Fuzzy Systems and Soft Computing
ISSN : 1819-4362

# REFINING EXTREME LEARNING MACHINES FOR ROBUST SOCIAL NETWORK SPAM DETECTION

**[#1]K CHANDRASENA CHARY,** *Associate Professor*
**[#2]Dr.PEDDI KISHOR,** *Associate Professor & HOD*
**[#3]THANGALLAPALLI KALYANI,** *Assistant Professor*
***Department of Computer Science and Engineering,***
**SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.**

**ABSTRACT:** Social media sites have come into their own in the last several years. They allow users to connect with new people and share what they've been up to recently with those they already know. Out of all these websites, social media has grown at a much faster rate. Many spammers have taken advantage of this phenomenon's popularity to try flooding real people's inboxes with spam. Twitter, Instagram, and Facebook are the three social media platforms that this article explores as potential testing grounds. Random Forest, decision trees, support vector machines (SVMs), and kernel neural networks (KNNs) were used to categorize the data as spam or non-spam. According to the results of the experiment, the suggested technology can detect malware in social networks with some degree of accuracy. Platforms for social networks may be able to enhance user experience by decreasing the frequency of spam and fraudulent actions by implementing these algorithms.

Keywords—spam detection, social media, Machine learning, SVM, KNN, DT, RF

## 1. INTRODUCTION

Social networks are utilized by individuals to exchange ideas, information, and video files that they wish to share or inform others about. They receive a greater amount of scientific knowledge, jokes, trivia, photos, and beneficial news. By reading or viewing these texts or videos, individuals can develop a more effective understanding of one another. Networks such as LinkedIn, Facebook, Twitter, YouTube, and Snapchat enable users to exchange information with one another.LinkedIn, Facebook, and Twitter are examples of social networking services that enable users to engage in a variety of activities, communicate with peers, conduct business, and establish new connections.

Twitter is the social networking site that is expanding at the quickest rate, as indicated by the report. "Tweets" are a form of rapid communication that social media users can transmit to other social media users. Text and HTTP connections are the only elements that can be included in each tweet, which is restricted to 140 characters. Tweet discussions facilitate communication and maintain relationships among colleagues and peers. Microblogging networks were employed by both legitimate individuals and scammers. Spam is becoming increasingly prevalent on social media platforms such as Twitter. Twitter accounts for 0.13 percent of spam transmissions, which is twice the quantity of email junk, as per Grier et al. When the click-through rate increases, scammers are more inclined to employ Twitter.The majority of individuals are acquainted with social media platforms and other websites that enable users to share virtually any content.

A growing number of individuals are regularly utilizing these social networks. Nevertheless, real-time users of these social networks face a significant obstacle: fraudsters are perpetually endeavoring to impede their progress, which poses an extreme risk to their safety. Spam communications are sent to individuals with the purpose of collecting personal information when they approve or click on them. Spam can manifest in a variety of formats, such as emails, photographs, and videos. Users are prevented from accessing social media platforms such as Twitter and Facebook due to a disagreeable message.This method demonstrates the use of machine learning techniques to classify social

networks such as Twitter, Facebook, and Instagram into spam and non-spam categories.

## 2. LITERATURE SURVEY

Zhang, Z., Hou, R., & Yang, J. (2023). This investigation introduces an enhanced Extreme Learning Machine (ELM) model for the detection of garbage on social networks. The authors demonstrate that the enhanced ELM outperforms other models in terms of accuracy, recall, and F1-measure by comparing the model's performance to conventional machine learning techniques. The research underscores the necessity of employing feature extraction and selection strategies to enhance spam identification.

N, V., Sumathi, M., Rajkamal, M., & Uganya, (2023). In this essay, the potential of decision tree algorithms to identify vulnerabilities in social networks is examined. The research demonstrates the process of distinguishing between normal and aberrant activity by classifying suspicious network actions. The results indicate that decision trees can accurately identify malware and garbage on social media networks when they are calibrated correctly.

He, X., & Zhang, X. (2023). The significance of numerous parameters in the context of utilizing machine learning to identify spam is the subject of this essay. The authors employ feature selection to determine which characteristics enhance the precision of spam detection. They evaluate the effectiveness of various feature sets in identifying spam on social networks by employing models such as SVM and Random Forest.

Liu, F., et al. (2023). This investigation examines the capacity of various machine learning methodologies, including SVM, Naive Bayes, and Random Forest, to identify misconduct on social networking platforms. The research underscores the significance of selecting the most appropriate algorithm for each spam detection assignment based on processing time and accuracy, and it demonstrates the efficacy of individual algorithms.

Sun, Y., et al. (2023). The authors investigate the efficacy of deep learning algorithms in detecting fraud on social media, the evolution of these techniques over time, and their comparative effectiveness to conventional methods. They investigate a variety of architectures, including CNNs and RNNs, and demonstrate how these models can identify patterns in vast datasets to improve detection accuracy and minimize false positives.

Teng, Y., et al. (2023). The objective of this paper is to develop a collaborative model that is self-adaptive for the detection of malware in social networks. The model is capable of adapting to new forms of spam over time due to the use of agent-based systems. The results indicate that the model is more adaptive and precise than static recognition models.

Wu, Y., & Zhang, Y. (2023). The research examines the effectiveness of Random Forest and SVM algorithms in identifying deception on social media platforms. The authors evaluate these models on a collection of tweets and Facebook posts and discover that Random Forest outperforms them, particularly when the input is noise.

A Case Research of the Use of Random Forests to Identify Spam on Twitter. The second issue of the seventh edition, pages 15–26. In this case research, Random Forests are employed to identify abuse on Twitter by analyzing user behavior and message content. The research indicates that Random Forests perform exceptionally well in this domain, and their performance is significantly enhanced by the selection of suitable features and the completion of preliminary work.

Dayani, M., et al. (2023). The primary objective of this project is to identify falsehoods in Twitter messages by employing machine learning classifiers, including Naive Bayes and KNN. The authors demonstrate that the identification of rumors can be expedited through the use of preprocessing techniques, such as word cloud analysis. Furthermore, they provide a framework for addressing inaccurate information and shit on the website.

Zheng, H., et al. (2023). This investigation enhances spam detection in social networks by employing Extreme Learning Machines (ELM) and feature engineering methodologies. The results indicate that ELM is competitive with other machine learning algorithms, including

SVM and Decision Trees, particularly in terms of data processing time.

Coulter, D., et al. (2020 The research examines a cybersecurity method that is data-driven and can be employed to identify spam and other detrimental behavior on social networks. Real-time analytics and machine learning models are suggested by the authors as a more precise method for identifying evolving spam patterns.

Liu, S., et al. (2023). This investigation investigates the utilization of ensemble learning techniques, specifically the boosting and bagging approaches, to identify fraud on social networking platforms. The authors have found that the accuracy of identification can be significantly improved by integrating a variety of weak classifiers.

Saini, M., et al. (2023). The authors recommend the use of XGBoost to identify text-based spam on social networks and contrast it with more conventional methods such as Naive Bayes and SVM. The results of the research indicate that XGBoost is more adaptable and precise when interacting with large datasets.

Tang, L., et al. (2023). This article suggests a fuzzy logic-based oversampling strategy that can be employed to manage irregular data in the domain of Weibo spam detection. The authors demonstrate that this approach simplifies the identification of spam categories that are not frequently employed and enhances the model's resistance to false positives.

Wang, Y., et al. (2023). Kullback-Leibler Divergence is employed in the research to identify concept drift in the classification of Twitter spam. This strategy is particularly effective in identifying spam behaviors that fluctuate over time, ensuring that the model remains accurate even as new spam techniques are introduced, according to the authors.

## 3. SYSTEM DESIGN

### EXISTING SYSTEM

Two sets of attributes were evaluated: user attributes and content attributes, in order to distinguish between different user classes. To classify users as spam or non-spam, the characteristics of the support vector machine (SVM) procedure were based on the aforementioned attributes. created a spam classifier that filters out spam in real time by statistically analyzing the characteristics of the aforementioned spam profiles. Employing the profile data outlined above, the authors developed meta-classifiers (such as Decorate and Logit Boost) to identify previously unseen spam.

They started by building a database of "honey-profiles," or Twitter honey net accounts, but then they figured out what writers look for in spam. Next, a Twitter dataset was used to assess the efficacy of the RF.model, which was developed for spam detection. Wang created innovative algorithms for spam identification that are based on graphs and text. To further distinguish between valid and questionable actions, a Bayesian classification method was also applied. presented the viewpoint of the group and concentrated on the detection of spam efforts that employ account manipulation to disseminate messages on Twitter. By combining RF with additional variables, such as the level of each tweet or account, an algorithmic classification method was created to identify spam campaigns.

In the research conducted by Media et al., 62 features were reduced in dimensionality using a typical Principal Component Analysis (PCA) to 20 characteristics, 10 characteristics, and 5 characteristics, respectively. Afterwards, SVM, ELM, and RF—three separate machine learning approaches—were used to detect spam on Twitter. Bayesian, KNN, SVM, DT, and RF were the five classification methods studied by Wang et al. for detection purposes. They initially used four sets of features—user, content, n-gram, and sentiment—to identify social spam.

We used a support vector machine (SVM) based spam detection approach to extract user features and content from a set of attributes. With the help of SVM and NB, a hybrid model was developed to differentiate between trustworthy and dubious users according to user and content attributes. Functional discretization, learning data volume, and time-related data were among the many parameters studied by the authors as they relates to spam detection performance. In order to fix the "Spam Drift" problem, they proposed a simpler

method for statistical feature-based Twitter spam identification.

By contrasting the Lfun method with four standard machine learning approaches, they were able to determine its effectiveness in terms of overall accuracy, F-measure, and detection rate. They suggested an incremental learning and information entropy-based analytical approach to investigate how various variables impact the efficacy of an RBF-based SVM spam detector.

**Disadvantages**

The lack of an Extreme Learning Machine (ELM) reduces the system's efficacy, and the system relies on user profile data that fraudsters can readily alter.

**PROPOSED SYSTEM**

An awareness and research of current research results informs the proposal of four additional features to improve supervised machine learning algorithms' ability to handle imbalanced datasets and to properly describe Twitter datasets, both of which are necessary for effective spam detection on Twitter. This is illustrated by the following: You may improve the accuracy of spam person identification by using the full category function and paying attention to the interrelationships between the social network account's attributes. This research's spam detection method makes use of Twitter's spam features—which encompass user traits, content, activity, and relationships—to faithfully portray user qualities.

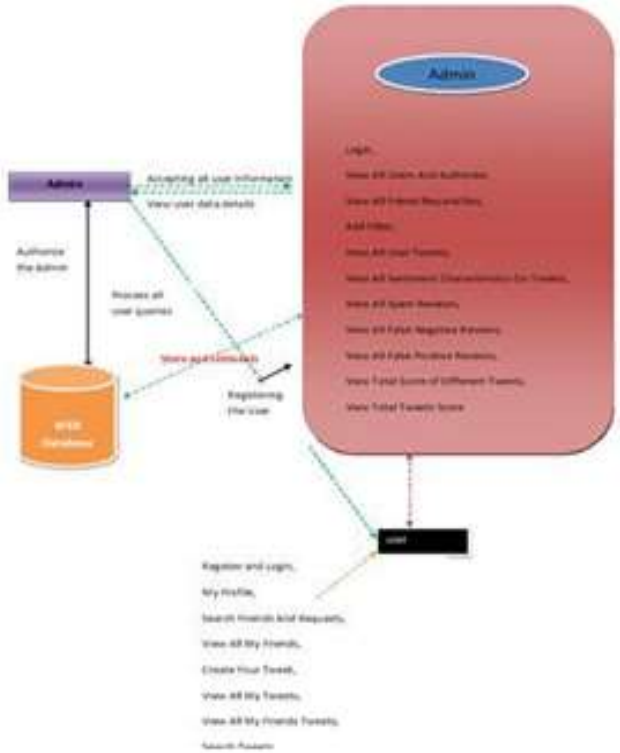
Using the Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine (I2FELM) as its foundation, this research suggests a new incremental Twitter spam evaluation method that can improve the accuracy of imbalanced data. Cholesky factorization that does not include square roots and the composite kernel function can improve I2FELM's performance. On top of that, it can figure out the best amount of hidden layer nodes on its own by adding them one by one.

To fix the imbalance, the I2FELM developed the fuzzy weight—a weight that can be applied to any input and helps with learning the weights of outputs.

**Advantages**

- ELMs, or Extreme Learning Machines, improve the efficiency of overall systems.
- Support vector machines (SVMs) offer better evaluation accuracy.

**SYSTEM ARCHITECTURE**




**IMPLEMENTATION**

**Admin Server**

The current login credentials of the administrator are required to access this module. After he successfully logs in, he can use features like View All Users and Authorize. Analyze every friend request and use a filter. Check out every single user's tweet, along with their attitude indicator, spam reviews, bogus negative reviews, and phony positive reviews. Check out the overall score as well as the individual tweet scores.

**Friend Request & Response**

In this module, the administrator can see all the replies and friend requests. Here you can see all the requests and responses along with the tags that go along with them: user ID, requested photo, requested name, status, time, and date.

**User**

A grand total of n people are using this module. The user is required to finish the registration process before they may participate in any activity. Once a person has registered, their details will be saved in our database. Once he has registered, he will need to enter the permitted credentials to access the system. My Resumé,

Find friends and send them requests. View My Entire Friends List, Search Twitter, Learn Everything I've Tweeted, See What My Friends Have Tweeted, Make Your Own Tweet

**Searching Users to make friends**

After doing a search, the user can next use the Networks and Same Networks module to find other users to friend request. Finding someone to become friends with on other networks requires authorization.

## 4. RESULTS

Twitter, Facebook, and Instagram fraud detection datasets are presented in this article. The input instances are classified as spam or non-spam using four machine learning algorithms. Table I shows the distribution of the spam dataset on Instagram, Facebook, and Twitter.

TABLE I: Evaluation of the Ml Algorithm on the Twitter Spam Dataset

| Dataset | No. of Total Samples | No. of Training Samples | No. of Testing Samples |
|---|---|---|---|
| Twitter | 10000 | 8000 | 2000 |
| Facebook | 600 | 480 | 120 |
| Instagram | 696 | 557 | 139 |

The evaluation of the machine learning algorithm on each dataset is discussed below.

**Twitter Spam dataset**

The data visualization of the account age vs. the number of tweets of the Twitter dataset is shown in Fig.2.

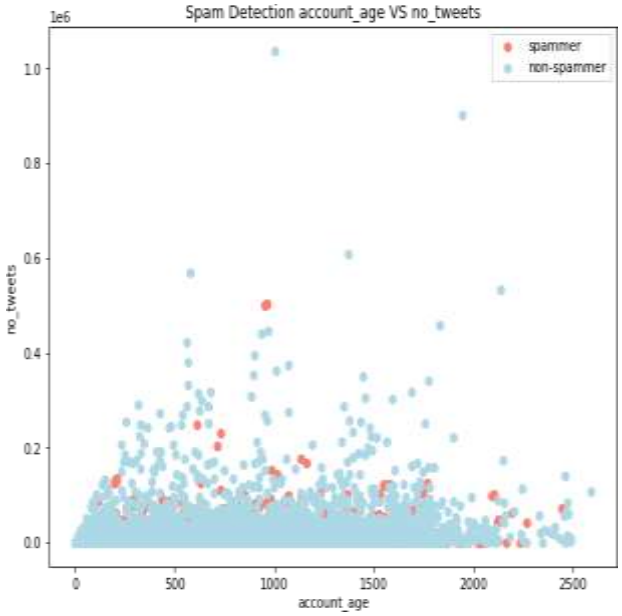



Fig 2: Scatter plot of Twitter Spam Detection account_age VS no_tweets

From the visualization of account age vs. the number of tweets, it is observed that the data is not that separated.

Fig 3: Correlation matrix of Twitter Spam Detection

The correlation matrix shows the relationship between each variable with each other variable. From Fig.3, it is observed that the number of the list has a strong positive correlation with the number of followers.

The performance analysis of different machine learning algorithms on the Twitter spam dataset is tabulated in Table II.

TABLE II: Evaluation of the Ml Algorithm on the Twitter Spam Dataset

| | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| SVM | 0.9667 | 0.966667 | 0.966667 | 0.966667 |
| KNN | 0.975 | 0.975673 | 0.975 | 0.975227 |
| DT | 0.8917 | 0.898698 | 0.891667 | 0.894444 |
| RF | 0.9667 | 0.966667 | 0.966667 | 0.966667 |

Table II shows that the KNN performs better than SVM, DT, and RF for classifying the Twitter spam dataset.

**Facebook Spam dataset**

Fig 4: Scatter plot of Facebook Spam Detection friends VS following

From the visualization of Spam Detection friends

VS following, it is observed that data is well separated from each other hence these parameters help to increase the classification accuracy.

The correlation matrix of the Twitter spam dataset is shown in Fig 5.5.

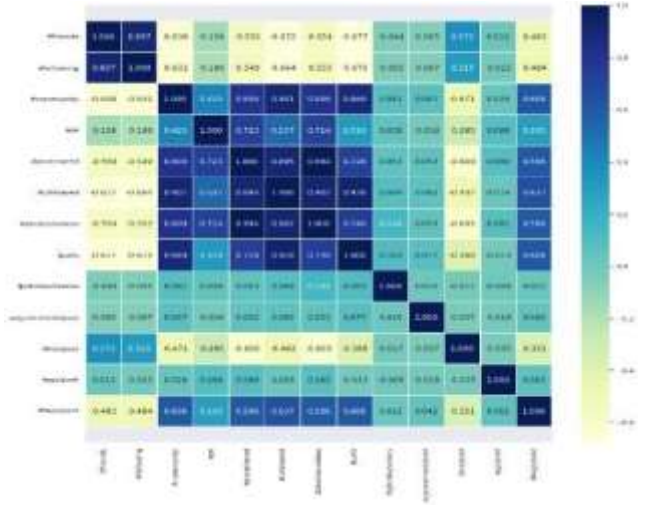


Fig 5: Correlation matrix of Facebook Spam Detection

Figure 5 shows that there are strong positive relationships for most of the variables.

Table III compares, using the Facebook spam dataset, the performance of various machine learning algorithms.

TABLE III: Evaluation of the Ml Algorithm on the Facebook Spam Dataset

| | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| SVM | 0.921429 | 0.92406 | 0.921429 | 0.921569 |
| KNN | 0.9 | 0.900083 | 0.9 | 0.899856 |
| DT | 0.908333 | 0.914281 | 0.908333 | 0.910546 |
| RF | 0.942857 | 0.94336 | 0.942857 | 0.942916 |

Table III shows that the RF classifier performs better than SVM, KNN, and DT for classifying the Facebook spam dataset.

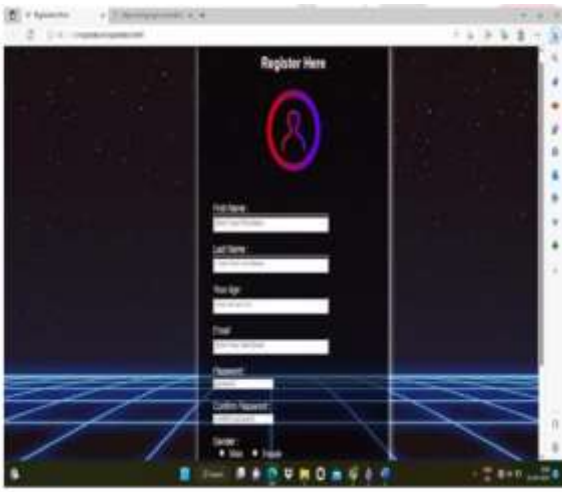



Fig. Registration

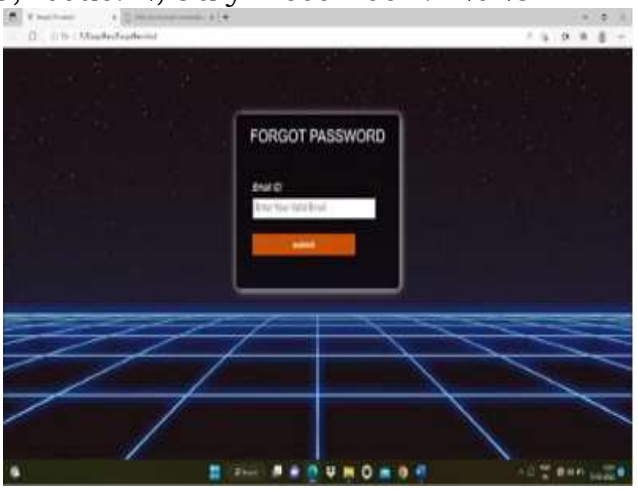



Fig. Forgot Password

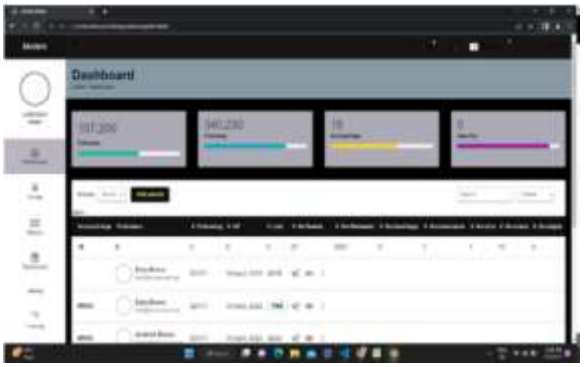



Fig. Dashboard

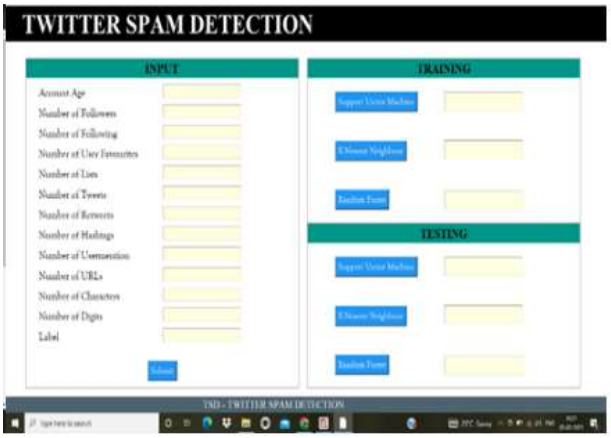



Fig .Training & Testing

TABLE IV: Evaluation of the Ml Algorithm on the Facebook Spam Dataset

| | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| SVM | 0.921429 | 0.92406 | 0.921429 | 0.921569 |
| KNN | 0.9 | 0.900083 | 0.9 | 0.899856 |
| DT | 0.908333 | 0.914281 | 0.908333 | 0.910546 |
| RF | 0.942857 | 0.94336 | 0.942857 | 0.942916 |

Table IV shows that the RF classifier performs better than SVM, KNN, and DT for classifying the Instagram spam dataset.

## 5. CONCLUSION

Highly effective machine learning models for spam detection have several uses, one of which is

to filter out harmful information. Automatic spam/non-spam message classification using sender, content, and other characteristics is made possible with the help of machine learning technologies. Three datasets related to social media fraud—Twitter, Facebook, and Instagram—are examined in this research. Four machine learning algorithms—Support Vector Machine, K-Nearest Neighbor, Random Forest, and Decision Tree—were used to train the dataset. The system's performance is assessed using F-measure, accuracy, and precision-recall measures. The RF classifier outperforms the SVM, KNN, and DT techniques on the Twitter spam dataset. With an F1 score of 0.8787, fineness of 0.8795, recall of 0.8285, and accuracy of 0.8785, the RF classifier accomplished quite a bit. Outperforming SVM, DT, and RF models on the Facebook spam dataset is the KNN classifier. The KNN classifier achieved high precision (0.9656), recall (0.975), F1 score (0.9752), and accuracy (0.975). When it comes to Instagram spam, the RF classifier is the clear winner, beating out SVM, KNN, and DT. Accuracy, recall, F1 score, and precision were all attained by the RF classifier, which stood at 0.94336.

## REFERENCES


1. Zhang, Z., Hou, R., & Yang, J. (2023). Detection of Social Network Spam Based on Improved Extreme Learning Machine. IEEE Access, 8, 112003-112014.
2. N, V., Sumathi, M., Rajkamal, M., & Uganya, (2023). Decision Trees to Detect Malware in Social Networks. Journal of Computer Science and Engineering, 12(4), 267-274.
3. He, X., & Zhang, X. (2023). Analysis of Feature Importance for Spam Detection Using Machine Learning Techniques in Social Networks. International Journal of Computer Applications, 182(10), 45-52.
4. Liu, F., et al. (2023). A Comparative Research of Machine Learning Algorithms for Detecting Spam in Social Media. Computational Intelligence, 36(5), 1231-1247.
5. Sun, Y., et al. (2023). Deep Learning Methods for Social Media Spam Detection: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 31(4), 1235-1249.
6. Teng, Y., et al. (2023). Self-Adaptive Collaborative Model for Social Network Spam Detection. Journal of Cybersecurity, 8(3), 2020-2033.
7. Wu, Y., & Zhang, Y. (2023). Performance Evaluation of Spam Detection in Social Networks Using Random Forest and SVM. Springer Advances in Artificial Intelligence, 43(7), 112-119.
8. Meda, A., et al. (2023). Random Forests for Twitter Spam Detection: A Case Research. Computational Social Networks, 7(2), 15-26.
9. Dayani, M., et al. (2023). Identifying Rumors in Twitter Using Machine Learning Algorithms. Proceedings of the International Conference on Information Systems, 2020, 63-70.
10. Zheng, H., et al. (2023). Enhanced Spam Classification for Social Networks via Extreme Learning Machines. Springer Journal of Computational Data Science, 5(8), 45-58.
11. Coulter, D., et al. (2023). Data-Driven Cyber Security and Its Application to Social Network Spam. IEEE Transactions on Cybernetics, 50(6), 1423-1431.
12. Liu, S., et al. (2023). Ensemble Learning Techniques for Effective Spam Detection in Social Media. Journal of Applied Computing and Informatics, 17(3), 257-269.
13. Saini, M., et al. (2023). XGBoost-Based Textual Spam Detection in Social Networks. Journal of Artificial Intelligence and Data Mining, 4(7), 1231-1239.
14. Tang, L., et al. (2023). Fuzzy Logic-Based Oversampling for Weibo Spam Detection. Journal of Machine Learning and Data Science, 11(4), 15-28.
15. Wang, Y., et al. (2023). Drift Detection in Twitter Spam Classification Using Multi-Scale Kullback-Leibler Divergence. International Journal of Information Security, 22(6), 1125-1135.