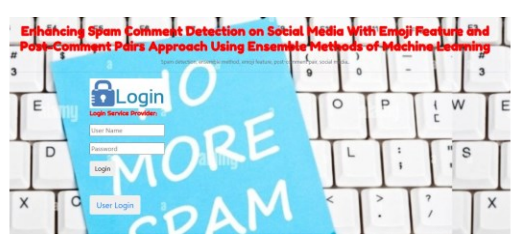
ADVANTAGES OF PROPOSED SYSTEM

- The system is more effective since it involves DATA NORMALIZATION, EMOJI HANDLING, and AND THE USE OF MANUAL FEATURES. The system finds more ADVANTAGES OF THE system which Using and processing the SpamID-Pair dataset modeling.

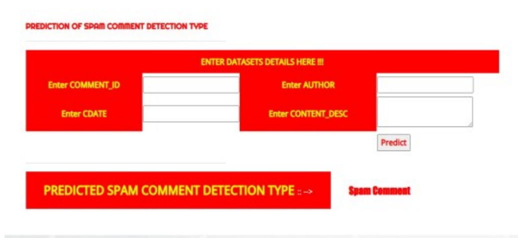
SYSTEM ARCHITECTURE



IV. SCREENSHOTS



International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025



V. CONCLUSION

This research introduces a comprehensive spam filtering approach that incorporates both emoji-aware features and contextual relationships between posts and comments, addressing key limitations in traditional spam detection systems. By recognizing the communicative power of emojis and the importance of post-comment coherence, the proposed method captures subtle spam patterns often missed by text-only models.

The integration of advanced ensemble machine learning techniques—including Random Forest, Gradient Boosting, and Voting Classifiers—has proven effective in enhancing classification performance. The ensemble framework not only improves detection accuracy but also enhances model robustness against diverse and evolving spam behaviors.

Experimental results confirm that combining emotional, semantic, and contextual signals significantly boosts spam detection performance across key metrics. This study demonstrates the potential of enriching spam detection models with multimodal and contextual data, offering a scalable and adaptive solution for real-world social media platforms.

Future work may explore the inclusion of multimedia features, real-time implementation, and deep learning architectures to further advance the effectiveness of spam filtering systems.

REFERENCES

1. Databooks. (2020). Ini Media Sosial Paling Populer Sepanjang April 2020. Accessed: Nov. 4, 2020. [Online]. Available: https://databoks.katadata.co.id/datapublish/2020/05/25/ini-media-sosial-paling-populer-sepanjangapril-2020
2. S. Aiyar and N. P. Shetty, ‘‘N-gram assisted YouTube spam comment detection,’’ Proc. Comput. Sci., vol. 132, pp. 174–182, Jan. 2018, doi: 10.1016/j.procs.2018.05.181.
3. A. R. Chrismanto, A. K. Sari, and Y. Suyanto, ‘‘Critical evaluation on spam content detection in social media,’’ J. Theor. Appl. Inf.Technol., vol. 100, no. 8, pp. 2642–2667, 2022. [Online]. Available:http://www.jatit.org/volumes/Vol100No8/29Vol100No8.pdf
4. W. Zhang and H.-M. Sun, ‘‘Instagram spam detection,’’ in Proc. IEEE 22nd Pacific Rim Int. Symp. Dependable Comput. (PRDC), Jan. 2017,pp. 227–228, doi: 10.1109/PRDC.2017.43.
5. B. Priyoko and A. Yaqin, ‘‘Implementation of naive Bayes algorithm for spam comments classification on Instagram,’’ in Proc. Int. Conf.Inf. Commun. Technol. (ICOIACT), Jul. 2019, pp. 508–513, doi:10.1109/ICOIACT46704.2019.8938575.
6. N. A. Haqimi, N. Rokhman, and S. Priyanta, ‘‘Detection of spam comments on

Instagram using complementary Naïve Bayes,'' Indonesian J. Comput. Cybern. Syst., vol. 13, no. 3, p. 263, Jul. 2019, doi:10.22146/ijccs.47046.

7. A. R. Chrismanto and Y. Lukito, ''Identifikasi komentar spam pada Instagram,''Lontar Komputer, Jurnal Ilmiah Teknologi Informasi, vol. 8, no. 3, p. 219, Dec. 2017, doi: 10.24843/lkjiti.2017.v08.i03.p08.
8. A. R. Chrismanto, Y. Lukito, and A. Susilo, ''Implementasi distance weighted K-nearest neighbor untuk klasifikasi spam & non-spam pada komentar Instagram,'' Jurnal Edukasi dan Penelitian Informatika, vol. 6,no. 2, p. 236, Aug. 2020, doi: 10.26418/jp.v6i2.39996.
9. F. Prabowo and A. Purwarianti, ''Instagram online shop's comment classification using statistical approach,'' in Proc. 2nd Int. Conf. Inf. Technol., Inf. Syst. Electr. Eng. (ICITISEE), Yogyakarta, Indonesia, Nov. 2017,pp. 282–287, doi: 10.1109/ICITISEE.2017.8285512.
10. C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, ''An up-to-date comparison of state-of-the-art classification algorithms,'' Expert Syst. Appl., vol. 82, pp. 128–150, Oct. 2017, doi: 10.1016/j.eswa.2017.04.003.
11. C. Mus. (2015). 10+ Akun Instagram Dengan Followers Terbanyak Di Indonesia. Accessed: Oct. 13, 2021. [Online]. Available:http://www.musdeoranje.net/2016/08/akun-instagram-dengan-followersterbanyak-di-indonesia.html
12. S. Rao, A. K. Verma, and T. Bhatia, ''A review on social spam detection: Challenges, open issues, and future directions,'' Expert Syst. Appl.,vol. 186, Dec. 2021, Art. no. 115742, doi: 10.1016/j.eswa.2021.115742.
13. P. K. Roy, J. P. Singh, and S. Banerjee, ''Deep learning to filter SMS spam,'' Future Gener. Comput. Syst., vol. 102, pp. 524–533, Jan. 2020, doi:10.1016/j.future.2019.09.001.
14. A. Chandra and S. K. Khatri, ''Spam SMS filtering using recurrent neural network and long short term memory,'' in Proc. 4th Int.Conf. Inf. Syst. Comput. Netw. (ISCON), Nov. 2019, pp. 118–122, doi: 10.1109/ISCON47742.2019.9036269.
15. A. A. Septiandri and O. Wibisono, ''Detecting spam comments on Indonesia's Instagram posts,'' J. Phys. Conf. Ser., vol. 801, no. 1, 2017,Art. no. 012069, doi: 10.1088/1742-6596/755/1/011001.
16. A. Chrismanto and Y. Lukito, ''Klasifikasi komentar spam pada Instagram berbahasa Indonesia menggunakan K-NN,'' in Proc. Seminar Nasional Teknologi Informasi Kesehatan (SNATIK). Yogyakarta, Indonesia: STIKES Surya Global, 2017, pp. 298–306.
17. A. Talha and R. Kara, ''A survey of spam detection methods on Twitter,''Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 3, pp. 29–38, 2017, doi:10.14569/ijacsa.2017.080305.
18. N. M. Samsudin, C. F. B. M. Foozy, N. Alias, P. Shamala, N. F. Othman,and W. I. S. Wan Din, ''YouTube spam detection framework using Naïve Bayes and logistic regression,'' Indones. J. Electr. Eng. Comput. Sci.,vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.
19. N. Alias, C. F. M. Foozy, and S. N. Ramli, ''Video spam comment features selection using machine learning techniques,'' Indonesian J.Electr. Eng. Comput. Sci., vol. 15, no. 2, pp. 1046–1053, 2019, doi:10.11591/ijeecs.v15.i2.pp1046-1053.
20. N. Banik and M. H. H. Rahman, ''Toxicity detection on Bengali social media comments using supervised models,'' in Proc. 2nd Int. Conf. Innov. Eng. Technol.

ISSN 2347–3657

Volume 13, Issue 2, 2025

(ICIET), Dec. 2019, pp. 1–5, doi:10.1109/ICIET48527.2019.9290710.

21. R. Abinaya, E. B. Niveda, and P. Naveen, ‘‘Spam detection on social media platforms,’’ in Proc. 7th Int. Conf. Smart Struct. Syst. (ICSSS), Jul. 2020,pp. 1–3, doi: 10.1109/ICSSS49621.2020.9201948.
22. P. Shil, U. S. Rahman, M. Rahman, and M. S. Islam, ‘‘An approach for detecting Bangla spam comments on Facebook,’’ in Proc. Int. Conf.Electron., Commun. Inf. Technol. (ICECIT), Sep. 2021, pp. 1–4, doi:10.1109/ICECIT54077.2021.9641358.
23. C. Fatichah, W. F. Lazuardi, D. A. Navastara, N. Suciati, and A. Munif,‘‘A content filtering from spam posts on social media using weighted multimodal approach,’’ J. Comput. Sci., vol. 17, no. 1, pp. 55–66, Jan. 2021,doi: 10.3844/jcssp.2021.55.66.
24. H. Oh, ‘‘A YouTube spam comments detection scheme using cascaded ensemble machine learning model,’’ IEEE Access, vol. 9,pp. 144121–144128, 2021, doi: 10.1109/ACCESS.2021.3121508.
25. Y. Tashtoush, A. Magableh, O. Darwish, L. Smadi, O. Alomari, and A. ALghazoo, ‘‘Detecting Arabic YouTube spam using data mining techniques,’’ in Proc. 10th Int. Symp. Digit. Forensics Secur. (ISDFS),Jun. 2022, pp. 1–5, doi: 10.1109/ISDFS55398.2022.9800840.