Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 260 (2025) 608–615

Procedia Computer Science

www.elsevier.com/locate/procedia

Seventh International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R-2025)

The Impact of SMS Phishing using Machine Learning Classifiers with Innovative Techniques

Anisha Asirvatham[1*],C.Meenakshi[2]

[1]Research Scholar, *Department of Computer Science, Vels Institute of Science Technology and Advanced Studies Pallavaram, Chennai,India*

[2]*Associate Professor, Department of Computer Science, Vels Institute of Science Technology and Advanced Studies Pallavaram, Chennai,India*

Abstract

These days, machine learning has an amazing ability to present the results efficiently. Network technology computing systems have long served as a critical framework to supply machine learning with deterministic computing resources. This technique can be advantageous to networking. The application of machine learning in detection of phishing of SMS in a network is the main topic of this article. This not only stimulates new network applications but also aids in the resolution of certain obsolete network issues. The fundamental approach to describe how machine learning technology works in network model is summarized in this article. The SMS spam can be detected using the proposed system, where the feature extraction can be done using stemming and the different machine learning classifiers are used to evaluate the accuracy of the proposed system. The different machine learning classifiers used in this system are Logistic regression, linear regression, K Nearest Neighbors and Naïve Bayes. The new system SpamSMS is the ensemble method using Naïve Bayes and Stacking Algorithm which gives the maximum accuracy towards the provided dataset to check for the spam messages.

Keywords: *Network Intrusion Detection; Neural Network; Artificial Intelligence; Computer Network; Machine learning classifiers.*

* Corresponding author.

E-mail address: anisha184u@gmail.com

10.1016/j.procs.2025.03.239

1. Introduction

As the purview of computer network technology continues to expand, various destructive attacks existing in the Internet field have caused serious damage to computer users and network resources. This study investigates the use of machine learning techniques in networked computers and makes an attempt to apply them there. This creates an enhanced Back Propagation neural network-based intrusion detection algorithm. Simulation scenarios can be used to discover system deceiving alarms by examining the various offensive concepts, examining the features of intervention techniques, collecting feature data, creating feature sets, and utilizing agent technology as an assistive technology. It also helps to prove the improvement effect rate, convergence velocity, and false - Negative rate where the rate reached 86.7%. From the research, we can conclude that this fast algorithm shortens network time limit, reduces network size, boosts classification performance, and improves the rate of detection in data breach.

ML, or machine learning, that solve issues and enable automation across a variety of fields have seen an unusual increase in applications. A spike in data availability, significant advancements in machine learning techniques, and breakthroughs in computing resources are the main causes of this. In fact, ML has been used to a wide range of straightforward and complex problems related to network management and operation. There are a number of surveys on ML for particular networking domains or network technologies. This survey is unique because it combines the deployment of multiple ML approaches in several important networking areas across various network technologies. An in-depth review of the numerous learning paradigms and machine learning (ML) approaches that are applied to basic Network problems, such as traffic forecasts, route planning and Classification, congestion control, resources and faults management, QoS and QoE management and network security will be beneficial to readers. The study also outlines the constraints, offers insights, identifies research difficulties, and points to probable future directions for ML in networking. This crucial impact on the networking implications of ML pushes ahead of the curve for automated network operation and management.

The number of technology with internet connections is growing daily owing to the internet of Things. This causes a significant volume of data to be shared online. It is one of the primary issues that may be distinguished by a careful data analysis between normal and aberrant traffic. This research uses machine learning techniques to conduct an analysis to identify whether the information obtained online was conventional or uncharacteristic. In order to accomplish this, the Naive Bayes (NB), Random Forest (RF), Multilayer Perception (MLP), and Sequential Minimal Optimization (SMO) algorithms are utilized to classify the KDD Cup 99 data set, which is often used in literature studies.

Bayesian learning and Support Vector Machine (SVM) based machine learning techniques were used to report the email spam filtering. As an exploratory step, a mobile application SMSAssassin was developed that can filter SMS spam messages based on the above methods to find the type of message and to blacklist the message. It uses crowd sourcing techniques to keep itself updated. [1]. The study of mathematical model-based algorithms that get better on their own via training is known as machine learning. Moreover, ML algorithms are capable of identifying network issues and recommending solutions. Thus, specialists in networks can decide how best to handle the network. Attackers now launch new attacks each day, making it difficult to spot them using conventional intrusion detection systems. Recent studies have also demonstrated how harmful ML is. Attackers and wretched users can potentially degrade machine learning systems by altering the training data and classification function of the algorithm, directly affecting a network's detection accuracy. Threats of this nature are extremely serious. To safeguard a network, new cyber security techniques must be created.

This article uses a BF neural network to analyze intrusion detection in computer networks because of its fast learning rate and well-defined architecture. This work uses the LM algorithm, batch processing technology, and variable learning rate algorithm to improve the BF approach. Less training data are used by this method. Theoretical analysis is done on how the improved BF algorithm and the traditional BF algorithm work and are used. The experimental results were examined using the evaluation parameters in accordance with the experimental data. When compared to the findings of traditional BF neural network detection, the improved technique's detection rate of 86.7%

indicates that the application of neural networks based on rapid training methods to the intrusion detection field has yielded positive results. [2]

2. Literature Review

As discussed about the Arabic region, [3] where it recognizes the features or factors related to their demographics which boost the likelihood that individuals might reach out to report spam or phishing attacks calls. The machine learning model CatBoost performs better than other models. This model forecasted the manner in which a person would report a phishing or spam threat. In the future, it may include expanding the approach to handle all types of call center complaints. [4].This article gives a quick rundown of the key machine learning tasks and highlights some of the most popular ML algorithms, including those for supervised, unsupervised, and reinforcement learning. The use of ML algorithms in various disciplines, including computer networking, has profited tremendously in recent years from cloud computing platforms and ready-to-use machine learning software packages. [5]. A malware detection and classification method is suggested in this paper based on the best encoder-decoder-driven LSTM networks. This suggested method contains numerous stages. The pre-processing for data normalisation is done in the first phase. The testing findings show that malware and benign classes may be classified with 97.14% and 98.33% accuracy, respectively, which is superior than the existing systems. A network traffic detection paradigm based on Modbus is suggested in order to improve the information contained in packets during feature extraction. The outcomes of the experiments and their explanation demonstrate the efficacy of this detection model. [6]

High accuracy and low latency spam SMS were the primary areas of interest for investigation. To reduce the amount of noise, the data was cleaned and pre-processed. Moreover, the distinctive characteristics were extracted using the TF-IDF and Bag of Words models. To pick the extracted features with the minimal probable latency, the Chi-Square feature selection method was used. Next, the machine learning models Bernoulli Naive Bayes, LightGBM, and XGBoost was utilized in addition to the conventional SVM and Random Forest basic models. The investigation's goals succeeded, and the spam messages were successfully scrutinized and the spam letters were effectively screened. The goals of the research were met, and the spam SMS were quickly and accurately screened. So, it is certainly feasible to declare the research successful. It yielded accuracy results of 95.4% and 96.5%, respectively. Also, the time taken by these models was 0.157 and 1.708 seconds which was significantly better than the other traditional models. [7]

SMS spam sieving, mitigation and other spam detection techniques as well as their limitations and future research directions are illustrated with other optimization techniques. A variety of SMS Spam techniques, with different datasets and comparisons are addressed in the given paper. [8] One of the most widely utilized and quickly expanding GSM value-added services globally is SMS which is named as Spam SMS [9]

The CHURN SMS abuse attribution mechanism was launched. With the use of this method, extensive SMS abuse campaigns' passive DNS records and auxiliary website attributes may be gathered and analysed [10]. They used CHURN to meticulously carry out attribution around the domain names and IP addresses used in such SMS spam campaigns across a five-year period. In this paper, Support Vector Machine was introduced to get the highest spam detected with the accuracy rate of 98.9% which was calculated using incredible accuracy, reduced processing times, increased kappa statistics, decreased error, and the lowest number of false positives. [11]

The system will initially look through SMS messages using an SMS spam detection mechanism at the SMS service provider level and call log databases to determine whether there is a direct or mutual relationship between the sender and recipient. If the message content is found to be spam, it will be handled as these. By leveraging this technique, issues like balance deduction and SMS memory usage that result from spam messages are resolved. [12]

It harnesses the accuracy to estimate the efficacy of classifiers in order to assess our constructed classification and compare it to existing methods. For SMS spam identification, the experiment was run using a variety of classifiers, including decision trees, KNN classifiers, and Naïve Bayes. Among the classifiers, the Naïve Bayes model had the best accuracy.[13]. The pre-processed messages are used in 70%-30% train and test split method with deep learning models using simpler architectures like Convolution Neural Network and a hybrid Convolution Neural Network along with Long Short-Term Memory Network for classification. The accuracy of this architecture BUNOW and GloVe word embedding techniques are incorporated with deep learning models. The optimal accuracy of 98.44% is achieved by the CNN LSTM BUNOW model on a 70% - 30% train-test split. [14]. This survey paper deals with the research analysis on the mobile spam happening due to SMS received from unwanted contacts and the messages are also unwanted. The study compares different machine learning algorithms like Naïve Bayes algorithm, Decision tree, Support vector machine, Logistic regression and random forest algorithms. [15]. The model used by this paper is the ensemble model grouping 4 into one method to give the maximum result of 99.91% [17].

3. Proposed Work

We have done a small-scale research to know the frequency of the unknown SMS received by the mobile users, where most of the messages are related to check for the status of the investment, update the account details, update your KYC, etc. which shows to the users as the genuine messages from the banks. And once the link is clicked the user's confidential financial bank details are asked like the credit card or debit card details, expiry date and name of the card holder, and CVV (Card Verification Value), number. It just replicates the normal banking processes where once the users card details are added, the OTP (onetime password) is sent to the registered mobile number of the user and the whole transaction appears so genuine, that the mobile users share the confidential OTP also to the malicious contact and the consecutive transaction happens in few minutes, leading to the bankrupt status of the account.

This leads to a novel idea to the users where, the users can be made aware of these contact details, once the SMS is delivered, so that the users can be protected from such scam. The proposed work SpamSMS shows the different steps of evaluation of the SMS received with different steps as illustrated in Fig. 1.

1. The SMS is first scrutinized with the text pre-processing methods like stemming which reduces the word to its source word which generates the keyword to create a spam message.
2. The presence of the keyword pattern in SMS is checked.
3. The SMS is checked for the sender's information with the bank authorized code.
4. When the source is known, the received message is ham message, otherwise the message is checked for the misspelled words, leet words, symbols and special characters.
5. Furthermore, the type of SMS is checked with the content of SMS using classification models to check for the spam or ham messages.

The given Fig.1 shows the flow of the analysis and the processes carried out to get the optimal results. The different ways performed are elaborated as follows:

3.1 Text Pre-Processing Phase: Text pre-processing is the phase where the user's input is tokenized in the machine-readable format (integers). The Stemming process helps to tokenize the collection of documents, creates its own vocabulary and then helps to find the base word which creates the spam messages.

3.2 Checking for SMS source pattern: The details of the sender of the message is taken, where, if the message is from the banking side, there is a special bank code which shows that the messages are from authorized bank and not from any unauthorized number to fool the people around.

3.3 Check for the authenticate contact number: The next step checks for the valid contact number, from where the spam message is delivered to the user. The proposed model tries to find the pattern of the contact number as genuine or not. If, the pattern is genuine and among the contact list, then the process id terminated. Otherwise, the model checks for the type of the contact number and informs

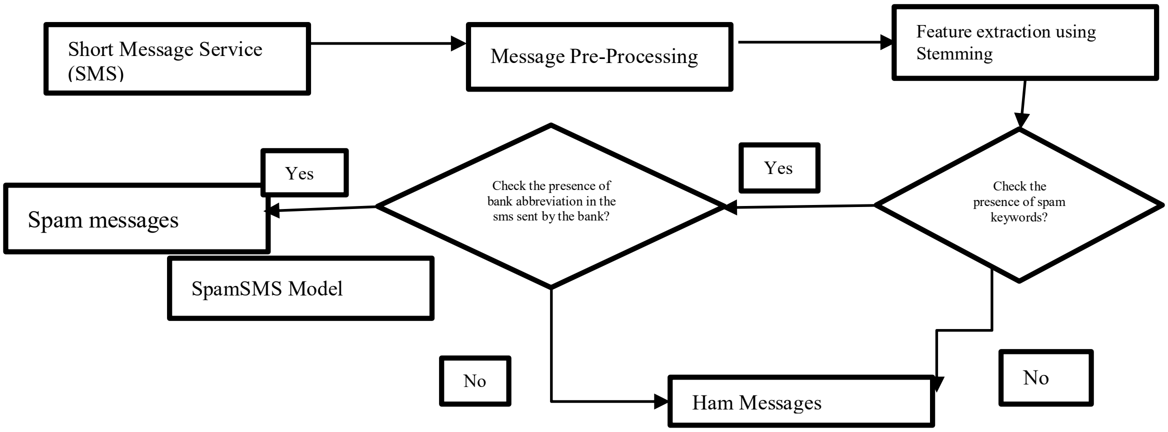
Short Message Service (SMS)
Message Pre-Processing
Feature extraction using Stemming
Check the presence of spam keywords?
Yes
Check the presence of bank abbreviation in the sms sent by the bank?
Yes
Spam messages
SpamSMS Model
No
Ham Messages
No

Fig. 1. Block Diagram of Proposed Work.

3.4 Leet words: Leet words are the words which are spelled incorrectly. For example, Sorry can be written as Sry, Srry. The words can also be entered with a combination of alphabets and numbers. For example, hac0R which can be pronounced as hacker.

3.5 Keywords: There are few keywords like Congratulations, Winner, Offer, etc. These words are included in the SMS, so that it prompts the users to click on the message to read and follow the steps. In banking messages also, promotional discounts and offers for loans are used to make the users.

This paper has proposed a new idea to reduce the SMS phishing model as SpamSMS to detect the spam message. The foremost objective of this study is to segregate the messages as spam and ham messages and then to also know the source of the message with the bank code.

4. Discussion and Result Evaluation

4.1 Data Collection: As per the Table 1, there are total 3500 SMS analysed for Spam and Legitimate messages. There were 280 spam messages and 1720 legitimate messages. With reference to the table data, the data was collected from different age group of mobile users. [16]. Fig. 2 shows the 14% of spam messages as negative messages and 86% of ham messages as positive messages.

Table 1: Text messages (Spam and Ham)

Total SMS	**Spam Message**	**Legitimate SMS**
3500	490	3010

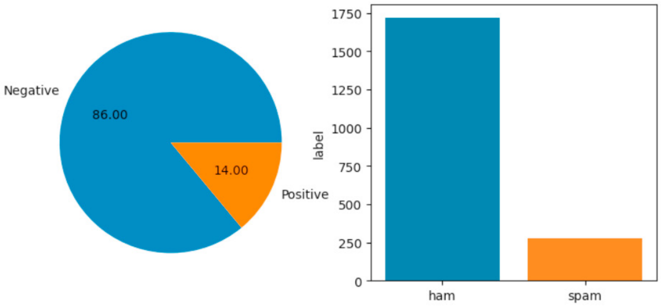

Fig. 2. Flowchart of the Proposed System.

4.2 Evaluation of Metrics: Accuracy, recall, precision, and F1 scores are utilized to evaluate the recommended approach and the algorithm used for the comparison study. These metrics and evaluated in percentage values. The Naïve Bayes algorithm uses the below formula to classify the data.

- Accuracy: The fraction of True Positive and True Negative over the total number of messages is applied to calculate the accuracy values as given by the formula below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision: Precision is calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

- Recall: Recall is calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \quad (3)$$

- F1-Score: F1-Score is the Harmonic mean of Precision and Recall.

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

- Bayes Theorem = P(A|B) = $\frac{P(B|A)P(A)}{P(B)}$, where A and B are events and P(B)≠0 (5)

4.3 Results

The feature extraction with text pre-processing is performed using stemming to get the action verbs to know the source words also, it uses the voting classifier to get the accuracy value by ensemble method of Naïve Bayes and stacking Classifiers. It also checks for the banking key words of the senders. The result shows the strong correlation between Character count and Word Count with 0.97, followed by Moderate correlations between sentence count and character count as between 0.64 and 0.68 and weak correlations between 0.26 to 0.38 which is shown in Fig.5 in the form of heatmap. The graph with spam and ham character count is shown in Fig. 3 and Fig. 4 shows the word count of spam and ham messages. Fig.6 shows the 320 times occurrence of most used spam word as "Call" and other keywords also in the graph.

To establish the significance of features employed in our system, the frequency of each characteristic in our dataset is determined. The 490 spam messages in our dataset's attributes are used to generate the frequency. The feature is tested by using python code to extract the features like mis-spelled words, Leet words and keywords from the results generated, the misspelled words feature is most used for detecting spam SMS. It exists 39.16% of misspelled words, 35.67% of leet words, 22.98% of keywords. Special characters fetch the least features among all the others. Table 2 shows the comparative study with different algorithm used in the spam detection.

When categorizing messages using the suggested SpamSMS Algorithm, our system's stated standards mentioned above yield the best results in terms of accuracy. Every algorithm's forecasting outcomes are derived from the recorded feature values. Each algorithm's predicted result is displayed in Table 2. The evaluation shows that the Logistic Regression algorithm gives 98.23% accuracy, The Naive Bayes algorithm (NB) also evaluated a decent performance with an accuracy of 96%, linear regression with 93.48% and KNN with 88.10%. Hence the best outcome is retrieved using SpamSMS methodology to achieve the maximum outcome.

Table 2: Performance of different Algorithm

Classifier	Precision	Recall	F1-scores	Accuracy
SpamSMS	0.97	0.99	0.96	98.97
Logistic regression	0.97	0.98	0.96	98.23
Naïve Bayes Classifier	0.95	0.99	0.97	96.00
linear regression	0.92	0.95	0.93	93.48
K Nearest Neighbors	0.87	0.89	0.92	88.10

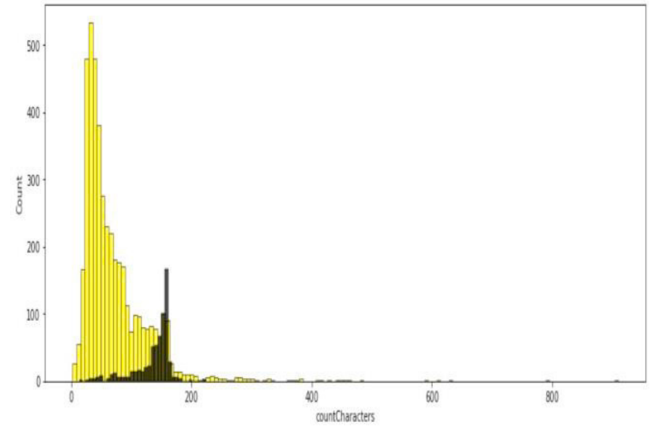

Fig. 3. Histogram showing Character Count with the Spam(Black color) and Non-Spam(Yellow color)

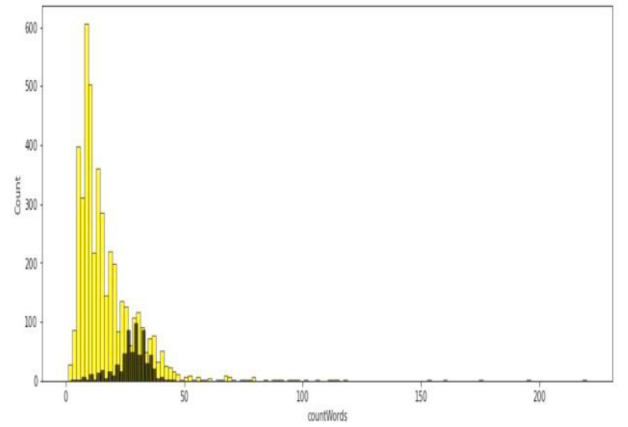

Fig. 4. Histogram showing the Word Count with Spam(Black color) and Non-Spam(Yellow color)

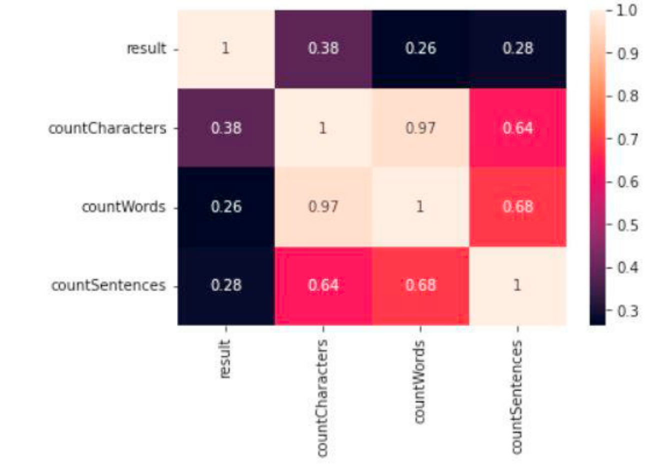

Fig. 5. Heat map to visualize the dataset with character count, word count and sentence count.

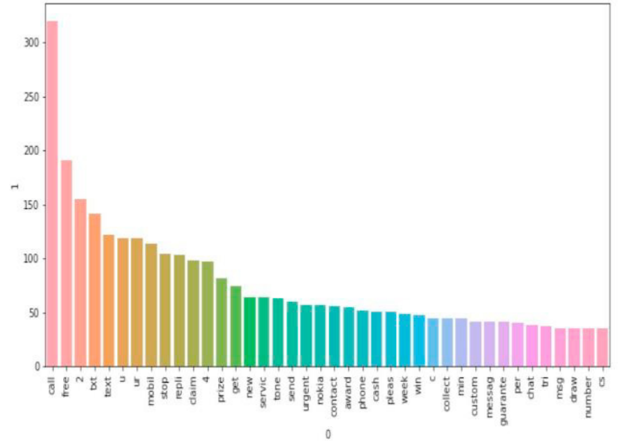

Fig. 6. Graph of most occurrence of spam words

The proposed system not only finds the SMS received in mobile user as spam SMS or legitimate SMS, but it also checks the source of SMS received by the client. Contents in the SMS like the URL, contact numbers as the spam or not is verified. Furthermore, the source of the message received is detected using the bank code to check for the spam SMS received by the common user.

5. Conclusion

Phishing is the more prominent fraud which is on-going in the current world leading to the financial losses to individual people and businesses too. The findings reveal that the SMS is basically classified into 2 different types like, legitimate and spam messages. This paper totally focused on the development of SpamSMS system. The evaluation and prediction is done with the 3200 dataset from authentic dataset sources and google forms, etc. This system had two phases of evaluation. The first phase is to check the type of the SMS received, where the contents of the SMS were checked for features around 18 in numbers, wherein few were URL, phone number, leet words, etc. The second phase is the checking the source of the bank with the bank code. There are many blacklisted numbers which are detected as spam, but these blacklisted numbers are not available in public. So, the database of the blacklisted numbers are listed in the system and the further evaluation of the system is carried out. The SMS dataset is scrutinized with algorithms like Naive Bayes algorithm, Linear regression, Logistic regression, K nearest neighbor, etc. Among the different algorithms evaluated, the SpamSMS model gave the accuracy of 98.97% of spam messages received in the mobile users. Natural Language processing techniques can also be used with its advanced procedures.

References

[1] K. Yadav, "SMSAssassin: Crowdsourcing Driven Mobile-based System for SMS Spam Filtering," 2015.

[2] Ahmad, Bilal et al., "Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions," Journal of Computer Networks and Communication, 2018.

[3] S. Al-Hussaini, "Are They Likely to Complain on Phish or Spam? A Prediction Model," 2020 7th International Conference on Behavioural and Social Computing (BESC), 2020.

[4] G. Alexander and S. Rieger, "A survey of Machine Learning applied to Computer Networks," in European Symposium on Artificial Neural Networks, Computational Intelligence, 2020.

[5] Alzubi, Omar et al., "Quantum Mayfly Optimization with Encoder-Decoder Driven LSTM Networks for Malware Detection and Classification Model," Springer, 2023.

[6] Namasudra,NamaSuyel et al., "The New Era of Computer Network by using Machine Learning," Springer, 2023.

[7] M. C. Ora Anchal, "Spam Detection in Short Message Service Using Natural Language Processing and Machine Learning Techniques," 2019.

[8] Muhammad Abdulhamid, Shafi'i,et.al, "A Review on Mobile SMS Spam Filtering Techniques," IEEE Access, 2017.

[9] O. S. Abayomi-Alli, "A deep learning method for automatic SMS spam classification: Performance of learning algorithms on indigenous dataset," 2022.

[10] Srinivasan, Bharat et al., "Understanding Cross-Channel Abuse with SMS-Spam Support Infrastructure Attribution".

[11] Amir Sjarif a*, Nilam Nur et al., "Support Vector Machine Algorithm for SMS Spam Classification in The Telecommunication Industry," International Journal on Advanced Science Engineering Information Technology, 2020.

[12] B. Alam and F. Zama, "Design SMS Spamming Detection System," International Journal For Engineering Applications and Technology, 2015.

[13] S. Selvapattu and P. Patil, "SMS Spam Detection," International Journal of Innovative Science, Engineering & Technology, 2020.

[14] Giri,Surajit et al., "SMS Spam Classification–Simple Deep Learning Models With Higher Accuracy Using BUNOW And GloVe Word Embedding," Journal of Applied Science and Engineering, pp. 1501-1511, 2022.

[15] S. Nagre, "Mobile SMS Spam Detection using Machine Learning Techniques," Journal of Emerging Technologies and Innovative Research, 2018.

[16] T. Almeida and J. Hidalgo, "SMS Spam Collection," UCI Machine Learning Repository.

[17] G. Abdallah and A.Manar, "Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning",Sensors,MDPI,2023, https://doi.org/10.3390/s23083861