RESEARCH ARTICLE OPEN ACCESS

A Review of Various Machine Learning Models for Email Spam Prediction

Shikha [1], Jatinder Singh Saini [2]

Student [1], Head of Department [2]
Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib - Punjab.

ABSTRACT

Email users often face an issue of number of spam emails coming from unfamiliar senders in their mailboxes daily. Spamming is also triggering online cyber fraud based on social engineering. Most of these frauds starts via an email from an unauthentic origin in which a URL is comprised and show compromised one's personal data after its opening. The email spam can be detected in numerous stages such to pre-process the data, extract the attributes and classify the emails. Researchers have constructed several ML (Machine Learning) algorithms in order to detect the email spam. This paper conducts a review on diverse methods used to detect the email spam.

Keywords: - Email Spam, Machine Learning, Supervised learning

I. INTRODUCTION

Email is a robust, effective, and private mode of communication. Spammers are interested in using this kind of communication to disseminate spam. Now that almost everyone has access to email, businesses must deal with the spam issue. Both users and Internet service providers struggle with spam (ISPs). The variables include the speed of electronic communication innovation from one perspective and the acceleration of spam innovation from another perspective [1]. Email is accessible, which puts it at risk for a number of dangers caused by hackers. Spam poses a serious threat to email and is a problem for all email clients worldwide. Unwanted email and messages sent to internet users' inboxes are referred to as spam. Email spam can thus be defined as the act of transmitting unrequested data to email boxes. Email spammers benefit greatly from being able to quickly and cheaply send a big number of messages to a large number of clients. It makes this issue relevant to everyone who uses the internet and frequently receives erratic email. Spam emails ultimately lead to lower productivity, take up space in letter boxes, transmit bugs, trojans, and materials containing possibly lethal data for a particular clientele, disrupt the stability of receiving mails, and as a result, customers waste their valuable time organizing incoming mail and deleting unpleasant messages.

One of the fundamental processes in semantic-based spam detection is the classification of the spam using a set of semantic qualities. Then, each set of semantic features provides the fundamental properties required to build a domain-specific classifier for spam identification. Semantic analysis takes place in two levels. Using a classification technique, emails from a huge training dataset are automatically segmented into the five categories being taken into account at the first level [2]. In the second level, a set of semantic features are automatically mined from each domain's dataset. The semantic properties are then used to build specialized classifiers for detecting spam specific to a given topic. To classify emails by domain, each email in the global training dataset is assigned a category. Be aware that suitable email pre-processing procedures must be followed before the information in an email's topic and content can be used successfully for classification. Figure 1 depicts the overall classification process for emails.

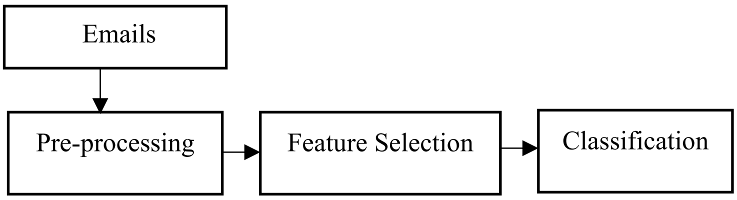

Figure 1: An outline of the different steps used for email classification

The majority of the real-time data that is currently accessible is imperfect and made up of mixed, noisy, and missing numbers. Before beginning the mining process, it is crucial to prepare the dataset for data mining [3]. The main objective of this stage is to exclude some terms from email structures that aren't crucial for classification, like combination words and articles. A few typical pre-processing activities are keyword identification, tokenization, stop-word removal, stemming, and spell checking. To reduce the amount of data after pre-processing, a subset of features is selected during feature selection. With this strategy, a certain cost function is minimized. Feature selection, as opposed to feature extraction, does not change the data and is used to clean the data before a classifier model is trained. Another name for this procedure is variable selection, which is also known as feature reduction and variable subset selection [4].

Some of the most useful characteristics for email spam identification are the mail body and subject, word count, word size, circadian rhythms, recipient age, gender, and nation, recipient reacted (indicates whether the recipient responded to the message), mature content, and bag of words from the mail content. Spam emails frequently include many semantic anomalies. The framework for email categorization by domain's last step involves learning classification algorithms while utilizing the features that were selected in the stage before. The learned models are used to classify the new email documents (test data) into one of the predefined categories, such as Health, Education, Money, Adult, or Computing. Several methods are compared in the experimental section in classifying emails into different domains [5].

The Bayes classifier, often known as naive bayes, is one of the most frequently used statistical spam classifiers. It is referred regarded as the "naive" technique because it ignores any dependencies or correlations among the inputs and breaks down a multivariate problem into a series of univariate problems. Spam emails can be categorized using this technique. Probabilities are used as the main operational strategy for these classifiers. If specific terms are regularly found in spam but not in ham, then this incoming email is most likely spam. The use of this classification approach has become very common in mail filtering software [6]. It is necessary to receive good Bayesian filter training. In its database, every word has a predefined probability of turning up in spam or trash email. Similar to a finite tree, a decision tree has branches that represent tests and leaves that represent categories. Tests are frequently Boolean formulas that relate to the term weights in the document. By starting at the base of the tree and working up and down its branches, one can categorise a document by choosing conditions that are believed to be true. Once a leaf is reached, successive assessments place the document in the category that was used to annotate the leaf. The learning tree is computed using a number of modern techniques, including ID3, C4.5, and C5 [7].

The k-nearest neighbour (K-NN) is an example-based classifier. In other words, this system compares training documents rather than explicitly describing categories. Often, there is no training phase with this method. To classify a new document, the k most similar documents are searched. Unless another class has been assigned to the bulk of these documents, the new document is likewise included in this group. Moreover, this strategy may discover the closest neighbours more quickly than traditional indexing techniques. The class of the messages that are closest to a communication while determining whether it is spam or ham is taken into consideration. Real-time vector comparison is possible.

1.1 Email Semantic Features Extraction

This stage involves extracting the semantic features from email text. A group of obscure ideas that characterize an email's content is referred to as email semantics. The ultimate goal is to create a semantic representation for spam identification that is extremely accurate. An effective method for automatically extracting semantic information in this situation is CN2-SD [8]. The classification rule learner CN2 and the Subgroup Discovery are the two most often employed techniques for sematic feature extraction (SD). The class labels are predicted using CN2's induction of classification rules, and the training data are inspected for intriguing patterns using SD. Finding a subgroup is different from classifying something since finding a subgroup is a descriptive work, but classifying anything is a prediction activity. These two algorithms are described as follows:

- Subgroup discovery algorithm: The subgroup discovery algorithm's descriptive induction feature makes it possible to look for patterns that most closely match the data [9]. The semantic ideas in email communications are explained using this technique. Condensing and making understandable the features of a target population (domain) into a set of patterns is a vital function of data mining's semantic concept description. The SD is a data mining technique for figuring out connections between different things (like emails) and particular characteristics of a target variable (class). These relations are encoded using the form rules:

$$r : \text{cond} \rightarrow y$$

where cond is a combination of properties of the form, and y is the target variable (in our case spam or ham). The objective of SD is not to generate a global model. Instead, it makes it possible to spot particular patterns of interest and extract knowledge that can then be analysed and evaluated for descriptive purposes.

- CN2 rule induction algorithm: The CN2 algorithm is one of the conventional rule-based learning methods for producing propositional classification rules. The algorithm is made up of two fundamental parts: a low-level component and a high-level component. A low-level component, usually referred to as a search strategy, searches for a single rule that applies to numerous circumstances [10]. A high-level component, also referred to as a control procedure, repeatedly executes the lower level to enforce a set of rules. Many heuristic metrics are used in the literature to assess the quality of an induced rule at the low level. The two high-level control processes that the CN2 algorithm can employ are a technique for producing an ordered list of rules and a way for producing an unordered list of rules. The low-level part generates an ordered list of rules by using heuristic metrics to choose the best rule in the training set. During each iteration of the search procedure, the high-level section deletes all cases covered by the induced (learned) rule until all examples in the training data are covered [11]. In order to learn the rules for each class separately in an unordered set of rules, the control approach (high-level) is repeated. With each learned rule, just the covered examples that are part of the rule class are deleted rather than all covered examples as is the case for an ordered list. CN2 removes the circumstances that learnt rules cover in order to stop the same rule from being injected in further rounds.

1.2 Generation of Domain-specific Classifiers

For the purpose of developing a domain-specific classifier for each distinct domain, the collection of semantic features that were extracted in the preceding stage are used as learning attributes [12]. The classification of email messages is a supervised learning activity. It seeks to create a probabilistic model of a function for email classification. The supervised learning of text in email messages presents a learning algorithm with a set of pre-classified, or labelled, patterns, where a whole email dataset serves as one example of a message to be classified. This is referred to as the practise set. Certain classified messages from the training set are eliminated before creating a model to be used for testing its efficacy. This collection serves as the testing set. Several models are created utilising different partitioning of the instances into training and testing sets in order to evaluate the classification accuracy of the obtained model [13]. After then, the categorization error for all models is averaged. The number of divisions of the instance set, "n," is the number of times this procedure is performed. Several models are created through this cycle for analysis and repeated cross validations. Once developed, the model can be used to classify incoming emails.

II. LITERATURE REVIEW

N. Saidani, et.al (2020) emphasized on analyzing a text semantic for enhancing the accuracy to detect the spam [14]. A two semantic level analysis based technique was investigated for detecting the spam. Primarily, the particular domains such as healthcare, educational and commercial sectors, were utilized for classifying the emails so that a separate conceptual view was separated for spams in every domain. Subsequently, a set of manual and automatic semantic attributes was incorporated to detect the spam in every domain. These features assisted in summarizing the email content into compact topics to distinguish the spam from authentic emails efficiently. The results depicted that the investigated technique offered higher efficiency as compared to the traditional techniques and provided more interpretability in results.

G. Andresini, et.al (2022) developed a novel technique known as EUPHORIA for distinguishing amid spam authentic reviews [15]. In this, MVL (multi-view learning) was integrated with DL (deep learning) for attaining more accuracy with regard to different information related to the content of reviews and behavior of reviewers. Two datasets of Yelp.com – Hotel and Restaurant employed to conduct the experiments. The results validated that the developed technique assisted in enhancing the efficacy of DL (deep learning) algorithm to detect the spam in reviews. Moreover, this technique offered AUC-ROC around 0.813 on initial dataset and 0.708 on second dataset.

C. Kumar, et.al (2023) formulated a hybrid mechanism called SMOTE-ENN (Synthetic Minority Oversampling Technique-Edited Nearest Neighbor) for detecting the spam on Twitter [16]. Both the algorithms were put together for generating the balanced data. Different DL (deep learning) methods were presented which made the deployment of this data for recognizing the tweet as spam or genuine. Moreover, classifiers namely DT (Decision Tree), SVM (Support Vector Machine), LR (Logistic Regression) etc. were implemented. The simulation and comparative analysis was conducted to quantify the formulated mechanism with respect to different parameters. The formulated mechanism performed well and the RF algorithm yielded an accuracy of 99.26%, recall of 99.07% and precision of 99.49%.

X. Liu, et.al (2021) suggested a modified Transformer algorithm in order to detect SMS spam messages [17]. SMS Spam Collection v.1 dataset and UtkMl's dataset were applied to simulate the suggested algorithm against diverse ML (machine learning) algorithms. The experimental results reported that the suggested algorithm was more effective and yielded an accuracy of 98.92%, recall up to 94.51%, and F1-Score of 96.13%. Moreover, the suggested algorithm offered higher performance on second dataset that represented its adaptability for dealing with other similar issues as compared to other methods.

Z. Zhang, et.al (2020) focused on analyzing Twitter spam attributes as the user attribute, content, activity and association [18]. A new algorithm of detecting the spam was introduced on the basis of RELM (regularized extreme learning machine) recognized as I2FELM (Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine), for detecting the Twitter spam in accurate manner. The experimental results revealed the effectiveness of the introduced algorithm for recognizing the balanced and unbalanced dataset. Additionally, based on some characteristics, the introduced algorithm was capable of detecting the spam more successfully in contrast to the conventional methods.

G. Al-Rawashdeh, et.al (2019) devised a hybrid approach of WC (Water Cycle) and SA (Simulated Annealing) implemented for optimizing the results and to detecting the spam [19]. The groundwork, introduction, enhancement, estimation and comparison quality were comprised in this approach. The data was trained and tested using the cross-validation and the devised approach was computed on 7 datasets for classifying the spam. This work exploited meta-heuristic called WCFS (water cycle feature selection) and 3 schemes of hybridization with SA as a technique of selecting features. The experimental results confirmed that the devised approach attained an accuracy 96.3%. This approach assisted in diminishing the amount of attributes.

S. A. A. Ghaleb, et.al (2022) designed a wrapper technique on the basis of MOGOA (multi-objective grasshopper optimization algorithm) to improve the efficiency of SDS (spam detection system) [20]. Hence, the attributes were extracted. Moreover, recently revised EGOA algorithm was utilized to train MLP (multilayer perceptron). SpamBase, SpamAssassin, and UK-2011 datasets were applied to evaluate the designed technique. The simulation outcomes demonstrated the supremacy of the designed technique over other methods. In addition, the accuracy of the designed technique was measured 97.5% on first dataset, 98.3% on second, and 96.4% on last dataset.

D. Liu, et.al (2020) projected an innovative detection technique in which the viewpoint of users was considered and screenshots of malevolent webpages were captured for invalidating the Web spams [21]. CNN (Convolutional Neural Network), form of DNN (deep neural network) was implemented as a classifier. The projected technique was quantified in the experimentation. Initially, this technique was compared with the other ML (machine learning) based methods. Subsequently, the testing of the projected technique was done for detecting the malicious websites in a real-time Web environment. The experimental outcomes revealed the applicability of the projected technique to a practical Web environment in contrast to the traditional methods.

J. D. Rosita, et.al (2022) recommended MOGA–CNN–DLAS (Multi-Objective Genetic Algorithm and a CNN-based Deep Learning Architectural Scheme) method to detect the Twitter spam [22]. The MO (multi-objective optimization) procedure was integrated with selection, mutation, and cross-layer to assist in classifying the tweets as genuine and malevolent spam tweets. The experimental outcomes proved that the recommended method was more efficient to enhance the accuracy up to 0.17, precision around 0.13, recall of 0.10 and F-score of 0.19 and mitigate the RMSE around 19%, MAD of 16%, and MAE of 21%

X. Tong, et.al (2021) established a CapsNet (capsule network) model in which LSA (long-short attention) mechanism was adopted for attaining higher efficacy to detect Chinese spam [23]. The text was represented using a MCS (multi-channel structure) on the basis of LSA mechanism for capturing the complex text attributes in spam and generating the contextual word vectors with more semantic information. The attributes were mined and classified when this model helped in enhancing the structure of the classic CapsNet (capsule network) and optimizing the dynamic routing algorithm. Hence, the established model offered higher accuracy at higher running speed. Experimental results reported the superiority of the stablished model over the existing methods for classifying and detecting the spam at accuracy of 98.72% on an unbalanced dataset and 99.30% on a balanced dataset.

A. S. Mashaleh, et.al (2022) introduced a new method in which HHO (Harris Hawks optimizer) algorithm was combined with the KNN (K-Nearest Neighbor) algorithm for classifying the spam [24]. HHO algorithm was based on cooperative relations of Harris' Hawks. The introduced algorithm assisted in handling the data of higher dimensionality. Moreover, its accuracy was counted higher in comparison with the traditional methods. According to the experimental results, the introduced method yielded an accuracy of 94.3% for classifying and detecting the spam.

2.1 Comparison Table

Author	Year	Technique Used	Results	Limitations
N. Saidani, et.al	2020	A two semantic level analysis-based technique	The results depicted that the investigated technique offered higher efficiency as compared to the traditional techniques and provided more interpretability in results.	The major task was of maintaining the efficacy to filter the spam in the long run.
G. Andresini, et.al	2022	EUPHORIA	The results validated that the developed technique assisted in enhancing the efficacy of DL (deep learning) algorithm to detect the spam in reviews. Moreover, this technique offered AUC-ROC around 0.813 on initial dataset and 0.708 on second dataset.	This technique had not any online learning phase due to which it was incapable of periodically augmenting the trained classifier after the recording of new reviews over time.
C. Kumar, et.al	2023	A hybrid mechanism called SMOTE-ENN	The formulated mechanism performed well and the RF algorithm yielded an accuracy of 99.26%,	When the amount of spam tweets was maximized, the efficiency of the formulated mechanism was affected.

			recall of 99.07% and precision of 99.49%.	
X. Liu, et.al	2021	Modified Transformer algorithm	The experimental results revealed the effectiveness of the introduced algorithm for recognizing the balanced and unbalanced dataset. Additionally, based on some characteristics, the introduced algorithm was capable of detecting the spam more successfully in contrast to the conventional methods.	The utilized datasets had only thousands of messages which led to provide false prediction in diverse scenarios.
Z. Zhang, et.al	2020	I2FELM (Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine)	The experimental results revealed the effectiveness of the introduced algorithm for recognizing the balanced and unbalanced dataset.	This algorithm was ineffective to analyze the semantic and emotional data.
G. Al-Rawashdeh, et.al	2019	a hybrid approach of WC (Water Cycle) and SA (Simulated Annealing)	The experimental results confirmed that the devised approach attained an accuracy 96.3%. This approach assisted in diminishing the amount of attributes.	The devised approach was not applicable on all the applications.
S. A. A. Ghaleb, et.al	2022	wrapper method	The accuracy of the designed technique was measured 97.5% on first dataset, 98.3% on second, and 96.4% on last dataset.	This method was not useful to detect malevolent attacks namely phishing and botnets
D. Liu, et.al	2020	an innovative detection technique	The experimental outcomes revealed the applicability of the projected technique to a practical Web environment in contrast to the traditional methods.	This technique worked slowly and inflexible to large-scale detection.
J. D. Rosita, et.al	2022	MOGA–CNN–DLAS (Multi-Objective Genetic Algorithm and a CNN-based Deep Learning Architectural Scheme) method	The experimental outcomes proved that the recommended method was more efficient to enhance the accuracy up to 0.17, precision around 0.13, recall of 0.10 and F-score of 0.19 and mitigate the RMSE	The multi-objective optimization was not possible using the recommended method.

			around 19%, MAD of 16%, and MAE of 21%	
X. Tong, et.al	2021	CapsNet (capsule network) model	Experimental results reported the superiority of the stablished model over the existing methods for classifying and detecting the spam at accuracy of 98.72% on an unbalanced dataset and 99.30% on a balanced dataset.	The employed dataset was relatively old and ineffective of reflecting the attributes of the latest spam.
A. S. Mashaleh, et.al	2022	A new method	According to the experimental results, the introduced method yielded an accuracy of 94.3% for classifying and detecting the spam.	Some of its metrics were not optimized due to which the performance was found poor.

CONCLUSION

A surge in the number of spammers and spam emails has been noticed in recent years, as the investment required for the spamming business is minimum. This has led to a system that finds each email suspicious, causing substantial investments in defence mechanisms. The most commonly used mail filtering schemes are Knowledge Engineering (KE) and Machine Learning (ML). The approaches based on KE generate a set of rules so as to classify messages as spam or genuine mail. The email spam detection has various phases like feature extraction and classification. The various schemes are analyzed in this paper for the email spam detection. It is analyzed that the machine learning algorithms are best performing algorithms as compared content filtering techniques.

REFERENCES

[1] K. Debnath and N. Kar, "Email Spam Detection using Deep Learning Approach," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 37-41

[2] S. Suryawanshi, A. Goswami and P. Patil, "Email Spam Detection: An Empirical Comparative Study of Different ML and Ensemble Classifiers," 2019 IEEE 9th International Conference on Advanced Computing (IACC), Tiruchirappalli, India, 2019, pp. 69-74,

[3] N. A. Farahisya and F. A. Bachtiar, "Spam Email Detection with Affect Intensities using Recurrent Neural Network Algorithm," 2022 2nd International Conference on Information Technology and Education (ICIT&E), Malang, Indonesia, 2022, pp. 206-211

[4] P. Thakur, K. Joshi, P. Thakral and S. Jain, "Detection of Email Spam using Machine Learning Algorithms: A Comparative Study," 2022 8th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2022, pp. 349-352,

[5] S. Nandhini and J. Marseline K.S., "Performance Evaluation of Machine Learning Algorithms for Email Spam Detection," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-4

[6] R. Amin, M. M. Rahman and N. Hossain, "A Bangla Spam Email Detection and Datasets Creation Approach based on Machine Learning Algorithms," 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 2019, pp. 169-172

[7] S. Shrivastava and R. Anju, "Spam mail detection through data mining techniques," 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2017, pp. 61-64

[8] W. Peng, L. Huang, J. Jia and E. Ingram, "Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 2018, pp. 849-854

[9] S. E. Rahman and S. Ullah, "Email Spam Detection using Bidirectional Long Short Term Memory with Convolutional Neural Network," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 1307-1311,

[10] R. P. Cota and D. Zinca, "Comparative Results of Spam Email Detection Using Machine Learning Algorithms," 2022 14th International Conference on Communications (COMM), Bucharest, Romania, 2022, pp. 1-5

[11] N. Nisar, N. Rakesh and M. Chhabra, "Voting-Ensemble Classification for Email Spam Detection," 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2021, pp. 1-6

[12] V. Vishagini and A. K. Rajan, "An Improved Spam Detection Method with Weighted Support Vector Machine," 2018 International Conference on Data Science and Engineering (ICDSE), Kochi, India, 2018, pp. 1-5

[13] T. Toma, S. Hassan and M. Arifuzzaman, "An Analysis of Supervised Machine Learning Algorithms for Spam Email Detection," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp. 1-5

[14] N. Saidani, K. Adi and M. S. Allili, “A semantic-based classification approach for an enhanced spam detection”, Computers & Security, vol. 11, no. 2, pp. 6594-6609, 9 January 2020

[15] G. Andresini, A. Iovine and A. Appice, “EUPHORIA: A neural multi-view approach to combine content and behavioral features in review spam detection”, Journal of Computational Mathematics and Data Science, vol. 7, no. 4, pp. 170003-170011, 22 April 2022

[16] C. Kumar, T. S. Bharti and S. Prakash, “A hybrid Data-Driven framework for Spam detection in Online Social Network”, Procedia Computer Science, vol. 218, pp. 124-132, 31 January 2023

[17] X. Liu, H. Lu and A. Nayak, "A Spam Transformer Model for SMS Spam Detection," in IEEE Access, vol. 9, pp. 80253-80263, 2021

[18] Z. Zhang, R. Hou and J. Yang, "Detection of Social Network Spam Based on Improved Extreme Learning Machine," in IEEE Access, vol. 8, pp. 112003-112014, 2020

[19] G. Al-Rawashdeh, R. Mamat and N. Hafhizah Binti Abd Rahim, "Hybrid Water Cycle Optimization Algorithm With Simulated Annealing for Spam E-mail Detection," in IEEE Access, vol. 7, pp. 143721-143734, 2019

[20] S. A. A. Ghaleb et al., "Feature Selection by Multiobjective Optimization: Application to Spam Detection System by Neural Networks and Grasshopper Optimization Algorithm," in IEEE Access, vol. 10, pp. 98475-98489, 2022

[21] D. Liu and J. -H. Lee, "CNN Based Malicious Website Detection by Invalidating Multiple Web Spams," in IEEE Access, vol. 8, pp. 97258-97266, 2020

[22] J. D. Rosita P and W. S. Jacob, “Multi-Objective Genetic Algorithm and CNN-Based Deep Learning Architectural Scheme for effective spam detection”, International Journal of Intelligent Networks, vol. 10, no. 2, pp. 5207-5222, 2 February 2022

[23] X. Tong et al., "A Content-Based Chinese Spam Detection Method Using a Capsule Network With Long-Short Attention," in IEEE Sensors Journal, vol. 21, no. 22, pp. 25409-25420, 15 Nov.15, 2021

[24] A. S. Mashaleh, N. F. B. Ibrahim and Q. M. Yaseen, “Detecting Spam Email with Machine Learning Optimized with Harris Hawks optimizer (HHO) Algorithm”, Procedia Computer Science, vol. 201, pp. 659-664, 27 April 2022