Nigerian Research Journal of Engineering and Environmental Sciences 7(2) 2022 pp. 486-494
p ISSN: 2635-3342; e ISSN: 2635-3350

Nigerian Research Journal of Engineering and Environmental Sciences
Journal homepage: www.rjees.com

Original Research Article

Artificial Intelligence Model for Email Spam Detection

***Adimora, K.C., Aru, O.E., Udo, E.U. and Ezeh, C.M-E.**

Department of Computer Engineering, College of Engineering and Engineering Technology, Michael Okpara University of Agriculture, Umudike, Umuahia, Abia State, Nigeria.
*kyrimanjero@gmail.com

http://doi.org/10.5281/zenodo.7496635

ARTICLE INFORMATION

Article history:
Received 22 Nov. 2022
Revised 15 Dec. 2022
Accepted 16 Dec. 2022
Available online 30 Dec. 2022

Keywords:
Email spam
Machine learning
Regression algorithm
Root mean squared error
Artificial intelligence

ABSTRACT

The continuous attack of email spam on internet users has geometrically increased and necessitated the need for a more robust and dependable anti-spam technology for filtering email spam. Presently, individuals and organizations often lose millions of dollars to fraud by mere opening or responding to email spam sent to their email inboxes despite the anti-spam software in existence. This has brought about major economic losses, email traffic problems, a shortage of memory space, and limits the system's computing power. This paper proposes an artificial intelligence (AI) model that trains, tests, and validates, email datasets using machine learning classification, regression, and clustering algorithms. The performance metric was the root mean squared error. The error value achieved was 0.02349, which indicated the effectiveness of the proposed AI model in filtering email spam. A web application was built to test the robustness, performance, accuracy, and reliability of the system. The results revealed an excellent performance at a minimal system error level of 0.0004.

1. INTRODUCTION

Email is one of the very important tools and fundamental means of communication. Generally, an email message is comprised of two important components; the header and the body. The header provides the technical details about the message such as the sender and receiver identities, the software used to compose it, and the email servers that it passed through on its way to the recipient. The body is seen as the heart of the email and does not contain pre-defined data information.

For effective communication on the internet using email as a tool spam detection is necessary. This is because email facilities are constantly attacked by spammers, fraudsters, hackers, and imposters. Therefore, several

anti-spam mechanisms are designed to identify incoming dangerous emails from hackers. Hackers most times pretend to offer a beneficial service in their emails, but they are really malicious attacks on the user's computer system meant to direct them to a dangerous site or with the sole aim of stealing their valuable information such as bank account details, login details, and other important sensitive information that may cost the user billions of dollars.

In the world today, almost everyone uses email for business, private, and educational activities. However, email spam is now a great concern and threat to internet users that must be curtailed or eliminated. Spam can be in the form of text messages sent indiscriminately to their mobile phone, often for commercial purposes. It can take the form of a simple message, a link to a number to call or text, a link to a website for more information, or a link to a website to download an application. The Naïve Bayesian Classifier algorithm is designed for separating spam and non-spam emails (Sharma and Jatana, 2014). Hackers attacking users' emails are the great threat in a social media network. Numerous problems have been created through spam creation which has adversely affected system performance and accuracy. AbdulNabi and Yaseen (2021) conducted a research that allows the effectiveness of words embedded in classifying spam emails through a single application of a natural language processing technique. Sao *et al.* (2017) carried out a research for text classification of email spam using the Naïve Bayesian classifier. Besides text, the work was not able to classify other forms of data. Email filtering techniques usually work based on the contents of the message, searching for specific phrases, words, and particular expressions. However, spammers started to avoid such methods by sending text messages containing no HTML codes and downloading images (Bhowmick and Hazarika, 2017).

Awad and Elseuofi (2011) reviewed, the newest concept of machine learning algorithms based on the issue of spam email classification and clustering. The work highlighted machine learning algorithms and their comparative performance and applicability to filtering email spam. Mahmoud *et al.* (2021) reviewed the performance metrics of ML techniques that are based on spam detection and classification. A modeling pipeline was developed as a review approach for email spam (Sethi *et al.* 2021). Meelony and Nikita (2021) reviewed a good number of spam filtration techniques that were designed by researchers and scholars. Aski and Sourati (2016) carried out a research that described three algorithmic concepts for filtering spam from valid emails with low error rates and high efficiency using a multilayer perceptron model. Alanazi and Ahmed (2021) conducted a research that employed data feature selection with classification for detecting illegitimate emails and temporal email addresses using natural language generation that hinges on a random forest approach. Akash *et al.* (2021) conducted a research that emphasized building a comprehensive model for email clustering and classification that focused on semantics-based text classification using NLP and URL-based filtering. Meng and Peng (2017) analyzed the variety of spam-detection methods to filter false or harmful knowledge of traditional systems such as email or short message service (SMS). A content-based approach called adaptive fusion for spam detection (AFSD) removes text features from an email's character filament, develops a spam detector for a double classification task (spam versus regular message), and explains promising accuracy in hostility email spam. Dada *et al.* (2019) reviewed related works that compare the merit and demerit of using a machine learning technique in combating and filtering email spam. Sesha *et al.* (2018) proposed a system that imports data from email accounts using linear regression and data mining techniques in preprocessing it.

The limitation of the reviewed works is the non-utilization of AI model of classification, regression, and clustering in building a formidable check for email spams. The primary aim of the work is to develop an artificial intelligence model for email spam detection on web application that trains, tests, and validates email data based on machine learning classification, regression, and clustering algorithms that was not used by the previous research work.

2. METHODOLOGY

Python, CSS, HTML, JavaScript, SQlite, MySQL, MATLAB, and Excel application are the technologies used in the research. An AI approach was deployed in the design, development, and deployment of the

proposed system. The three points attributes and instances of the new model are analyzed separately in this section.

2.1. Classification Algorithm for Spam Filtration

The classification algorithm helps in classifying a set of data into email spam and none email spam. In this paper, email spam was classified as high which denotes a one (1), and none spam email as low which denotes a zero value (0). This classification can be in different categories based on their behavioral patterns, historical pattern, and web browsing patterns. In this study, an algorithm was trained to recognize spam emails by learning the properties of spam and non-spam email. The classification AI model is a function that maps an email text to spam or non-spam classification. The trained model was used to filter new incoming emails as spam or non-spam as depicted in Figure 1. The binary classification was employed model to predict either spam or non-spam on every email received.

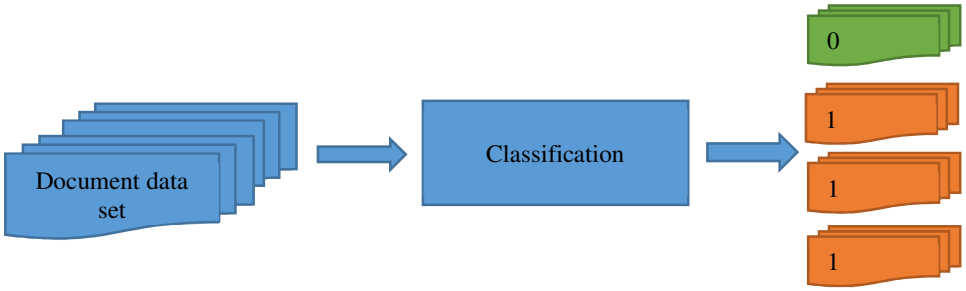

Figure 1: Spam filtration classification framework

Anomalous pattern detection in financial transactions may indicate fraud and can be regarded as spam. In this case, the spam filtering algorithms should be able to detect rare cases that lie outside the training distribution which then can be taken as fraud (spam). Equation (1) is the decision rule for the spam classification algorithm.

$$\hat{y} = \begin{Bmatrix} 1, if\ h(x) \geq t \\ 0,\ otherwise \end{Bmatrix} \quad (1)$$

Where $\hat{y}$ is the spam classification outcome, t is the threshold of the classification, and $h(x)$ is the response surface.

2.2. Regression Algorithm for Email Spam Detection

This algorithm estimates the relationships among variables that predict the output values based on input features of the data fed into the system. In this study, the trained data was used to predict spam features based on the available data. Considering an AI model with one independent feature (r) and trained data (t). The best 100 attributes were considered for linear regression email data classifications. Equation 2 illustrated spam detection regression algorithms.

$$H(r) = \theta_a + \theta_b r_b + \theta_c r_c + \cdots + \theta_n r_n \quad (2)$$

Where θ_a, θ_b, θ_c, and θ_n are the regression coefficients.

These coefficients represent the weights of the email spam, and r_b, r_c, and r_n, represent the attributes, and n represents the number of instances. $H(r)$ is the relationship among variables that fits the line which filters the spam. These weights are calculated from the extracted email training datasets and the training instances are presented in Equation (3).

$$\theta_a {r_a}^{(n)} + \theta_b {r_b}^{(n)} + \theta_c {r_c}^{(n)} + \cdots + \theta_n {r_n}^{(n)} = \sum_{a=0}^{n} \theta_a {r_a}^{(n)} \quad (3)$$

$^{(n)}$ Represents several training instances. In order to reduce the squared error on training data a specific weight θ is chosen and the process is illustrated in Equation (4).

$$\sum_t^r (H(r) - \sum_{a=0}^n \theta_a r_a{}^{(n)})^{\wedge 4} \quad (4)$$

Here, $H(r)$ is the training instances and $\sum_{a=0}^n \theta_a r_a{}^{(n)}$ represents the predicted value of the nth training instances.

The University of California Irvine (UCI) machine learning repository, Kaggle, and Amazon Web Service (AWS) datasets split was 60% for training and 40% for testing. Therefore, the number of instances considered for training and testing is shown in Table 1. The training and testing correlation coefficient were 0.92305 and 0.80114 respectively. The mean absolute error was 0.12341 while the root means squared error for the training and testing data set were 0.00157 and 0.10312 respectively. These values were found to be relatively good for spam detection. Figure 2 depict the spam filtration regression framework.

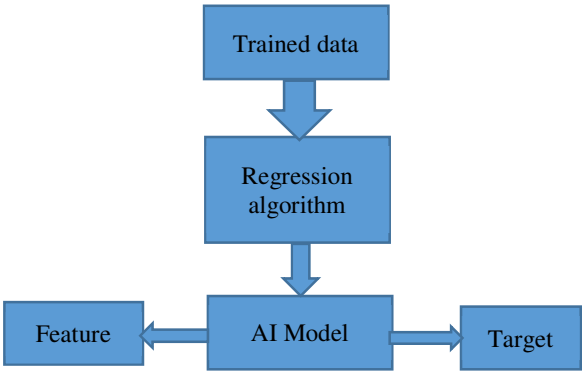

Figure 2: Spam filtration regression framework

2.3. Clustering Algorithms for Spam Detection

This algorithm only interprets the trained input data and finds cluster in feature space. In this paper, clustering based on spam and non-spam emails for a function (f) were investigated.

$$f_c = \max d(t_i, t_j) - \min \min d(t_i, t_j) \quad (5)$$

The minimization of f_c means the maximization of the minimum distance between any two clusters. The criterion f_c can be interpreted as the compactness of the clustering. Where d is the dissimilarity between two objects (spam and non-spam), t_i and t_j are the vertices.

2.4. Root Mean Squared Error (RMSE)

The statistical tests are very important for performance evaluation, accuracy, and error determination of artificial intelligence models. In this paper, the root mean squared error is employed. The models were trained with the training and testing datasets respectively. For each dataset sample $a \in T$ of the test set and $b \in R$, were computed, classified based on the defined features of the new model. Equation 6 illustrates the how RMSE was applied to this research to determine system's error in detecting email spam.

$$V = \sum R_b + T_a + \cdots + R_{nb} + T_{na} \quad (6)$$

Here, V is the dataset, R is the training data, T is the testing data, a and b are the datasets content being trained and tested.

$$RMSE = \sqrt{\frac{V}{N}} \quad (7a)$$

$$RMSE = \sqrt{\frac{\sum R_b + T_a + \cdots + R_{nb} + T_{na}}{N}} \quad (7b)$$

Here, N is the number of non-missing data points.

2.5. Datasets Analysis

Table 1 shows the training and testing email spam dataset for this study. The analysis covers datasets from 1999 to 2022. The datasets were got from UCI machine learning repository, AWS and KAGGLE datasets.

Table 1: Spam versus legitimate mail dataset

Training	Testing	Total	Ratio	Date
3672	1500	5975	1:3	1999 - 2004
4361	1496	5857	1:3	2005 - 2010
4012	1500	5512	1:3	2011 - 2015
4500	1500	6000	1:3	2016 -2020
3675	1500	5175	1:3	2021 - 2022

The datasets were split into training and testing in the ratio of one by three (1:3). According to Table 1, from 1999 to 2004 a total of 3672 datasets were trained while 1500 datasets were tested. Similarly, from 2005 to 2010 a total of 4361 datasets were trained while 1496 were tested. Also, from 2011 to 2015 4012 were trained while 1500 were tested. From 2016 to 2020 a total of 4500 datasets were trained and 1500 were tested. Finally, from 2021 to 2022 a total of 3675 datasets were trained and 1500 were tested. These enormous datasets training, testing and validation analysis was done to achieve research set goal. The algorithm for the proposed model is shown as follows.

AI model spam detection algorithm

Start
for each email text received:
Analyze the header and the body of each email.
Is header information indicating trusted source?
$Yes = non - spam = 0\ and\ No = spam = 1$
Determine email text size.
$if\ (textSize = emailBody)$
$if(TextSize\ !\ emailBody)\ then\ spam\ exists$
$for\ Ae = \hat{y} + H(r) + f_c$
$if\ (Ae = 0)$ // Non-Spam exists
$else\ if\ (semantic\ relationship\ (H(r) = 0)$// non-spam
$else\ if\ (pattern\ recognition = 0)$// non-spam
$else$
$spam\ is\ detected$
$end\ for$
$for\ each\ emailText$
$if\ emailText = 1$// spam exists
$else\ \ emailText = nsp$
$nsp = emailText - spam(sp)$
$return\ nsp$
$end\ if$
$end\ for$
$end\ if$
$return\ spam\ or\ non - spam$
end

3. RESULTS AND DISCUSSION

The study datasets analysis was carried out, which were trained and tested. Table 2 presents the impact of spam emails on various internet services without the application of the proposed AI model. It is found from the table that about 69% of spam emails are email services and about 53% of spam emails are from the social

network. The remarks explained the impact level (high) of spam email on several internet activities when the new system was not applied. Also, Table 3 depicts how the application of the new system has minimized the high spam email of Table 2. Table 3 demonstrated the efficacy of the new system with every spam email minimized to a harmless level

Table 2: Impact of spam emails without the new system

Activities	Spam email (SE) (%)	Non-spam email (NSE) (%)	Remark
Cloud space for storage	52	48	S.E. high
E-Payment	64	36	S.E. high
E-Commerce	60	40	S.E. high
Email	69	31	S.E. high
Social network	53	47	S.E. high
Education	30	70	S.E. high
others	26	74	S.E. high

Table 3: Impact of spam emails with the new system

Activities	Spam email (SE) (%)	Non-spam email (NSE) (%)	Remark
Cloud space for storage	12	88	S.E. minimized
E-Payment	18	82	S.E. min.
E-Commerce	11	89	S.E. min.
Email	23	77	S.E. min.
Social network	16	84	S.E. min.
Education	10	90	S.E. min.
others	20	80	S.E. min.

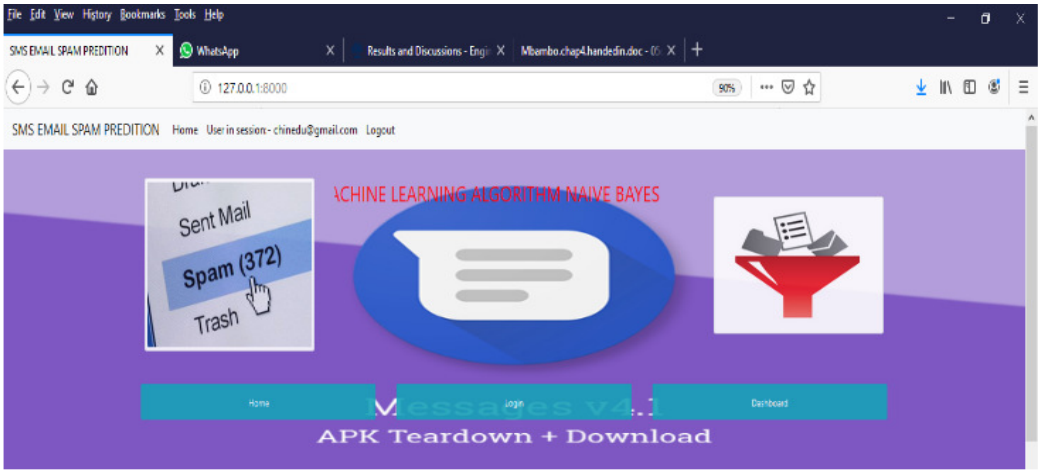

Figure 3: Spam email detection via web application

Figure 3 depicts spam email detection of the proposed system via the web applications developed for evaluating the robustness, efficiency, and reliability of the new system. Figure 4 and 5 are the regression model of the new system. Figure 4 revealed the training accuracy of 0.95466, validation accuracy of 0.95523, testing accuracy of 0.98893. The testing accuracy is optimium with error threshold of 0.18. The fitted regression line for all reveals the efficiency of the proposed system. Data training of Figure 4 was not done to 100 percent but that of Figure 5 was carried out to 100 perecent. Here, we observed that the training, validation, and testing accuracy values are represented by a number one (1) that indicates 100 percent datasets trained, tested, and validated. The testing accuracy is 1 with minimal error value of 0.0004. Figure 6 demostrated the state of the proposed model to validate data after training. It is the state that sieves

acceptable data from unacceptable data after dataset training and testing through data validation. The performance evaluation of the new system is depicted in Figure 7. Figure 7 showed the best valid performance of 0.87026 for the proposed system. The mean squared error (MSE) at the training and a validation intersection point is 1 and is negligible. The MSE value for data sets testing was below zero (0) which expresses the efficiency and dependability of the proposed model. Figure 8 is the mean squared error analysis in the histogram. Figure 8 showed 0.02349 error value at all instancess with the highest bar of the histogram. This value is approximately zero (0) and negligible. Therefore, the new system is considered zero error tolarnce in filtering spam in email communications.

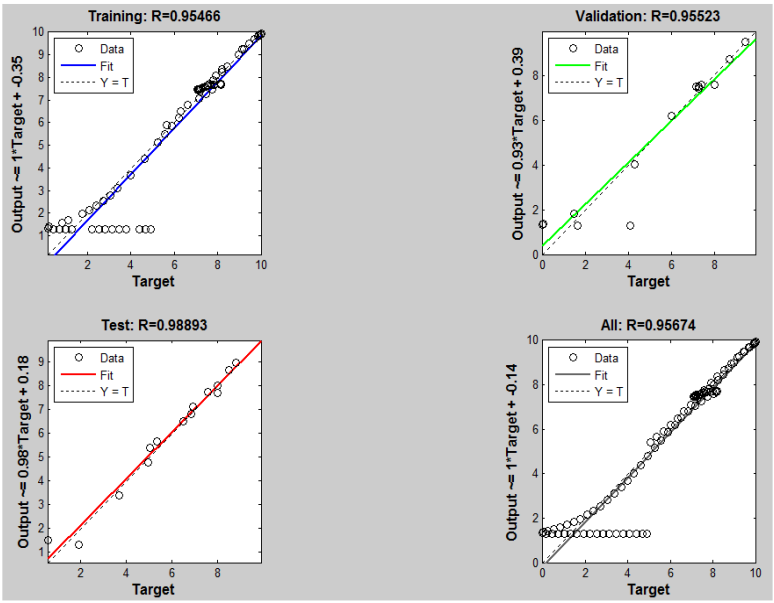

Figure 4: Regression model on selected datasets training

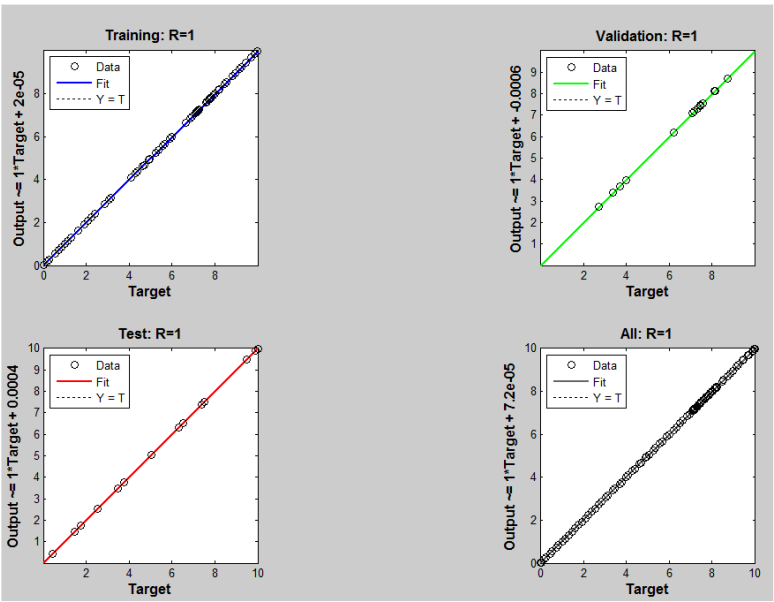

Figure 5: Regression model on full datasets training

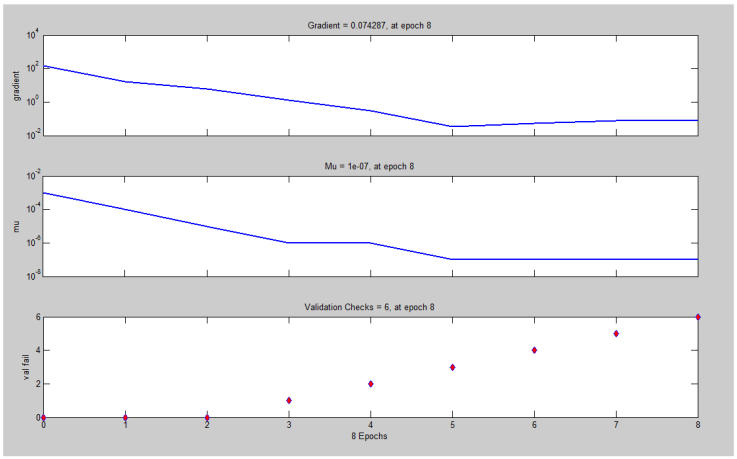

Figure 6: Training state of the proposed model

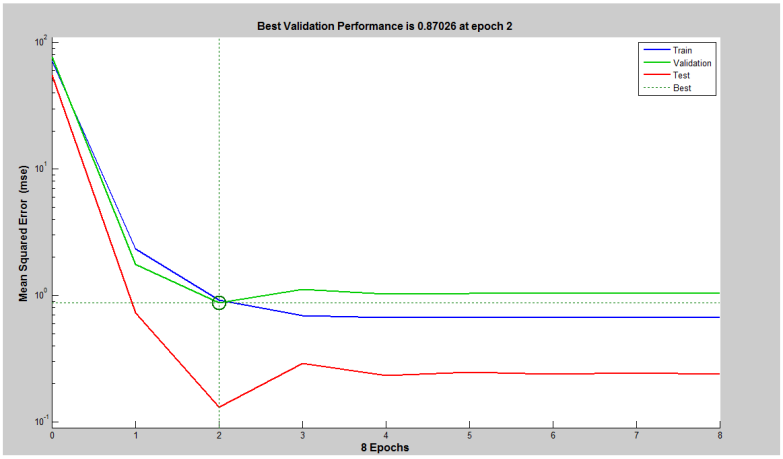

Figure 7: Performance evaluation of the AI model

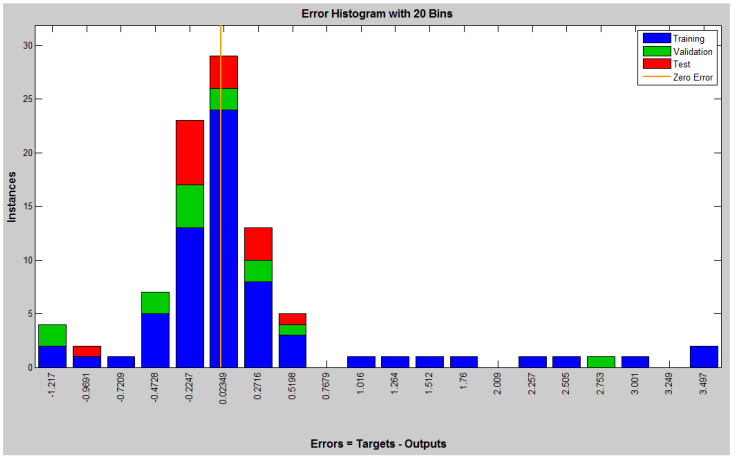

Figure 8: Mean squred error analysis

4. CONCLUSION

The issue of spam on the internet, and users of email communication is a big concern and a great threat. This study proposed an AI model for filtering email spam. In this paper, several datasets were trained and tested to solve the issue of email spam. The computed root mean squared error of the new system is minimal and negligible with 100 percent accuracy on fully trained and tested data. This indicated the effectiveness of the new system in detecting email spam compared to other existing anti spam. The future direction of this research should consider data security with the application of cryptography, steganography, and computer vision.

5. CONFLICT OF INTEREST

There is no conflict of interest associated with this work.

REFERENCES

AbdulNabi, I. and Yaseen, Q. (2021). Spam Email Detection using Deep Learning Techniques, Proceedings of the 2nd *International Workshop on Data-Driven Security,* Warsaw Poland, 184(2), pp. 853 -858.

Akash, J. Siddhant, A., Jainam, F., Priya, C. and, Deepak, K. (2021). Email Spam classification via Machine Learning and Natural Language Processing. Proceedings of *the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Network (ICICV)*, 3(2), pp. 70-84

Alanazi, R., and Ahmed, I. T. (2021). Detection of Email Spam Using Natural Language Processing Based Random Forest. *Engineering Application of Artificial Intelligence,* 189(10), pp. 1234 – 1247.

Aski, A. S. and Sourati, N.K. (2016). Proposed efficient algorithm to filter spam using machine learning techniques. *Pacific Science Review: Natural Science and Engineering*, 18 (20), pp. 145 -149.

Awad, W. A. and ELseuofi, S. M. (2011). Machine learning methods for spam e-mail classification. *International Journal of Computer Science & Information Technology* (IJCSIT), 3(1), pp. 173-184. Bhowmick, A. and Hazarika, S. M. (2017). E-Mail Spam Filtering: A Review of Techniques and Trends. In: Lecture Notes in Electrical Engineering, *Springer Singapore*, pp. 583–590.

Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S.M., Adtumbi, A. O. and Ajibuwa, O.E. (2019). Machine Learning for email sapm filtering: review, approaches and Open research problems. *Heliyon,* 5(3), pp. 1 -23.

Mahmoud, J. Rasheed, F. Y. and Derar, E. (2021). Evaluation of Machine Learning Techniques for Email Spam Classification. *International Journal Education and Management Engineering*, 4 (1), pp. 35 - 42.

Meelony, M. and Nikita, S. (2021). Email Spam Filtering Techniques: A Review. *Design Engineering,* 3(9), pp. 8327 – 8338.

Meng, J. and Peng, C. (2017). Suspicious Behavior Detection: Current Trends and Future Direction. *IEEE ACCESS*, 6(2), pp. 123 – 132.

Sao, P., Chaubey, S. and Katailiha, S. (2017). Text Classification for Email Spam using Naïve Bayesian classifier. *International Journal of Advances Research in Science and Engineering,* 6(2), pp. 1-9.

Sesha, A. R.., Avadhani, P.S. and Nandita, B.C. (2018). Detecting Targeted Malicious E-Mail using Linear Regression Algorithm with Data Mining Techniques. *Computational Intelligence in Data Mining,* 43(2), pp. 1-13.

Sethi, M., Chandra, S. and Chaudhary, V. (2021). Email Spam Detection using Machine Learning and Neural Networks. *International research journal of engineering and technology*, 8(4), pp. 1-7.

Sharma, K. and Jatana, N. (2014). Bayesian Spam Classification: Time Efficient Radix Encoded Fragmented Database Approach. *IEEE*, pp. 939 – 942.