LAUTECH Journal of Computing and Informatics (LAUJCI) – ISSN: 2714-4194
Volume 3 Issue 1, June 2023 – www.laujci.lautech.edu.ng

SPAM EMAIL DETECTION SCHEME BASED ON RANDOM FOREST ALGORITHM

Oyelakin A.M
Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria
amoyelakin@alhikim.edu.ng

Salau Ibrahim T.T
Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria
ttsalau@alhikim.edu.ng

Ogidan B.S
Adjunct Lecturer, Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria.
bolaji.gidan@alhikim.edu.ng

Olufadi H.I
Part Time Lecturer, Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria

Yusuf S.A
Part Time Lecturer, Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria
yusufsuliyat@gmail.com

Adeinji I.A
Lecturer, Center for Part Time and Professional Studies, Al-Hikmah University, Ilorin, Nigeria
herdeyni@gmail.com

ABSTRACT
Emails are used for communication purposes in different sectors of the economy such as education, health, businesses, manufacturing, agriculture. People with malicious intent have been using emails accounts for different spam email attacks. Spam email refers to as unsolicited bulk email. It is the practice of sending large frequent, unwanted e-mail messages with commercial content to indiscriminate set of recipients. Spam emails expose users to challenges such as time wastage, high usage of computing resources and stealing of valuable information. Machine learning approaches have been widely accepted to be better than traditional approaches for the identification of spam emails. For this reason, several machine learning techniques have been proposed in the literature for the classification of spams in emails. This paper proposed a Random Forest-based scheme for email spam detection. A fairly large spam email dataset named spam base was collected from UCI machine learning repository. The dataset was pre-processed based on the feature encoding. Then, promising features were selected using feature importance technique. The feature selection yielded 12-feature subsets that were arrived at based on the feature scores. The Random Forest (RF) spam email detection model that was built achieved 99.65% Accuracy, 99.21% Precision, 99.46% of Recall and F1-score of 99.33%. The study concluded that the RF-based spam email detection model performed better than some of the approaches in similar studies

Keywords: Email, Spam Email Attacks, Detection Accuracy, Ensemble Algorithm

1. INTRODUCTION

Spam attacks are of various types. For instance, we have SMS spam attacks, social media spam attacks and spam attacks in the email platforms. Aside social media spam attacks such as twitter spam attacks (Rao, Verma & Bhatia, 2010; Ameen, Oyelakin, Ajiboye, Olatinwo, Obiwusi & Ogundele, 2022) that have been very popular in the internet space, email spam attacks are also very prevalent. One of the reasons for the prevalence of email spam threats is because of the fact that emails are used for communication purposes in different sectors of the economy. Spam email is a type of email that can be used to harm any user by wasting time, computing resources, and stealing valuable information (Ahmed, Amin, Aldabbas, Koundal, Alouffi & Shah, 2021). The general use of emails for communication can be traceable to the large penetration of the internet. The Internet has become a part of daily life and has been found to be indispensable in all sectors of the economy (Chih-Fong, Yu-Feng, Chia-Ying & Wei-Yang, 2009). This is because it aids people in every aspect of life. Spam email is dangerous due to the fact that it include malicious links that can infect the computer with malware. Machine learning approaches have been widely accepted to be better than conventional detection approaches in the identification of spam emails (Shaukat, Luo, Chen & Liu, 2020). Thus, several machine learning techniques have been proposed in the literature for the classification of spam emails (Kashapov, Wu, Abuadbba, & Rudolph, 2022; Sethi, Chandra, Chaudhary, & Dahiya, 2022). Alpaydın (2010) explained how machine learning algorithms can be used to build predictive models from training data. Alpaydin (2010) further argued that stored data becomes useful only when it is

analysed, pre-processed and transformed into useful information that can be used in machine learning models. Emails are common forms of communication in the internet space globally (Hariharan, Kamaraj, Ramanuja, 2021). Due to the popularity of emails as forms of communication, one of the techniques used by cyber attackers to launch phishing based attacks is the use of spam emails. People with malicious intent use links or the attachments in the unsolicited email to release malware in the system and related devices on innocent cyber users.

The use of machine learning approach for spam email attack identification is investigated in this study. The need for this approach is based on the argument of Oyelakin, Salau-Ibrahim, Ogidan, Azeez. and Ajiboye (2019) who identified that there is a particular kind of security threat named botnet in the internet space that is currently evading existing threat detection techniques. Some other internet attcaks/threats like spam attacks, social media spam attacks are also very powerful and can evade signature-based detection techniques. Thus, machine learning detection approaches for malware or attacks are getting popular each day because of the fact that they are promising for detecting unknown attacks. This study used an ensemble machine learning algorithm named Random Forest to build spam email detection scheme that can detect spam emails more efficiently based on the promising results in the chosen metrics.

2. RELATED STUDIES

Jazzar, Yousef and Eleyan (2021) performed evaluation of some machine learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), J48, and Naïve Bayes for email spam classification in Urdu language. In conclusion, support vector machine performed better than any individual algorithm in term of accuracy. The authors argued that the LSTM model outperforms other models with a highest score of 98.4% accuracy. Ahmed, Amin, Aldabbas, Koundal, Alouffi and Shah (2021) surveyed the machine learning techniques used for spam filtering techniques in email and IoT platforms. The study emphasised that the machine learning techniques were used for classifying them into suitable categories. Thereafter, a comprehensive comparison of the machine learning-based techniques using some identified performance metrics such as accuracy, precision, recall was carried out. It was argued that the study provided comprehensive insights and future research directions were also discussed. However, no experimentation was carried out in respect of email spam detection. Furthermore, Kothapally and Vijayalakshmi (2021) classified spam messages by making use of Random Forest Algorithm. The study only focused on short messaging platforms as against email messages being considered in this study. Also, Gaikwad and Halkarnikar (2013) built a spam email detection model with the use of Random Forest Algorithm. The authors argued that the study produced excellent results based on the chosen metrics. Nandini and Jeen (2020) built machine learning based models for the classification of spam emails. The authors used a popular UCI Machine Learning Repository spam dataset called Spambase. The performance of five selected machine learning classification algorithms namely Logistic Regression, Decision Tree (DT), Naïve Bayes, K-Nearest Neighbor (KNN) and SVM were evaluated using some metrics. Weka tool was used for training and testing the data set. It was reported that Decision Tree and KNN algorithms have the best overall performances. Unfortunately, the study did not report whether the model built was computationally costly or not. Hariharan et al. (2021) proposed a Deep Learning approach for the classification of Email spam evidence. Authors argued that their focus was to find an effective solution to filter possible spam e-mails. The authors used a hybrid solution that combines Deep Neural Network and Convolution Neural Network algorithms to produce an improved result and efficiency compared to existing system. The experimental results showed that the proposed algorithm has 92.8% accuracy. However, authors were silent on the computational complexities that may set in as a result of the hybridisation of deep learning methods in the spam email detection. However, the approach was considered very expensive computationally. Dar, Luo, Chen and Liu Dongxi (2020) used three selected machine learning techniques namely deep belief network (DBN), DT and SVM for building threat detection models. Authors reported the investigation of the performances of the learning algorithms in the spam detection, intrusion detection and malware detection using different benchmark intrusion datasets. For the spam detection, the authors argued that Decision Tree has the best overall performance in the selected metrics. Sharma and Bhardawaj (2018) proposed a machine learning based hybrid approach by combining Naïve Bayes and J48 DT based algorithm for the spam email detection. In this process, dataset was divided into different sets and given as input to each algorithm. Three experiments were carried out in which the first two experiments involved the use of Naive Bayes and J48 algorithm separately while the third experiment was hybrid. Authors concluded that support vector machine performs better than any individual algorithm in term of accuracy.

Based on the gap identified in some of the existing but relevant studies, this study sets out to investigate the performance of the selected ensemble learning algorithm in the classification of spam email evidence.

3. METHODOLOGY

The methodology adopted in this study is machine learning-based. The processes involved in building the machine learning-based spam email detector is: Dataset Collection, Dataset Exploration, Pre-processing, Feature Selection and Spam Email Classification. All the experimentations are carried out in Anaconda Python 3.72 environment. In the experimentations, the selected algorithm was used for the training and testing purposes. The dataset was split into 85 percent training set and 15% test set respectively. The results obtained in the split of 85 to 15 were found to be very promising across all the metrics used. Homogeneous Random Forest Algorithm is chosen being an ensemble of Decision Tree algorithms and its parameters are carefully varied and selected during the result validation stage.

Email Spam Detection Model using the Random Forest Learning Algorithm

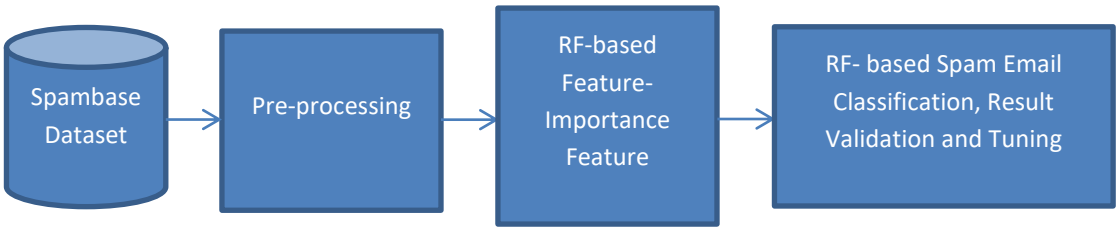

Figure 1: Architecture of Random Forest-based Spam Email Detection Scheme

The problem at hand is a classification problem that can be formulated as follows: Given a spam email dataset with set of features in the form x1; x2 ... xn, where n represents the number of features. To detect the presence of spam email evidence in the chosen dataset, the classification is a mapping of function f between the input features and target class of spam evidence.

3.1Dataset Description

The dataset used in this study was obtained from UCI Machine Learning Repository. The dataset is named spambase dataset. The target class is binary in nature as the dataset contains spam and non-spam evidence in adequate proportion. The dataset is available at Index of /ml/machine-learning-databases/spambase (uci.edu) as released by Hopkins, Reeber, Forman and Suermondt (2009).The dataset has 4601 instances and 57 input features as well as one target class. The exploratory data analysis carried out on the dataset further indicated that 55 of the input features are integer and real data type while the target class is categorical. This further described the need for minimal pre-processing so as to make the features useful for model building. The class distribution in the dataset is Spam-1813 which constitutes 39.4% and Non-Spam 2788 which constitutes 60.6%. This dataset is considered to be a good representative dataset for spam email detection studies.

3.2 Random Forest Algorithm

Random Forest is a supervised machine learning algorithm that has support for both classification and regression tasks. The base estimators used for building the ensemble in this study is a DT Algorithm. The spam email detection model was built from Random forest algorithm which contains forest of Trees as described by Breiman (2001). It is equally argued that Random forest algorithm works well because it aggregates many decision trees and therefore reduces the effect of noisy results, (Breiman, 2001). In the case of a random forest classification model, each decision tree votes and then produces the final result and the most popular prediction class is selected. The proposed email spam detection model works as described in the below algorithm.

Algorithm 1: Algorithm for Random Forest-based Spam Email detection model

(a) Input: Spambase input features
(b) Output: Email Spam classification based on Random Forest Algorithm

1. Load the dataset in csv format
2. Select random samples from the given dataset.
3. Construct a decision tree for every sample.
4. Obtain the prediction result from every decision tree.
5. Perform voting for every predicted result
6. Select the most voted prediction result as the final prediction result
7. Stop

4. RESULTS OF EXPERIMENTAL ANALYSES

4.1 Explorative Data Analysis

The various exploratory data analyses carried out revealed values shown in figures 1, 2, 3 and 4.

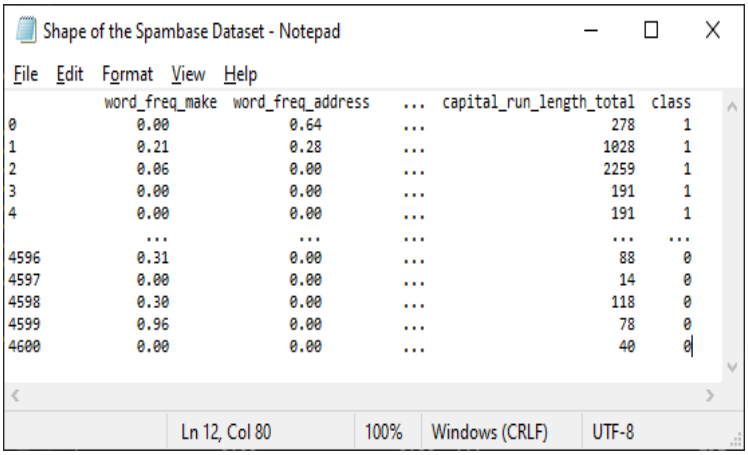

Figure 1: Screenshot of the Dataframe

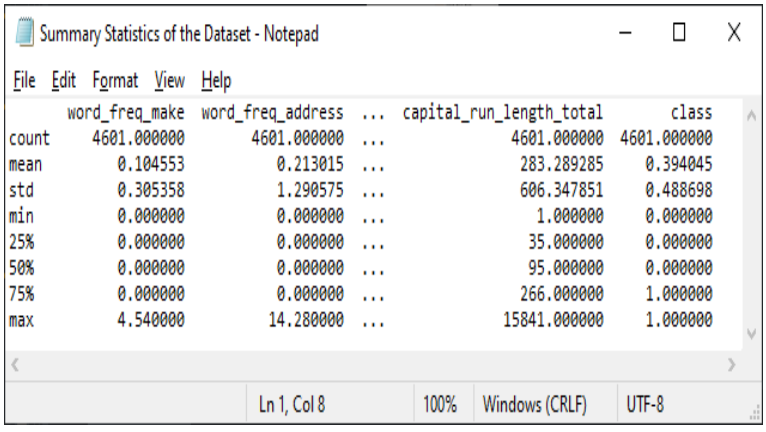

Figure 2: Statistical Summary of the Dataset

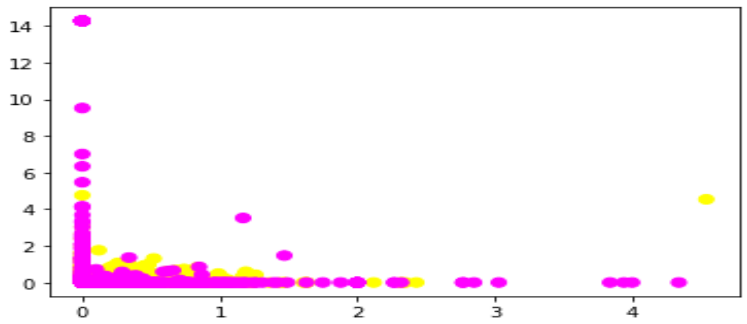

Figure 3: Scatter Diagram for the Spam Email Dataset

Figure 4: Feature Distribution Histogram

All the techniques used for the exploratory analysis or visualisation are aimed at gaining better understanding of the data distributions in the dataset. It could be seen that the features in the dataset range from word_frequency_make to capital_run_length_total.

4.2 Dataset Pre-processing

Based on the data exploration, some of the features in the dataset are already in pre-processed forms. For instance, it was deduced that some of the dataset features are in numeric forms already. This is the reason why some stages of data preprocessing steps such as tokenization, back of words handling are not required. However, majority of the input features are in floating point forms and they were handled. Also, it was discovered from exploratory data analysis that there are no missing values in the spambase dataset used in this study. Hence, no missing values were handled.

4.3 Features Selection Technique

For every machine learning task, it is important to identify and decide the type of features to use and which machine learning technique as the features selected will shape the type of model that is formed (Miller & Busby-Earle, 2017). One

can argue that this is one of the reasons why Markou and Singh (2003) identified that the benefits of performing feature selection before modeling your data include: reduces over fitting, improves accuracy and reduces training time. Confirming this, Salau-Ibrahim and Jimoh (2020) pointed out that the use of feature selection in data preprocessing leads to improved performance, reduced training time and enhanced model understanding. In this [-study, the feature selection technique used is RF-based Feature Importance. The technique ranks the features based on their importance as a result of scores computed for each. In all, twelve features were selected and were used in the model building.

4.4 Model Evaluation Metrics

The built model was evaluated using the metrics whose formulae are as shown in equations (1), (2), (3), and (4).

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

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

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

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

The interpretation of the TP, TN, FP and FN in equations 1 to 4 are given below.

(i) True positives (TP): These are cases in which email spam is predicted and truly the spam is present.

(ii) True negatives (TN): These are situations in which the algorithm predicted non-spam, and the spam is not present in the dataset samples.

(iii) False positives (FP): These are situations in which the model predicted the presence of email spam, but the spam is not present.

(iv) False negatives (FN): These are situations in which the model predicted no, but email spam is actually present.

The performance of the spam email detection model was validated using holdout technique. The holdout validation approach refers to creating the training and the holdout sets, also referred to as the 'test' or the 'validation' set. The training data is used to train the model while the unseen data is used to validate the model performance. In this technique of cross-validation, the whole dataset is randomly partitioned into a training set and validation set. In this paper, 85% of the whole dataset was used as a training set and the remaining 15% was used as testing set. The models' parameters were adjusted severally until better results were achieved.

4.5 Model Classification Results

Table1: Results of Random Forest-based Email Spam Detection Model

Metrics\ML Algorithm	Random Forest
Accuracy	99.54%
Precision	99.21%
Recall	99.46%
F1-score	99.33%.

As shown in table 1, the results of RF-based model performances are recorded. Best results were obtained for the four metrics using the 85% of training set and 15% of testing set. The results for accuracy, precision, recall and F1-score are as shown in table 1. being the most promising. The hyper parameters of the Random Forest algorithm were altered severally to achieve the values obtained

Table 2: Benchmark Comparison with Existing Study

Author	Approach	Accuracy (%)	Precision %	Recall %	F1-score %
Nandini and Jeen (2020)	The study used linear classifiers namely Logistic Regression, Decision Tree (DT), Naïve Bayes, K-Nearest Neighbor (KNN) and SVM for spam email detection. This study focuses on comparative analyses of selected single learning algorithms	The average accuracy of the five algorithms is fair enough. However, Only SVM and Decision Tree have high predictive accuracy similar to the one in this study	The average precision is not as promising as the one found in this study.	The average recall is not as promising the one found in this study.	The average F1-score is not as promising the one found in this study.
This study by Oyelakin et al.	Random Forest-based ensemble technique was used in building the spam email detection model in this study. The algorithm was well tuned to achieve good results.	The result is very promising as 99.54 was arrived at for the accuracy.	The precision is very promising as 99.21 was arrived at.	The result is very promising as 99.46 was arrived at.	The result is very promising as 99.33 was arrived at.

Benchmark comparison was carried out in respect of a similar study that used the spambase dataset for its machine learning-based spam email detection. The results of the chosen work and this current study are as shown in table 2. It was revealed that the proposed ensemble learning approach was more promising in all the four selected evaluation metrics.

5. DISCUSSIONS OF RESULTS

The results of the model obtained from the study are shown in table 1. The results in this study are in the respect of the transformed dataset, selected features and spam email classification results. The summary statistics of the dataset revealed basic patterns of the features in the dataset. Similarly the scatter diagram shows the distribution of the data in the dataset. With the summary statistics, the authors obtained a good idea about statistical measures in these features. The statistical measures include: count, average, standard deviation and quartiles. The dataset was pre-processed and 12 subset features were selected based on feature importance technique. The features were used for building the spam email detection model. The performance of the model was evaluated based on the calculation of values for the selected metrics. Experimental results from the Random Forest (RF) based spam email detection model built achieved 99.54% Accuracy, 99.21% Precision, 99.46% of Recall and F1-score of 99.33%. The results were validated using hold-out validation method. The benchmark comparison with the study that used the same dataset revealed that the ensemble technique has improved performance in all the metrics used for the evaluation.

6. CONCLUSION

This study investigated how to use ensemble machine learning approach for improved classification of spam email. Specifically, an ensemble learning algorithm named Random Forest was chosen for building the spam email model. The study made use of a publicly available dataset named Spambase that was collected from UCI Machine Learning Repository. First of all, exploratory data analysis (EDA) was carried out on the dataset with a view to gaining better insights on the features and samples in the dataset. The EDA results were presented in various ways and guided the pre-processing stage. Then, the collected dataset was pre-processed and used for the RF-based model building. Generally, the

model performed excellently across the four selected metrics for the spam email detection. The algorithm was well tuned to achieve good results.The study emphasised the promising and strength of ensemble machine learning-based model for email spam identification. It is recommended that future studies may consider another set of learning algorithms and different feature selection techniques that may give improved performances across in spam email classification.

7. ACKNOWLEDGEMENT

Authors acknowledge the anonymous reviewers who reviewed the paper and provided sound feedbacks which added value to this research.

8. REFERENCES

[1] Ahmed Naeem , Amin Rashid , Aldabbas Hamza, Koundal Deepika, Alouffi Bader & Shah Tariq (2021). Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges, *Security and Communication Networks*, Volume 2022, https://doi.org/10.1155/2022/1862888

[2] Alpaydın, E. (2010). An Introduction to Machine Learning. 2nd Ed.The MIT Press Cambridge, Massachusetts London, England

[3] Ameen A. O., Oyelakin A. M., Ajiboye I. K., Olatinwo I. S., Obiwusi K. Y., & Ogundele T. S. (2022).Evaluating the performance of heterogeneous and homogeneous ensemble-based models for twitter spam classification,UMT Artif. Intell. Rev., 2(2), 01-16, 2022, doi: https://doi.org/10.32350.icr.22.01

[4] Breiman L. (2001). Random Forests, Machine Learning, 45(1), 5-32, 2001. Available at:https://doi.org/10.1023/A:1010933404324

[5] Gaikwad Bhagyashri U. & Halkarnikar P. P. (2013) Spam E-mail Detection by Random Forests Algorithm, International Journal of Advanced Computer Engineering and Communication Technology (IJACECT), : 2278-5140, Volume-2, Issue – 4

[6] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin & Wei-Yang Lin (2009).Intrusion detection by machine learning: A review, *Expert Systems with Applications*,36, 11994–12000, available at https://tarjomefa.com/wp-content/uploads/2017/11/8000-English-TarjomeFa.pdf

[7] Dar Shaukat Kamran , Luo Suhuai , Chen Shan & Liu Dongxi (2020).Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective, *IEEE International Conference on Cyber Warfare and Security: Islamabad*, Pakistan, DOI: 10.1109/ICCWS48432.2020.9292388

[8] Hariharan N, Kamaraj G, & Ramanuja Babu R D (2021). An Collaborative and Early Detection of Email Spam Using Multitask Learning, *International Journal of New Technology and Research (IJNTR)*, 7(4), 35-40, https://doi.org/10.31871/IJNTR.7.4.11

[9] Hopkins Mark, Reeber Erik, Forman George, Suermondt Jaap (2009). Spambase.Available at the UCI Machine Learning Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html

[10] Jazzar Mahmoud, Yousef Rasheed F., Eleyan Derar (2021). Evaluation of Machine Learning Techniques for Email Spam Classification, *International Journal of Education and Management Engineering (IJEME)*, 11(4), 35-42, 2021. DOI: 10.5815/ijeme.2021.04.0

[11] Kashapov, A., Wu, T., Abuadbba, A., & Rudolph, C. (2022). Email Summarization to Assist Users in Phishing Identification. *arXiv preprint arXiv:2203.13380*

[12] Kothapally N R, & Vijayalakshmi K (2021) Classification of Spam Messages using Random Forest Algorithm, *Journal of Xidian University*, ISSN No:1001-2400

[13] Markou, M. & Singh S. (2003). Novelty detection: a review-part 2: Neural Network Based Approaches, *Signal Processing*, 83, 2499-2521

[14] Miller, S., & Busby-Earle, C. (2017). The role of machine learning in botnet detection. 2016, 11th*International Conference for Internet Technology and Secured Transactions, ICITST 2016*, (December), 359–364. https://doi.org/10.1109/ICITST.2016.7856730

[15] Nandini S. & Jeen M. K. S. (2020). Performance Evaluation of Machine Learning Algorithms for Email Spam Detection, *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*

[16] Oyelakin A.M., Salau-Ibrahim T.T., Ogidan B.S., Azeez R.D. & Ajiboye I.K. (2019).Peer-to-Peer Botnets: A Survey of Propagation, Detection and Detection Evasive Techniques, *Fulafia Journal of Science and Technology, a Tetfund-funded Journal of Federal University, Lafia, Nassarawa State, Nigeria, 5(3):13-18*

[17] Rao, Justin M.; Reiley, David H. (2012). Economics of Spam, Journal of Economic Perspectives, **26** (3): 87–110, *doi*:*10.1257/jep.26.3.87*

[18] Rao S., Verma A. K., and Bhatia T. (2010). A review on social spam detection: Challenges, open issues, and future directions, Expert Sys. Application Review, 33 (1–2): 1–39, 2010, doi: https://doi.org/10.1007/s10 462-009-9124-7

[19] Salau- Ibrahim, T. T., & Jimoh, R. G. (2020). Negative Selection Algorithm Based Intrusion Detection Model. In *20th IEEE Mediterranean Electrotechnical Conference, MELECON 2020 - Proceedings* (pp. 202–206). https://doi.org/10.1109/MELECON48756.2020.9140562]=q\=

[20] Sethi, M., Chandra, S., Chaudhary, V., & Dahiya, Y. (2022). Spam Email Detection Using Machine Learning and Neural Networks. In *Sentimental Analysis and Deep Learning* (pp. 275-290). Springer, Singapore.

[21] Sharma P. & Bhardawaj U. (2018). Machine Learning based Spam Email Detection, *International Journal of Intelligent Engineering and Systems*, 11(3), DOI: 10.22266/ijies2018.0630.01

Author's Brief Profile

Dr. Oyelakin A. M. is currently a lecturer in the Department of Computer Science Al-Hikmah University, Ilorin, Nigeria. He was educated at Federal Polytechnic, Offa and University of Ilorin where he bagged National Diploma (Distinction Classification) and First Degree (Second Class Upper) respectively. He obtained M.Sc. Computer Science from University of Lagos, Akoka and PhD in Computer Science at University of Ilorin. Oyelakin currently serves as the Sub-Dean of Postgraduate School and Departmental postgraduate Coordinator in the University. He has over 30 articles in notable peer reviewed journals and conference proceedings. He is a member of Nigeria Computer Society (NCS) and Internet Society. He is the current Vice-chairman of NCS in Kwara State. His current areas of research interest include Security/Privacy, Machine Learning, Networks and Data Communications, Mobile Computing as well as Health Data Analytics. He is happily married with kids. He can be reached via amoyelakin@alhikmah.edu.ng

Salau-Ibrahim T. T. (PhD) lectures in the Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria. She holds First Degree in Computer Science, M.Sc. Computing and Information Systems and PhD Degree in Computer Science from Al-Hikmah University, Ilorin, Queen Mary University of London and University of Ilorin, Ilorin, Nigeria respectively. She is a member of Nigeria Computer Society. Her areas of research interest include Information Security, Machine Learning IoT and Software Engineering. She is happily married with kids. She can be contacted via ttsalau@alhikmah.edu.ng.

Ogidan B. S. obtained his First and Master Degrees in Computer Science from University of Ilorin, Ilorin, Nigeria. He is the current Acting Director of ICT at Al-Hikmah University, Ilorin. He also works as adjunct lecturer in the Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria. His area of research interests include: Data Mining and Information Systems. He is happily married with kids.

Olufadi H. I. Yusuf S. A. obtained her Bachelor's Degree and Master Degree in Computer Science from University of Ilorin, Ilorin, Nigeria. She has about four articles to her credit. She works as a part-time lecturer in the Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria. She has passion for teaching and has been practicing on part time basis for some years now. Her area of research interests include: Data Mining and Cyber Security. She is happily married with kids. She can be contacted via olufadi.halimat@gmail.com.

Yusuf S. A. obtained her Bachelor's Degree and Master Degree in Computer Science from University of Ilorin, Ilorin, Nigeria. She has three articles to her credit. She works as a part time lecturer in the Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria. Her area of research interests are Software Engineering and Text Mining.

Adeniji I. A. obtained Bachelor's Degree in Computer Science from Al-Hikmah University, Ilorin, Nigeria. He equally obtained master's degree in Computer Science from University of Ilorin, Ilorin, Nigeria. He has published peer reviewed articles in local and international research outlets. His research interests include: Data Science Data Mining, Cyber Security and Information Systems. He is a member of International Association of Engineers as well as Computer Science and Engineering Society. He is happily married with kids. He can be contacted via herdeynig@gmail.com