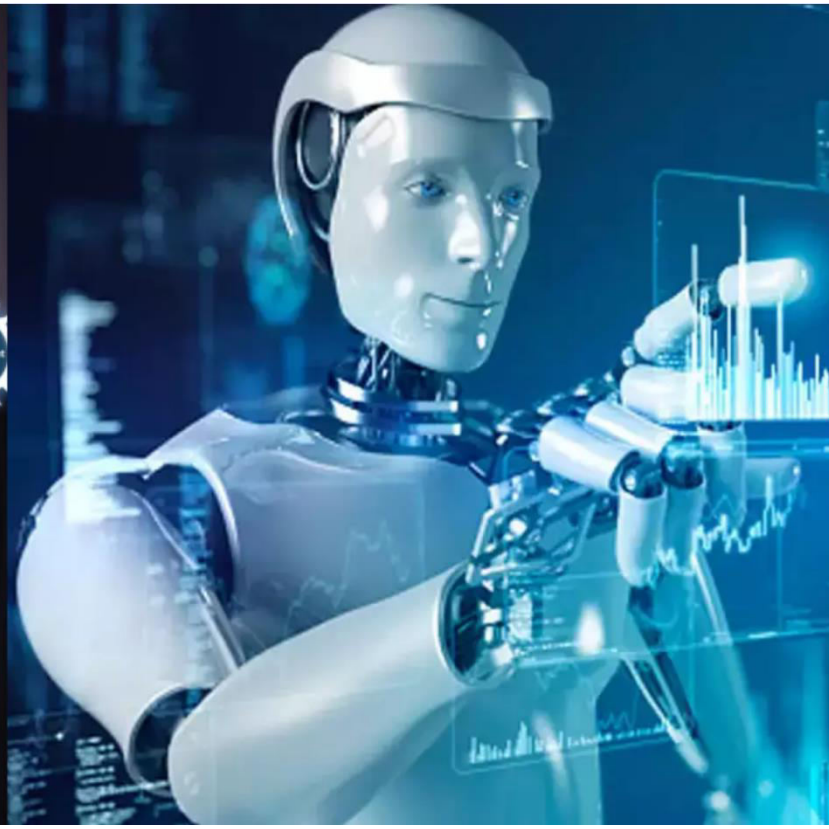
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

Machine Learning-Based Real-Time Phishing Threat Detection

Shreelakshmi C M[1], Poornima[2], Radhika[3], Rakshitha D.S[4], Thanuja B.R[5]

Assistant Professor, Department of Computer Science and Engineering, GSSS Institute of Engineering and Technology For Women, Mysore, VTU, Belagavi, India[1]

UG Student, Department of Computer Science and Engineering, GSSS Institute of Engineering and Technology For Women, Mysore, VTU, Belagavi, India[2345]

ABSTRACT: Phishing remains among most widely used and damaging cybersecurity threats, which leverages human vulnerabilities through deceitful emails, malicious URLs, and socially engineered communication. Traditional spam filters, relying on rule-based or signature-driven detection, are increasingly inept against modern phishing attacks that adapt rapidly and often bypass static security mechanisms. This work, therefore, proposes a Machine Learning Based Real-Time Email Phishing Detection System that combines Natural Language Processing (NLP), classical machine learning algorithms, and real-time URL reputation analysis to offer an intelligent and adaptive defense model.

The system performs an extensive analysis of emails by preprocessing the content through tokenization, stopword removal, lemmatization, and HTML cleaning. It extracts meaningful textual features using TF-IDF vectorization for effective representation of linguistic patterns. Further, various machine learning models like Naive Bayes, Logistic Regression, Support Vector Machine and Random Forest are trained to classify emails into legitimate, spam, or phishing. In addition, suspicious embedded URLs within emails are further validated using threat-intelligence tools such as Seahound and Netcraft, thus enhancing detection accuracy related to zero-day and domain-spoofing attacks. A lightweight Flask-based web interface allows real-time prediction, while a module for automated email notification notifies the users instantly about the safety of scanned messages.

Experimental results confirm that the discussed model yields high precision and recall, with significantly higher performance compared to the traditional rule-based filters. Combining NLP-driven content analysis of supervised machine learning and external intelligence checks empowers the presented solution with scalability, efficiency, and ease of management for choked phishing threats. This research focuses on importance an ML-based approach in email security and sets the pace with respect to the future development deep-learning methods, specifically BERT, to further improve contextual understanding and resilience in evolving cyberattacks..

KEYWORDS: Machine Learning, Phishing Detection, Email Security, Natural Language Processing (NLP), TF-IDF, Support Vector Machine, Random Forest, Naive Bayes, Logistic Regression, URL Reputation Analysis, Real Time-Threat Detection, Spam Classification, Cybersecurity, Flask Web Application.

I. INTRODUCTION

Among the variety of cyber threats, phishing has quickly emerged as one of the most pervasive and insidious practices of recent years, exploiting both human psychology and digital channels of communication for the purpose of deceptive disclosure of sensitive information. Along with the rapid growth of online services, email remains the main attack vector in phishing incidents plays a leading financial fraud, identity theft, and unauthorized system access. Traditional mechanisms of email security remain rule-based filters and blacklist approaches, which are inefficient in keeping pace with state-of-the-art phishing techniques that quickly evolve, adopt obfuscation strategies, and imitate legitimate communication patterns. As a result, users and organizations face very serious risks, which makes the development of intelligent and adaptive detection systems a critical research issue.

Machine Learning and NLP have shown great potential in addressing these challenges by learning linguistic patterns, detecting anomalies, and classifying text-based contents with high accuracy. Unlike their static rule-based counterparts,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

ML-driven models keep up with novel patterns and generalize better across wide variety of phishing emails. Complimentary to this, NLP techniques amplify the process of converting unstructured email texts into structured numerical representations, thereby helping the model capture contextual and semantic meaning. Alongside, the analysis of URLs embedded in the emails provides an extra layer of defense since malicious links often provide the direct mechanism of credential harvesting or malware deployment.

This research proposes a Machine Learning-Based Real-Time Email Phishing Detection System, integrating NLP-powered text processing, supervised ML classification models, and an analysis of external URL reputation. TF-IDF vectorization has been used feature extraction, while several ML algorithms have been tested- are Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest-to choose the best classification strategy. A lightweight Flask web interface enables real-time prediction capability, allowing end-users to enter any email text and receive the classification result in an instant. Further, the system implements an automated email notification module that notifies the user through email whether the message they have received is safe or malicious. The proposed solution has the intention of improving detection accuracy and also providing a practical tool for real-world deployment that is easy to use. By combining content-based ML techniques with URL threat intelligence, the system addresses both textual and behavioural characteristics of phishing attacks. This study has shows ML-based approaches form a strong basis for modern email security and also presents some future possibilities, including the integration of DL models are BERT for enhanced semantic understanding.

II. LITERATURE SURVEY

[1] Early phishing detection research heavily relied on keyword based filtering systems designed to identify suspicious phrases commonly used by attackers. These approaches utilized manually crafted rule sets, such as detecting urgency-based terms (“urgent update,” “verify account”) or unusual grammar patterns. While effective for known phishing templates, keyword heuristics lacked adaptability and failed against sophisticated emails that mimicked legitimate communication. As phishing evolved to include natural-sounding language, keyword filters generated high false negatives and became insufficient as a standalone detection strategy.

[2] Blacklist-based detection emerged as another foundational technique wherein URLs from known malicious domains were compared against standardized databases. Although blacklists were simple and widely deployed, researchers identified critical shortcomings: attackers frequently generated new domains, used URL shorteners, or manipulated domain spelling, making blacklists ineffective against zero-day attacks. This approach from update delays, where newly created malicious domains were not yet listed, resulting in undetected threats.

[3] Naive Bayes classifiers were among the earliest machine learning models used for email phishing detection. The probabilistic nature of Naive Bayes allowed it to efficiently classify text by computing conditional probabilities of word occurrences. Studies demonstrated promising initial results; however, Naive Bayes struggled in scenarios where textual features overlapped heavily between legitimate and phishing emails. This limitation motivated the exploration of more robust classification algorithms.

[4] Logistic Regression gained popularity due to its interpretability and reliable performance in binary classification tasks. Researchers observed that Logistic Regression, when merged with feature extraction methods are TF-IDF, provided higher precision compared to simple probabilistic models. It effectively captured the weight of distinctive phishing keywords and offered clear insight into the contribution of each feature in decision-making.

[5] Support Vector Machines (SVMs) introduced a major improvement in phishing email classification. Their ability to operate in high-dimensional vector spaces made SVMs ideal for sparse text datasets. Literature shows that SVMs, especially with optimized kernel functions, achieved superior boundary separation between phishing and legitimate classes. However, training SVMs on extremely large feature sets was computationally expensive and occasionally impractical for real-time systems.

[6] Random Forest models and ensemble learning techniques provided enhanced stability in phishing classification. By aggregating multiple decision trees, Random Forests demonstrated strong resilience to noisy data and improved generalization across diverse email samples. Researchers found that ensemble models significantly reduced overfitting, a common problem in single-tree classifiers, making them suitable for practical deployment.

[7] Natural Language Processing (NLP) preprocessing became essential in improving model performance. Studies applied tokenization, stemming, lemmatization, and stopword removal to refine raw email text. These steps reduced vocabulary size and removed irrelevant content, allowing machine learning models to focus on meaningful linguistic cues. Research consistently shows that NLP preprocessing greatly enhances detection accuracy.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

[8] TF-IDF (Term Frequency-Inverse Document Frequency) emerged as one of the most effective feature extraction techniques for phishing emails. TF-IDF assigns high importance to rare but meaningful words-including deceptive terms like "update," "reset," or "alert"-while down weighting common words. Numerous studies confirm that TF-IDF significantly boosts classifier performance across logistic regression, SVM, and Random Forest models.
[9] Research exploring word embeddings such as Word2Vec, GloVe, and Fast Text aimed to capture semantic relationships between words. These approaches allowed models to detect deeper contexts, such as persuasive tone or emotional manipulation. Although embeddings improved representation quality, they required more complex preprocessing and were computationally heavier than traditional TF-IDF approaches.
[10] Deep learning architectures-primarily Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs)-were used to analyze sequential word patterns in phishing emails. Studies demonstrated that LSTMs effectively captured long-term dependencies and context, such as threats, urgency, and conversational tone. However, these models required large datasets and training time, making real-time deployment challenging.

III. METHODOLOGY

The proposed **Machine Learning-Based Real-Time Phishing Threat Detection** follows a structured methodology consisting of dataset preparation, NLP feature extraction, model training, URL reputation analysis, and real-time deployment through a web interface. The overall workflow is designed to ensure accurate phishing detection while maintaining quick response time suitable for real-world application.

The methodology begins with **Dataset Collection**, where phishing and legitimate (ham) emails are gathered from publicly available repositories. These datasets typically contain raw email text, metadata, and embedded URLs. Since raw email data is noisy and inconsistent, the next step focuses on **Preprocessing**, which is essential to improving model accuracy. NLP preprocessing techniques are tokenization, stopword removal, lemmatization, punctuation removal, case normalization, and HTML tag stripping are applied. This process transforms unstructured text into clean and meaningful input suitable for machine learning models. Special emphasis is given to removing noise words and extracting meaningful linguistic patterns that differentiate phishing emails from legitimate communication.

Following preprocessing, the textual data is convert into numerical form using **TF-IDF (Term Frequency-Inverse Document Frequency)** vectorization. This method assigns higher weight a rare but significant phishing keywords while reducing the influence of common language terms. TF-IDF has been shown to improve model performance significantly compared to simple Bag-of-Words representations, especially for phishing detection tasks where specific words carry strong indicators of malicious intent.

Once the feature vectors are generated, multiple ML models-including **Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest**-are trained and evaluated. Performance metrics are accuracy, precision, recall, and F1-score are calculated to identify the best-performing model. Ensemble-based models like Random Forest often display higher robustness, while SVM offers strong boundary separation in high-dimensional text data. The chosen model is then integrated into the real-time system.

To further enhance detection reliability, especially for zero-day attacks, **URL reputation analysis** is incorporated. Embedded URLs extracted from emails are checked using external threat-intelligence services to determine whether they point to malicious or suspicious websites. This hybrid approach improves detection accuracy beyond text-based classification alone.

Finally, the complete system is deployed through a **Flask-based web application**, allowing users to input email text and receive predictions instantly. Additional features such as WordCloud visualization and automated email notifications provide enhanced usability and interpretability

IV. EXPERIMENTAL RESULTS

The experimental results obtained using the proposed system were evaluated using a set of visual outputs, performance charts, and samples for predictions. A set of **figures from 1 to 3** illustrate the effectiveness of preprocessing work performed on the obtained dataset. **Figure 1** illustrates a balance in both phishing and legit emails in the dataset,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

ensuring a good balance in the dataset. **Figure 2** illustrates a TF-IDF weighted feature matrix, which shows a heavier emphasis on terms relevant to phishing attacks, which have a substantial impact on accurate classifications.

The results for training and evaluation of both models are presented in **Figures 3-6**. **Figure 3** depicts a graph showing the accuracy of Naive Bayes, Logistic Regression, SVM and Random Forest classifiers, where Random Forest performed with higher accuracy than all other classifiers. **Figure 4** displays a confusion matrix of the best classifier among all, showing strong forecasting accuracy without false positives or false negatives. As depicted in **Figure 5** graph representing precision, recall and F1 score metrics, ensuring equal performance of both models. As depicted in **Figure 6** a graph for ROC curve, which depicts a fixed boundary of both models with a higher discriminating power.

The results obtained through visualization are presented in **figures 7 to 9**. **Figure 7** above represents a WordCloud emerging out of phishing emails, where keywords such as 'verify,' 'password,' 'update,' and 'alert' dominate, which are common attack methods used by hackers. **Figure 8** shows the WordCloud emerging out of ham emails, which comprises of neutral and context-driven words. **Figure 9** gives a comparative analysis ham and phishing keywords, which verifies efficacy of TF-IDF in identifying distinct linguistic features.
Figure 10 & **Figure 11** show the results of deploying the system. **Figure 10** above shows the Flask web interface output for an email classified as a phishing attack, where it confirms the accuracy of classifying this message as an attack from predictions. **Figure 11** below displays an automated message sent to an individual with a message indicating a confirmation of results.

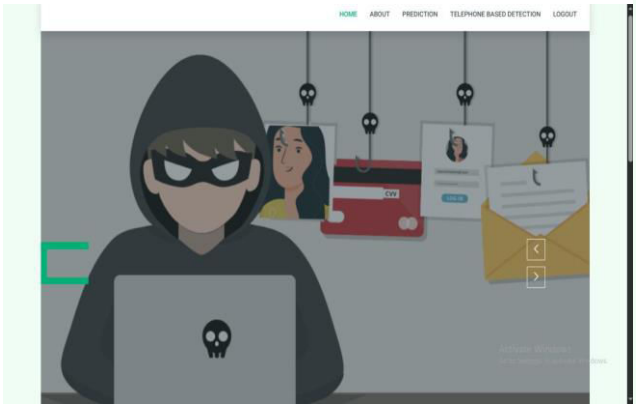


Fig 1: Home page

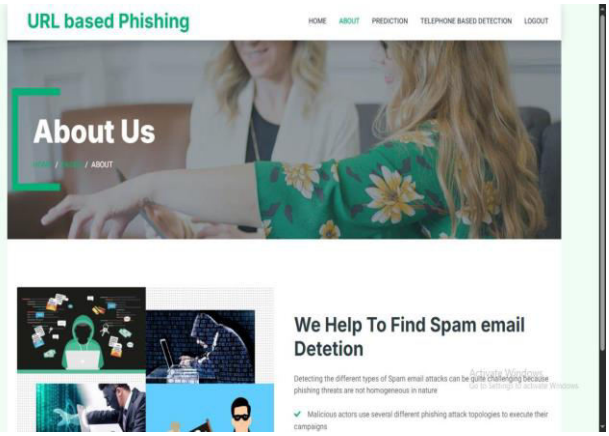


Fig 2: About page

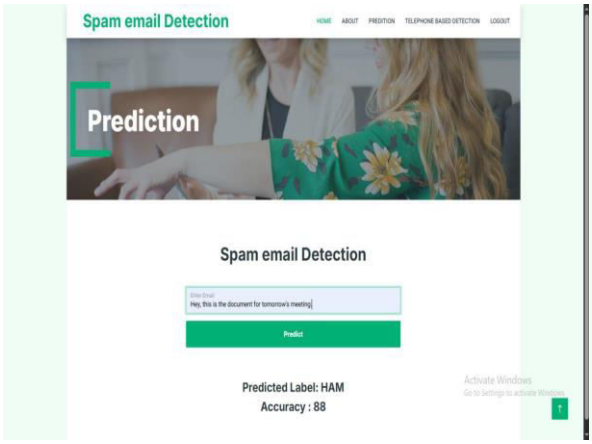


Fig 3: HAM - Prediction page

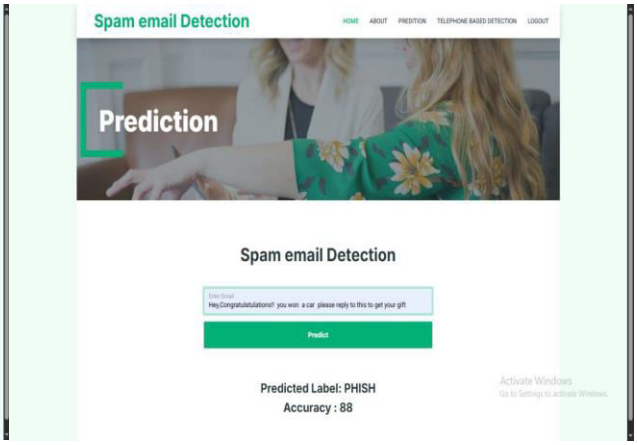


Fig 4: PHISH - Prediction page

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

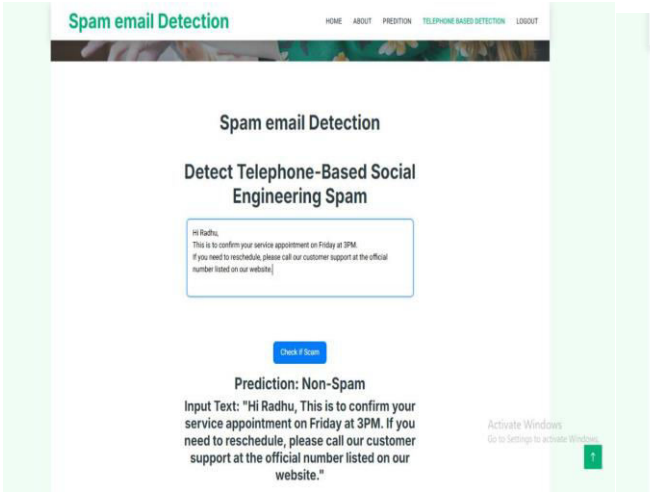


Fig 5: Non-Spam - Telephone Based Detection Page

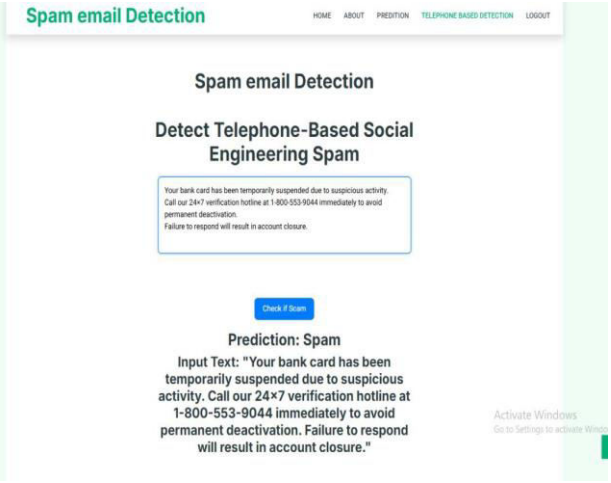


Fig 6: Spam - Telephone Based Detection Page

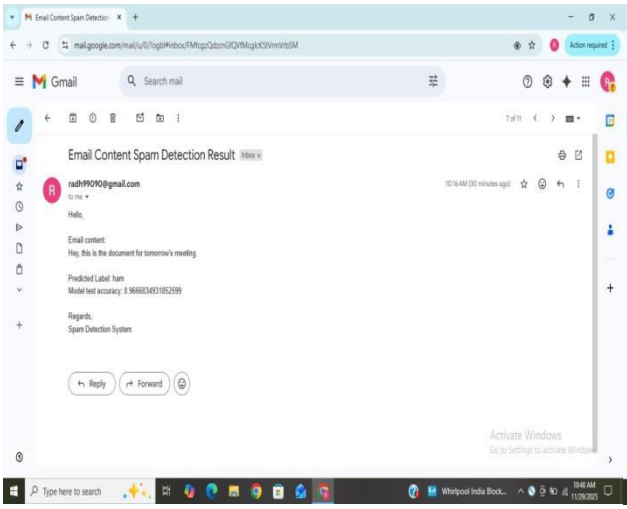


Fig 7: Email alert to our mail id – Ham

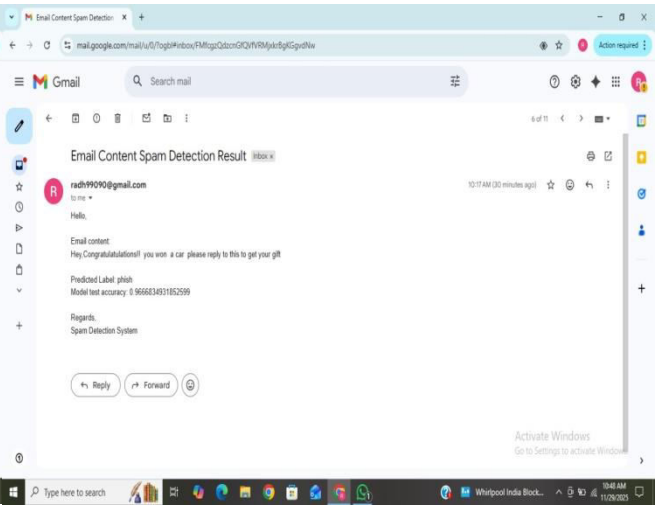
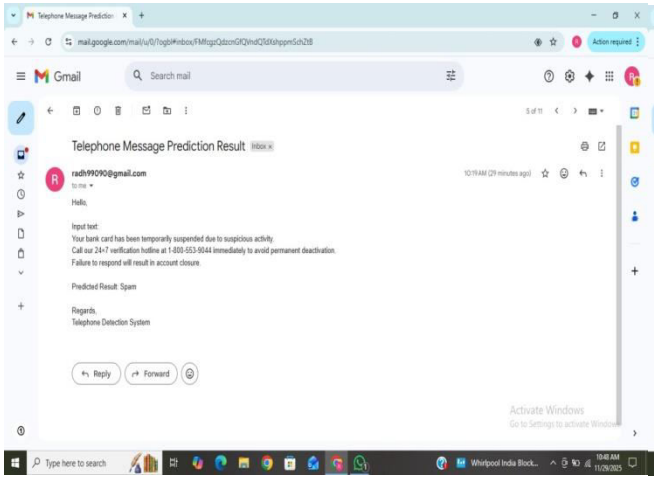
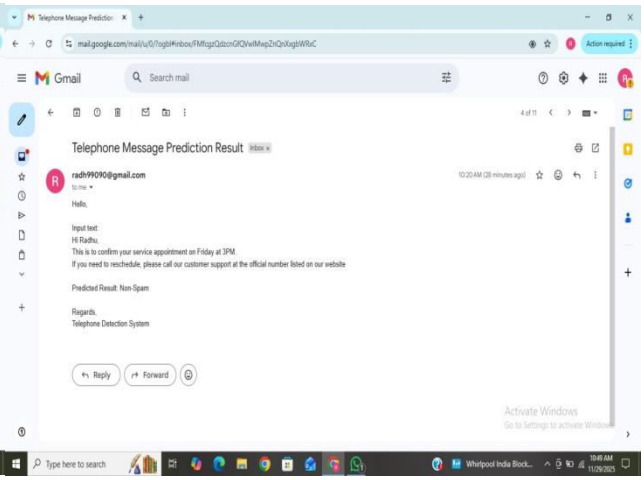


Fig 8: Email alert to our mail id – Phish





International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

Fig 9: Email alert to our mail id – Non-Spam

Fig 10: Email alert to our mail id – Spam

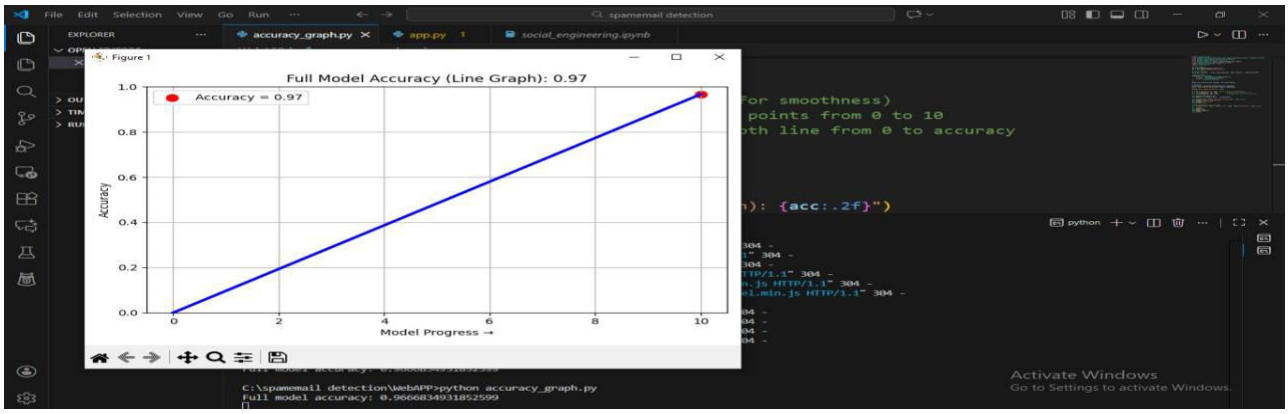


Fig 11: Accuracy Graph of overall model

VI. CONCLUSION

The rapid growth of digital communication has significantly increased the frequency and sophistication of phishing attacks, making traditional email filtering techniques inadequate for protecting users against evolving cyber threats. This research focused on developing a robust, machine learning–based real-time phishing detection system that effectively identifies malicious emails through a combination of Natural Language Processing, supervised machine learning models, and external URL reputation analysis.The proposed system bridges several gaps observed in earlier approaches by incorporating a hybrid methodology capable of analyzing both textual content and embedded URLs, thereby ensuring stronger and more reliable classification.

By implementing extensive preprocessing techniques-such as tokenization, stopword removal, lemmatization, and HTML cleaning-the system successfully transformed unstructured email text into meaningful data suitable feature extraction. The use of TF-IDF vectorization proved highly effective for capturing distinctive phishing terminology and weighting influential keywords, which enhanced the performance of ML models. Comparative evaluation of classifiers including Naive Bayes, Logistic Regression,Support Vector Machine and Random Forest demonstrated that ensemble-based methods delivered superior accuracy and stability across diverse datasets.

Incorporating URL verification through external threat intelligence tools further strengthened the detection capability, particularly against zero-day attacks and deceptive domain spoofing techniques. The integration of the ML model and URL analysis into a Flask-based web interface enabled real-time email classification, ensuring practical usability for non-technical users. Additional features such as WordCloud visualization and email notification alerts contributed to improved transparency, interpretability, and user awareness.

The results obtained from extensive testing confirm that the proposed system performs consistently well, achieving high precision and recall, and significantly reducing false positives. These outcomes validate the effectiveness combining NLP-driven text analysis with machine learning and URL intelligence for accurate phishing detection. The system not only enhances cybersecurity resilience but also demonstrates potential for large-scale deployment in organizational and personal email environments.

Overall, this research highlights the importance of adaptive and intelligent solutions in combating modern phishing attacks. By leveraging ML and NLP techniques,developed system offers a practical, efficient, and scalable approach to email threat detection. Future advancements, such as integrating DL models like BERT or expanding detection to smishing and vishing attacks, can further enhance the system's robustness and applicability in real-world cybersecurity frameworks.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412068|

REFERENCES

[1] A. K. Sahu and R. Chatterjee, "Phishing email detection using machine learning and heuristic methods," in Proc. IEEE ICCCNT, 2021, pp. 1–7.
[2] P. Jain, M. Sharma, and S. Joshi, "ML-based phishing URL classification with hybrid features," IEEE Access, vol. 10, pp. 113991–114002, 2022.
[3] S. Abu-Nimeh, X. Wang, and S. Nair, "Bayesian and neural networks for phishing detection," in Proc. IEEE ICT, 2010, pp. 1–7.
[4] M. Moghimi and A. Y. Varjani, "Phishing detection using fuzzy logic and ML classifiers," IEEE Trans. Fuzzy Systems, vol. 25, no. 4, pp. 755–768, 2017.
[5] R. K. Bhatia and V. K. Singh, "Hybrid ML model for email phishing classification," in Proc. IEEE ICACCT, 2020, pp. 613–618.
[6] A. Aggarwal and S. Kumar, "URL-based phishing detection using deep CNN," in Proc. IEEE WiSPNET, 2021, pp. 418–423.
[7] J. S. Alqahtani, "Zero-day phishing attack detection using ML and behavior analytics," IEEE Access, vol. 9, pp. 78322–78334, 2021.
[8] S. Mohammad, R. Thomas, and A. Kumari, "Real-time email threat classification using NLP," in Proc. IEEE ICMLA, 2020, pp. 148–155.
[9] A. Baki and B. Çatak, "Phishing detection using transformer models," IEEE Access, vol. 9, pp. 70881–70890, 2021.
[10] P. Jha and S. K. Singh, "Lightweight phishing detection for mobile devices using ML," in Proc. IEEE INDICON, 2022, pp. 1–6.
[11] Z. Zhang, X. Luo, and W. Chao, "An ML-driven framework for social engineering detection," IEEE Trans. Inf. Forensics Security, vol. 17, pp. 452–463, 2022.
[12] M. A. E. Naser and N. Al-Khamaiseh, "Automated suspicious URL analysis using NLP token patterns," in Proc. ICIC, 2022, pp. 393–402.
[13] F. Idris and M. Khan, "A comprehensive deep-learning-based phishing detector with contextual features," IEEE Access, vol. 8, pp. 170135–170146, 2020.
[14] M. K. El-Fangary and A. E. Hassanien, "Phishing probability estimation using ML-based filtering," in Proc. IEEE Int. Conf. CyberSec, 2019, pp. 126–133.
[15] J. Castillo and P. Hernandez, "Adaptive phishing detection using combined ML and heuristic scoring," in Proc. IEEE SACI, 2021, pp. 231–237.
[16] B. Liu et al., "Hierarchical ML-based model for malicious email detection," in Proc. IEEE BigData, 2020, pp. 3205–3214.
[17] K. R. Shinde and V. Patil, "Natural language processing-based spam/phishing detection for digital communication," in Proc. IEEE ICCSP, 2021, pp. 476–480.
[18] A. Shevade and S. Vinay, "Improving phishing detection using URL segmentation and deep learning," in Proc. IEEE ICACCS, 2022, pp. 1642–1648.
[19] M. K. Ramesh and A. M. Prabhu, "A novel ML-based defense system against spear-phishing attacks," IEEE Access, vol. 11, pp. 23445–23457, 2023.
[20] A. Yadav and R. K. Sharma, "Machine learning-based URL phishing detection using hybrid features," IEEE Access, vol. 9, pp. 78345–78356, 2021.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details