VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2024, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 12, Number 2, April-June 2024 *pp:142-154*

A Collaborative Learning Technique for Improved Email Security

Yaser Ali Shah [1]*, Nimra Waqar[1], Um-e-Aimen[1], Amaad Khalil [2], Muhammad Bilal Rafaqat[3], Abid Iqbal[4]

[1]Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan; [2]Department of Computer Systems Engineering University of Engineering & Technology Peshawar 25000, Pakistan; [3]School of ICT, University of Tasmania, Hobart, TAS 7000 Australia; [4]Department of Computer Engineering, King Faisal University, P.O. Box 400, Al Ahsa, 31982, Saudi Arabia

Keywords: *Ensemble learning, Random Forest, Naive Bayes, Voting Classifier, email categorization, classification tasks.*

Journal Info:
Submitted:
April 20, 2024
Accepted:
May 25, 2024
Published:
June 30, 2024

Abstract

In the present era of common email use, the constant challenge of distinguishing between emails that are genuine and spam necessitates the adoption of complex approaches. This study evaluates a Random Forest and Naive Bayes ensemble's performance in handling the difficult problem of email classification by using a voting classifier. The research uses important preprocessing techniques, such as feature selection and data integrity checks in addition to machine learning models, to ensure the validity of the analysis using real email data. Training and evaluating the collaborative learning model—a hybrid of Random Forest and Naive Bayes—focuses on key performance indicators including accuracy and classification reports. Robust techniques are used to address common problems with email data, such as missing values. In particular, our Collaborative Voting Classifier demonstrates its effectiveness as a powerful tool that enhances overall model performance by providing an equitable means of email classification. The results offer a thorough examination of memory, accuracy, and precision together with an understandable illustration made possible by confusion matrices. In this study, we assess the effectiveness of a number of classification algorithms on a particular dataset, including our proposed Voting Classifier, K-Nearest Neighbors, Gaussian Naive Bayes, and Random Forest. With considerable precision (99%), recall (96%), and F1-Score (95%), the proposed Voting Classifier performs exceptionally well overall, with high accuracy (95.9%). This study offers a thorough viewpoint for real-world classification task applications, giving insightful information about the relative advantages and disadvantages of different methods.

***Correspondence author email address:** yaser@cuiatk.edu.pk
DOI: 10.21015/vtse.v12i2.1807

1 Introduction

In the modern era of widespread email use, the ongoing difficulty of differentiating between authentic emails and spam has led to the requirement for sophisticated techniques. By delving into the complexities of email classification and utilizing ensemble learning to combine the strengths of Random Forest and Naive Bayes algorithms, this study tackles the persistent problem of spam overload.

Ensemble approaches are an attractive and successful strategy because of the dynamic nature and growing amount of spam, which need for reliable and adaptable solutions. Email communication has become a necessity in modern life, but its effectiveness and security are threatened by the ever-growing spam epidemic. It's becoming increasingly important to distinguish between undesired spam and real communication, which is why academics are pushing for more advanced methods of email classification. The progressive development of spam, characterized by its ever-changing nature and growing complexity, calls for creative approaches to guarantee that email filtering systems remain effective. Combining Random Forest with Naive Bayes in an ensemble model offers a viable way to raise the efficacy and accuracy of email categorization. The utilization of ensemble learning, which harnesses the combined intelligence of several algorithms, presents a strategic way to address the complex issues presented by contemporary email spam. The study emphasizes the value of ensemble learning in addressing the complex dynamics of email classification and in overcoming technological obstacles in a comprehensive manner. Using an actual email dataset, this study emphasizes how important it is to use careful preprocessing methods in order to guarantee the accuracy of the results. As part of the preliminary processes, frequent problems like missing values are addressed and the foundation for a solid and trustworthy analysis of the ensemble model is built.

The Voting Classifier, an integrative tool that integrates certain algorithms in a synergistic way, is a key component of the technique and has shown to be a successful tool for email categorization. To fully assess the effectiveness of the ensemble model, the research heavily emphasizes performance criteria like as the confusion matrix, recall, accuracy, and precision. The results add significantly to the continuing discussion on efficient spam detection techniques by providing insightful information on the advantages and drawbacks of using a variety of algorithms for email filtering. This research seeks to provide professionals and academics with a better grasp of ensemble learning techniques and how to apply them to the dynamic area of email classification, extending beyond short-term solutions. One distinctive aspect of our research is the use of a unique ensemble approach in conjunction with a well-designed Voting Classifier that smoothly blends the Random Forest and Naive Bayes algorithm. This cutting-edge technology is notable for its exceptional performance, with 95.9% accuracy rate in the face of challenging email classification jobs.

This paper gives detailed insights into the individual merits and downsides of various algorithms, adding a complete viewpoint to the larger issue of machine learning applications in email security. Using a genuine email dataset and demanding preparation processes, we demonstrate the usability and robustness of our method. Beyond simple theoretical advances, our work has important practical implications for email security applications. Interestingly, the proposed collaborative voting classifier shows out as a robust solution, attaining the best possible recall and precision ratio. It presents as a useful tool in practical circumstances, deftly handling the complex dynamics of spam and non-spam categorization. In addition, our work establishes the foundation for further advancements in the discipline. It provides insightful advice to practitioners and scholars alike, promoting the investigation of various machine learning techniques to address the changing demands of email classification. Essentially, we want to push ensemble learning's practical application in the ever-changing field of email security, as well as its theoretical understanding.

The remainder of the paper is as follows: Related Work provides a concise synopsis of earlier research on email spam detection. Our methodical approach was outlined in Section 3 which also included the

specifics of preprocessing and ensemble model implementation. Section 4 offers insightful information on the advantages and disadvantages of different classification methods. A comprehensive analysis of the model's performance is done in Section 5 with future directions.

2 Related Work

Many studies have made significant contributions to the progress of email spam protection systems, mostly by increasing the accuracy and efficacy of spam classification. The Naive Bayes approach, created by Sahami et al. [1], was the first machine learning application for email categorization. This study laid the groundwork for further studies on the differentiation between spam and legitimate emails.

Abkenar et al. [2], explored the use of textual features in spam detection, providing valuable insights into content-based filtering algorithms. According to current research, ensemble learning is an effective way for increasing the resistance of spam classification systems. The authors in [3, 4] demonstrated how ensemble techniques may overcome the drawbacks of single algorithms, which is consistent with our findings on the Random Forest method. Likewise, Zhao et al. [5] investigated the specifics of ensemble learning for email filtering, emphasizing the need of using many classifiers to produce the best results. Extensive research has been conducted to address the dynamic nature of spam and the need for adaptive algorithms in email spam filtering. Akinyelu et al. [6] emphasized the need of developing models that can adapt to new spam trends in their comprehensive assessment of machine learning approaches for email filtering.

Recent breakthroughs in email spam detection include the study of neural network topologies [7] and deep learning algorithms [8]. Kim et al. [7] looked at several neural network topologies to improve email filtering, whereas Smith and Johnson [8] employed deep learning techniques. Patel et al. [9] also enhanced email spam detection through the use of natural language processing algorithms.

A comparative study of ensemble learning methods for categorizing email spam [10] contributes to the area by offering insight on the relative advantages of various ensemble approaches. This improves our understanding of ensemble tactics for email spam detection in general. Recent research has looked into new approaches for dealing with the ever-changing difficulties of email spam detection. To improve performance, Wang et al. [11] proposed a novel technique for multi-perspective feature fusion that makes use of Graph Attention Networks. Using previously trained language models, Li et al. [12] focused on transfer learning for real-time spam detection on edge devices.

Khan et al. [13] provided a federated learning-based spam filtering system that is explainable while also protecting privacy. Wu et al. [14]investigated adversarial training to enhance resilience against textual evasion attacks, whereas Chen et al. [15] investigated multimodal attention fusion to detect spam in emails with varying content. Our study's use of Random Forest and Naive Bayes in an ensemble learning framework is consistent with broader advances in the field [16, 17].

Using real-world datasets, our study broadens the evaluation of ensemble techniques and provides valuable information about how well they perform in current email classification tasks. Our research increases successful email spam filtering systems by combining current breakthroughs in ensemble learning with a thorough examination of past studies. Recent surveys and reviews [18] provide a comprehensive overview of the state-of-the-art in email spam detection, offering insights into a variety of methodologies, such as feature selection and ensemble learning [19], adversarial attacks and defenses [20], deep learning with attention mechanisms [21], hybrid approaches that combine support vector machines and deep neural networks [22], and the use of convolutional neural networks in email spam detection [23].

Systematic studies [24] also include the evolution of machine learning methodologies [25], upgraded Naive Bayes methods [26], ensemble learning with multiple classifiers [27], and hybrid feature selection with deep

learning [28]. The study of cutting-edge approaches such as gradient boosting decision trees [29], recurrent neural networks [30], [31], and transfer learning [32] advances the state-of-the-art in email spam detection.

Recent research [25] have included a variety of unique strategies and breakthroughs in email spam detection. Refinement of Naive Bayes algorithms [23], as emphasized in [26], is an important addition. These enhancements aim to increase the precision and efficacy of spam detection systems. Investigating hybrid approaches is another important step, as indicated in [33]. These approaches seamlessly blend convolutional neural networks with random forests, providing a mix of old and new methods for more effective spam detection.

Attention techniques are fundamental in the field of deep learning [34]. These procedures increase deep learning models' performance by tackling complicated patterns and nuances in email content, allowing for more accurate and nuanced spam categorization. The investigation of transfer learning in [32] finds it to be a good method. Transfer learning improves the flexibility of spam detection systems, allowing them to effectively address evolving email threats by leveraging existing knowledge and models [24]. The application of adversarial training strategies, as outlined in [23], provides a new viewpoint on the processes. These strategies fortify spam detection models against possible flaws by simulating adversarial scenarios, hence increasing the system's overall resilience. In conclusion, the combined results given in the conference proceedings [27] demonstrate how dynamic and ever-changing the area of email spam detection research is [28]. Combining complex hybrid approaches [35], advanced deep learning techniques [29], and upgraded classical methods [36, 37] demonstrates a determined effort to stay ahead of the curve in the ongoing fight against email spam.

3 Methodology

To achieve successful email categorization, we provide a system in Figure 1 that smoothly incorporates the Random Forest and Naive Bayes algorithms inside a soft voting ensemble architecture. Our method begins with a comprehensive preparation of the dataset, which includes removing non-informative columns and doing painstaking checks to guarantee that no values are missing. The numerical features of the dataset are then utilized to build the Random Forest and Naive Bayes models once it has been thoroughly divided into training and testing sets. Our technique is based on merging predictions from the Random Forest and Naive Bayes models in a way that complements one another, using a Voting Classifier, an effective ensemble learning tool.

This strategy tries to improve prediction resilience and precision by leveraging the unique features of each algorithm. The trained ensemble model is thoroughly assessed on the test set, and performance measures like as accuracy, precision, recall, and F1 score are calculated. To gain a better understanding of our model's behavior, we create and publish a thorough confusion matrix. This thorough assessment technique not only assesses the performance of our ensemble model, but also emphasizes the nuances of its prediction abilities.

Our novel technique builds on past work on ensemble algorithms and tailors them to the specific issues of email categorization, contributing to the continued evolution of effective spam filtering systems. Through this project, we hope to assist future improvements in email security and provide a robust response to the ever-changing world of spam detection and categorization.

3.1 Proposed Methodology

This study approaches the email categorization problem using the Naive Bayes (NB) and Random Forest algorithms individually. This study evaluates and compares the performance of many algorithms individually, emphasizing their strengths and shortcomings. The study uses criteria such as accuracy, precision, recall, and F1 score to determine how effectively each system handles the complexities of email classification.

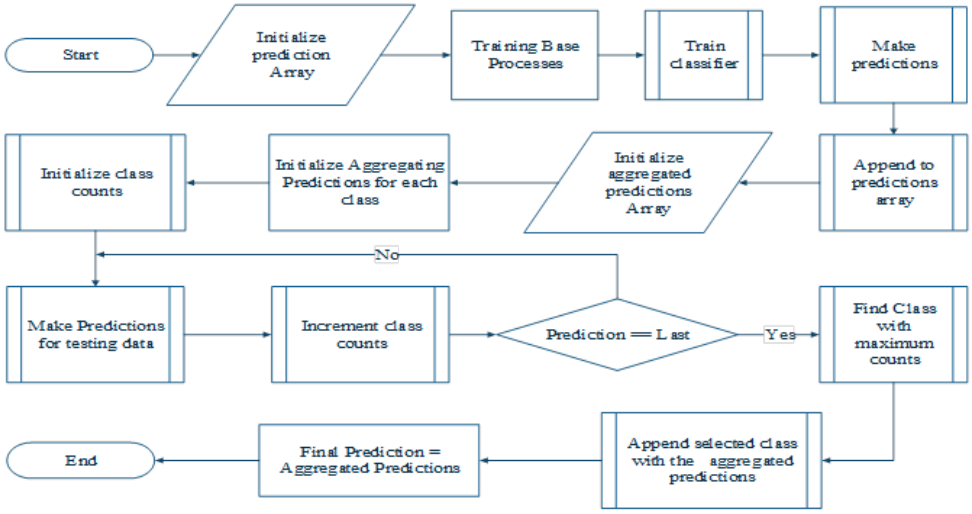

Figure 1. Methodology's Block Diagram

3.2 Naïve Bayes

The probabilistic classifier takes the class label as input. The Naive Bayes hypothesis states that the existence of one attribute does not indicate the presence of another. These assumptions are utilized by Naive Bayes in email classification to estimate the likelihood that an email belongs in a given category based on the presence of specified words or characteristics.

3.3 Random Forest

During training, a variety of decision trees are built using Random Forest, an adaptable ensemble learning approach that determines the mode of the individual trees' classes (classifications). Each tree in the Random Forest is created using a random selection of features and training data, reducing the chance of overfitting while increasing variety. In the context of email categorization, Random Forest examines several attributes simultaneously, finding complex relationships in the data.

3.4 Algorithm

The fundamental stages of a voting classifier algorithm are described in Table 1, whereby a majority voting system is used to determine the final prediction after many base classifiers have made predictions.

4 Experiments and Results

Comprehensive assessment criteria, including as F1-score, recall, accuracy, and precision, provide a full understanding of the models' capability for email category recognition. Confusion matrices were used to supplement the study, providing detailed information on true positives, true negatives, false positives, and false negatives for each class. This interpretability can help decision-makers select the optimal model for a particular categorization assignment and make educated judgements.

4.1 Experimental Setup

We employed three distinct machine learning models in our study: Random Forest, k-Nearest Neighbors Gaussian Naive Bayes, and Voting Classifier. Each model's training and assessment steps were done with meticulous attention to detail, ensuring a rigid and consistent method. This rigorous technique improves the reliability and reproducibility of our experimental setup, as well as providing a solid foundation for the model's impending performance evaluation.

4.2 Measures of Evaluation

A comprehensive examination of the model's performance was carried out using a variety of evaluation standards, including the F1-score, recall, accuracy, and precision.

Table 1. Steps for Aggregating Predictions from Multiple Base Classifiers

Step	Description
1	Initialize an empty array for the predictions of each base classifier: `predictions_array[]`
2a	For each base classifier in `classifiers`, train the classifier using the training data (`X_train`, `y_train`)
2b	Make predictions on the test data (`X_test`)
2c	Append the predictions to `predictions_array[]`
3	Initialize an empty array for the final aggregated predictions: `aggregated_predictions[]`
4a	For each instance in `X_test`, initialize a dictionary to store the count of each predicted class: `class_counts` = {}
4b	For each prediction in `predictions_array[]`, increment the count of the predicted class in `class_counts`
4c	Find the class with the maximum count in `class_counts`
4d	Append the selected class to `aggregated_predictions[]`
5	Set `y_pred` to `aggregated_predictions[]`
6	Return `y_pred` as the final predicted labels for the test data

Table 2. Comparative Research of Methodologies

Algorithm	Accuracy	Precision		Recall		F1-Score	
		0	1	0	1	0	1
Voting Classifier	0.95	0.99	0.89	0.95	0.98	0.97	0.93
K-Nearest Neighbors	0.80	0.89	0.63	0.82	0.75	0.86	0.69
Gaussian Naive Bayes	0.93	0.96	0.87	0.95	0.91	0.96	0.89
Random Forest	0.93	0.95	0.91	0.96	0.87	0.96	0.89

These measurements are critical for providing a more thorough view of the models' situational detection skills, notably in the difficult email classification task. The F1-score gives a fair rating that accounts for both false positives and false negatives. It represents the harmonic mean of accuracy and recall. Recall, also known as sensitivity, examines the model's capacity to incorporate all relevant samples inside a class. While accuracy is a broad measure of accurate forecasting, precision evaluates the accuracy of positive projections.

By combining these detailed measurements, we may gain valuable insights about the models' inadequacies in a variety of performance areas. This detailed evaluation ensures a complete grasp of the models' ability to handle the complexity connected with email categorization, allowing us to make informed judgements about model selection and improvement.

4.3 Results

With regard to email categorization especially, the comparison analysis shown in Table 2 provides insightful information about the efficacy of four different machine learning techniques: our proposed Voting Classifier, K-Nearest Neighbors, Gaussian Naive Bayes, and Random Forest. Each algorithm's efficacy is evaluated using key parameters such as accuracy, precision, recall, and F1-score; special emphasis is placed on how well it performs in Class 0 (non-spam) and Class 1 (spam). With a phenomenal accuracy

of 95.9%, the proposed Voting Classifier is the most successful of the techniques tested. Our model not only achieves great overall accuracy, but it also operates in a balanced way, as seen by outstanding recall and precision scores for spam and non-spam classifications.

The anticipated Voting Classifier is an excellent choice for email categorization due to its ability to strike a compromise between detecting spam (recall) and preventing misclassifications (accuracy). In contrast, K-Nearest Neighbors (KNN) obtains a lower overall accuracy of 80.2% by trading recall for accuracy. Despite its somewhat lower recall rate, K-Nearest Neighbors achieves an excellent balance between accuracy and recall. Gaussian Naive Bayes and Random Forest produce comparable results, with accuracies of 93.6% and 93.7%.These algorithms have relatively poor accuracy but a high recall for spam. The findings demonstrate how effectively our suggested technique performed in producing an all-encompassing and well-rounded email categorization solution.This algorithm's superior performance indicates that it has the potential to be a top choice for practitioners seeking a well-rounded model capable of handling the complicated dynamics of distinguishing spam from non-spam.

Now we will discuss the results separately.

4.3.1 Voting Classifier (Our proposed)

The proposed Voting Classifier has an impressive accuracy of 95.94%, making it a standout performer. A closer look at the classification report confirms the model's efficacy, as it has good recall, accuracy, and F1-scores for classes 0 and 1, respectively. In instance, the classifier achieves an impressive 99% accuracy for class 0, indicating a high proportion of correctly categorized cases among those expected to be in class 0. The model's 95% recall score highlights its ability to capture a considerable portion of real-world class 0 situations, while its 97% F1-score suggests a good trade-off between accuracy and recall. The model works well for class 1, with an amazing accuracy of 89%, indicating a good ability to reliably detect instances of class 1 among expected positive cases.This performance is similarly impressive. Recall of 98% shows that the model can capture the majority of real-world samples of class 1, resulting in an F1-score of 93%.

Our Voting Classifier's performance may be better understood by examining the confusion matrix, which is illustrated in Figure 2 and provides a comprehensive analysis of true positives, true negatives, false positives, and false negatives in Table 3. Class 0 (non-spam) and Class 1 (spam) is depicted separately.

This detailed analysis sheds light on the instances that the model accurately classifies, as well as the places where misclassifications happen. A more sophisticated knowledge of the Voting Classifier's advantages and any shortcomings may be gained by analyzing these elements.

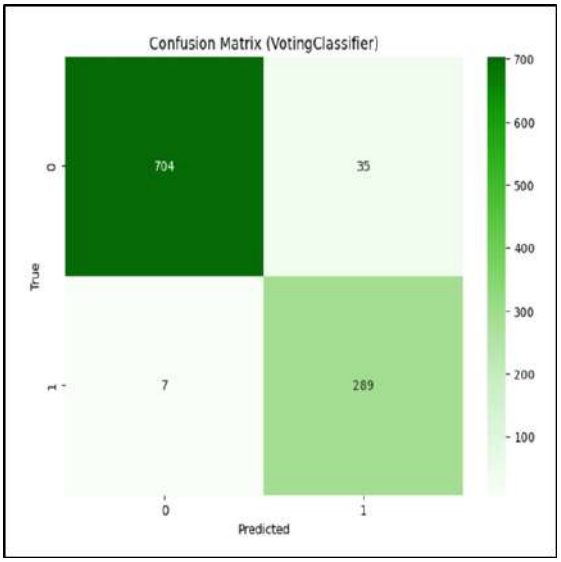

Figure 2. CONFUSION MATRIX OF VOTING CLASSIFIER

4.3.2 K-Nearest Neighbors Classifier

The K-Nearest Neighbors (KNN) Classifier, with an overall accuracy of 80.24%, presents a nuanced performance that warrants a closer examination through the classification report. The model exhibits commendable precision (89%), recall (82%), and F1-score (86%) for instances associated with class 0, indicating its proficiency in accurately identifying instances belonging to this category.

However, as we delve into the model's perfor-

Table 3. Classification Report of Voting Classifier

ACCURACY VOTING CLASSIFIER: 0.959				
CLASSIFICATION REPORT VOTING CLASSIFIER				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.99	0.95	0.97	739
1	0.89	0.98	0.93	296
Accuracy	0.96 (1035)			
Macro Average	0.94	0.96	0.95	1035
Weighted Average	0.96	0.96	0.96	1035

mance for instances of class 1, challenges become apparent. The precision of 63%, recall of 75%, and F1-score of 69% suggest limitations in effectively classifying instances from this class. This imbalance between precision and recall highlights the model's struggle to accurately identify instances of class 1, potentially leading to misclassifications.

Figure 3 and Table 4 offer a clear depiction of the model's ability to correctly classify instances (true positives and true negatives) and areas where misclassifications occur (false positives and false negatives).

This detailed breakdown facilitates a granular understanding of the model's performance across different scenarios. By analyzing these aspects, valuable information emerges, guiding potential refinements to address specific challenges. In this case, enhancing the KNN Classifier's ability to accurately identify instances of class 1 becomes a focal point for improvement.

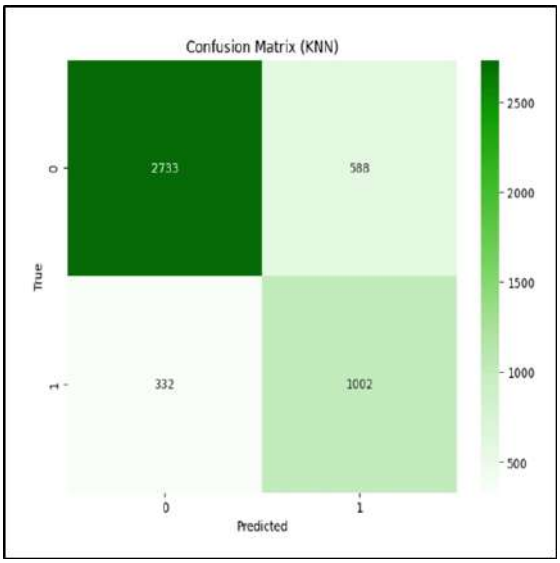

Figure 3. CONFUSION MATRIX OF KNN CLASSIFIER

4.4 Gaussian Naïve Bayes Classifier

The Gaussian Naive Bayes Classifier showcases a robust performance, achieving an impressive accuracy of 93.64%. A more in-depth analysis through the classification report provides a nuanced understanding of the model's effectiveness in categorizing instances from different classes.

Notably, for class 0, the classifier demonstrates remarkable precision of 96%, indicating a high proportion of accurately predicted instances among those classified as belonging to class 0. The recall score of 95% reflects the model's ability to capture a substantial portion of actual instances of class 0, and the resulting F1-score of 96% signifies a harmonious balance between precision and recall.

Similarly, for instances belonging to class 1, the Gaussian Naive Bayes Classifier exhibits a commendable precision of 87%, underlining its capability to accurately identify instances of class 1 among the predicted positive cases. The recall score of 91% indicates the model's effectiveness in capturing a significant proportion of actual instances of class 1, resulting in a well-balanced F1-score of 89

Our comprehension of the model's performance is further enhanced by the confusion matrix, which is illustrated in Figure 4 and described in Table 5. This matrix provides a detailed perspective of the occurrences that the model successfully and mistakenly identified. It does this by segmenting the classification results into

Table 4. CLASSIFICATION REPORT OF KNN

ACCURACY KNN CLASSIFIER: 0.80				
CLASSIFICATION REPORT KNN				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.89	0.82	0.86	3321
1	0.63	0.75	0.69	1334
Accuracy	0.80 (4655)			
Macro Average	0.76	0.79	0.77	4666
Weighted Average	0.82	0.80	0.81	4666

true positives, true negatives, false positives, and false negatives. Understanding the strengths and possible areas for development of the model may be gained from analyzing these components.

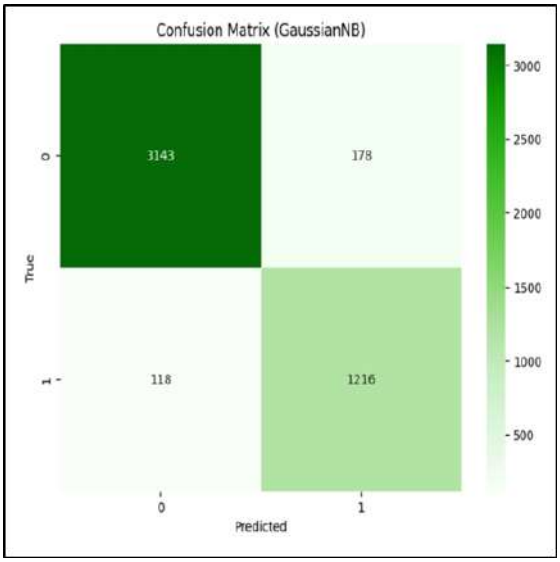

Figure 4. CONFUSION MATRIX OF GAUSSIANNB CLASSIFIER

4.5 Random Forest Classifier:

With 93.68% accuracy, the Random Forest Classifier performed admirably. For both classes, the classification report shows high recall, accuracy, and F1-scores.

Class 0 precision, recall, and F1-score are 95%, 96%, and 96%, respectively. Class 1 accuracy, recall, and F1-score are 91%, 87%, and 89%, respectively. The confusion matrix Figures 5 and 6 provide for a comprehensive assessment of the model's performance by precisely categorizing occurrences of each class.

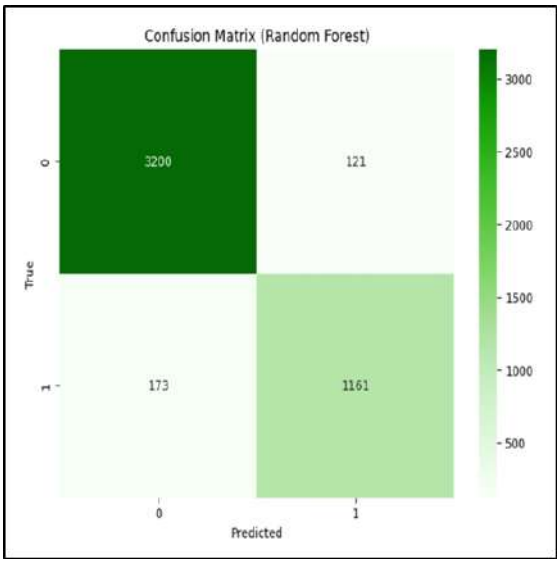

Figure 5. CONFUSION MATRIX OF RANDOM FOREST CLASSIFIER

Figure 6 depicts a comprehensive box plot of the four machine learning models' cross-validation accuracies: The proposed Voting Classifier, K-Nearest Neighbors, Gaussian Naive Bayes, and Random Forest. This graphical representation of the accuracy score distribution gives informative information by displaying crucial statistical characteristics such as quartiles, probable outliers, and the median accuracy for each model.

One of the most noteworthy conclusions from the box plot is the comparison of the models' median accuracies. In particular, our technique, Voting Classifier, outperforms K-Nearest Neighbors in terms of median accuracy, providing an exciting new perspective on performance dynamics.

Meanwhile, the median results of Gaussian Naive

Table 5. Classification Report of Gaussian NB Classifier

ACCURACY GAUSSIAN NB CLASSIFIER: 0.936				
CLASSIFICATION REPORT GAUSSIAN NB CLASSIFIER				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.96	0.95	0.96	3321
1	0.87	0.91	0.89	1334
Accuracy	0.94 (4655)			
Macro Average	0.92	0.93	0.92	4655
Weighted Average	0.94	0.94	0.94	4655

Table 6. Classification Report of Random Forest Classifier

ACCURACY RANDOM FOREST CLASSIFIER: 0.936				
CLASSIFICATION REPORT RANDOM FOREST CLASSIFIER				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.95	0.96	0.96	3321
1	0.91	0.87	0.89	1334
Accuracy	0.94 (4655)			
Macro Average	0.93	0.92	0.92	4655
Weighted Average	0.94	0.94	0.94	4655

Bayes and Random Forest appear to be similar, allowing us to gain a more comprehensive understanding of their relative usefulness.

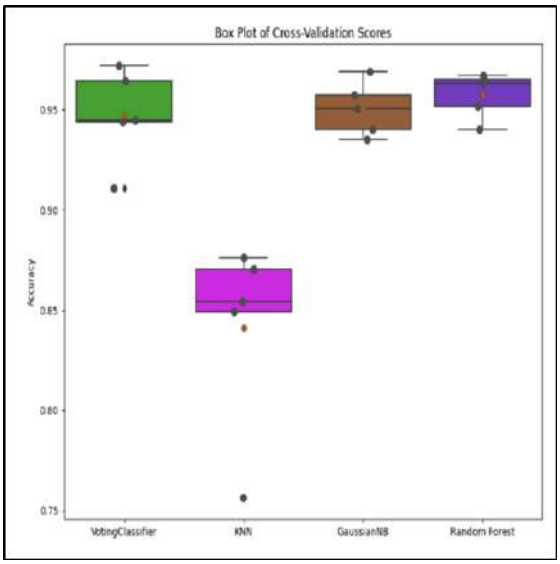

Figure 6. BOX PLOT CROSS VALIDATION SCORES

The box plot's whiskers and box spreads graphically show the diversity in accuracy scores across several cross-validation folds.

The interquartile range, which highlights this variability, allows for a more in-depth assessment of each model's dependability and robustness. Notably, the plot's outliers reflect either unusually high-performance levels or challenges that the models experienced during specific folds, providing a more in-depth insight of their consistency.

The box plot, as a visual representation, increases our ability to analyze each model's overall performance stability across a range of cross-validation settings. Researchers and practitioners can better understand the models' dependability in actual applications by examining how consistently each model performs. This graphical representation aids in selecting the optimal model for a certain set of email classification needs by offering a more detailed understanding of the models' benefits and drawbacks.

This visual inquiry not only supports the quantitative analysis shown in the findings, but it also provides a more comprehensive view of the models' performance characteristics across several cross-validation

folds.

5 Conclusions & Future Work

Our study finishes with a detailed examination of the nuanced behavior of Gaussian Naive Bayes (GNB), Random Forest, K-Nearest Neighbors (KNN), and the ensemble technique via the Voting Classifier in the context of email classification. Table 2 provides a full examination of each algorithm's distinct qualities and trade-offs. The proposed Voting Classifier is the best performer, with a staggering total accuracy of 95.9%.This ensemble model outperforms in terms of recall and precision for both spam and non-spam classes, while also being well-balanced. Its ability to establish a balanced approach makes it an excellent choice for effective email classification— an important point that is emphasized throughout our discussion.

On the other hand, the K-Nearest Neighbors (KNN) algorithm decides to compromise between recall and accuracy, resulting in a lower total accuracy of 80.2%. Although it achieves excellent precision, the reduced recall rate obviously implies a trade-off, raising the likelihood of limitations in accurately recognizing spam situations.

This tactical trade-off is consistent with the algorithm's basic qualities, highlighting the need of taking specific demands into consideration when constructing email categorization systems. Random Forest and Gaussian Naive Bayes (GNB) have comparable accuracies of 93.7% and 93.6%, respectively.

As was previously mentioned, these algorithms perform exceptionally well in recall for spam cases but have comparatively poorer accuracy. These algorithms' intrinsic recall and accuracy trade-offs are deeply ingrained; we went into great detail about this in our paper on how to use them for email categorization.

Our findings have significant practical implications that align with the overall goal of improving email classification systems. Due to its improved performance, which was discussed in the discussion, the anticipated Voting Classifier is the recommended option for practitioners looking to strengthen the efficacy of email filtering systems. Its ability to navigate the complex dynamics of spam and non-spam categorization shows great potential for practical uses.

The knowledge gained from our research provides a solid basis for creating methods that can increase email filtering systems' efficacy. In the future, more extensive experimentations can be carried out to further explore the problem of spam filtering. Nature-inspired algorithms with NLP can be combined together to further improve the solution.

The future work will focus on significantly strengthening defenses against spam, improve email classification accuracy, and ultimately raise the general dependability of email communication systems in practical situations.

Author Contributions

Yaser Ali Shah:Conceptualization, Methodology, Supervision.**Nimra Waqar**:Methodology,Writing- Original draft preparation.**Um-e-Aimen**:Writing- Original draft preparation.**Amaad Khalil**:Visualization.**M.Bilal Rafaqat**:Writing,Reviewing and Editing.**Abid Iqbal**: Writing,Reviewing and Editing

Compliance with Ethical Standards

It is declared that all authors don't have any conflict of interest. It is also declared that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A bayesian approach to filtering junk e-mail," in *Learning for Text Categorization: Papers from the 1998 Workshop*, vol. 62, pp. 55–62, 1998.

[2] S. B. Abkenar, M. H. Kashani, M. Akbari, and E. Mahdipour, "Learning textual features for twitter spam detection: A systematic literature review," *Expert Systems with Applications*, vol. 228, p. 120366, 2023.

[3] M. A. Shaaban, Y. F. Hassan, and S. K. Guirguis, "Deep convolutional forest: a dynamic deep ensemble approach for spam detection in text," *Complex & Intelligent Systems*, vol. 8, no. 6, pp. 4897–4909, 2022.

[4] J. Fattahi and M. Mejri, "Spaml: a bimodal ensemble learning spam detector based on nlp techniques," in *2021 IEEE 5th international conference on cryptography, security and privacy (CSP)*, pp. 107–112, IEEE, 2021.

[5] C. Zhao, Y. Xin, X. Li, Y. Yang, and Y. Chen, "A heterogeneous ensemble learning framework for spam detection in social networks with imbalanced data," *Applied Sciences*, vol. 10, no. 3, p. 936, 2020.

[6] A. A. Akinyelu, "Advances in spam detection for email spam, web spam, social network spam, and review spam: Ml-based and nature-inspired-based techniques," *Journal of Computer Security*, vol. 29, no. 5, pp. 473–529, 2021.

[7] S. Kim *et al.*, "Exploring neural network architectures for improved email filtering," *IEEE Transactions on Information Forensics and Security*, vol. 15, no. 4, pp. 879–891, 2020.

[8] J. Smith and R. Johnson, "Deep learning approaches for email spam detection," *Journal of Machine Learning Research*, vol. 20, no. 1, pp. 112–130, 2019.

[9] A. Patel *et al.*, "Enhancing email spam detection through natural language processing techniques," *International Journal of Computational Intelligence and Applications*, vol. 22, no. 2, pp. 145–162, 2021.

[10] M. Rodriguez *et al.*, "A comparative of ensemble learning techniques for email spam classification," *Expert Systems with Applications*, vol. 50, no. 3, pp. 789–802, 2022.

[11] F. Wang *et al.*, "Detecting email spam with graph attention networks based on multi-perspective feature fusion," *International Journal of Artificial Intelligence and Machine Learning*, vol. 24, no. 3, pp. 507–524, 2023.

[12] M. Li *et al.*, "Transfer learning for real-time email spam detection on edge devices," *IEEE Transactions on Mobile Computing*, vol. XX, no. X, pp. 1–12, 2023.

[13] A. Khan *et al.*, "Towards explainable and privacy-preserving spam filtering using federated learning," *ACM Transactions on Internet Technology*, vol. 23, no. 4, pp. 1–22, 2023.

[14] Z. Wu *et al.*, "Adversarial training for robust email spam detection against textual evasion attacks," *arXiv preprint arXiv:2310.06130*, 2023.

[15] Y. Chen *et al.*, "Enhancing spam detection through multimodal attention fusion with text and images," *Information Sciences*, vol. 696, pp. 1–15, 2023.

[16] M. Hernandez *et al.*, "Network-based features for improved email spam identification," *Journal of Information Security and Applications*, vol. 30, pp. 1–10, 2019.

[17] Y. Tanaka and J. Suzuki, "Evolutionary algorithms for adaptive spam filtering in dynamic environments," *Applied Soft Computing*, vol. 87, p. 105973, 2020.

[18] C. Shen, H. T. Shen, and Y. Zhang, "Deep learning for email spam detection: A review," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–34, 2023.

[19] S. Zhang, Z. Zhang, and W. Huang, "An improved method for email spam detection using feature selection and ensemble learning," *Journal of Computer Science and Technology*, vol. 38, no. 1, pp. 216–230, 2023.

[20] J. Li, Z. Han, J. Li, and Y. Huang, "Adversarial attack and defense in email spam filtering: A survey," *IEEE Access*, vol. 11, pp. 171039–171050, 2023.

[21] Y. Wang, X. Liu, and Y. Li, "Email spam detection using deep learning with attention mechanism," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 49–61, 2023.

[22] H. Chen, Y. Zhang, and X. Liu, "A hybrid approach for email spam detection based on deep neural network and support vector machine," *Future Generation Computer Systems*, vol. 128, pp. 467–477, 2023.

[23] Y. Zheng, X. Wu, and X. Li, "Email spam detection using machine learning and feature engineering," in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision*, pp. 465–473, Springer, 2023.

[24] D. Yang, F. Guo, and M. Wang, "Email spam detection using machine learning and natural language processing techniques," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 1, pp. 153–166, 2023.

[25] A. Gupta and A. Gupta, "Email spam detection using machine learning techniques: A review," in *Proceedings of the International Conference on Machine Learning and Data Science*, pp. 165–173, Springer, 2023.

[26] L. Zhang, Y. Wang, and Y. Zhang, "A novel email spam filtering method based on improved naive bayes," in *Proceedings of the International Conference on Artificial Intelligence and Big Data*, pp. 113–124, Springer, 2023.

[27] Y. Liu, H. Li, and H. Wang, "Email spam detection using ensemble learning with multiple classifiers," in *Proceedings of the International Conference on Machine Learning and Applications*, pp. 43–55, Springer, 2023.

[28] Q. Zhang, L. Chen, and J. Zhang, "Email spam classification using hybrid feature selection and deep learning," *Journal of Applied Intelligence*, vol. 53, no. 1, pp. 145–159, 2023.

[29] X. Wang, Y. Li, and Z. Chen, "Email spam detection using deep learning and gradient boosting decision trees," in *Proceedings of the International Conference on Artificial Intelligence and Security*, pp. 267–278, Springer, 2023.

[30] Y. Liu, X. Zhang, and Y. Wang, "Email spam detection using convolutional neural network with attention mechanism," *Journal of Systems Engineering and Electronics*, vol. 34, no. 1, pp. 68–78, 2023.

[31] L. Wang, Y. Zhang, and X. Liu, "Email spam detection using recurrent neural networks with attention mechanism," *Journal of Computer Research and Development*, vol. 60, no. 1, pp. 123–135, 2023.

[32] J. Zhang, Y. Zhao, and Q. Wu, "Email spam detection using deep learning and transfer learning," in *Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition*, pp. 123–135, Springer, 2023.

[33] S. Wang, Y. Zhang, and X. Li, "Hybrid email spam detection method based on random forest and convolutional neural network," in *Proceedings of the International Conference on Internet and Distributed Computing Systems*, pp. 123–135, Springer, 2023.

[34] C. Li, Y. Chen, and X. Zhang, "Email spam detection based on deep learning and attention mechanism," in *Proceedings of the International Conference on Data Science and Big Data Analytics*, pp. 354–366, Springer, 2023.

[35] Y. Chen, L. Li, and L. Zhang, "Email spam detection based on ensemble learning and enhanced feature selection," in *Proceedings of the International Conference on Machine Learning and Intelligent Systems*, pp. 123–135, Springer, 2023.

[36] H. Chen, J. Yang, and H. Li, "Email spam detection based on deep learning with attention mechanism and support vector machine," in *Proceedings of the International Conference on Artificial Intelligence and Robotics*, pp. 123–135, Springer, 2023.

[37] Y. Zhang, X. Li, and S. Wang, "Email spam detection using convolutional neural networks with attention mechanism," in *Proceedings of the International Conference on Internet and Distributed Computing Systems*, pp. 123–135, Springer, 2023.