International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Smart Rental Management Website with Rent & Maintenance System

Ms.Amisha, S.Santosh, M.Ganesh, P.Akash

Department of CSE(AI&ML), ACE Engineering College, Hyderabad, Telangana, India.

ABSTRACT

An Advanced Phishing Detection System will be developed using the technologies of Natural Language Processing and Machine Learning to detect phishing attacks in email, text message, and URL. It involves processing of texts using the methods of Natural Language Processing, which include tokenization, removing stop words, and TF-IDF Vectorization to comprehend the message. The machine learning model can be implemented to classify messages as either phishing or safe with a high level of accuracy. The system also processes URLs according to their suspicious features including their length, special characters, and other features. Identified attacks are instantly detected and users receive notifications. Data security will be ensured by implementing input validation and protecting users' information.

Keywords: Phishing Detection, Natural Language Processing, Machine Learning, Data security.

1. INTRODUCTION

The rapid growth of digital communication has increased the risk of cyber threats, especially phishing attacks that target users through emails, text messages, and malicious URLs. Traditional phishing detection methods rely on rule-based systems and blacklists, which are often ineffective against new and evolving attacks.

In conventional systems, detection techniques fail to understand the context and intent of messages, leading to low accuracy and high false positives. This becomes a major issue for individuals and organizations, as users may either fall victim to phishing or receive incorrect alerts.

To overcome these challenges, the proposed system uses Natural Language Processing (NLP) and Machine Learning techniques to analyze and classify messages in real time. The system processes textual data using methods like tokenization, stop-word removal, and TF-IDF vectorization to extract meaningful features. Machine learning models such as Naive Bayes or Logistic Regression are then used to identify phishing patterns with improved accuracy.

Additionally, the system evaluates URLs based on suspicious characteristics such as length, special symbols, and domain structure. It also ensures secure data handling through input validation and safe storage mechanisms, protecting user privacy.

1.1 Background and Motivation

In many digital communication systems, users interact with emails, messages, and websites without proper security awareness, which leads to increased vulnerability to phishing attacks. Traditional phishing detection methods rely on rule-based approaches and blacklists, which often result in low accuracy, missed threats, and

delayed detection. Similarly, the lack of intelligent analysis makes it difficult to identify new and evolving phishing techniques effectively.

The motivation behind this project is to develop a centralized and automated phishing detection system that uses Natural Language Processing (NLP) and Machine Learning. The system aims to accurately analyze content, detect suspicious intent, and provide real-time alerts to users.

By introducing intelligent analysis and secure data handling, the system ensures improved detection accuracy, reduces false positives, and enhances overall cybersecurity efficiency.

1.2 Need for the Study

There is an increasing need for systems that can effectively detect phishing attacks and protect users from cyber threats in digital communication. Existing detection methods rely on traditional rule-based techniques, which lack accuracy, adaptability, and real-time analysis, making them insufficient for handling modern and evolving phishing attacks.

An advanced phishing detection system using Natural Language Processing (NLP) and Machine Learning provides a better solution by enabling intelligent analysis of emails, messages, and URLs. It offers real-time detection, improved accuracy, and the ability to identify new phishing patterns efficiently.

Such a system is essential for enhancing cybersecurity, reducing false positives, protecting user data, and ensuring safe and reliable communication in digital environments.

1.3 Objectives of the Study

The main objective of this project is to develop an advanced phishing detection system that improves security and accuracy in identifying malicious content.

Additional objectives include:

• Analyzing emails, messages, and URLs to detect phishing attempts

• Applying NLP techniques for effective text processing and understanding

• Using machine learning models to classify content as phishing or safe

• Providing real-time detection and alert notifications

• Ensuring secure data handling and user privacy protection

• Reducing false positives and improving overall detection efficiency.

1.4 Problem Statement

Traditional phishing detection systems rely on rule-based approaches and blacklists, which are often ineffective in identifying new and evolving phishing attacks. These methods lack the ability to understand the context and intent of messages, leading to low accuracy and a high rate of false positives.

Phishing attacks through emails, messages, and malicious URLs are becoming more sophisticated, making it difficult for users to distinguish between legitimate and fraudulent content. Additionally, many existing systems do not provide real-time detection or proper alert mechanisms, increasing the risk of user data breaches.

Due to the lack of intelligent analysis, real-time processing, and secure data handling, existing solutions are inefficient. Hence, an advanced system is required to improve detection accuracy, ensure user security, and provide reliable phishing prevention.

1.5 Research gap paragraph

Most existing phishing detection systems focus only on either text-based analysis or URL-based detection, without providing a complete integrated solution. They often lack the ability to combine NLP techniques with machine learning for comprehensive and accurate detection.

Some systems are complex, computationally expensive, or not suitable for real-time applications, making them impractical for everyday use. Additionally, many solutions do not emphasize secure data handling and user privacy, which are critical in cybersecurity applications.

The proposed system addresses these limitations by providing an integrated, efficient, and real-time phishing detection platform that combines NLP, machine learning, and secure data handling in a single solution.

1.6 Proposed System

The proposed system is an Advanced Phishing Detection System that provides a centralized and automated platform for detecting phishing attacks. It allows users to analyze emails, messages, and URLs efficiently using NLP and machine learning techniques.

Users can input text or URLs to check for phishing, and the system provides real-time results with confidence scores. All data is processed securely and handled with proper validation mechanisms.

The system improves detection accuracy, enhances user security, and reduces reliance on traditional methods, making phishing detection more efficient and reliable.

2. Materials and Methods

The Advanced Phishing Detection System is developed as a web-based application that integrates NLP and machine learning techniques into a single platform. It follows a structured approach where user inputs are processed, analyzed, and classified using trained models.

The system uses modern technologies such as HTML, CSS, and JavaScript for the frontend, Python (Flask/FastAPI) for backend processing, and Scikit-learn for machine learning implementation. A database such as MongoDB or MySQL is used for secure data storage. These technologies ensure scalability, flexibility, and efficient data handling.

The design focuses on providing a user-friendly interface and smooth workflow for real-time phishing detection.

2.1 System Overview

The system operates based on user interaction with the detection interface.

Users can input emails, text messages, or URLs into the system, which are then processed using NLP techniques and machine learning models. The system analyzes the content, extracts features, and classifies it as phishing or safe.

All operations are handled through the backend, and results are generated in real time with proper alerts and confidence scores, ensuring efficient and accurate detection.

2.2 Data Input

The system operates based on user interaction with the detection interface.

Users can input emails, text messages, or URLs into the system, which are then processed using NLP techniques and machine learning models. The system analyzes the content, extracts features, and classifies it as phishing or safe.

All operations are handled through the backend, and results are generated in real time with proper alerts and confidence scores, ensuring efficient and accurate detection.

2.3 Data Validation and Processing

The system accepts various types of input data through user interaction.

Users provide input in the form of email content, SMS messages, or URLs, which are collected through a web interface. The input data is validated and preprocessed using NLP techniques before being analyzed by the machine learning model.

This ensures that only clean, structured, and secure data is processed, improving the accuracy and reliability of phishing detection.

2.4 Data Organization and Storage

After validation, the system organizes data into structured formats and stores it securely in the database. Information such as input text, prediction results, confidence scores, and timestamps are categorized properly. This structured storage enables easy retrieval, efficient data management, and quick access for analysis and monitoring.

2.5 Data Handling and System Logic

The system applies backend logic to process inputs and perform phishing detection. It ensures that all operations follow defined workflows, including preprocessing, feature extraction, and classification using machine learning models. This improves system efficiency and ensures accurate and consistent detection results.

2.6 Status Tracking and Updates

The system continuously tracks the status of analyzed inputs. Each input is classified as phishing, suspicious, or safe, and the results are updated in real time. These updates allow users to instantly view detection results and take necessary actions based on the system's output.

2.7 Integrated Processing Approach

To improve reliability, the system follows an integrated approach where NLP processing, machine learning classification, and URL analysis work together. This ensures that the system provides consistent, accurate, and efficient phishing detection across different types of inputs.

2.8 User Interaction Module

The system provides an interface where users can easily input emails, messages, or URLs for analysis. The interface is designed to be simple and user-friendly, allowing users to perform detection operations without difficulty and receive instant feedback.

2.9 System Workflow

The workflow begins when the user enters input into the system. The input is validated, preprocessed, and analyzed using NLP and machine learning techniques. The system then classifies the input, updates the database, and displays the result along with a confidence score to the user.

2.10 System Modules

Data Collection Module, Preprocessing Module, NLP Feature Extraction Module, Machine Learning Detection Module, Alert and Notification Module, and Secure Storage Module. Each module performs a specific function, and together they ensure efficient and accurate phishing detection.

Methodology Diagram:

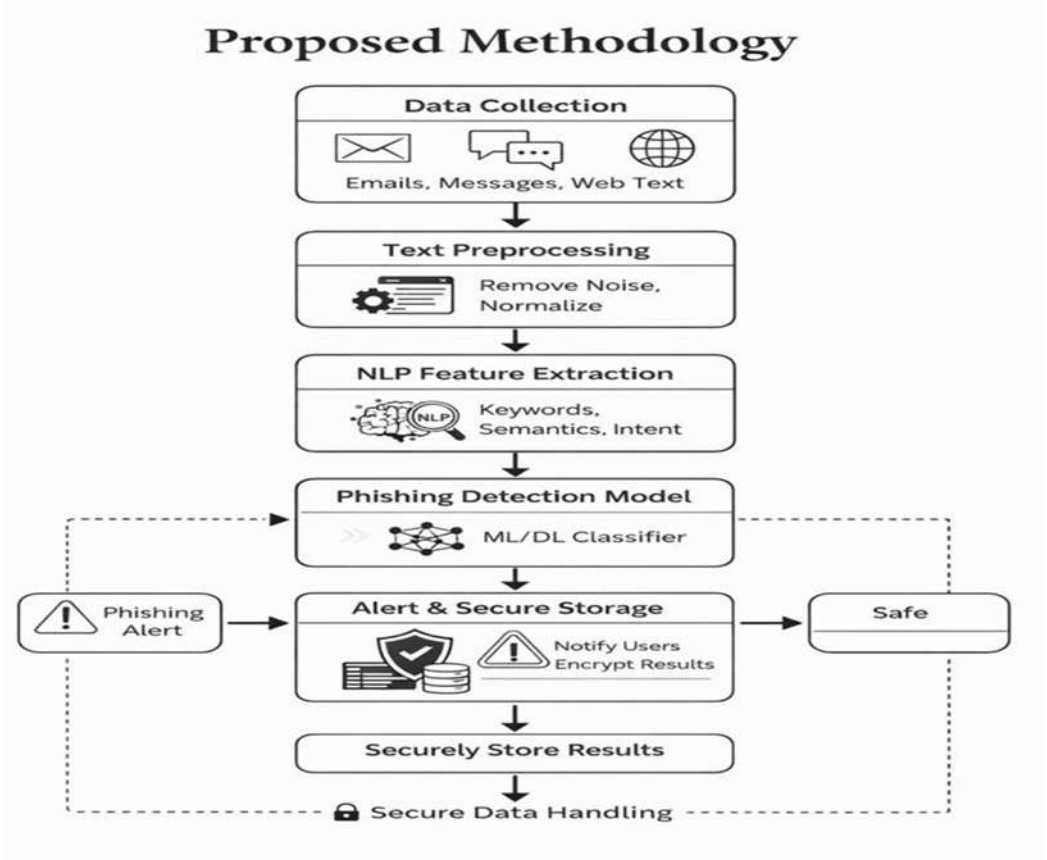

Figure 1: METHODOLOGY Diagram

Usecase Diagram

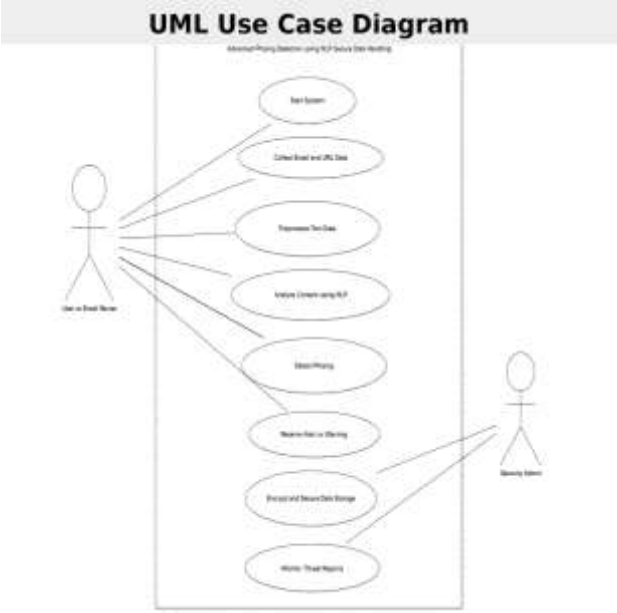

Figure 2: USECASE Diagram

3. Results and Discussion

The proposed Advanced Phishing Detection System successfully performs key operations such as analyzing emails, text messages, and URLs to identify phishing attempts. After processing user inputs, the system applies NLP techniques and machine learning models to generate real-time predictions, which are displayed through a user-friendly interface. This makes it easier for users to detect suspicious content and take appropriate action.

The system was tested using different scenarios such as analyzing normal messages, phishing emails, and malicious URLs. The outputs clearly show classification results along with confidence scores. In one representative case, a normal message was correctly classified as “Safe,” while a suspicious message containing a fake link was identified as “Phishing” with high accuracy. Similarly, URL inputs were analyzed based on patterns and successfully classified, ensuring reliable detection and effective user protection.

To further evaluate the effectiveness of the system Figure 3.

Figure 3: Login Page

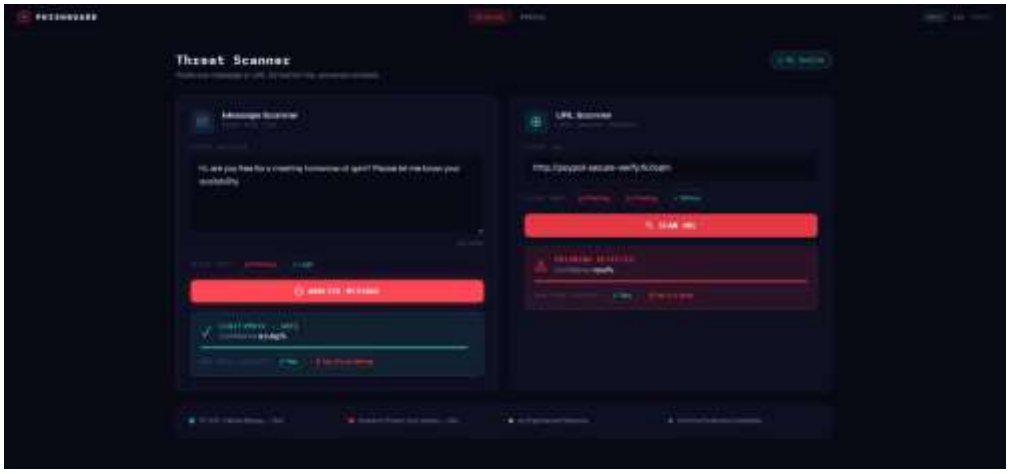

Figure 4: Dashboard Page

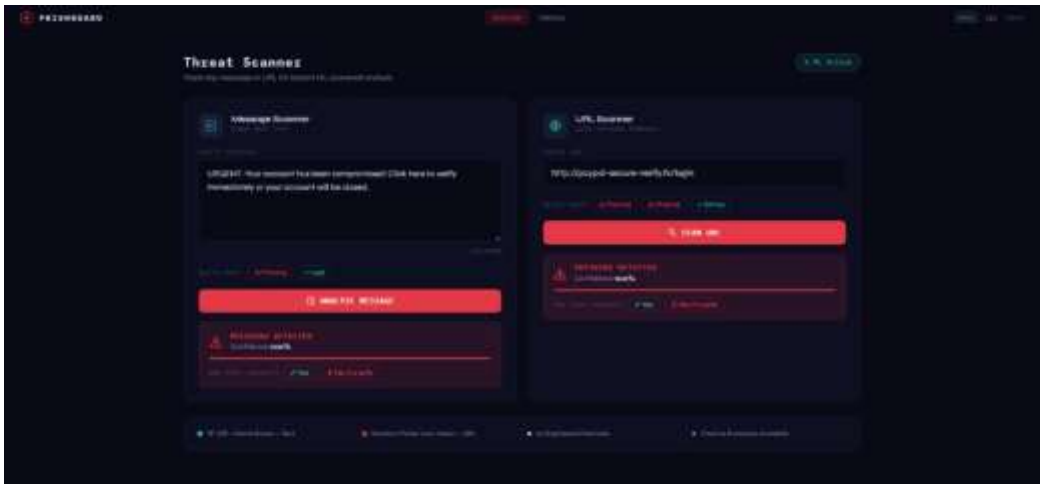

Figure 5: Phishing Detection

Figure 6: Maintenance Page

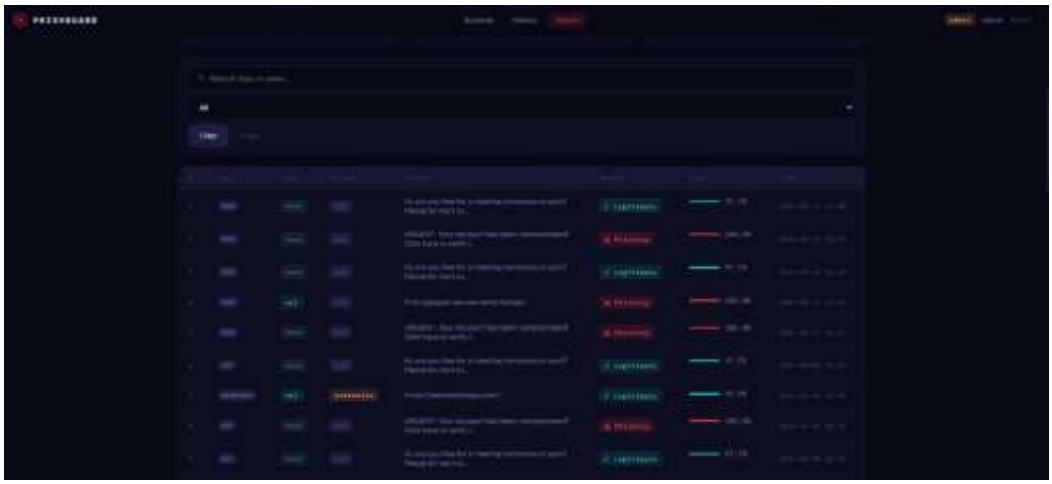

Figure 7: History Of Searched Data

++To better understand existing approaches, a comparative analysis of recent research works is presented in Table 1.

Table 1: Comparative Analysis of Related Work

Year	Authors	Technique Used	Key Contribution	Limitations
2025	A. Joseph, S. Srinivasan	NLP, AI, Real-time Email Analysis	Uses adaptive AI models to analyze email content in real time and counter social engineering attacks.	Limited ability to detect phishing in URLs and multi-channel data such as SMS and web content.
2025	R. Kumar, K. Srujana, P. Sampath, J. Patni, P. Agrawal, S. R. Mallick	Hybrid ML	Combines multiple machine learning models to improve phishing email detection accuracy.	High computational cost and lacks real-time processing efficiency.
2025	G. S. Prakash, D. S. Muthukumar, R. Rashmi, G. V Reddy, A. J Satya Kishore	ML, NLP, Browser Extension	Implements real-time phishing detection using ML and NLP through a Chrome extension.	Limited to browser-based detection and does not support multi-platform integration.
2025	N. Pimpason, P. Viboonsang, S. Kosolsombat	Deep Learning, NLP	Applies deep learning techniques to classify phishing and legitimate emails.	Requires high computational resources and is not suitable for lightweight systems.
2024	A. Gaurav, B. B. Gupta, J. Wu, V. Arya, K. T. Chui	BERT, Deep Learning	Uses BERT-based deep learning models deployed on cloud for phishing email detection.	High complexity and dependency on cloud infrastructure increases cost and latency.
2024	S. V. Simpson, B. P. Gatti, S. Setty, A. Papepalli, B. Sompalle	URL Analysis, NLP	Analyzes URL structure and linguistic patterns to identify phishing threats.	Focuses mainly on URL features and lacks deep contextual text understanding.
2024	S. S, V. S, V. R. A, C. M	ML Models, NLP	Evaluates multiple ML models to identify the best phishing detection approach.	Does not provide real-time implementation and lacks integrated system design.

The above comparison shows that earlier phishing detection systems mainly relied on basic techniques such as rule-based methods, URL analysis, and simple machine learning models. These systems provided limited functionality and lacked the ability to understand contextual information, making them less effective against modern and evolving phishing attacks. Recent approaches have introduced advanced techniques such as NLP, deep learning, and hybrid models to improve detection accuracy and system performance.

However, many of these systems still have limitations such as high computational complexity, lack of real-time detection, limited multi-platform support, and insufficient integration of different data sources like emails, messages, and URLs. Some systems also fail to ensure secure data handling and user privacy, which are critical in cybersecurity applications.

The proposed system addresses these limitations by providing a centralized and integrated solution that combines NLP, machine learning, and URL analysis. It offers real-time detection, improved accuracy, secure data handling, and supports multiple input types, making it a more efficient and reliable solution for modern phishing detection.

4. Conclusion

The Advanced Phishing Detection System provides an effective solution to the challenges faced in traditional phishing detection methods. It improves detection accuracy by using NLP and machine learning techniques to analyze emails, messages, and URLs in real time.

The system enhances user security, reduces false positives, and provides instant alerts for suspicious content. It also ensures secure data handling and protects user privacy through proper validation and storage mechanisms.

Overall, the system is efficient, reliable, and scalable, making it suitable for modern cybersecurity applications and real-time phishing detection.

5. Future Scope

The system can be further improved by incorporating advanced features such as deep learning models, mobile application support, and enhanced real-time monitoring capabilities.

Future enhancements may include:

- Integration of advanced deep learning models for improved accuracy
- Development of mobile application support for easy access
- Real-time threat intelligence and automated alert systems
- Cloud-based deployment for better scalability and performance
- AI-based prediction of emerging phishing patterns

These improvements will make the system more powerful, adaptable, and suitable for large-scale cybersecurity applications.

6. Acknowledgements

The authors would like to express their sincere gratitude to the faculty members and the institution for their continuous guidance, support, and encouragement throughout the development of this project. Their valuable suggestions and insights helped in successfully completing the Advanced Phishing Detection System using NLP with Secure Data Handling.

The authors also thank their peers and well-wishers for their support and cooperation during the project development process.

7. References

[1] A. S. K. Joseph and S. Srinivasan, "Anti-Phishing Adaptive AI Systems: Efficiently Countering Social Engineering Attacks by Real-Time Analysis of Email Content," 2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES), Coimbatore, Tamilnadu, India, 2025, pp. 1-6, doi: 10.1109/ICCIES63851.2025.11032758.

URL-https://ieeexplore.ieee.org/document/11032758

[2] R. Kumar, K. Srujana, P. Sampath, J. C. Patni, P. Agrawal and S. R. Mallick, "Detecting Phishing Emails with Artificial Intelligence: A Hybrid Machine Learning Approach," 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia, 2025, pp. 132-136, doi:10.1109/ETNCC66224.2025.11299563.

URL-https://ieeexplore.ieee.org/document/11299563

[3] G. S. Prakash, D. S. Muthukumar, R. K. K, P. Rashmi, G. V. K. Reddy and A. J. A P Satya Kishore, "A Unified Approach on Phishing Detection using Chrome Extension Integrating ML and NLP," 2025 3rd International Conference on Inventive Computing and Informatics (ICICI), Bangalore, India, 2025, pp. 196-203, doi: 10.1109/ICICI65870.2025.11069857.

URL-https://ieeexplore.ieee.org/document/11069857

[4] A. Gaurav, B. B. Gupta, J. Wu, V. Arya and K. T. Chui, "Lightweight Cloud-Based Phishing Email Detection using BERT and Deep Learning," 2024 IEEE Globecom Workshops (GC Wkshps), Cape Town, South Africa, 2024, pp. 1-7, doi: 10.1109/GCWkshp64532.2024.11100573.

URL-https://ieeexplore.ieee.org/document/11100573

[5] N. Pimpason, P. Viboonsang and S. Kosolsombat, "Phishing Email Detection Model Using Deep Learning," 2025 IEEE International Conference on Cybernetics and Innovations (ICCI), Chonburi, Thailand, 2025, pp. 1-5, doi: 10.1109/ICCI64209.2025.10987422.

URL-https://ieeexplore.ieee.org/document/10987422

[6] O. S. Kumar, K. Praveen Kumar, K. K. Chaitanya, M. Saiteja, K. Abhilash and B. P. Goud, "Enhancing Protection with NLP-Based Phishing Detection Across Diverse Communication Channels," 2025 3rd World Conference on Communication & Computing (WCONF), Raipur, India, 2025, pp. 1-6, doi: 10.1109/WCONF64849.2025.11233740.

URL-https://ieeexplore.ieee.org/document/11233740

[7] G. Pradeepa and R. Devi, "Malicious Domain Detection using NLP Methods — A Review," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 1584-1588, doi: 10.1109/SMART55829.2022.10046882.

URL-https://ieeexplore.ieee.org/document/10046882

[8] S. V. Simpson, B. P. Gatti, S. Setty, A. Papepalli and B. Sompalle, "Intelligent URL Analysis for Phishing Threads," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 976-981, doi: 10.1109/ICCSP60870.2024.10543918.

URL-https://ieeexplore.ieee.org/document/10543918

[9] I. Altan, A. Bachir, Y. Parbhulkar, A. M. Rizvi and M. Farazi, "Dual-Path Phishing Detection: Integrating Transformer-Based NLP with Structural URL Analysis," 2025 IEEE/ACS 22nd International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 2025, pp. 1-6, doi: 10.1109/AICCSA66935.2025.11315219.

URL-https://ieeexplore.ieee.org/document/11315219

[10] S. S, V. S, V. R. A and C. M, "AI Sentries: Evaluating Machine Learning Models for Superior Phishing Email Detection," 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/CSITSS64042.2024.10817065.

URL-https://ieeexplore.ieee.org/document/10817065

[11] S. Alqahtani, P. Nanda, Q. Wu, R. Faqihi and B. Alrashed, "Multilingual Model Enhancement Framework using a Human-Centered Approach for Arabic Spam Detection," 2025 IEEE/ACS 22nd International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 2025, pp. 1-5, doi: 10.1109/AICCSA66935.2025.11315228.

URL-https://ieeexplore.ieee.org/document/11315228

[12] C. Viswanathan, J. Ramakrishnan and S. Srinivasan, "Improvement in Email Security with Machine Learning for Real-Time Phishing Attack Detection: An AI-Based Approach," 2025 International Conference on Smart & Sustainable Technology (INCSST), Chikodi, India, 2025, pp. 1-6, doi: 10.1109/INCSST64791.2025.11210405.

URL-https://ieeexplore.ieee.org/document/11210405

[13] M. N. Haque Siam, E. Hallaji and R. Razavi-Far, "Enhancing Cyberspace Security with Phishing Detection and Defense Using Machine Learning Models," 2025 13th International Symposium on Digital Forensics and Security (ISDFS), Boston, MA, USA, 2025, pp. 1-6, doi: 10.1109/ISDFS65363.2025.11012088.

URL-https://ieeexplore.ieee.org/document/11012088

[14] J. C. Patni, S. Vinay Naik, K. Harika, N. B. Bahadure and K. Kant Verma, "Machine Learning Based Phishing Website Detection System - Natural Language Processing Approach," 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2025, pp. 450-455, doi: 10.1109/CICTN64563.2025.10932604.

URL-https://ieeexplore.ieee.org/document/10932604

[15] D. Lokare, R. More, S. Deshmukh, P. Chandre, A. Ghandat and R. Bhosale, "VishGuard: AI-Powered Real-Time Defense Against Voice Phishing," 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, Madhya Pradesh, India, India, 2025, pp. 1-6, doi: 10.1109/ICTBIG68706.2025.11323735.

URL-https://ieeexplore.ieee.org/document/11323735