Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Phishing Email Detection With Machine Learning Algorrithms

Aalia Shaukat
MS (Cyber Security), SZABIST University, Islamabad, Pakistan

Tayyaba Javed
Department of Software Engineering, BIIT, Rawalpindi, Pakistan

Saquib Nazir
MS (Cyber Security), SZABIST University, Islamabad, Pakistan

ABSTRACT

Phishing emails are considered one of the greatest serious cybersecurity threats due to the exploitation of human vulnerabilities to leak private data or make compromises to an online infrastructure. Four conventional algorithm include NB, LR, DT, and RF are used in the suggested lightweight ML architecture for phishing email identification. TF-IDF based feature extraction, text normalization, and methodical data pretreatment are all included. A collection of phishing emails with both positive and negative class labels was used to train and evaluate individual algorithms. The accuracy, precision, recall, F1 score, ROC curve, and confusion matrix will all be used to measure each model's performance. With accuracy levels exceeding 98%, outstanding precision, and recall, experimental results demonstrate that RF and LR are the best evaluated models. While the NB model performed good at a very cheap computing cost, the Decision Tree showed good scores and a modest accuracy that was still interpretable. The comparison analysis further demonstrates that the proposed lightweight models performance competitiveness with deep learning based approaches with significantly lower computational resource consumption. Overall, this research highlights the fact that classical machine learning algorithms, when combined with effective preprocessing and TF-IDF representation, form an efficient and practical solution for phishing email detection to suit real time cyber-security applications.

Keywords— cybersecurity, machine learning, email classification, phishing email detection, and zero-day phishing

Introduction

Phishing emails are becoming among the most prevalent and destructive forms of cyberattacks because they allow threat actors to acquire private information, compromise accounts, and launch large-scale invasions. Because phishing techniques are getting increasingly sophisticated, traditional rule-based detection can no longer be deemed enough; hence, this motivates researchers to develop ML and DL models for more accurate detection. A number of related studies have identified that data-driven models provide much better performance compared to manual or signature-based

Online ISSN 3007-3197
Print ISSN 3007-3189

approaches while analyzing the content and metadata of e-mails. [1].

According to recent studies, CNNs, RNNs, LSTMs, and hybrid neural architectures can use deep learning techniques to extract semantic patterns from email text, increasing detection accuracy. Deep learning-based phishing email detection systems have a lot of potential to handle complex language manipulation and dynamic phishing keywords in contemporary cybersecurity environments [1]. Similar to this, additional studies using machine learning classifiers like Random Forest and SVM on phishing email datasets demonstrate that these models will undoubtedly detect the pattern suggesting malicious intent and perform well even on noisy or imbalanced datasets [2].

Because deep learning models extract contextual and sequential information from email bodies, headers, and embedded links, they are generally superior to traditional machine learning techniques, according to systematic reviews of deep learning-based phishing detection. [3]. A global trend toward automation and intelligent systems that can understand linguistic structures and attacker behavior is brought forth by such reviews. Such a change becomes extremely important in terms of reducing cybersecurity risk because email communications are becoming more and more common within enterprises. [3]

Many supervised machine learning algorithms detect phishing emails and commonly achieve higher accuracy once several feature types are combined, like lexical, syntactic, and structural email attributes. Methods with Decision Trees, Random Forest, and SVM, are amongst the methods that manifest high detection rates on benchmark datasets, underpinning the role of feature engineering and data preprocessing for successful model performances [4]. Some comparative works also conclude that integrating NLP techniques significantly facilitates classifying with higher accuracy by enabling models to analyze contextual meaning and phrasing associated with phishing attempts [5].

Studies on the combination of machine learning and natural language processing have demonstrated that more dependable phishing detection systems are produced when textual feature extraction techniques like TF-IDF, Bag of Words, and word embedding are combined with ML classifiers. Combining these methods allows a system to capture the subtle linguistic distinctions that distinguish a phishing email from a valid one, which can be challenging to accomplish with any manual method. [6]

This is further supported by more thorough research on the methods used for phishing and spam filtering, which has found that ML and DL models are important in widespread email threats. These studies have also shown that machine learning (ML) is beneficial because it ensures scalability, automation, and adaptability in filtering all of which are critical for managing high email traffic volumes in businesses [7]. Additionally, as demonstrated by cloud-based phishing email detection frameworks [8], shifting ML or DL calculations to the cloud may improve scalability and real-time performance to make the detection systems successful in dynamic threat landscapes.

Using the advantages of various learning methodologies, recent hybrid models that combine ensembles of neural networks with machine learning algorithms achieve improved detection accuracy. Recent research has combined deep learning and machine learning techniques; hybrid models exhibit improved generalization and robustness for intricate phishing patterns. [9] Deep learning-based email analysis systems carry on this

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

work by automatically learning hierarchical representations from unprocessed email text, reducing the need for manual feature engineering [10].

Detecting BEC attacks, which are also more targeted and context-sensitive than typical phishing emails, has been the topic of several recent research. Systematic reviews on BEC detection support the effectiveness of machine learning models using behavioral and linguistic characteristics in identifying high risk communication patterns across companies [11]. Similar in-depth research on deep learning-based phishing detection methods supports the advantages of deep models in comprehending semantic complexity across phony emails [12].

Since many spam-filtering approaches overlap with dangerous email identification, machine learning research on email spam detection also advances our understanding of phishing detection. [13]. The research which employs URL based machine learning models shows that the analysis of URLs embedded within phishing emails adds another dimension to their detection and significantly improves the performance against malicious redirection attacks [14]. Dedicated studies concerning the classification of phishing emails based on Random Forest show that ensemble tree-based methods can outperform individual classifiers owing to their capability in modeling nonlinear relationships and reducing overfitting. Similarly, related deep learning techniques for spam email detection have demonstrated that neural networks are able to automatically learn hidden patterns in message content that contribute to high detection rates on various datasets. It has also been determined that, in NLP based systematic reviews, extracting linguistic, semantic, and syntactic features from emails significantly improves the accuracy of phishing detection. This finding tends to showcase that language-based features remain relevant for deception detection [17]. Phishing detection has also made extensive use of old ML techniques like NB, DT, and SVM that provide strong classification results when paired with pertinent email properties [18]. Recent research on spear-phishing detection demonstrates once more how machine learning can assess context-driven and customized phishing communications that are impossible to identify using a straightforward rule-based method. [19]

Literature review

Deep learning methods have been intensively researched for phishing email recognition. The majority of studies assert that neural architectures like CNNs and LSTMs can automatically extract semantic and structural patterns from email text. Although they typically demand large datasets and lengthy training durations, many of these models perform better than traditional machine learning techniques because they typically capture contextual and sequential links better.[1] The traditional methods of machine learning have also been investigated, and the majority of them still function well as long as the right preprocessing procedures are followed, such as tokenization, feature extraction, and TF IDF representations. According to a comparative analysis, Naive Bayes, Random Forests, Decision Trees, and SVM can be used for lightweight phishing detection systems Due to their exceptional accuracy and computational economy. [2]

Comparative investigation reveals that Naive Bayes, Random Forests, Decision Trees, and Support Vector Machines all exhibit excellent accuracy while maintaining

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

computational efficiency, creation them suitable for lightweight phishing detection systems. [2]

Various deep learning based approaches have been holistically reviewed, and research has highlighted the advantages of learning high-level linguistic features directly from raw text. Surveys emphasize that deep models are promising but face limitations in explain ability, resource requirements, and dataset inconsistencies across studies, which make performance comparisons difficult. [3]. The traditional machine learning pipelines based on feature engineering still show a very strong baseline for phishing detection. These works demonstrate that text preprocessing, n-gram representation, and handcrafted linguistic indicators are very important in enhancing the performance of classical models without using any deep learning. [4]. The linguistic cues, semantic embedding, and context-aware features greatly improve the ML models' ability to discern between legitimate and phishing emails, according to additional comparative analysis of the NLP-based features and machine learning classifiers. Results like these support the mixture of NLP and ML due to the improved prediction accuracy. [5]

Recent research on graph-based deep learning shown that word-document graphs processed by GCNs can encode structural correlations that are not achievable with conventional techniques. Despite achieving extremely high accuracy, these approaches have limited interpretability, are computationally demanding, and are susceptible to the quality of graph building. These features point to the significance of a more comprehensive comparison with traditional algorithms. [6]. A review of research on phishing and spam detection reveals that fraudulent emails have a number of traits in common, therefore models must be adaptable to deal with new attack patterns. Although machine learning is successful, it must incorporate techniques to deal with hostile content and extremely unbalanced datasets. [7]. To get beyond the computational limitations of deep learning models, several academics have also investigated cloud-based frameworks. These studies demonstrate that real-time processing while managing resource-intensive algorithms can be made possible by shifting detection jobs to cloud servers, albeit such systems may raise privacy and latency issues. [8]

While deep networks are typically better at identifying subtle phishing indications, classical models retain their advantages in speed, simplicity, and interpretability, according to the few comparative studies that assess ML and DL models jointly. As a result, hybrid systems which incorporate the best aspects of both paradigms are suggested as a comprehensive solution. [9]. When it comes to identifying sequential dependencies in email body processing, pure deep learning models continue to perform quite well. While keeping in mind the demanding training requirements they require, these show considerable promise for situations requiring high accuracy. They immediately pick up patterns that conventional feature engineering might not have found. [10]. Content-only detection methods are insufficient for context-specific, customized attacks, according to recent research on high-risk threats like BEC. These shown that increasing the number of environmental, behavioral, or organization-specific variables significantly improves detection. [11]

Comprehensive analyses of deep learning techniques also show that inconsistent pre-processing and split procedures cause gaps in comparison validity by inflating accuracy

Online ISSN 3007-3197
Print ISSN 3007-3189

estimates, which results in the lack of consistent methodology for evaluation across datasets. [12] Research on machine learning in spam detection has shown that traditional algorithms are still useful as long as they include robust feature engineering. This finding supports the notion that robust pre-processing pipelines are beneficial for phishing detection as well. [13]. Email phishing detection systems are strengthened when URL-level lexical and host-based elements are included, according to proof of concept in URL-based phishing detection studies. It has been demonstrated that combining these characteristics with email-body analysis improves overall robustness. [14].

Because the Random Forest classifier can handle noisy, high-dimensional feature spaces while taking interpretable feature rankings into account, multiple trials validate its applicability for phishing detection. [15]. Deep learning techniques for spam identification have proven to be quite successful in capturing email semantics. Using DL techniques, it is demonstrated that they are quite independent of manually created characteristics, albeit at the expense of time. [16]. In NLP-based systematic reviews, the linguistic characteristics of the emails such as vocabulary richness, sentiment signals, syntactic patterns, and semantic anomalies are thought to be crucial for distinguishing phishing emails and incorporating them into detection models. [17] The effectiveness of TF-IDF features, n-grams, and metadata properties in improving classifier performance is still demonstrated by general machine learning research in phishing detection. These techniques are suitable for real-time applications and are lightweight. [18]

Spear phishing requires more than simply textual patterns, according to research on targeted phishing detection, and will benefit from the use of sender-receiver context, communication history, and user role to identify highly tailored attacks. [19] ML-based comparative studies verify that feature engineering, hyper-parameter tuning, and appropriate validation techniques have a substantial impact on model performance and necessitate a methodical assessment. [20]. Phishing classification accuracy has increased thanks to hybrid feature sets that incorporate lexical, structural, and metadata cues. The ensemble classifiers use several decision bounds to minimize mis-classification. [21]. In order to prevent improper learning from noise, intelligent DL-based detection systems require strict regularization, even though they show promise for adaptive learning, in which the model continuously updates for new phishing trends. [22]. When used with ensemble learning, under sampling effectively addresses the issue of severe class imbalance in phishing datasets, increasing recall while minimizing precision loss. [23]

It has been demonstrated by sophisticated deep learning frameworks that employ multi-task learning that shared representations between email categorization tasks enhance phishing detection and category identification, increasing the models' efficacy. [24] ML-based phishing detection study further supports the significance of various feature family combinations, including n-grams, statistical indicators, and embedded semantics, in maximizing classifiers' conclusions. [25]. Research on lean machine learning pipelines provides more proof of the need for scalable and lightweight real-time phishing detection solutions, demonstrating that computationally efficient models may nevertheless achieve high accuracy. [26] Spam detection studies have underlined

Online ISSN 3007-3197
Print ISSN 3007-3189

the importance of preprocessing quality by showing that the structure and cleanliness of the input text make a big difference in the performance of the models. [27]. The new architectures, such as transformers, outperform classic neural networks due to the complex semantic structure and long-range interaction captured in e-mails, according to the study that compared deep learning approaches. [28]. In exhaustive comparisons of traditional machine learning methods, ensemble models remain hard to beat, often outperforming single classifiers, especially on noisy or heterogeneous phishing datasets. [29]

It remains a very paramount area of study at the URL level. It is supported by the fact that malicious URLs in phishing emails could be successfully detected by ML classifiers trained on lexical and structural URL patterns. [30]. NLP based phishing surveys reveal that the integration of syntactic, semantic, and metadata features provides holistic detection systems. Such integrated approaches have typically outperformed single feature approaches. [31]. Machine learning based phishing detection research continues to reinforce the importance of realistic dataset splitting approaches. That is, random shuffling often overestimates performance and fails to represent real world conditions. [32] Classical supervised algorithms are still highly competitive and reliable for large-scale spam detection problems, which is parallel to phishing detection challenges, as can be shown in studies analyzing supervised ML models. [33]. Hybrid ensemble learning based phishing systems demonstrate that combining diverse models contributes to much higher robustness against different attack strategies and reduces error rates across the datasets. [34]. RL based neural models have introduced adaptive phishing detection frameworks which can change and evolve with new patterns. These approaches highlight some future research directions, but they need further validation because of the operational complexity. [35].

Research Methodology

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

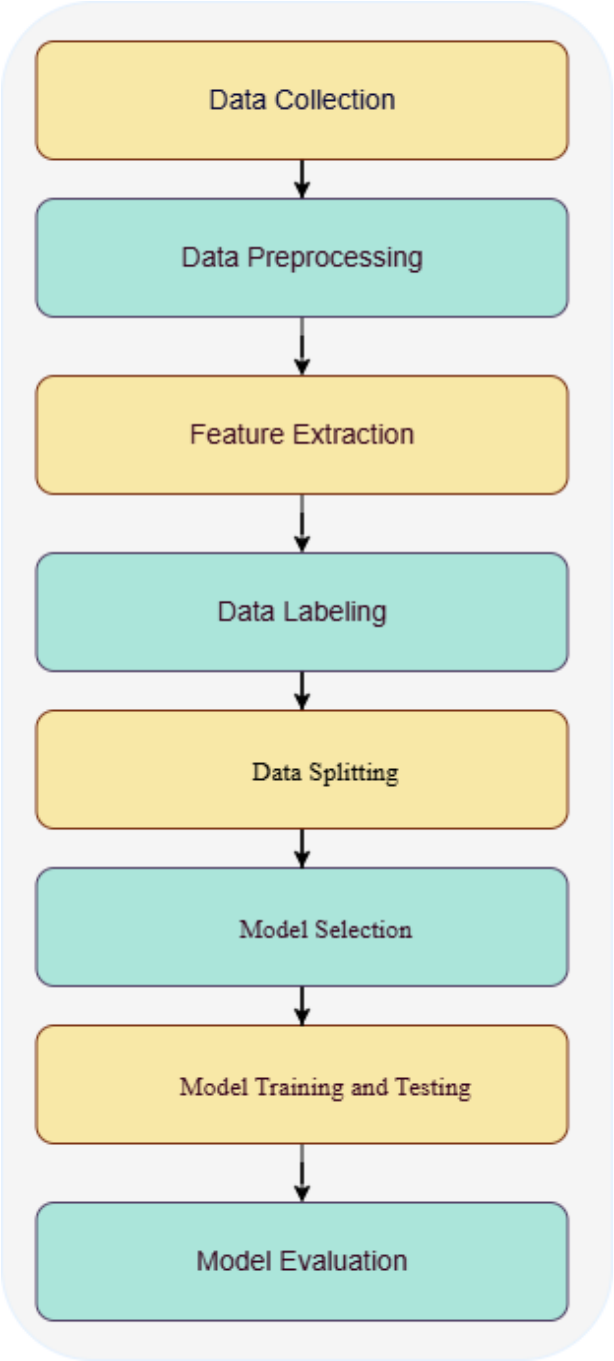


Figure 1: Architecture of the Proposed Framework

Data Collection

Data collection, which involves acquiring an appropriate dataset for phishing email recognition. This study use Phishing Email Dataset, that contains thousands of emails categorized as either phishing or legitimate. It provides diversity that will enable the model to learn different attack patterns by covering a wide range of emails using different phishing techniques. The dataset is of exceptional quality with very few errors,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

and the pre-labeled emails ensure reliability for algorithms training and evaluation. The proposed research suggests a ML based phishing email recognition system that is able to detecting zero-day phishing attempts while maintaining a lightweight methodology.

Data preprocessing

The following step is data preprocessing, which is necessary to clean up the dataset in order for modeling to be successful. Duplicate or superfluous records are removed, and missing values are corrected, to maintain the data's quality. Categorical features will be converted into their mathematical representations so that models such as NB, DT, RF, and LR can handle the data efficiently. Text-based characteristics have been tokenized, stop words removed, and expressed numerically using TF-IDF vectorization. Numerical features are standardized to ensure consistent scaling across all properties. In order for models to identify significant trends in the data, the dataset will then be divided training and testing, and stratified sample will be used to account for the proportion of phishing and authentic emails.

Feature Extraction

The goal of feature extraction, is to put the cleaned text on numerical foam that ML algorithms can use. In this case, the TF-IDF vectorization allocates weights to terms based on how important they are in each email in relation to the total dataset. Each email will be represented as a numerical vector of dimension with the TF-IDF score for each word to allow the classifier to spot patterns that indicate phishing emails. This step, therefore, ensures that machine learning algorithms can take these kinds of textual data into consideration in an efficient manner in classifying emails as either phishing or legitimate.

Data Labeling

Each email is assigned a binary label depending on its class. Phishing emails are labeled as 1 and valid emails are labeled as 0. This is because the machine learning models are able to identify the pattern of features which are associated with the target class. This data set is ideal for training and testing of models since labeled data is saved in the form of feature vectors with matching labels in a CSV file. Just to make sure that classifiers are properly trained and assessed, this phase will be crucial.

Data splitting

This labeled dataset will be divided into a training and test set to assess the model's performance. 30% of the data for testing, and the remaining 70% will be used for training.

Model Selection

Instead, it focuses more on lightweight machine learning models rather than complex models such as GCN to save computation and model size. The different classifiers selected like: NB which is a probabilistic model that has done very well in text classification; RF is an ensemble learning technique, DT is a tree-based model that captures feature interactions; and LR is a linear classifier that performs well for binary classification. Selection of these models is because of their simplicity, their effectiveness, and their suitability for phishing email detection with fewer parameters.

Model Training and Testing

The selected models, in this regard, are fitted to the training dataset using the TF-IDF feature vectors. The hyper-parameters, in turn, can be tuned for better performance, and

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

the models can be trained on the phishing versus legitimate email classification. Later, such trained models will be tested on the test dataset for their correctness in classifying unseen emails. Additionally, it allows the comparison of multiple classifiers to regulate which one is most effective for detecting phishing emails.

Evaluation Metrics

Performance of each classifier is calculated using the binary classification metrics of accuracy, precision, recall, and F1-score. F1-score provide a balance among accuracy and recall, precision show the ratio of correctly recognized phishing emails, and recall show the number of correctly recognized actual phishing emails. Accuracy show the overall correctness of prediction.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Let's represent true positive, true negative, false positive, and false negative by TP, TN, FP, and FN, respectively. These signs altogether indicate the model's reliability in distinguishing between malicious and legitimate traffic.

Experiment and simulation

This paper is aimed at assessing the performance of some machine learning algorithms for the identification of phishing emails on a benchmark dataset. In this work, the dataset used was obtained, which contains phishing and legal emails. The languages used in this project are Python and Jupyter Notebook. Data science tools that will be used in this project include Pandas for data manipulation, Scikit-learn, and NLTK. In this paper, the collected dataset will go through preprocessing such as stemming, noise reduction, text cleaning, and TF-IDF feature extraction. To avoid bias when carrying out the evaluation on the models, the dataset will first be preprocessed and then divided into training and test sets. Four models have been selected for the experiment by considering their performance regarding email categorization tasks: NB, LR, DT, and RF.

Each of these selected machine learning models will be executed by the simulation on the preprocessed dataset, and its classification performance will be investigated. Each algorithm was fed exactly the same TF-IDF feature vectors for fairness throughout the trial. The performance measures for each model were determined by comparing the expected class of every email coming from the simulation to the actual labels. For each model, sklearn's in-built functions were used to train the model, make predictions, and estimate probabilities. Confusion matrices, accuracy scores, precision, recall, and F1-scores of every classifier were generated by the simulation. These will enable the comparison of which algorithm provides better detection of phishing emails with reduced false positives. Moreover, the simulation environment allowed the hyperparameter tuning so as to observe the changes that affect the fidelity of the model. This will prove the advantages and disadvantages of each strategy and how machine

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

learning can automate and enhance phishing email detection.

Results and discussion

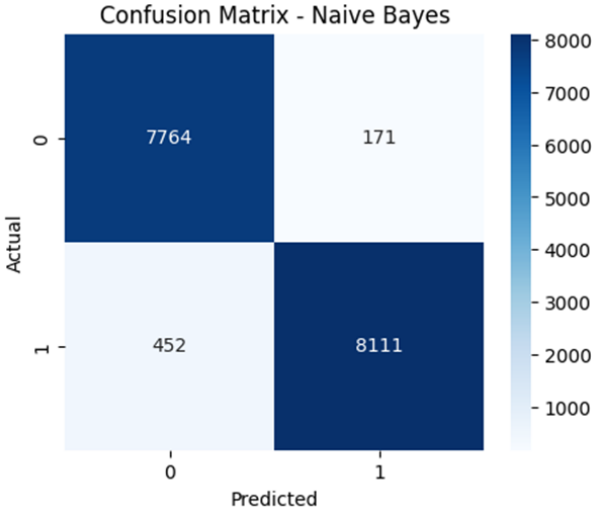
Naïve byes

The fact that the NB model had an accuracy of 96% means that a large portion of phishing and legitimate emails were correctly recognized. The recall was 95%, which means that 95% of all phishing emails were correctly recognized, with very few FN. The F1 score of 96% show a very strong balance among recall and precision. 98% precision means that when this model recognized an email as phishing, it was right 98% of the time. There are two classes in the classification report, 0 denoting legal and 1 representing phishing. The model predicted correctly 98% of all 7935 real legitimate emails with a recall of 98%, catching all the real legitimate emails. In the case of phishing, it has a precision of 98% but a recall of 95% out of 8563 phishing emails, which means most were correctly classified. The support for these classes is 7935 for class 0 and 8563 for class 1.

The confusion matrix from the Naïve Bayes classifier classification is shown in Fig. 2, which illustrates the model's ability to distinguish between phishing and legitimate classifications. Figure 2 demonstrates the model's excellent discriminative capacity by properly classifying legitimate emails and phishing emails. However, the classifier produced 452 false negatives, and 171 false positives.

Table 1: Report on NB classification.

	Precision	Recall	F1 score	Support
0	94	98	96	7935
1	98	95	96	8563
Accuracy	-	-	96	16498
Macro avg	96	96	96	16498
Weight avg	96	96	96	16498



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Figure 2: Confusion matric naïve Bayes

Logistic regression

The LR model good performed as compare to NB with an accuracy of 98%, indicating a superior overall classification of emails as phishing or legitimate. 98% precision show, this model was accurate 98% of the time when it identified an email as phishing. A 99% recall rate indicates that nearly every phishing email was detected. The model's reliability is further demonstrated by the F1-score of 98%, which shows a very good balance among precision and recall. For class 0, the model had 98% precision and 98% recall, thus correctly classifying most of the legitimate emails. For class 1, precision is 98%, and recall is 99%, so this model will catch nearly all phishing emails with few false alarms. Support for class 0 is 7935 and class 1 is 8563.

Figure 3: show model performance in differentiating between the legitimate and phishing classes is displayed in the confusion matrix from the categorization using the logistic regression model. Figure 3 demonstrates the model's excellent discriminative capacity in correctly classifying both phishing and legal emails.

On the other hand, the classifier generated 190 false positives, classifying, and 126 false negatives.

Table 2: Repot on LR classification.

	Precision	Recall	F1 score	Support
0	98	98	98	7935
1	98	99	98	8563
Accuracy	-	-	98	16498
Macro avg	98	98	98	16498
Weight avg	98	98	98	16498

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

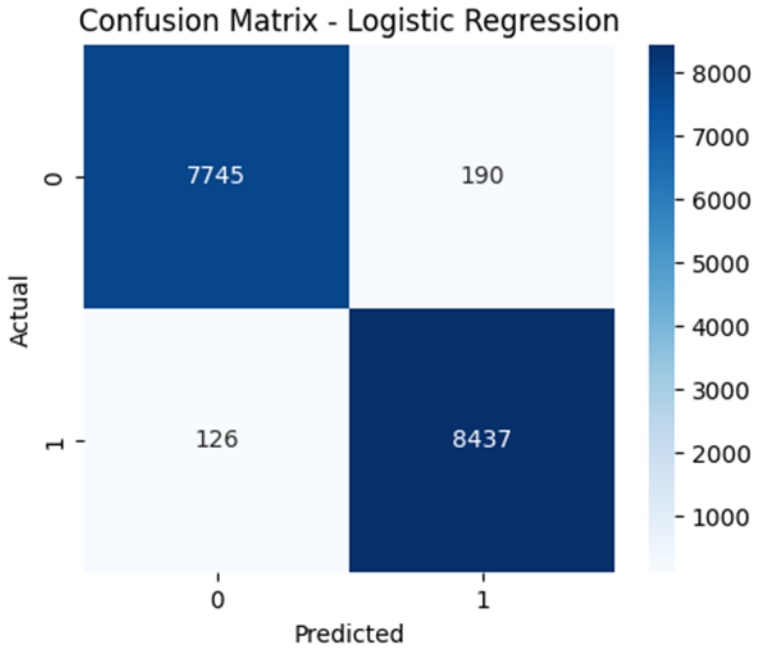


Figure 3: LR confusion matrix

Decision Tree

Similar to NB, which properly classified the majority of emails, DT model had 96% accuracy. Precision is 96%, the model was accurate 96% of the time when it predicted phishing. The model identified 97% of real phishing emails, according to the recall of 97%. The F1-score of 96% shows a balanced performance. For class 0, which is legitimate, the precision and recall are 96%, which means that most of the legitimate emails fell into this class correctly. For class 1, recall is 97% and precision is 96%; the majority of phishing emails were identified with few FP. Support value for class 0 is 7935 and class 1 is 8563.

The classification's confusion matrix is displayed in Figure 4. Show the model ability in differentiating among the authentic and phishing classes using Decision Tree classifier. Figure 4 show the model good discriminative capabilities in accurately classifying phishing and authentic emails. The classifier produce 282 FN and 313 FP

Table 3: Report on DT classification.

	Precision	Recall	F1 score	Support
0	96	96	96	7935
1	97	97	97	8563
Accuracy	-	-	96	16498
Macro avg	96	96	96	16498
Weight avg	96	96	96	16498

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

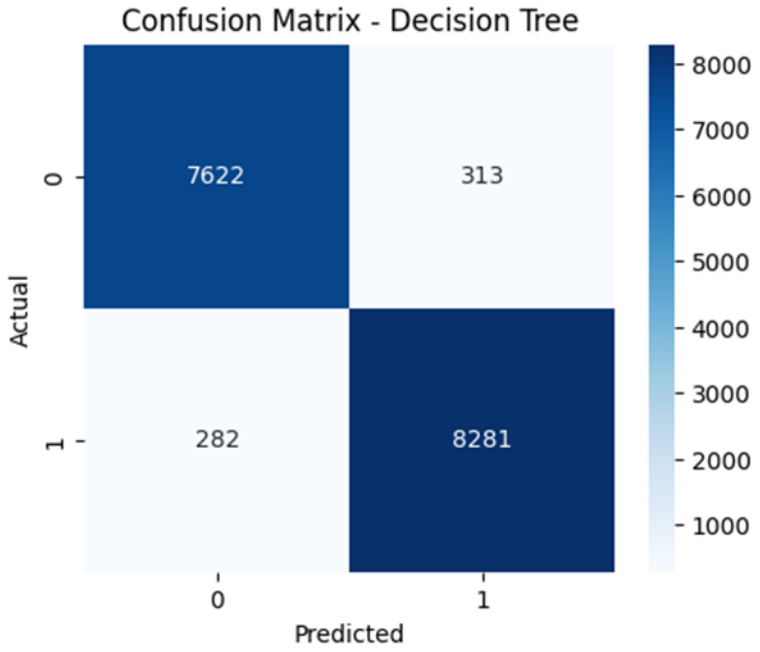


Figure 4: confusion matrix Decision tree

Random forest

With a 99% accuracy rate and outperforming all other models, the Random Forest model is very dependable in identifying phishing and authentic emails. 99% Precision indicates almost entire emails were correctly identified as such. With a recall score of 99%, almost all phishing emails were identified. The best model found in this investigation has F1-score 99% that indicates a good balance among recall and precision. Precision and recall of 99% for class 0 indicate nearly all valid emails were accurately detected. Class 1 precision and recall values are 99%, indicating that almost all phishing emails were identified through the models with insufficient FP or FN. Support scores for classes 0 and 1 are 7935 and 8563, respectively, demonstrating a balanced dataset evaluation.

The confusion matrix from the logistic regression classifier classification in Figure 5 illustrates how good the model differentiates among the phishing and authentic classes. The model successful identifying phishing and authentic emails are show in Figure 5. Model produce 105 FP and 106 FN.

Table 4: Report of RF classification

	Precision	Recall	F1 score	Support
0	99	99	99	7935
1	99	99	99	8563
Accuracy	-	-	99	16498
Macro avg	99	99	99	16498
Weighted avg	99	99	99	16498

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN
3007-3197

Print ISSN
3007-3189

http://amresearchreview.com/index.php/Journal/about

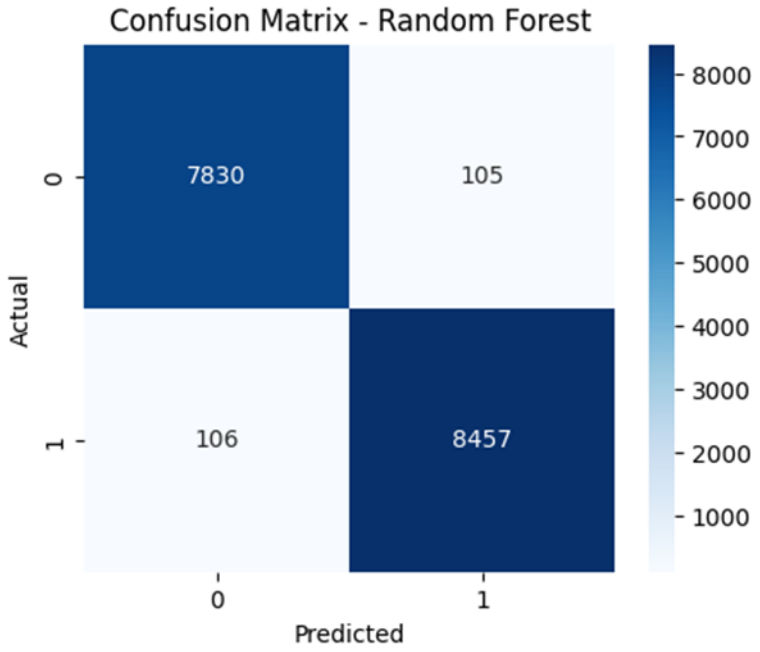


Figure 5: RF Confusion matrix

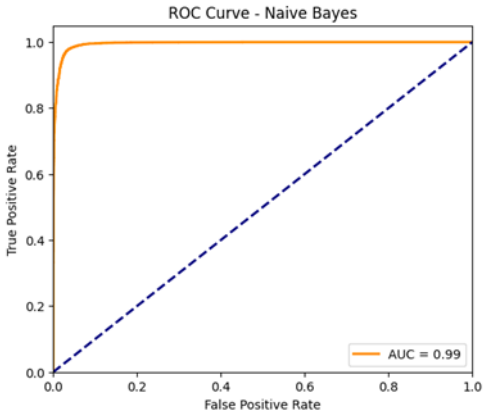
ROC Curve: It provides a graphic representation of the machine learning models' categorization performance. The ROC curves and AUC values for the LR, RF, DT, and NB classifiers are shown in Figure 6. All four classifiers have very high discriminatory power, according to the overall results, with RF and LR yielding the best outcomes. With an AUC of 1.00, the RF and LR model offers flawless classification performance with neither FP nor FN. This suggests that RF does an excellent job of distinguishing between the classes in the dataset. The DT classifier also performed well; its equivalent AUC was 0.96 when compared to RF and LR.

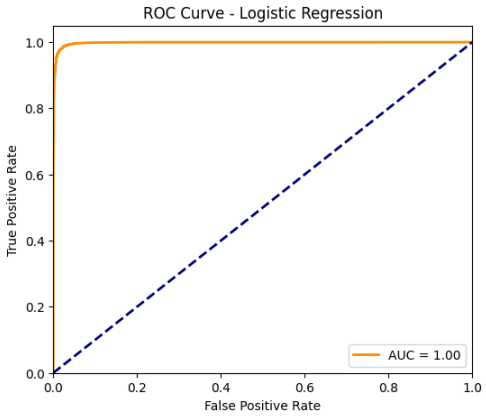
Nonetheless, DT still performs excellently in predictive accuracy and has a high capability to identify TP at low FP rates. At the same time, the NB classifier records an AUC of 0.99, this already proves that NB also performs at an almost optimal level. Its curve stays toward the upper-left corner of the ROC space, indicating accurate predictions with minimal misclassification. Comparing ROC curves generally demonstrates that, out of the four, RF and LR are the best, closely followed by NB and DT. Each of the four models produces high ROC regions, indicating strong performance in the dataset's categorization.

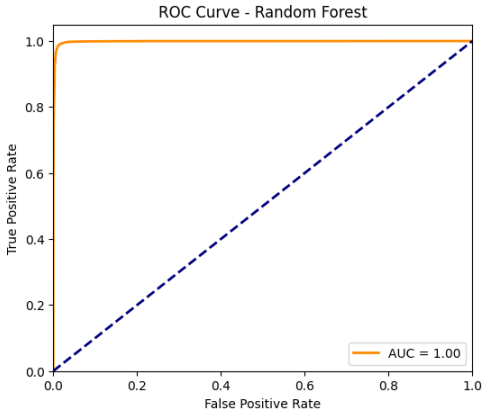
Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about







Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

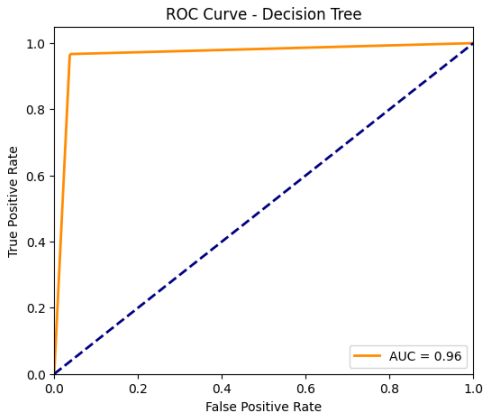


Figure 6: The ROC curve of NB, LR, DT, RF

Comparative Analysis of ML models

The total performance of all four models is shown in Table 5.

Table 5: Comparison of performance measures.

Metric	Naïve byes	Logistic regression	Decision tree	Random forest
Accuracy	96.22%	98.08%	96.42%	98.72%
Precision	97.94%	97.80%	96.36%	98.77%
Recall	94.72%	98.53%	96.71%	98.76%
F1 score	96.30%	98.16%	96.53%	98.77%

Figure 7 compares the accuracy of all models. This fig show that all models perform well. Random forest show high accuracy as compare to all other models but other models also achieve best performance not bed. Fig 8 show f1 score comparison of different models. Random forest show highest f1 score. Fig 9 show the comparison between accuracy, precision, recall and f1 score.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

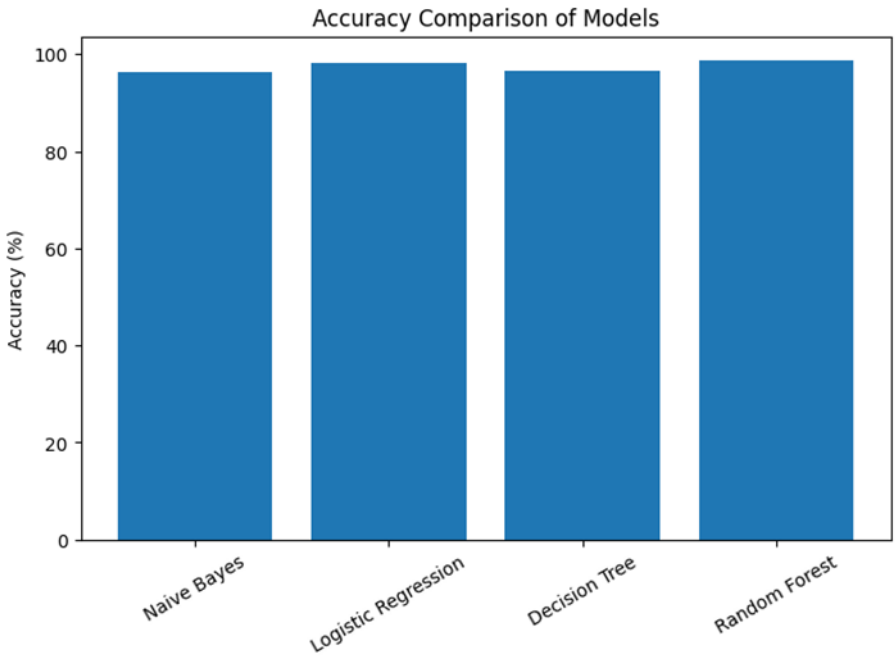


Figure 7: Accuracy comparison

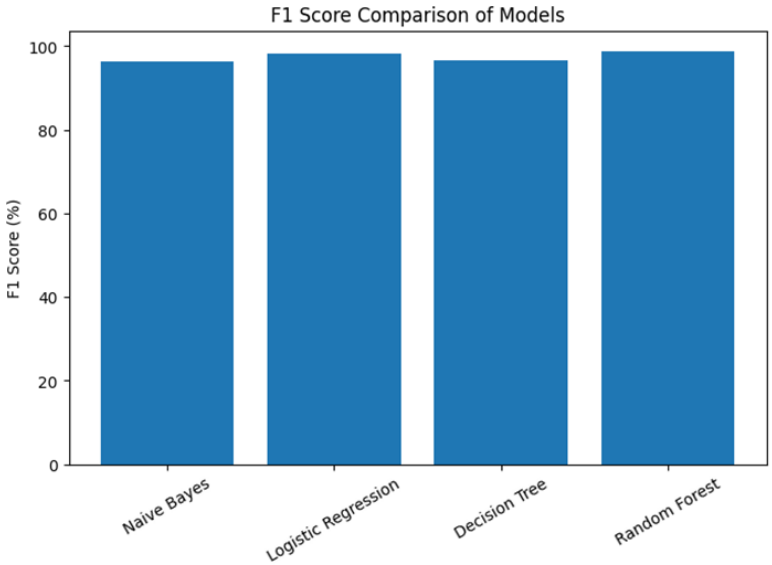


Figure 8: F1 score comparison

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

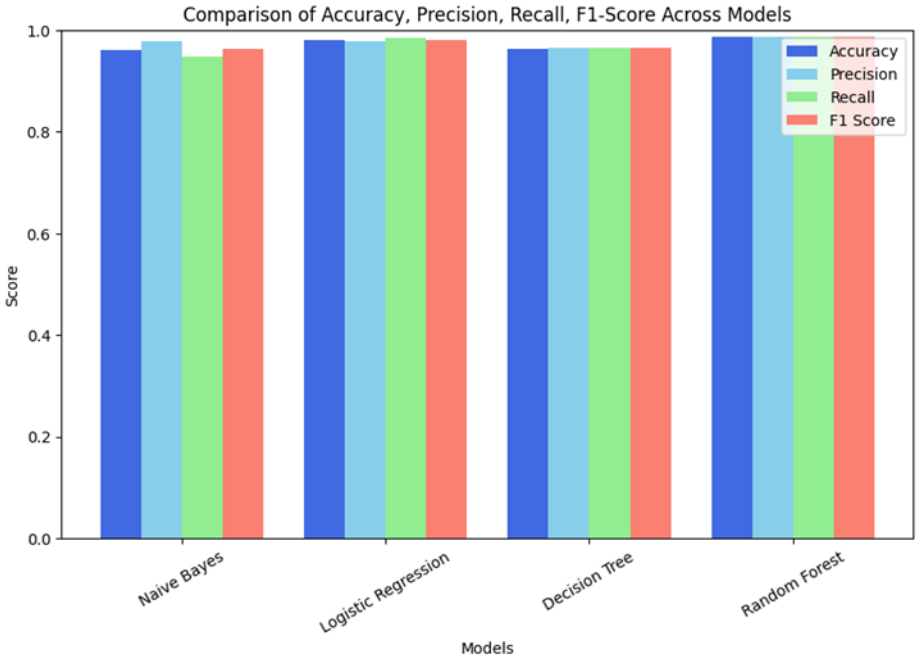
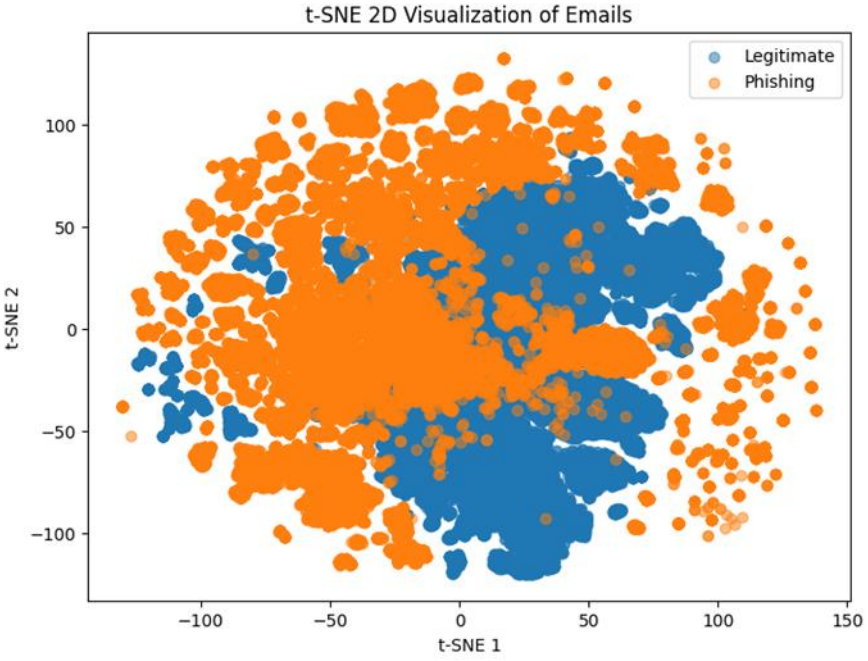


Figure 9: Accuracy, Precision, Recall, and F1 score comparison graph

Data visualization

Figure 10 show the 2D visualization of emails using TD-IDF feature extraction. Legitimate emails are show in blue and phishing emails are show in red. Maximum phishing and legitimate emails from well separated clusters, which in turn supports good separability with in the feature space.



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Figure 10: 2D visualization of emails.

Comparative Analysis with Related Work

Table 6 compares the effectiveness of the proposed lightweight machine learning models with a prior study that used a Graph Convolutional Network (GCN) in combination with natural language processing techniques for phishing emails. 98.2% accuracy, 98.5% precision, and 98.3% recall were attained with the GCN model discussed here. [22] This was really successful, although it uses a lot more processing power due to its deep learning design. The proposed lightweight machine learning models, on the other hand, achieved competitive and often even better performance while maintaining a low computational power. The Naïve Bayes classifier demonstrated strong baseline performance with very low resource needs, achieving 96.22% accuracy, 97.94% precision, and 94.72% recall. Despite utilizing many less parameters, the Logistic Regression model achieved 98.08% accuracy, 97.80% precision, 98.53% recall, and an F1-score of 98.16%, roughly matching the performance of the GCN model. With an F1-score of 96.53% and an accuracy of 96.39%, a decision tree classifier demonstrated interpretability in addition to a reliable detection capability. With 98.72% accuracy, 98.77% precision, 98.76% recall, and an F1-score of 98.77%, the Random Forest classifier outperformed the GCN-based method among all the suggested models. This demonstrates that on the phishing email detection challenge, an ensemble-based lightweight model may at least match, if not surpass, complex deep learning architectures.

Table 5: Comparative Analysis with Related Works

Reference	Technique	Accuracy	Precision	Recall	F1 score
Ref [22]	GCN	98.2%	98.5%	98.3%	----
My models	Naïve byes	96.22%	97.94%	94.72%	96.30%
	Logistic regression	98.08%	97.80%	98.53%	98.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 4 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

						16%
	Decision tree		96.39%	96.36%	96.71%	96.53%
	Random forest		98.72%	98.77%	98.76%	98.77%

Conclusion

This study demonstrates how lightweight machine learning models can efficiently and with low computational cost identify phishing emails with very high accuracy. By combining four conventional machine learning classifiers, TF-IDF feature extraction, and preprocessing, the proposed method effectively separates phishing from authentic emails. With an accuracy of 98.72% and an F1-score of 98.77%, Random Forest outperformed all other models. However, Logistic Regression has also shown excellent outcomes. Naive Bayes and Decision Tree, though a little less accurate, reliably performed and proved suitable for resource-constrained environments.

Furthermore, the comparison with cutting-edge deep learning techniques reveals that while architectures like GCN achieve strong performance, the lightweight machine learning models employed in this study achieve comparable or even higher accuracy without the computational overhead associated with deep learning. This confirms the fact that optimized classical ML models remain one of the strongest and practical candidates for phishing detection systems, particularly for real-time deployment in organizations.

References

[1] V. Shahrivari, M. M. Darabi and M. Izadi, "Phishing Detection Using Machine Learning Techniques," *IEEE,* pp. 1-9, 2020.

[2] A. Basit, M. Zafar, A. R. Javed and Z. Jalil, "A Novel Ensemble Machine Learning Method to Detect Phishing Attack," in *IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, 2020.

[3] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar and E. A. Elsoud , "An intelligent cyber security phishing detection system using deep learning techniques," *Springer,* vol. 25, p. 3819–3828, 2022.

[4] Q. Qi, Z. Wang , Y. Xu , Y. Fang and C. Wang , "Enhancing Phishing Email

Online ISSN 3007-3197
Print ISSN 3007-3189

Detection through Ensemble Learning and Undersampling," *MDPI,* vol. 13, pp. 1-19, 2023.

[5] R. Brindha, . S. Nandagopal, H. Azath, V. Sathana, G. P. Joshi and S. W. Kim, "Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification," *Computers, Materials & Continua,* vol. 74, no. 3, pp. 5901-5914, 2022.

[6] S. Rawal , B. Rawal, A. Shaheen and S. Malik, "Phishing Detection in E-mails using Machine Learning," *International Journal of Applied Information Systems (IJAIS) ,* vol. 12, no. 7, pp. 21-24, 2017.

[7] E. Gandotra and D. Gupta, "An Efficient Approach for Phishing Detection using Machine Learning," *Springer,* p. 239–253, 2021.

[8] P. Sharma and U. Bhardwaj, "Machine Learning based Spam E-Mail Detection," *International Journal of Intelligent Engineering and Systems,* vol. 11, no. 3, pp. 1-10, 2018.

[9] N. Altwaijry, I. A.-. Turaiki, R. Alotaibi and F. Alakeel , "Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models," *MDPI,* vol. 24, pp. 1-19, 2024.

[10] J. Kolla, S. Praneeth, M. S. Baig and G. R. Karri, "A comparison study of machine learning techniques for phishing detection," *Journal of Business and Information Systems ,* vol. 4, no. 1, pp. 21-33, 2022.

[11] J. James, S. L. and C. Thomas, "Detection of phishing URLs using machine learning techniques," in *International Conference on Control Communication and Computing (ICCC)*, 2013.

[12] S. Salloum, T. Gaber, S. Vadera and K. Shaalan, "Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey," *Procedia Computer Science,* vol. 189, p. 19–28, 2021.

[13] N. A. Unnithan, H. NB, A. S, V. R and S. KP, "Machine Learning Based Phishing E-mail detection," *Security-CEN@ Amrita,,* pp. 1-5, 2018.

[14] T. Toma, S. Hassan and M. Arifuzzaman, "An Analysis of Supervised Machine Learning Algorithms for Spam Email Detection," in *International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, Rajshahi, Bangladesh, 2021.

[15] P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," *Journal of network and computer applications,,* vol. 2012, p. 103545, 2023.

[16] S. Smadi, N. Aslam and L. Zhang, "Detection of online phishing email using dynamic evolving neural network based on reinforcement learning," *Decision Support Systems,* vol. 107, pp. 88-102, 2018.

[17] R. Hassanpour, E. Dogdu, R. Choupani, O. Goker and N. Nazli, "Phishing E-mail Detection By Using Deep Learning Algorithms," in *ACM Southeast Regional Conference*, 2018.

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

[18] Y. S. Murti and P. Naveen, "Machine Learning Algorithms for Phishing Email Detection," *Journal of Logistics, Informatics and Service Science,* vol. 10, no. 2, pp. 249-261, 2023.

[19] K. Thakur, M. L. Ali, M. A. Obaidat and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," *MDPI,* pp. 1-26, 2023.

[20] H. NB, V. R and S. KP, "A Machine Learning approach towards Phishing Email Detection," *Proceedings of the anti-phishing pilot at ACM international workshop,* pp. 1-8, 2018.

[21] P. Bountakas, , K. Koutroumpouchos, and h. Xenakis, "A Comparison of Natural Language Processing and Machine Learning Methods for Phishing Email Detection," in *Proceedings of the 16th International Conference on Availability, Reliability and Security*, 2021.

[22] A. Alhogail and A. Alsabih, "Applying machine learning and natural language processing to detect phishing email," *Computers & Security,* vol. 110, p. 102414, 2021.

[23] T. Gangavarapu, C. . D. Jaidhar and B. Chanduka , "Applicability of machine learning in spam and phishing email filtering: review and approaches," *Springer,* vol. 53, p. 5019–5081, 2020.

[24] U. A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi and A. Ahmadian , "Cloud-based email phishing attack using machine and deep learning algorithm," *Springer,* vol. 9, p. 3043–3070, 2022.

[25] D. Rathee and S. Mann, "Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning," *International Journal of Computer Applications,* vol. 183, no. 47, pp. 1-7, 2022.

[26] N. Pimpason, P. Viboonsang and S. Kosolsombat, "Phishing Email Detection Model Using Deep Learning," in *IEEE International Conference on Cybernetics and Innovations (ICCI)*, Chonburi, Thailand, 2025.

[27] H. F. Atlam and O. Oluwatimilehin, "Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review," *MDPI,* pp. 1-28, 2022.

[28] P. H. Kyaw, J. Gutierrez and A. Ghobakhlou, "A Systematic Review of Deep Learning Techniques for Phishing Email Detection," *MDPI,* vol. 13, pp. 1-20, 2024.

[29] N. Kumar, S. Sonowal and Nishant, "Email Spam Detection Using Machine Learning Algorithms," in *Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2020.

[30] O. K. Sahingoz, E. Buber, O. Demir and a. Diri , "Machine learning based phishing detection from URLs," *IEEE,* vol. 117, pp. 345-357, 2019.

[31] A. A. Akinyelu and A. O. Adewumi, "Classification of Phishing Email Using Random Forest Machine Learning Technique," *Journal of Applied Mathematics,* pp. 1-6, 2014.

Online ISSN 3007-3197
Print ISSN 3007-3189

[32] I. ’. AbdulNabi and Q. Yaseen, "Spam Email Detection Using Deep Learning Techniques," *Elsevier,* vol. 184, pp. 853-858, 2021.

[33] S. SALLOUM, T. GABER, S. VADERA and K. SHAALAN, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," *IEEE,* vol. 10, pp. 65703-65727, 2022.

[34] N. A. Unnithan, H. NB,, V. R and S. KP, "Detecting Phishing E-mail using Machine learning techniques," pp. 1-7, 2018.

[35] X. Ding, B. Liu, Z. Jiang and Q. Wang, "Spear Phishing Emails Detection Based on Machine Learning," in *IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Dalian, China, 2021.