Spam Detection in Connected Networks Using Particle Swarm and Genetic Algorithm Optimization: Youtube as a Case study

Amel Ali Alhussan [*1], Hassan K. Ibrahim Al-Mahdawi[2], Ammar Kadi[3]

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

[2]Electronic and Computer Center, University of Diyala, Baqubah MJJ2+R9G, Iraq

[3] Department of Food and Biotechnology, South Ural State University; pr. Lenina 76, Chelyabinsk, 454080 Russia

Emails: aalhussan@pnu.edu.sa; hssnkd@gmail.com; ammarka89@gmail.com

Abstract

Although there are many networks security tools, both wire and wireless connected networks are still suffering from many types of attacks. YouTube's meteoric rise to prominence as a social platform speaks for itself. The sheer volume of comments on YouTube has made it an ideal medium for spammers to spread their malicious software. Phishing attacks, in which anyone who clicks on a bad link might be a victim, have contributed to this problem. Classification systems may be used to examine spam for its unique characteristics and identify it. This is why it is suggested that YouTube already has built-in mechanisms for identifying spam. A YouTube Spam detection framework was designed with the five stages of data collection, pre-processing, features extraction, classification, and detection, allowing for the execution of the tests. To analyze and validate each stage of the YouTube detection methodology presented in this study, two metaheuristic optimization methods are employed to optimize the parameters of a new voting ensemble classifier. These methods are the particle swarm optimization (PSO) and the Genetic Algorithm (GA). The ensemble model is based on three classifiers: neural. Results indicate that the proposed approach is accurate. In addition, statistical analysis is performed to emphasize the superiority and effectiveness of the proposed methodology.

Keywords: Connected Networks; Spam detection; Voting ensemble; Neural network; Support vector machine; Decision tree.

1. Introduction

YouTube is a popular and well-known platform for sharing and distributing videos online. The purpose of YouTube is to allow users to upload and share videos of interest easily. The video is available for viewing by any Internet user anywhere globally. Users of YouTube may upload, share, and discuss videos directly from the platform. Users' comments often include spam [1] or electronic messages delivered in mass to a set of recipients that are either undesired or uninvited and have no relevance to the topic at hand. Spam has various adverse effects, such as wasting time and resources for users and the network. The potential for monetary loss due to spam is a serious concern for businesses and individual users [2]. Advertisements, computer viruses, and attempts to steal users'

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

financial information are just a few of the spammy purposes of the YouTube comments section [3]. Malicious mail, if opened, redirects users to phishing websites [3, 4], and the spread of malware [5, 6] poses the most significant security concerns. According to Table 1 from Gandra [4], spam outnumbers legitimate videos on YouTube by a factor of 100 to 1. The severity of the spam assault on YouTube is seen here.

Table 1: Report on Popular Social Platforms for Spam [4]

Description	Data
Social media apps that are spammy	5%
Spammy social media apps that are brand-owned	20%(that is 1% overall)
The average number of social profiles contacted by a spamming account	23
Number of new spam accounts created	5 out of every new accounts
Most Popular social platform for spammers	Facebook & You Tube
Percentage of spam that contains URL	15%
The overall number of social media messages	1 out of every 200

The architecture in Figure 1 shows the overview of this research paper.

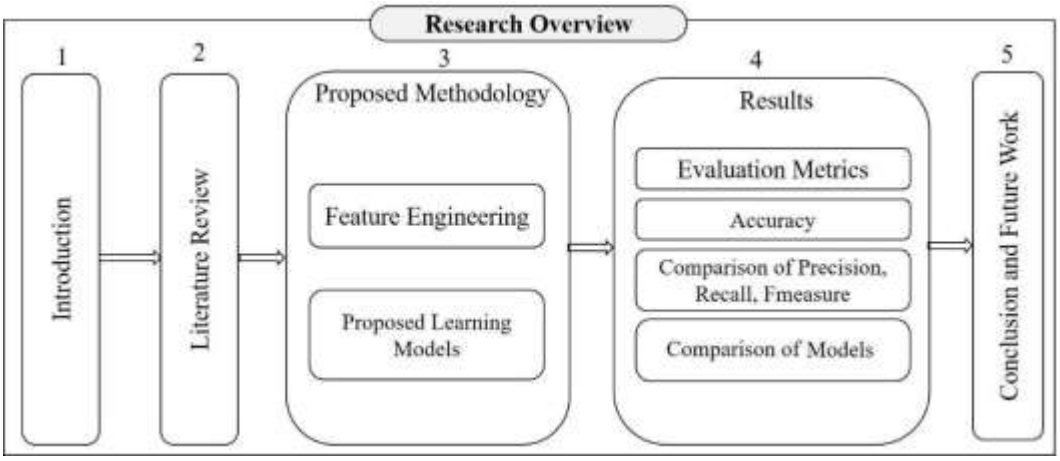

Figure 1: Research Paper Overview

2. Literature Review

The news of the Spam assault quickly spread. SMS and emails are included in this, along with the more prominent social media channels like Facebook, Twitter, YouTube, and blogs. According to [5], spam emails are defined as those that are sent repeatedly from several individuals. Tran et al. [6] define email spam as unsolicited mass emails received by users. Nonetheless, according to Stone [7], spam is any form of a commercial email that the recipient does not request. Spam is any unsolicited or unwelcome communication sent to a mobile device over the Short Message Service (SMS) [8]. Web spamming, also known as SEO-optimization fraud, refers to manipulating search engine results to boost the visibility of a particular website artificially. Spam comments are defined as those that contain commercial content that is irrelevant to the debate and includes demands for information that are not sought [1]. According to Yusof and Sadoon [1], video spam is defined as having irrelevant and annoying material compared to the video's title.

The news of the Spam onslaught had quickly spread. Facebook, Twitter, YouTube, weblogs, text messages, and electronic letters. According to [5], unwanted email, or spam, is messages received at least once daily from many individuals. Email spam, as defined by Tran et al. [6], is any unsolicited, mass email sent to a user. Though according to Stone [7], spam is any form of commercial communication that is not requested. SMS spam refers to any unsolicited or unwelcome text message sent to a mobile phone [8]. Spam has come to describe any practice that is designed to trick a search engine into giving a particular web page a better rating than it deserves [9], [10]. Comments with commercial content irrelevant to the conversation and comments with unwelcome

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

content or demands have been detected as spam [1]. Additionally, Yusof and Sadoon [1] describe video spam as irrelevant or undesirable material compared to the video's title.

Table 2: Comparison of YouTube Spam Framework

Author	Title	Framework				
		Data	Pre-processing	Features	Classification	Evaluation
[1]	Detecting Video Spammers in YouTube Social Media	√	√	√	√	√
[11]	TubeSpam: Comment Spam Filtering on YouTube	√	√		√	
[12]	Detecting spammers in YouTube: A study to find spam content in a video platform.	√		√	√	
[13]	A Data Mining-Based Spam Detection System for YouTube	√		√	√	

According to Table 2, most studies that attempted to create a framework for detecting YouTube spam included data collection, feature selection, and classification. Data collection, preprocessing, feature selection, classification, and detection are all components of the framework employed in this study that is similar to those employed by Yusof and Sadoon [1].

As it stands, researchers are gathering two different kinds of information. YouTube videos and user comments make up the two dataset kinds. The YouTube comments dataset was utilized by Alberto et al. [11] which was obtained from the UCI Machine Learning Repository [28]. In all, there are 1005 spam comments and 951. However, Kiran [12] utilized data from a Crawling Algorithm applied to YouTube videos, including 473 spam videos and 119 ham videos. In addition, Yusof and Sadoon [1] employ a web page-extracted YouTube video dataset. A total of 30621 movies, both spam, and ham, make up the dataset. Chowdury et al. [13], on the other hand, get their data from TubeKit, where they find 685 spam films and 1115 ham videos.

Table 3 displays the kind and quantity of datasets utilized in previous studies. Since the YouTube spam comments dataset used by Alberto et al. [11] is publicly available through the UCI Machine Learning Repository [28], it serves as the basis for this study.

Table 3: Dataset Collection

Author	Dataset Type	Total Number of Dataset
[11]	YouTube Comment	1005 Spam, 951 Ham
[12]	YouTube Video	473 Spam, 119 Ham
[1]	YouTube Video	30621 Spam and Ham
[13]	YouTube Video	685 Spam, 1115 Ham

Three separate research groups—Yusof and Sadoon [1], Alberto et al. [11], and Kiran [12]—have chosen to focus on the problem of YouTube video spam by using three distinct sets of characteristics. One of many different feature types might be selected to act as a research parameter. Yusof and Sadoon [1] found that the Edge Rank Algorithm is more effective in selecting features. The reason for using this algorithm is that it is identical to the one used by Facebook to identify spam. Regarding identifying spammers, the following study by Kiran [12] employed a trifecta of video-based, user-based, and social network information. The number of users, number of

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

comments, number of unique users, number of rating counts, and number of separate categories are some features employed by Chowdury et al.

Comment spam on YouTube is still an emerging field of study, hence typical spam characteristics are rarely recognized. As a result, research is being conducted into making it possible to leave comments on websites using social media platforms like Twitter and text messaging. On the other hand, the most common spam terms are employed as a feature by Alberto et al. [11] to identify spam in YouTube comments. AlSaleh and AlArifi [14] researched comments and extracted features such as post-comment similarity, the interval between posts and comments, the number of words in the comments, the number of sentences in the comments, the length of the comments, contact information (phone number, email address, URL), a black word list, a stop word ratio, and a word duplication ratio. Uysal et al. [15] then employed characteristics such as message length, number of phrases, percentage of capital characters, percentage of non-alphanumeric characters, percentage of alphanumeric characters, and presence of URL in comments. Using negative word counts, negative word counts, a URL, positive word counts, and a positive word ratio, Perveen [16] performed research.

Based on the data in Table 4, most researchers employed a combination of heuristic, keyword, and URL link searches. Only Alberto et al. [11] utilized keyword features, whereas the other three researchers used Heuristic and URL links as features. As a result, we choose and extract from the datasets a collection of characteristics based on a mix of heuristics, keywords, and the existence of URL linkages.

Table 4: Comparison of Comments Features

Author	Title	Framework		
		Heuristic	Keyword	URL Links
[9]	TubeSpam: Comment Spam Filtering on YouTube	√	√	√
[14]	Combating Comment Spam with Machine Learning Approaches	√	√	
[15]	The Impact of Feature Extraction and Selection on SMS Spam Filtering	√		√
[16]	Sentiment-Based Twitter Spam Detection	√		√

Research into detection methods relies heavily on the classification process. At this stage, the classifier that will be used to process the characteristics picked by the researchers and produce the desired output is decided upon. Many studies, including those by Yusof and Sadoon [1], Chowdury et al. [13], Alberto et al. [11], and Kiran [12], employ categorization methods. The study by Yusof and Sadoon [1] employs a nine-classifier system. Functional Tree (FT), J48, and Random Forest (RF) are the three types of classifiers that make up the Decision Tree (DT). Class Bayesian is made up of Bayes Network (BN) and Naive Bayes (NB), whereas Class Function includes LibLINEAR (LL), LibSVM (LSVM), Logistic (LR), Multilayer Perceptron (MLP), and NB (NB). Alberto et al. [11] used six (6) different classifiers to determine which was best at identifying spam comments on YouTube. K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, Logistic Regression, and Naive Bayes Extreme Learning Machine are all examples of classifiers that may be used by a user (LR). With a Support Vector Machine (SVM) classifier, Kiran [12] could identify criminal users on YouTube. In addition, Chowdury et al. [13] employ the more sophisticated methods of Naive Bayes, Clustering, and Decision Tree.

3. Proposed Methodology

The proposed methodology is shown in Figure 2. In this figure, three base classifiers were employed. These classifiers are decision trees (DT), multilayer perceptron (MLP), and support vector machines (SVM). These classifiers are used in a voting ensemble model, where the votes are optimized in a hybrid metaheuristic optimization algorithm composed of the particle swarm optimization (PSO) algorithm and the genetic algorithm (GA).

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

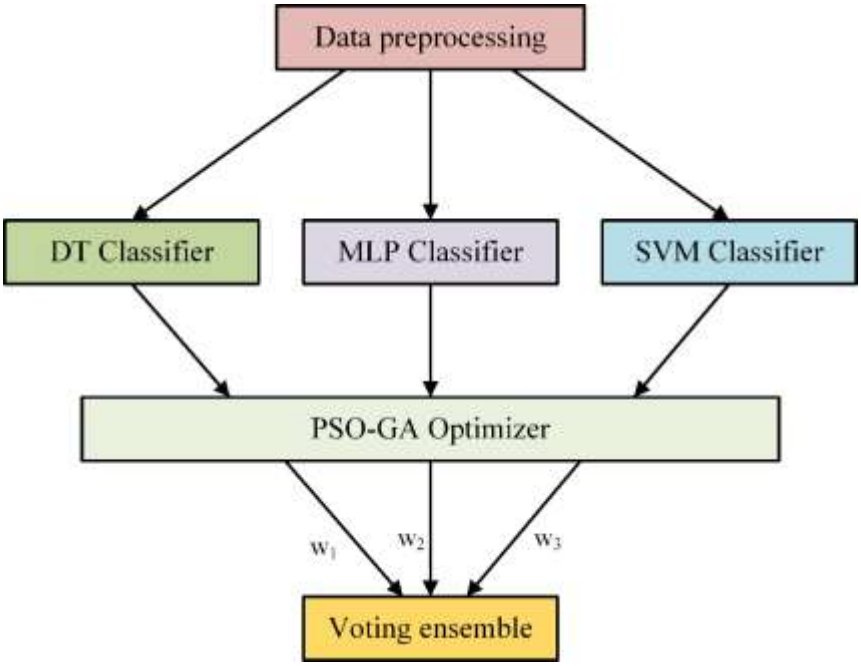

Figure 2: Overview of research methodology

A. Dataset

Existing researchers are gathering two sorts of information. The datasets consist of either user comments or videos uploaded to YouTube. UCI Machine Learning Repository [28] was accessed for the YouTube comments dataset used by Alberto et al. A total of 1005 spam comments, and 951 ham comments are included in the sample. But Kiran [12] utilized data from YouTube videos obtained using a crawling algorithm, which included 473 spam videos and 119 ham videos. Another web-extracted YouTube video dataset is utilized by Yusof and Sadoon [1]. There are a total of 30621 movies, both ham, and spam, in the collection. Chowdury et al. [13] instead use a dataset they get from TubeKit that includes 685 spam videos and 1115 ham films. Since the YouTube spam comments dataset used by Alberto et al., [11] is publicly available through the UCI Machine Learning Repository [28], it serves as the basis for this study. Table 5 displays the kind and quantity of datasets utilized in previous studies.

Table 5: Dataset Collection

Reference	Dataset Type	Total Number of Dataset
[1]	YouTube Video	30621 Spam and Ham
[11]	YouTube Comment	1005 Spam, 951 Ham
[12]	YouTube Video	473 Spam, 119 Ham
[13]	YouTube Video	685 Spam, 1115 Ham

B. Support Vector Machines (SVM)

In classification, a support vector machine (SVM) is described as a family of closely comparable supervised learning techniques. To put it another way, if it is given a collection of training samples that have been labeled as belonging to one of two classes. For each possible classification of a given set of examples, the SVM training algorithm constructs a model to make predictions. To do classification, support vector machines (SVM) create a hyperplane or a series of hyperplanes in a high-dimensional space [25].

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

C. Multilayer Perceptron (MLP)

We have a neural network when a group of nodes, or neurons, are connected via synapses. Comparable to the human nervous system, artificial neural networks are used as estimation models. Every synthetic neural network consists of three distinct layers: an input layer, a hidden layer, and an output layer. The input layer comprises a set of nodes that take in data, and the technique's output is an activation function. In between the input and output layers is a hidden layer that assigns weights to the inputs. The output layer provides the final results. The structure of a multilayer neural network is shown in Figure 3.

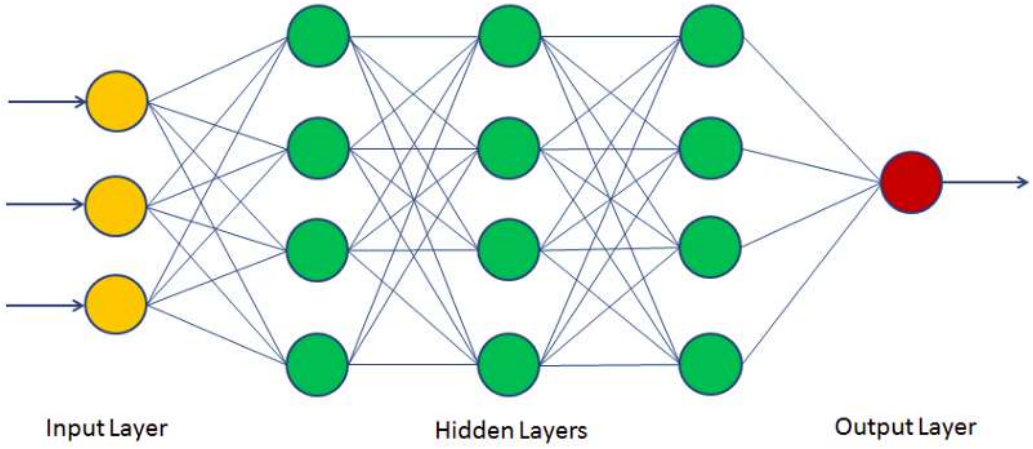

Figure 3: Structure of a multilayer neural network

D. Decision Trees (DT)

Using a greedy search, decision tree learning finds the best possible branching points inside a tree to divide and conquer the problem. This partitioning is then iterated from the top down until all or the vast majority of entries fit neatly into predefined categories. The intricacy of the decision tree heavily influences the likelihood that all data points will be assigned to homogeneous groupings. Pure leaf nodes, or data points belonging to a particular class, can be more easily obtained in smaller trees. However, as a tree expands, it is harder to preserve this purity, and this often leads to too little information fitting within a specific subtree. This is known as data fragmentation, and it frequently results in overfitting. Therefore, decision trees favor tiny trees, which aligns with Occam's Razor's parsimony concept that "entities should not be multiplied beyond necessity." That is to say, decision trees should only get more intricate if required, as the simplest explanation is sometimes the most convincing. Pruning removes branches that divide on attributes of low value and is commonly used to accomplish both goals (reducing complexity and preventing overfitting). Cross-validation then provides an assessment of the model's fitness. The random forest technique is another method for maintaining the accuracy of decision trees; when the trees in the ensemble are uncorrelated, the classifier produces more accurate predictions. The structure of a decision tree is depicted in Figure 4.

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

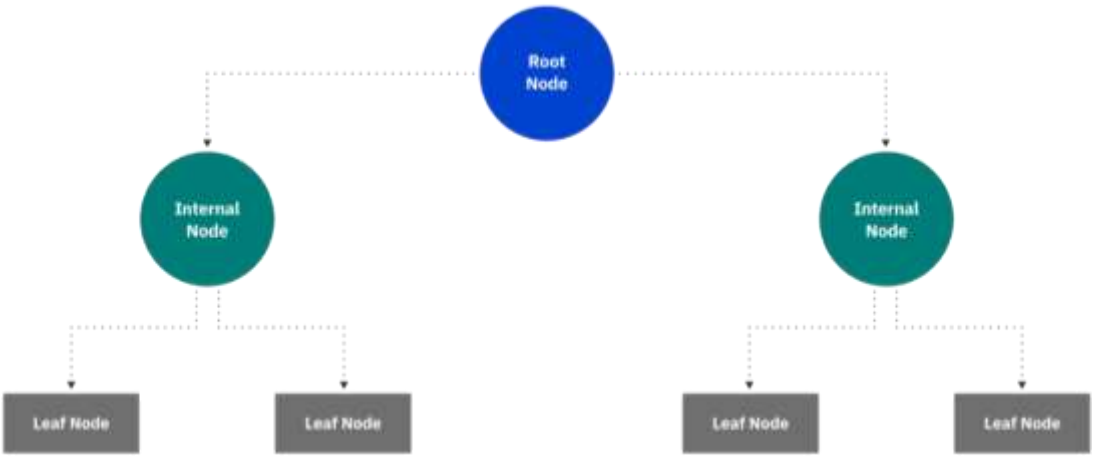

Figure 4: Structure of a decision tree

E. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a potent meta-heuristic optimization technique based on natural swarm behaviors like fish and bird schooling. The PSO is a model of a hypothetical social order with a few basic rules. In its first iteration, the PSO algorithm was developed to graphically replicate a flock of birds' fluid yet unpredictable dance. Depending on the conditions, a bird's field of view might be somewhat restricted in the wild. But a swarm of birds has the advantage of being able to perceive the bigger fitness function surface collectively.

F. Genetic Algorithm (GA)

The genetic algorithm is an example of a conventional evolutionary algorithm. In the context of the genetic algorithm (GA), randomness refers to applying random mutations to existing solutions to produce new ones and move closer to a solution. It's worth noting that, compared to other evolutionary algorithms, GA may be considered Simple GA (SGA). In GA, Darwin's theory of evolution serves as the foundation. It's a method that requires time and effort, as it relies on modest, incremental shifts. Furthermore, GA iteratively improves upon its solutions over time.

4. Results

A comparison between the proposed approach and other machine learning models is presented in Table 6. This table shows the accuracy, sensitivity, specificity, pvalue, nvalue, and F-score better using the proposed optimized voting ensemble classifier.

Table 6: Classification results using the proposed method compared to other methods

	Accuracy	Sensitivity (TRP)	Specificity (TNP)	Pvalue (PPV)	Nvalue (NPV)	F-score
NN	0.8947	0.9412	0.8571	0.8421	0.9474	0.8889
SVM	0.8293	0.9412	0.7500	0.7273	0.9474	0.8205
DT	0.9024	0.9412	0.8750	0.8421	0.9545	0.8889
PSO-GA	0.9487	0.9412	0.9545	0.9412	0.9545	0.9412

The statistical analysis presented in Table 7 shows the superiority of the proposed voting ensemble classifier. These results are better when the proposed optimized voting ensemble is employed.

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

Table 7: Statistical analysis of the results recorded by the proposed method

	NN	SVM	DT	PSO-GA
Number of values	10	10	10	10
Minimum	0.8747	0.8293	0.9024	0.9487
25% Percentile	0.8922	0.8293	0.9024	0.9487
Median	0.8947	0.8293	0.9024	0.9487
75% Percentile	0.8947	0.8318	0.9049	0.9487
Maximum	0.8947	0.8493	0.9124	0.9487
Range	0.02	0.02	0.01	0
10% Percentile	0.8757	0.8293	0.9024	0.9487
90% Percentile	0.8947	0.8483	0.9124	0.9487
95% CI of median				
Actual confidence level	97.85%	97.85%	97.85%	97.85%
Lower confidence limit	0.8847	0.8293	0.9024	0.9487
Upper confidence limit	0.8947	0.8393	0.9124	0.9487
Mean	0.8917	0.8323	0.9044	0.9487
Std. Deviation	0.006749	0.006749	0.004216	0.0000
Std. Error of Mean	0.002134	0.002134	0.001333	0.0000
Lower 95% CI of mean	0.8869	0.8274	0.9014	0.9487
Upper 95% CI of mean	0.8966	0.8371	0.9075	0.9487
Coefficient of variation	0.7569%	0.8110%	0.4662%	0.000%
Geometric mean	0.8917	0.8322	0.9044	0.9487
Geometric SD factor	1.0080	1.0080	1.0050	1.0000
Lower 95% CI of geo. mean	0.8869	0.8275	0.9014	0.9487
Upper 95% CI of geo. mean	0.8966	0.8371	0.9074	0.9487
Harmonic mean	0.8917	0.8322	0.9044	0.9487
Lower 95% CI of harm. mean	0.8868	0.8275	0.9014	0.9487
Upper 95% CI of harm. mean	0.8966	0.837	0.9074	0.9487
Quadratic mean	0.8918	0.8323	0.9044	0.9487
Lower 95% CI of quad. mean	0.887	0.8274	0.9014	0.9487
Upper 95% CI of quad. mean	0.8965	0.8371	0.9075	0.9487
Skewness	-2.277	2.277	1.779	
Kurtosis	4.765	4.765	1.406	
Sum	8.917	8.323	9.044	9.487

On the other hand, the Wilcoxon signed rank test is used to study the difference between the proposed approach and the other set of algorithms. The results of this test are listed in Table 8. In this table, the p-value proves this difference.

Table 8: Wilcoxon signed rank test of the recorded results of the proposed method

	NN	SVM	DT	PSO-GA
Theoretical median	0	0	0	0
Actual median	0.8947	0.8293	0.9024	0.9487
Number of values	10	10	10	10
Wilcoxon Signed Rank Test				
Sum of signed ranks (W)	55	55	55	55

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

Sum of positive ranks	55	55	55	55
Sum of negative ranks	0	0	0	0
P value (two tailed)	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact
P value summary	**	**	**	**
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
How big is the discrepancy?				
Discrepancy	0.8947	0.8293	0.9024	0.9487

The plot shown in Figure 3 demonstrates the accuracy achieved by the proposed optimized voting ensemble classifier compared to the base models. As shown in this figure, the proposed approach is more effective.

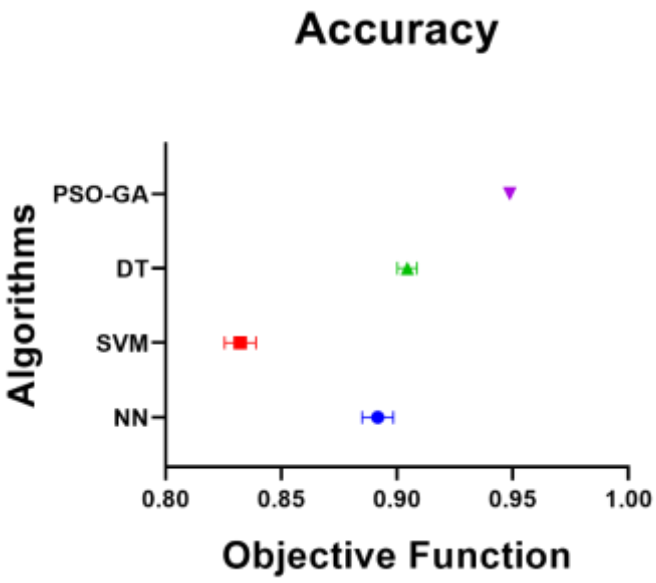

Figure 3: The accuracy of the proposed method compared to other methods

5. Conclusion

In this study, a spam comment detection system was built utilizing machine learning methods. With the Internet today as a warning sign of security flaws, measures must be taken to increase safety [29]. Despite the abundance of research on preventing attacks and safeguarding user privacy, few of these methods have been adapted for use on social media [30-34]. In addition, this study intends to contribute by investigating the appropriate characteristics based on the actual comment from social media sites for constructing spam comment detection systems. Data collection, preprocessing, feature extraction, and classification are steps in creating this system. We have experimentally verified each of these stages by employing machine learning methods. Before running any experiments, you will want to obtain the Data Collection from UCI Machine Learning and run Preprocessing. The findings demonstrate that the proposed methodology is effective. Still, to create an effective Spam detection tool in the future, it is necessary to establish a framework for identifying spam comments. It is crucial to have a tool to recognize spam comments to prevent the user from clicking the dangerous link; thus, it may be incorporated into future works and tested with different classifiers.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

References

[1] Y. Yusof and O. H. Sadoon, "Detecting Video Spammers In Youtube Social Media," no. 082, pp. 228–234, 2017.
[2] U. K. Sah and N. Parmar, "An approach for Malicious Spam Detection In Email with comparison of different classifiers," IRJET, vol 4, i.8, pp. 2238–2242, 2017.
[3] I. Daugher and R. Antoun, "Ham- Spam Filtering Using Different PCA Scenarios," IEEE Int. Conf. Comput. Sci. Eng. IEEE Int. Conf. Embed. Ubiquitous Comput. Int. Symp. Distrib. Comput. Appl. to Business, Eng. Sci., pp. 542–545, 2016.
[4] S. Gandra, "Implementation Of Prototype To Detect Spam In YouTube Using The Application TubeKit And Naïve Bayes Algorithm," 2014.
[5] M. Esmaeili, et al., "An Anti-Spam System using Naive Bayes Method and Feature Selection Methods," International Journal of Computer Applications, vol. 165, no. 4, pp. 1–5, 2017.
[6] K. Tran et al., "Towards a Feature Rich Model for Predicting Spam Emails containing Malicious Attachments and URLs," in Proceedings of the 11-th Australasian Data Mining Conference, 2013, pp. 161–171.
[7] T. Stone, "Parameterization of Na ¨ ıve Bayes for Spam Filtering," 2003.
[8] M. Shafie et al., "A Review on Mobile SMS Spam Filtering Techniques," vol. 5, 2017.
[9] H. Garcia-molina, "Web Spam Taxonomy," pp. 1–9.
[10] J. Zhang and G. Gu, "Neighbor Watcher : A Content-Agnostic Comment Spam Inference System," no. 2.
[11] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "TubeSpam: Comment spam filtering on YouTube," Proc. - 2015 IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015, no. 2012, pp. 138–143, 2016.
[12] P. S. Kiran, "Detecting spammers in YouTube : A study to find spam content in a video platform.," IOSR Journal of Engineering (IOSRJEN), vol. 05, no. 07, pp. 26–30, 2015.
[13] R. Chowdury, N. M. Adnan, G. A. N. Mahmud, and R. M. Rahman, "A Data Mining Based Spam Detection System for YouTube," pp. 373–378, 2013.
[14] M. Alsaleh and A. Alarifi, "Combating Comment Spam with Machine Learning Approaches," 2015.
[15] N. Abdel Samee, E. M. El-Kenawy, G. Atteia, M. M. Jamjoom, A. Ibrahim et al., "Metaheuristic optimization through deep learning classification of covid-19 in chest x-ray images," Computers, Materials & Continua, vol. 73, no.2, pp. 4193–4210, 2022.
[16] A. A. Abdelhamid and S. R. Alotaibi, "Optimized two-level ensemble model for predicting the parameters of metamaterial antenna," Computers, Materials & Continua, vol. 73, no.1, pp. 917–933, 2022.
[17] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "The Impact of Feature Extraction and Selection on SMS Spam Filtering," Elektronika Ir Elektrotechnika, pp. 67–72, 2013.
[18] N. Perveen, "Sentiment Based Twitter Spam Detection," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 7, no. 7, pp. 568–573, 2016.
[19] A. Patwari, "Identifying Undesireble Behaviour in Social Media : Towards Automated Fact-Checking and YouTube Meta-Data Spam Detection," Purdue University, 2017.
[20] S. R. Gomes, S. G. Saroar, M. A. Telot, B. N. Khan, A. Chakrabarty, and M. Mostakim, "A Comparative Approach to Email Classification Using Naive Bayes Classifier and Hidden Markov Model," in Proceedings of the 2017 4th International Conference on Advances in Electrical Engineering (ICAEE), 2017, pp. 28–30.
[21] T. Verma, "Tokenization and Filtering Process in RapidMiner," International Journal of Applieed Information Systems, vol. 7, no. 2, pp. 16–18, 2014.
[22] T. Yang and K. Qian, "Spam Filtering using Association Rules and NaIve Bayes Classifier," pp. 638–642, 2015.
[23] W. Hijawi, H. Faris, J. Alqatawna, A. M. Al-zoubi, and I. Aljarah, "Improving Email Spam Detection Using Content Based Feature Engineering Approach," 2016.
[24] R. Cristina, "Identification of Spam Comments using Natural Language Processing Techniques," pp. 29–35, 2014.
[25] R. E. Mercer, R. Shams, and R. E. Mercer, "Classifying Spam Emails Using Text and Readability Features Classifying Spam Emails using Text and Readability Features," no. December, 2013.
[26] S. Raschka, "Introduction and Theory," pp. 1–20, 2014.

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022

[27] J. Badresiya, Ashok; Vohra, Saifee; Teraiya, "Performance Analysis of Supervised Techniques for Review Spam Detection," Int. J. Adv. Netw. Appl., pp. 21–24, 2014.
[28] C. Visani and N. Jadeja, "A Study on Different Machine Learning Techniques for Spam Review Detection," no. August, 2017.
[29] K. Zainal, N. F. Sulaiman, and M. Z. Jali, "An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka," Int. J. Comput. Sci. Inf. Secur., vol. 13, no. 3, pp. 66–74, 2015.
[30] Lichman, M., "UCI Machine Learning Repository", [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
[31] Salleh, S. N. M., Din, R., Zakaria, N. H., & Mustapha, A., "A Review on Structured Scheme Representation on Data Security Application,". Indonesian Journal of Electrical Engineering and Computer Science, 11(2), pp. 733-739, 2018.
[32] Umapathy, K., & Khare, N., "An Efficient & Secure Content Contribution and Retrieval content in Online Social Networks using Level-level Security Optimization & Content Visualization Algorithm," "Indonesian Journal of Electrical Engineering and Computer Science, 10(2), pp. 807-816, 2018.
[33] Abdelhamid, A.A.; El-Kenawy, E.-S.M.; Khodadadi, N.; Mirjalili, S.; Khafaga, D.S.; et al., Classification of Monkeypox Images Based on Transfer Learning and the Al-Biruni Earth Radius Optimization Algorithm. Mathematics 2022, 10, 3614.
[34] Eid, M.M.; El-Kenawy, E.-S.M.; Khodadadi, N.; Mirjalili, S.; Khodadadi, et al., Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases. Mathematics 2022, 10, 3845.

DOI: https://doi.org/10.54216/IJWAC.060101
Received: September 18, 2022 Accepted: November 17, 2022