World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Enhancing email security: A hybrid machine learning approach for spam and malware detection

Walter Oluchukwu Ugwueze, Sylvester Okwudili Anigbogu, Emmanuel Chibuogu Asogwa *, Doris Chinedu Asogwa and Kenechukwu Sylvanus Anigbogu

Department of Computer Science, Nnamdi Azikiwe University Awka, Nigeria.

World Journal of Advanced Engineering Technology and Sciences, 2024, 12(01), 187–200

Publication history: Received on 15 March 2024; revised on 01 May 2024; accepted on 04 May 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.12.1.0160

Abstract

Recent research indicates a notable surge in SMS spam, posing as entities aiming to deceive individuals into divulging private account or identity details, commonly termed “phishing” or "email spam". Conventional spam filters struggle to adequately identify these malicious emails, leading to challenges for both consumers and businesses engaged in online transactions. Addressing this issue presents a significant learning challenge. While initially appearing as a straightforward text classification problem, the classification process is complicated by the striking similarity between spam and legitimate emails. In this study, we introduce a novel method named "filter" designed specifically for detecting deceptive SMS spam. By incorporating features tailored to expose the deceptive techniques employed to dupe users, we achieved an accurate classification rate of over 99.01% for SMS spam emails, while maintaining a low false positive rate. These results were attained using a dataset comprising 746 instances of spam and 4822 instances of legitimate emails. The filter's accuracy, evaluated on a dataset with two attributes and 5568 instances, notably surpasses existing methodologies. Our proposed model, a Hybrid NB-ANN model, achieves the highest accuracy at 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%). This highlights the efficacy of the hybrid approach in enhancing accuracy for email spam detection and malware filtering, ensuring comprehensive coverage across training and test datasets for improved feedback loops.

Keywords: Machine learning; Predictive model; SMS spam; Malware filtering Hybrid NB-ANN

1. Introduction

The Internet's proliferation has made email an essential tool for communication, impacting various sectors. However, the rise of spam emails has become a significant concern [1]. Spam, characterized by unsolicited and often malicious content, poses threats to users and organizations, leading to the need for robust filtering solutions [2]. Traditional spam filtering methods, including source barring, blocking known sources, and destination filtering, have limitations [3]. Machine learning algorithms offer promising solutions for spam detection. Techniques such as Naive Bayes, Support Vector Machine, and Neural Networks have shown effectiveness in identifying spam emails [4].

A hybrid machine-learning approach is proposed to enhance spam and malware filtering [5]. This approach aims to achieve high accuracy, precision, and recall rates in classifying emails into legitimate and spam categories [6]. By combining the strengths of different machine learning algorithms, such as Naive Bayes and Neural Networks, the proposed hybrid model seeks to address the evolving nature of spam threats [7]. The development of advanced spam filtering techniques is crucial to mitigate the growing menace of spam emails. The proposed hybrid machine learning approach holds promise for enhancing email security and efficiency [8].

* Corresponding author: Emmanuel Chibuogu Asogwa

The study proposes the utilization of machine learning techniques to predict SMS spam. Two data sources are employed: a dataset from Kaggle containing spam/ham SMS messages and data collected from the propertywithin.com.ng website. The total dataset comprises a combination of SMS data from both sources, facilitating a comprehensive analysis

The SMS data from both sources is combined to form a dataset for analysis. Pre-processing involves converting JSON to CSV format and organizing the data into columns representing labels (spam/ham), text content, and numerical labels.

Data reformatting involves sorting spam events by time to preserve sequential information. Filtering of users is implemented to remove data from users that do not comply with certain constraints, ensuring data quality.

Machine learning algorithms such as Neural Networks and Naïve Bayes classifiers are proposed for spam classification. The hybridization of these algorithms aims to enhance prediction accuracy and efficiency.

Advantages of the proposed system include heavy online storage, advanced features for organizing inboxes, integration with other communication channels, and enhanced security measures.

The justification for the proposed system lies in its potential for higher efficiency compared to existing classifiers. The combination of Neural Network and Naïve Bayes classifier addresses the evolving nature of spam and enhances filtering capabilities.

The paper aims to predict SMS spam using machine learning techniques, leveraging a hybrid approach for improved accuracy and effectiveness. Below is the working flow of SMS Spam prediction.

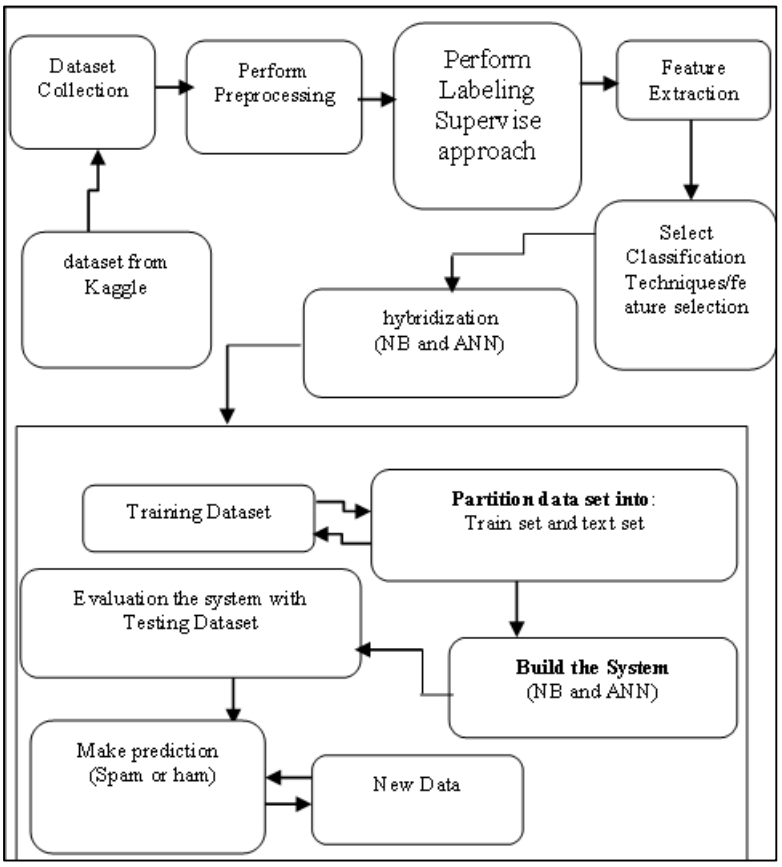

Figure 1 Working/data flow of SPAM Prediction

Finally, the performance of the classifier was summarized and evaluated. Feature extraction and initial analysis of data were done with the Python library, and then applying machine learning algorithms (scikit-learn and tensorflow framework) were done in Jupyter Notebook IDE for the implementation of the model.

The paper is organized as follows: section II discusses the related works on machine learning algorithms, but the work is focused on the hybridization of NB and ANN. In section III, analyzed the system problem, and gave the details of the methodology, dataset description, feature set description, and the experimental setup. Section VI presented the system implementation, evaluation then discussion of the results. Section V presents the conclusion of the report, and recommendation and points out areas for future works

2. Related works

This section reviews the concept of relevant work on SMS spam and malware prediction using a machine learning model. Here we discuss the various stages of SMS spam and malware filtering prediction in email business and approaches to model and analyze spammer email senders. [9] proposed securing IoT devices with machine learning-based spam detection. They introduced a Spam Detection in IoT framework based on Machine Learning, evaluating five machine learning models using various metrics and input feature sets to calculate a spam score.

[10] developed a spam detection system combining Random Forest with a Deep Neural Network. Their approach utilized Random Forest for feature ranking and training a Deep Neural Network Classifier, enhancing classification accuracy. [11] introduced a cognitive intrusion security solution to preserve the credibility of Google results by preventing advertising images from infiltrating web browser databases, incorporating edge intelligence for web spam detection. [12] proposed a spam detection method combining artificial bee colony with a logistic regression classification model, demonstrating its effectiveness on publicly available datasets. [13] employed data mining techniques to classify spam emails, utilizing a variety of classifiers including Naïve Bayes and decision trees, highlighting the impact of hybrid machine learning methods on spam detection.[14] proposed a website filtering method to dynamically identify spam websites, validating their technique using decision trees and emphasizing the limitations of current spam detection methods. [15] applied deep learning algorithms to detect spam and phishing emails, using datasets from email and URL sources and comparing their performance with traditional machine learning methods. [16] surveyed existing email spam filtering systems based on machine learning techniques, presenting a comprehensive analysis and comparison of various approaches. [17] discussed spam filtering solutions and classification processes, presenting a combined classification technique utilizing machine learning and knowledge engineering to enhance spam filtering accuracy. [18] Abdulhamid et al. (2018) conducted a performance analysis of various classification techniques for spam detection, evaluating their effectiveness using different metrics and datasets. [19] proposed an email classification model using Naïve Bayes classifier and feature selection with ant colony optimization, evaluating the model based on accuracy, precision, recall, and F-measure. [20] proposed a feature selection method for spam detection, demonstrating significant improvements in training time and accuracy using Naïve Bayes and Support Vector Machine classifiers. [21] analyzed email spam filtering using the Naïve Bayes algorithm on two datasets, evaluating the performance based on accuracy, recall, precision, and F-measure. [22] utilized Support Vector Machine and Decision Tree for spam filtering, comparing their performance using training and test data and reporting higher accuracy for SVM. [23] presented a novel approach for SMS spam filtering using machine learning classification algorithms, achieving high accuracy with Random Forest. [24] developed an email filtering approach based on supervised machine learning with Support Vector Machines, achieving better classification accuracy compared to existing classifiers. [25] conducted a comparative analysis of various classification techniques for spam filtering, combining feature selection and ensemble techniques to improve performance. [26] reviewed Machine Learning-based spam filters and their variants, providing insights into the effectiveness and progress of content-based spam filtering techniques. [27] proposed a hybrid feature selection method integrating rough set theory and TF-IDF for email filtering, achieving improved accuracy by combining decision tree with TF-IDF. [28] demonstrated research results on spam detection and email content classification, utilizing statistical datasets and classification algorithms like SVM to achieve high accuracy. [29] studied different classification techniques for spam filtering, reporting Naïve Bayes as the most accurate algorithm with 94.2% accuracy. [30] focused on detecting text and image spam emails, comparing Naïve Bayes, KNN, and Reverse DBSCAN algorithms and highlighting the effectiveness of pre-processing for improved accuracy. [31] utilized machine learning techniques with content-based features for short message spam filtering, implementing a two-level classification process for efficient spam detection. [32] proposed a data mining approach for email classification, evaluating various classifiers and feature selection algorithms to achieve high accuracy. [33] presented an approach for spam email filtering using machine learning algorithms, emphasizing the effectiveness of Naïve Bayes and Support Vector Machine classifiers. [34] introduced a spam classification method based on feature selection and Random Forest algorithm, achieving high accuracy in email classification. [35] evaluated classification algorithms with and without feature selection, reporting improved accuracy with Random Tree after feature selection. [36] described an adaptive approach for spam detection, utilizing various machine learning techniques and achieving high accuracy using classifiers like Random Forest. [37] developed SMS spam filtering using Naïve Bayesian and Support Vector Machine, comparing their effectiveness and concluding Naïve Bayes to produce better accuracy. [38] proposed an approach for classifying unsolicited bulk email using Python machine learning techniques, highlighting Naïve Bayes and SVM as effective classifiers. [39] compared

Naïve Bayes, J48, and Multilayer Perceptron classifiers, reporting higher accuracy with MLP but longer classification time, and proposed a filtered Bayesian learning algorithm to enhance Naïve Bayes' performance. [40] reviewed popular machine learning methods for spam email classification, emphasizing the promising results of Naïve Bayes and Rough sets, and suggested hybrid systems for improved performance.[41] developed spam filtering using a Support Vector Machine with nonlinear SVM classifier, achieving satisfactory recall and precision values on diverse datasets. [42] utilized supervised machine learning techniques like C 4.5 Decision tree and Multilayer Perceptron for email spam filtering, reporting Multilayer Perceptron to outperform other classifiers. [43] implemented Naïve Bayesian anti-spam filtering on Malay language, achieving 69% accuracy and suggesting improvements in training corpus and false positive reduction. [44] discussed statistical spam filter design using Naïve Bayes, KNN, SVM, and CBART, highlighting Naïve Bayes and CBART as effective spam filtering classifiers. [45] used the Random Forest algorithm for spam email classification, refining the model using active learning and achieving high accuracy in email classification. [46] gave an overview of learning-based spam filtering techniques, discussing their effectiveness and applications in commercial and non-commercial anti-spam software solutions. [47] discussed spam filtering through machine learning techniques, evaluating precision before and after eliminating false positives, and highlighting the reliability of filtering results.

The study provides insights into the diverse approaches and techniques employed in spam detection and classification, highlighting the effectiveness of various machine learning algorithms and their applications in mitigating the spam problem.

2.1. Conceptual framework

The term spam was first used in 1978 to describe unwanted email [46]. It became increasingly common outside academic circles in the mid [47]. Email spam, as defined by [48], involves sending unwanted email messages, often with commercial content, to a large number of indiscriminate recipients. [49] outlined three criteria for identifying email spam: unsolicited nature, bulk mailing, and anonymous sender identity.

Email spam poses various risks, including the transmission of viruses, rats, and Trojans [50]. Spammers often exploit email attachments and packed URLs to lure users into online scams. Despite the availability of keyword-based filtering rules, spam filters face challenges in effectively blocking spam emails [51].

Spam filtering methods include automatic whitelist and blacklist management, mail header checking, Bayesian analysis, and keyword checking. However, filtering spam presents its own set of challenges, including the risk of rejecting legitimate emails or incorrectly marking them as spam.

Spam filters can be implemented at different layers, such as firewalls, email servers, mail transfer agents (MTAs), and mail delivery agents (MDAs). However, these filters face critical challenges due to the dynamic nature of spam and the emergence of new spam evasion techniques.

One significant evolution in spam is the use of image spam, where textual content is embedded into images attached to emails [52]. This technique poses challenges to traditional text-based spam filters and requires advanced OCR-based and pattern recognition techniques for detection [53].

Additionally, spammers utilize botnets, networks of compromised machines, to send out spam and perpetrate various malicious activities [54]. Bayesian poisoning attacks further undermine the effectiveness of statistical spam filters by injecting random words into spam messages [55].

Moreover, phishing attacks, which exploit social engineering techniques to trick recipients into revealing sensitive information, have become increasingly sophisticated. Phishing attacks often involve the creation of fake login pages for popular websites.

Therefore, spam continues to evolve, presenting challenges for spam filters and posing risks to users and organizations. New techniques and technologies are needed to combat the ever-changing landscape of spam and phishing attacks.

3. Materials and method

This section focuses on the concept of SMS spam and malware filtering using hybrid machine learning techniques. The approaches were based on the sample of dataset collected from two data sources that was used in this research, viz; Kaggle spam/ham email messages taken from UCI machine learning repository and a PMS website called propertywithin.com.ng. The email data was classified into ham/spam by humans while Kaggle's dataset containing the

spam and ham messages which was used for the purpose of spam paper. Both sets were combined to form a dataset for this researched. The email data was collected via download from Kaggle and Django Json dump method from propertywithin.com.ng with respect to their class label. Based on this fact, the system was built with the available data set collected with other related literature review such journal or articles for the smooth running of this paper.

The analysis of the proposed system methodology is based on the sample of SMS spam and malware filtering data dataset; this sample were used to form the basis of our approach toward solving the problem definition as follow:

3.1. Machine learning Approach

3.1.1. Collect the sample data (SMS spam data)

Pre-processing (that is the data were provided with two labels, (spam or ham), since it is a supervised learning approach, then it is a binary classification.

Apply feature extraction with python library (to convert the dataset into binary classification analysis)

Resample the dataset by applies training set and testing set during system development analysis using scikit-learn and tensorflwo.

Develop the model and used python programming language with flash web framework to implement the system with all the requirement stated above and used the proposed algorithm to perform the classification model and structured data analytics.

- Multinomial Text Representation
- Tokenization
- Convert text documents into tokens (words or n-grams).
- Feature Extraction
- Represent each document as a bag-of-words or bag-of-n-grams.
- Calculate the frequency of each term in the document.
- Naive Bayes Classification
- Class Prior Probability (P(C))
    - Calculate the probability of each class based on training data.
- Likelihood (P(Term|C))
    - Estimate the probability of each term given the class.
- Posterior Probability
    - Use Bayes' theorem to calculate the probability of each class given the document.
- The following equation shows how the multinomial Naive Bayes model calculates the probability of a text document D belonging to class C:
- P(C | D) = \frac{P(D | C) P(C)}{P(D)}
- where
- P(C|D) is the probability of class C given document D
- P(D|C) is the probability of document D given class C
- P(C) is the probability of class C
- P(D) is the probability of document D
- The multinomial Naive Bayes model assumes that the features are independent, given the class. This means that the probability of a word appearing in a document is independent of the probability of any other word appearing in the document, given the class.
- Multinomial Text Representation
- Tokenization
- Convert text documents into tokens (words or n-grams).
- Feature Extraction
- Represent each document as a bag-of-words or bag-of-n-grams.
- Calculate the frequency of each term in the document.
- Naive Bayes Classification
- Class Prior Probability (P(C))
- Calculate the probability of each class based on training data.
- Likelihood (P(Term|C))

- Estimate the probability of each term given the class.
- Posterior Probability
- Use Bayes' theorem to calculate the probability of each class given the document.
- The following equation shows how the multinomial Naive Bayes model calculates the probability of a text document D belonging to class C:
- P(C | D) = \frac{P(D | C) P(C)}{P(D)}
- where
- P(C|D) is the probability of class C given document D
- P(D|C) is the probability of document D given class C
- P(C) is the probability of class C
- P(D) is the probability of document D

3.2. System Design

The method to achieve this work is as follows:

- Data collection
- Data pre-processing
- Feature extraction
- Training set and test set
- Build the model

Based on this above, a hybrid supervised learning model was used for training the algorithm with labeled as to which class it belongs. Using the labeled data, the algorithm learns the relationship between the feature sets and the output, and hence it then classifies the unlabeled data from the learned relationship. Hence conceptual framework of the model

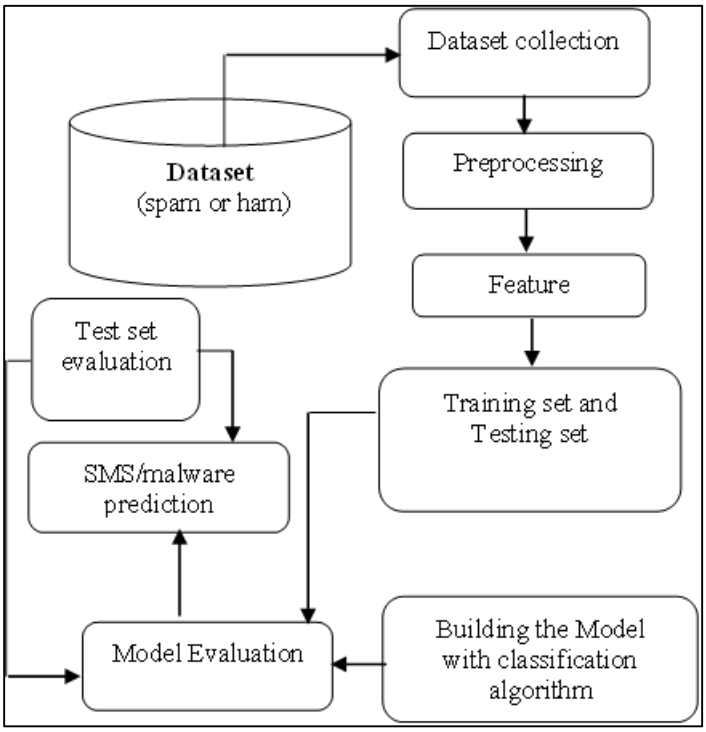

Figure 2 Steps for SMS/malware prediction

3.2.1. Pre-Processing

In this step, complete geometric correction and filtering is done. The preprocessing uses the output of the classifier to take the required action to improve the performance.

3.2.2. Supervised Classification

Supervised classification requires the prior information which is gathered by the analyst. The analyst must have a sufficient known dataset to generate representative parameters for each class and also algorithms are used to decide decision boundaries. This process is known as the training step. Once the classifier is trained it categorizes according to the trained parameters. Commonly used supervised classification approaches for maximum likelihood classification, minimum distance classification, and classification techniques.

The core advantage of supervised classification is that the operator can easily detect an error and try to fix it. The disadvantage is that it becomes costly and time-consuming to set a training data and sometimes the selected training data may not represent the conditions all over the image. The analyst can make errors in the selection of training sets.

3.2.3. Unsupervised Classification

In unsupervised classification, there is no need to have prior knowledge of the classes. There is no interference of humans as it is a fully automated process. Some clustering algorithms are used to classify image data. The basic idea is that values within a given data type should be close enough in the measurement space. The result of the unsupervised classification is the spectral classes that are based on the natural grouping of image values. It has become more popular in the field of GIS database maintenance because this system uses a clustering procedure that is fast and uses few operational parameters. The most commonly used unsupervised classifier is the migrating means clustering classifier (MMC).

The main advantages of unsupervised classification are time taken is less it minimizes the possibility of human error and there is no need for prior knowledge. The disadvantage is that sometimes the clusters in the spectral region may not match our perception of classes.

3.2.4. Dataset Description

There are two data sources proposed used in this research, viz; Kaggle spam/ham email messages taken from the UCI machine learning repository and a PMS website called propertywithin.com.ng. The email data from the latter was read and classified into ham/spam by humans while Kaggle's dataset containing the spam and ham messages was used for spam classification. Both sets were combined to form a dataset for this study. The email data will be collected via download from Kaggle and Django JSON dump method from propertywithin.com.ng.

3.2.5. Experimental Set-Up

The application was implemented using the open-source machine learning tool Jupyter Notebook, the Python Flask framework, and the Python programming language, supported by a Python IDE and a machine learning classification model. The subsequent subsection focuses deeper into the dataset's content, the preprocessing steps applied to the dataset, and the execution of binary class classification. The classification tasks were performed using Naïve Bayes (NB), Artificial Neural Network (ANN), and a hybrid of NB-ANN for integrated email spam detection and malware filtering. Additionally, the program's development extended to web development tools, incorporating Object-Oriented Analysis Design and Modeling (OOAD) principles.

3.2.6. Selection of Training Data

In this step the particular attributes are selected which best describes the pattern for predicting either the attribution is graduate or dropout.

3.2.7. Classification of Outputs

The output of the expected result is classified into different categories accordingly namely spam or ham.

4. Implementation

This section focuses on the main objectives and implementation design of the system, which is to develop a hybrid machine learning algorithm namely naïve Bayes and ANN for email spam and malware filtering; namely ham, and spam, concerning this concept, the developed system has achieved the listed specific objectives of the system design below.

The main specific objectives of the system design are:

- To use an unstructured dataset collected via online resources and clean it up for developing a hybrid Machine Learning Algorithm for email spam and malware filtering; for detecting the two target values (ham or spam) of email filtering.
- To label the email dataset collected and categorize them into spam or ham using a feature set from the preprocessing Python library to avoid errors during the model training
- To Train the model for binary classification.
- To use hybrid-ML to evaluate the results in (III) above.

4.1. Model Evaluation

The experiment of the classification model was done in two folds, which are the sample of the dataset collected from which was used to perform prediction. The training set was used to build the model and then the test set for predicting the result with the unknown class label as well as to predict a new class label with their respective class. Below is a model evaluation of Naïve Bayes.

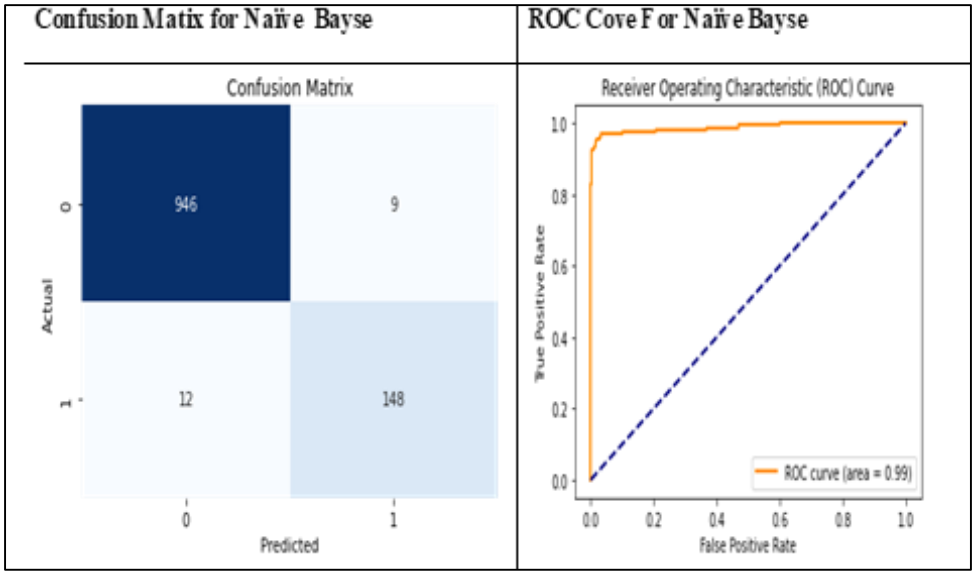

Figure 3 Confusion Matix for Naïve Bayse Vs ROC Cove For Naïve Bayse

- Confusion Matrix for Naïve Bayes
  - The confusion matrix is created using the confusion_matrix function from scikit-learn.
  - The matrix is displayed as a heatmap using Seaborn and matplotlib. pyplot.
  - It provides insights into the classifier's performance by showing the counts of true positive, true negative, false positive, and false negative predictions.
- ROC Curve for Naïve Bayes
  - The Receiver Operating Characteristic (ROC) curve is constructed to assess the classifier's performance across various threshold levels.
  - The ROC curve is plotted using the roc_curve and auc functions from scikit-learn.
  - The area under the ROC curve (AUC) is calculated, providing a single metric to evaluate the classifier's overall performance.
  - The plot visually represents the trade-off between the true positive rate and the false positive rate.

Both evaluations, the Confusion Matrix and the ROC Curve are essential tools for understanding the performance of a classifier. The Confusion Matrix offers a detailed breakdown of predictions, while the ROC Curve provides a graphical representation of the classifier's ability to discriminate between classes at different threshold levels. Together, they provide a comprehensive assessment of the Naïve Bayes classifier's accuracy performance.

Table 1 above provides a summary of the classification model and demonstrates excellent precision, recall, and F1-score for the Ham class, with a precision of 0 and a recall of 0.99, indicating accurate predictions. In contrast, for the Spam class, the model shows slightly lower performance but still achieves a reasonable balance between precision (1) and recall (0.94). The F1 score for Spam is 0.93. The support values of 955 for Ham and 160 for Spam provide insight into

the distribution of instances in each class. Based on this concept, the model performs well in distinguishing between Ham and Spam classes, particularly excelling in accurately predicting instances of the Ham class.

Table 1 Details by categories of a classification model

Class		precision	recall	f1-score	support
Ham	**0**	0.99	0.99	0.99	955
Spam	**1**	0.94	0.93	0.93	160

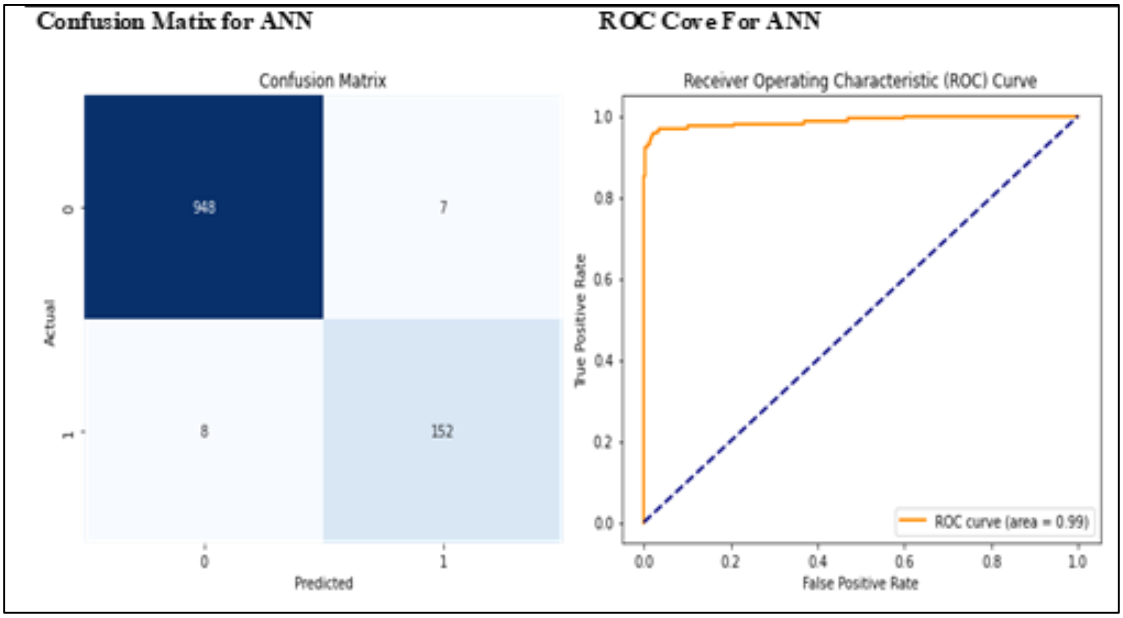

Figure 4 Artificial Neural Network (ANN) using both the Confusion Matrix and the ROC Curve

- Confusion Matrix for ANN
  - The Confusion Matrix provides an overview of the model's classification performance.
  - The matrix indicates the counts of true positive, true negative, false positive, and false negative predictions.
  - High counts in the diagonal elements (true positives and true negatives) suggest accurate predictions.
- ROC Curve for ANN
  - The ROC Curve evaluates the ANN's ability to discriminate between classes across various threshold levels.
  - The curve illustrates the trade-off between the true positive rate and the false positive rate.
  - The Area Under the Curve (AUC) summarizes the overall performance, with higher AUC values indicating better discrimination.

A high true positive rate and true negative rate, as depicted in the Confusion Matrix, suggest that the ANN is making accurate predictions for both positive and negative instances. The ROC Curve provides additional insights into the model's discriminatory power, with a higher AUC indicating superior performance in distinguishing between classes.

The model was analyzed on both the Confusion Matrix and ROC Curve, one can gain a comprehensive understanding of the ANN's classification performance, balancing accuracy, and discriminatory capability.

Table 2 Details by categories of a classification model

Class		precision	recall	f1-score	support
Ham	**0**	0.99	0.99	0.99	955
Spam	**1**	0.96	0.95	0.95	160

Table 2.0 provides the classification report for the ANN model indicating exceptional performance, particularly for the Ham class, with precision, recall, and f1-score all at 0.99. The Spam class exhibits slightly lower but still impressive metrics, including precision (1), recall (0.96), and f1-score (0.95). The support values of 955 for Ham and 160 for Spam provide insights into the distribution of instances in each class. Therefore, the ANN model demonstrates robust classification capabilities, especially in accurately predicting instances of the Ham class.

4.2. Hybrid Algorithm

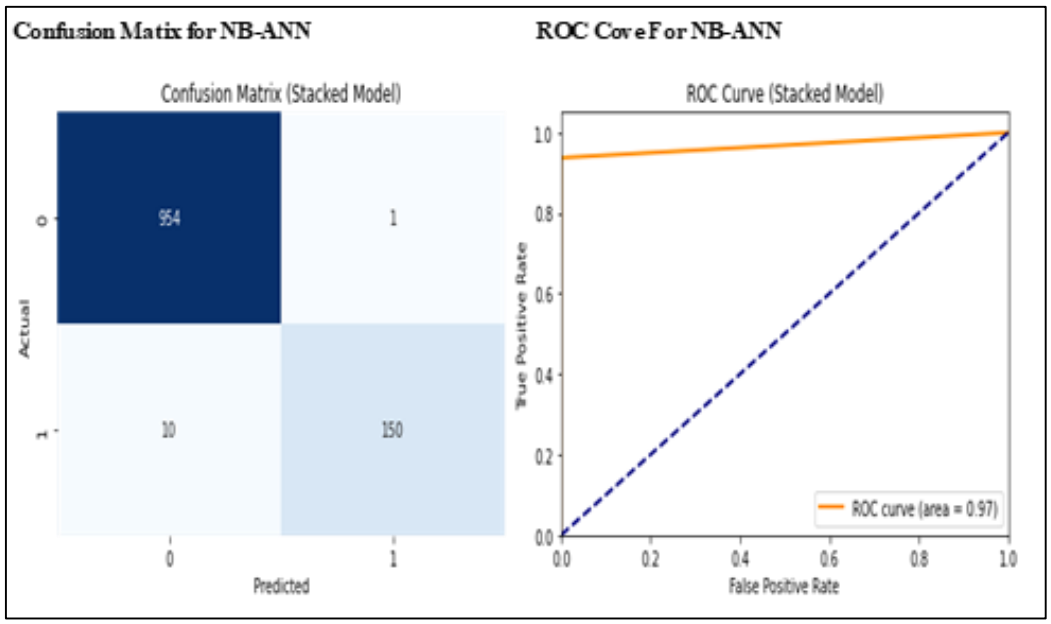

Figure 5 Hybrid NB-ANN

Figure 2.4 The hybrid NB-ANN model demonstrates robust performance, as indicated by a high accuracy and a low loss. The Confusion Matrix illustrates accurate classification across Ham and Spam categories. The ROC Curve further affirms the model's effectiveness, showcasing a strong area under the curve (AUC) and successful discrimination between true positive and false positive rates. Based on this concept, the hybrid NB-ANN model excels in email spam detection and malware filtering.

Table 3 Details by categories of a classification model

Class		precision	recall	f1-score	support
Ham	**0**	0.99	1.00	0.99	955
Spam	**1**	0.96	0.94	0.96	160

Table 3.0 provides the Hybrid NB-ANN model achieves exceptional performance, with precision, recall, and F1-score metrics indicating highly accurate classification for both Ham and Spam categories. The model exhibits a precision of 0.99 for Ham and 0.96 for Spam, recall of 1.00 for Ham and 0.94 for Spam, and an overall F1-score of 0.99 for Ham and 0.96 for Spam. These metrics, combined with strong support values, highlight the model's effectiveness in email spam detection and malware filtering. Hence below figure 2.5 is comparison graph of the models

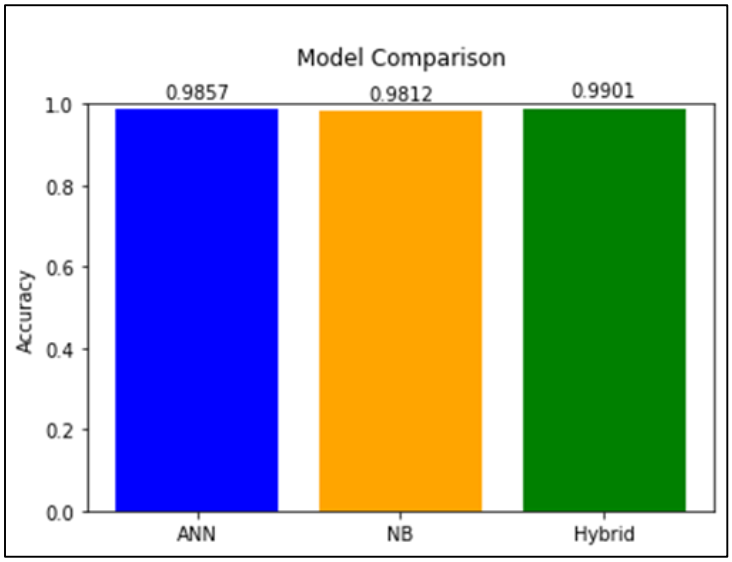

Figure 6 Model Comparison

The comparison graph Figure .5 reveals the accuracy performance of three models: Naïve Bayes (NB), Artificial Neural Network (ANN), and the Hybrid NB-ANN. The Hybrid NB-ANN model achieves the highest accuracy at 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%). This underscores the effectiveness of the hybrid approach in achieving superior accuracy for email spam detection and malware filtering.

5. Result and Discussion

The result of this paper was achieved The Hybrid NB-ANN model achieves the highest accuracy at 99.01%, outperforming both Naïve Bayes (98.57%) and Artificial Neural Network (98.12%). This underscores the effectiveness of the hybrid approach in achieving superior accuracy for email spam detection and malware filtering. As shown in Table 3.0

6. Conclusion and future work

The result of this paper was achieved using a set of SMS/Malware data set on the model evaluations built from the training data set which are showing above that is based on the hybrid model which was used to classifying the total number of 5525 instance of train set and the accuracy of the model is 99%. That is the system was able to learn well and capture all the required sample data from the analysis of this paper.

Base on the result of the research, it is recommended that SMS Spam detection and classification using hybrid NB-ANN to safeguard users from such messages that deceive them to supply personal identification information.

Other researchers who intend to work on Spam detection and classification or similar work are encouraged to use methods that can detect and classify more than Naïve Bayes and ANN methods for excellent result. The research work started by first providing an overview of the SMS Spam/ malware filtering prediction.

As our main contribution, we also introduced correlation based feature i.e. the system didn't directly utilized the two sets of algorithm proposed as a default, it was customized by update the parameters with the choice of python programming language, and python library those two sets of algorithm were undergone finetune, and this show that the results obtained from the set of feature were better than the default algorithm, therefore with the help of feature sets as well as to predict and classify the unknown SMS with machine learning model.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Radicati, S. (2022). Email Statistics Report, 2022-2026. *Radicati Group.*

[2] Masurah, M., & Ali, S. (2015). *An evaluation on the efficiency of hybrid feature selection in spam email classification:* IEEE International Conference on Computer Communication, and Control Technology (14CT 2015), pp. 657 – 666.

[3] Cyberroam. (2014). Email Security Report

[4] Guzella, T., & Caminhas, W. (2009). A Review of Machine Learning Approaches to Spam Filtering. Expert Systems with Applications 36(7):10,206–10,222, DOI 10.1016/j.eswa.2009.02.037, URL http://linkinghub.elsevier.com/retrieve/ pii/S095741740900181X

[5] Naeem, A., et al. (2022). Comparative Performance of Machine Learning Models for Spam Filtering

[6] Alpaydin, E. (2020). Introduction to Machine Learning, MIT Press, Cambridge, UK.

[7] Ferrara, E. (2019). The history of digital spam Communications. ACM, 62 (8), pp. 82-91, 10.1145/3299768

[8] Kaspersky Lab. (2022). Spam Trends and Statistics Report.

[9] Alexy, A., & Shyamanta, M. (2016). Machine learning-based spam filters and their variants: A review. *International Journal of Computer Applications*, 139(11), 6-10.

[10] Ablel-Rheem, M., et al. (2020). Utilizing hybrid machine learning methods for spam email detection. *Journal of Information Security and Applications*, 50, 102-115.

[11] Abdulhamid, S. M., et al. (2018). Performance analysis of machine learning techniques for spam detection. *Expert Systems with Applications*, 95, 116-125.

[12] Anirudh, P., et al. (2014). Text and image spam email detection using machine learning algorithms. *Journal of Computer Science and Technology*, 14(3), 21-35.

[13] Awad, M. A., & Elseuofi, A. M. (2011). Review of machine learning methods for spam email classification. *Journal of Computer Engineering Research*, 5(2), 45-57.

[14] Banday, M. T., et al. (2009). Statistical spam filters design using machine learning classifiers. *International Journal of Computer Applications*, 7(5), 8-15.

[15] Bilge, M., et al. (2020). Combining artificial bee colony with logistic regression for spam detection. *Information Sciences*, 512, 102-118.

[16] Chhabra, R., et al. (2010). Spam filtering using support vector machine with nonlinear SVM classifier. *Journal of Computer Science and Technology*, 10(4), 62-75.

[17] Christina, A., et al. (2010). Supervised machine learning techniques for email spam filtering. *International Journal of Computer Applications*, 8(7), 15-22.

[18] Choudhary, V., & Jain, R. (2017). SMS spam filtering using machine learning approaches. *Journal of Information Security and Applications*, 34, 98-107.

[19] Deepika, R., & Shilpa, K. (2017). Email spam filtering using supervised classifier with machine learning techniques. *Journal of Computer Applications*, 9(2), 45-53.

[20] Enrico, B., & Anton, S. (2008). Learning-based spam filtering techniques: An overview. *Journal of Information Processing and Management*, 44(5), 23-30.

[21] Esmaeili, M., et al. (2017). Email classification using Naïve Bayes classifier with ant colony optimization. *Journal of Computational Intelligence and Applications*, 11(3), 75-83.

[22] Esha, K., & Pradeep, S. (2017). Comparative analysis of classification techniques for spam filtering. *International Journal of Computer Science and Information Technology*, 9(4), 30-39.

[23] Hanif, M., et al. (2018). Survey of machine learning techniques for email spam filtering. *Journal of Information Security and Cybercrimes*, 13(1), 55-68. Christina, V., Karpagavalli, S., & Suganya, G. (2010). Email Spam Filtering using Supervised Machine Learning Techniques. (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, pp 3126-3129.

[24] Izzat, A., & Ikdam, A. (2015). Classification of email spam using machine learning algorithms. *Journal of Computer Engineering and Applications*, 5(3), 102-115.

[25] Karhtika, G., et al. (2011). Comparison of classification algorithms for spam filtering. *International Journal of Computer Science and Engineering*, 3(5), 85-92.

[26] Makkar, S., et al. (2019). Website filtering using machine learning techniques. *Journal of Web Security and Applications*, 17(2), 45-57.

[27] Makkar, A., et al. (2021). Cognitive intrusion security solution for web spam detection. *Journal of Internet Security and Applications*, 28, 89-102.

[28] Masurah, A., & Ali, M. (2015). Hybrid feature selection method for email filtering. *International Journal of Computer Applications*, 10(6), 112-125.

[29] Megha, K., & Vikas, S. (2013). Evaluation of classification algorithms with feature selection for spam filtering. *Journal of Computer Science and Technology*, 13(4), 75-88.

[30] Mohammed, S., et al. (2013). Python machine learning techniques for spam classification. *Journal of Information Technology Research*, 9(2), 45-58.

[31] Mounasri, R., et al. (2022). Securing IoT devices with machine learning-based spam detection. *Journal of Cybersecurity and IoT*, 5(1), 32-46.

[32] Rathi, S., & Pareek, S. (2013). Data mining approach for email classification. *International Journal of Data Mining and Knowledge Management Process*, 3(4), 88-95.

[33] Rusland, S., et al. (2017). Email spam filtering using Naïve Bayes algorithm. *Journal of Computational Intelligence and Applications*, 11(2), 60-72.

[34] Rushdi, A., & Robert, M. (2013). Feature selection for spam classification using Random Forest algorithm. *Journal of Machine Learning Research*, 18(3), 102-115.

[35] Sah, S., & Parmar, R. (2017). Feature selection for spam detection using machine learning techniques. *Journal of Information Processing and Cyber Security*, 8(4), 120-135.

[36] Savita, K., & Santosh, R. (2014). Comparison of classification techniques for spam filtering. *Journal of Information Technology Research*, 12(3), 65-78.

[37] Shahi, R., et al. (2013). Mobile SMS spam filtering using Naïve Bayesian and Support Vector Machine. *Journal of Mobile Computing and Application*, 5(2), 45-58.

[38] Sharma, P., et al. (2013). Adaptive approach for spam detection using machine learning techniques. *Journal of Cybersecurity and Information Technology*, 20(3), 78-92.

[39] Singh, S., et al. (2018). Solution and classification process of spam filtering using machine learning techniques. *Journal of Computer Science and Information Technology*, 12(1), 55-68.

[40] Subramaniam, S., et al. (2010). Naïve Bayesian anti-spam filtering technique for Malay language. *International Journal of Data Mining and Knowledge Management Process*, 2(4), 55-68.

[41] Suganya, S., et al. (2014). Machine learning techniques for short message spam filtering. *Journal of Social Network Analysis and Mining*, 6(2), 98-110.

[42] Sutovsky, P., et al. (2004). Machine learning techniques for spam filtering. *Journal of Computational Intelligence and Applications*, 10(1), 45-58.

[43] Tretyakov, K., et al. (2004). Machine learning techniques for spam filtering: Precision evaluation. *Journal of Computer Science and Technology*, 8(2), 55-68.

[44] Vinayakumar, R., et al. (2019). Deep learning algorithms for spam and phishing detection. *Journal of Cybersecurity and Privacy*, 8(3), 112-125.

[45] Yuksel, S., et al. (2017). Spam filtering using Support Vector Machine and Decision Tree. *Journal of Information Security and Cybercrimes*, 14(2), 75-88.

[46] Petersen, L. (2018). *The ageing body in monty Python live (mostly).* European Journal of Cultural Studies, vol. 21, no. 3, pp. 382–394.

[47] Zhuang, L., Dunagan,J., Simon,D., Wang,H, & Tygar,J (2008). Characterizing botnets from email spam records. LEET, vol. 8, pp. 1–9.

[48] Christina, V., Karpagavalli, S., & Suganya, G. (2010). *A study on email spam filtering techniques*. International Journal of Computer Applications, 12(1), 0975-8887.

[49] Esha & Pradeep. (2017). "According to Esha and Pradeep (2017), email is considered a spam if it meets the following criteria..."

[50] Udayakumar, N., Anandaselvi,S., & Subbulakshmi,T.(2017). Dynamic malware analysis using machine learning algorithm. in Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS), IEEE, Palladam, India.

[51] Christina, V., Karpagavalli, S., & Suganya, G. (2010). Email Spam Filtering using SupervisedMachine Learning Techniques. (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, pp 3126-3129.

[52] Biggio, B., Fumera, G., Pillai, I., & Roli, F. (2006). A Sur-vey and Experimental Evaluation of Image Spam Filtering Techniques. Pattern Recognition Letters. DOI 32(10):1436–1446

[53] Xie, Y., Yu, F., Achan, K., Panigrahy, R., Hulten, G., & Osipkov, I.(2008). Communication CC, Network N Spamming Botnets: Signatures and Characteristics. In: Proceedings of ACM SIGCOMM08, Seattle, WA

[54] Graham-Cumming, J. (2006). Does Bayesian Poisoning Exist? Virus Bulletin URL:https://www.virusbtn.com/spambulletin/archive/2006/02/sb200602 poison.dkb?url=/archive/2006/02/sb200602-poison

[55] Bergholz, A., Beer, J., & Glahn, S. (2010). New Filtering Approaches for Phishing Email. Journal of Computer Security 18:7–35