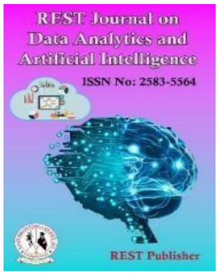
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/73



AI-Based Classification Models for Phishing and Intrusion Detection

***Shilpa Shesham, N. Venkat Sai, Y. Ayyappa Reddy, T. Bhavana**
Anurag University, Hyderabad, Telangana, India.
*Corresponding Author Email: Shilpaai@anurag.edu.in

Abstract: *In the modern digital era, phishing and intrusion attacks pose significant threats to cybersecurity. Traditional detection methods struggle to keep pace with the evolving attack landscape. This paper presents an AI driven approach utilizing classification models to detect phishing attempts and intrusions effectively. Various machine learning algorithms, including Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Random Forests, are explored to enhance detection accuracy. The proposed model integrates Natural Language Processing (NLP) techniques for improved phishing identification. Performance evaluation is conducted using key metrics such as accuracy, precision, and recall. Experimental results demonstrate the effectiveness of the AI-based approach in mitigating cyber threats and improving digital security.*

Key words: *Cybersecurity, Phishing Detection, Intrusion Detection, Machine Learning, AI-Based Security.*

1. INTRODUCTION

The rapid growth of the internet and digital platforms has led to an increase in cyber threats, particularly phishing attacks and network intrusions. Phishing involves tricking users into revealing sensitive information, such as login credentials or financial details, by masquerading as a legitimate entity. Intrusion attacks target network systems, attempting unauthorized access to compromise data integrity, confidentiality, and availability. These cyber threats continue to evolve, making traditional detection mechanisms less effective in identifying new and sophisticated attack techniques. Existing cybersecurity solutions rely on rule-based systems and signature-based detection, which struggle to keep pace with modern attack vectors. Rule-based systems require constant updates to maintain effectiveness, while signature-based methods fail to detect zero-day threats and polymorphic attacks. As a result, the cybersecurity community has turned to artificial intelligence (AI) and machine learning (ML) to enhance detection and mitigation strategies. This paper proposes an AI-driven approach that leverages classification models to detect phishing and intrusion attempts with greater accuracy. By integrating Natural Language Processing (NLP) for phishing detection and machine learning algorithms for intrusion detection, the proposed system aims to improve threat identification while reducing false positives. The research focuses on training and evaluating different classification models, such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Random Forests, to assess their effectiveness in real-world scenarios. The rest of the paper is structured as follows: Section 2 discusses related work and existing methodologies, highlighting their limitations. Section 3 introduces the proposed AI-driven framework, detailing its architecture and implementation. Section 4 explains the methodology, including data collection, feature engineering, and model training. Section 5 presents experimental results and analysis, followed by the conclusion in Section 6, which discusses future research directions.

2. BACKGROUND

AI-Based Classification Models for Phishing and Intrusion Detection: The rise in digital communication and connectivity has given cyber attackers a broad surface to exploit through phishing schemes and network intrusions. Traditional static security systems struggle to identify complex and rapidly evolving attack vectors. As a result, Artificial Intelligence (AI), particularly Machine Learning (ML), is being employed to dynamically and intelligently identify and classify threats such as phishing attempts and network-based intrusions. This project presents an AI-based classification system that effectively detects phishing (via emails and URLs) and intrusion patterns (focusing on denial-of-service attacks), thereby enhancing cyber defense mechanisms through predictive and data-driven techniques. Machine Learning for Cybersecurity Applications: The core of the system lies in the use of supervised learning models trained on labeled datasets containing real-world phishing and intrusion records. Techniques like Random Forest, Naive Bayes, Support Vector Classifier (SVC), and Multi-Layer Perceptron (MLP) are utilized to learn patterns from phishing URLs, email content, and network traffic indicators. These models are capable of capturing subtle correlations and anomalies that typically elude rule-based systems. Through model evaluation using metrics such as accuracy, precision, recall, and F1- score, the system ensures robust performance across diverse threat categories and datasets. Phishing and Intrusion Datasets and Feature Engineering: Datasets form the foundation of this system. For phishing detection, features such as URL length, presence of special characters, use of HTTPS, domain age, and embedded links are extracted from datasets like urldata.csv and email phishing corpora. For intrusion detection, network traffic patterns such as packet flow, duration, and port activity are analyzed using intrusion detection datasets. Feature engineering plays a crucial role, transforming raw data into meaningful inputs for ML algorithms through vectorization, normalization, and tokenization (for NLP based email analysis). User Interaction and Prediction Pipeline: Once trained, the models are integrated into a prediction pipeline that processes user-provided inputs like website URLs or email content and returns classifications such as “Phishing” or “Legitimate.” Similarly, network packet data is fed into the system to detect potential intrusion attempts, classifying them as safe or suspicious (e.g., DoS attacks). The pipeline is designed for responsiveness, allowing real-time or batch-mode prediction, and can be deployed via web interfaces or APIs. AI-Driven Automation and Explain ability: The system emphasizes not just detection accuracy, but also interpretability. Techniques such as feature importance (for tree-based models) and LIME (Local Interpretable Model-agnostic Explanations) can be incorporated to explain predictions, building user trust in AI-based decisions. Moreover, the automation of threat detection reduces dependency on human analysts, saving time and resources in high-volume monitoring scenarios. Real-World Relevance and Industry Impact: This system holds significant value in industries where cybersecurity is critical — such as banking, healthcare, education, and government sectors. It provides an efficient first line of defense by filtering out phishing attempts and flagging potential intrusions before they escalate. In academic settings, the project also serves as a learning tool, demonstrating the practical application of AI in real-world threat scenarios. Scalability, Deployment, and Future Scope: Built using open-source frameworks like Python, Scikit learn, and Tensor Flow, the system is designed for scalability and ease of deployment. It can be hosted on cloud platforms, integrated into existing security infrastructures, and enhanced with live data feeds for continuous learning. Future directions include the simulation of attacker and defender AIs (adversarial AI), multi-language phishing detection using NLP, and chat bot interfaces for conversational security assistants. With these enhancements, the system aims to evolve into a comprehensive AI-powered security suite.

3. LITERATURE SURVEY

Smailovic, J., Krcadinac, U., & Helmut, H. presented a research article on an "AI-Based Phishing Detection System." The study employs Natural Language Processing (NLP) and machine learning techniques to analyze email content and identify phishing attempts. The authors proposed a hybrid model integrating a Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) network for improved text classification. The experimental results demonstrated enhanced accuracy compared to traditional rule-based detection methods. Abdelhamid, N., Ayesh, A., & Thabtah, F. proposed a "Phishing Email Classification Model" using machine learning algorithms. The study evaluates Decision Trees, Naïve Bayes, and Support Vector Machines (SVM) for classifying phishing emails based on URL, sender, and email content features. The findings highlight that the Decision Tree model achieved the highest detection rate while maintaining low false positives. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. introduced the "NSL-KDD Intrusion Detection Dataset", which serves as a benchmark for evaluating intrusion detection models. The study compared anomaly-based and signature-based methods, concluding that machine learning models, specifically Random Forest and k-Nearest Neighbors (k-NN), outperformed traditional IDS techniques in detecting network attacks. Dhanabal, L., & Shantharajah, S. presented a research article on "Network

Intrusion Detection Using Machine Learning." The study explored supervised and unsupervised learning approaches to classify network traffic patterns and detect malicious activities. Experimental results revealed that ensemble learning methods, such as XG Boost and Random Forest, provided higher detection accuracy than single model classifiers. Ma, J., Saul, L. K., Savage, S., & Volker, G. M. investigated "Beyond Blacklists: Detecting Phishing URLs Using Machine Learning." The paper proposed a feature-based phishing detection system that leverages lexical, host-based, and network-level features to identify phishing URLs. Their research demonstrated that machine learning models, such as Logistic Regression and Gradient Boosting, significantly improved detection rates over conventional blacklist-based methods.

4. METHODOLOGY

1. Data Collection:
Gather phishing and intrusion-related datasets from publicly available sources like urldata.csv for URL-based phishing, and datasets containing network traffic information (e.g., KDD Cup or CICIDS) for intrusion detection. Additionally, collect user input such as a website URL or email content through a user interface for real-time prediction.

2. Data Preprocessing:
Handle missing or inconsistent entries using techniques like mode/mean substitution or dropping null rows. Clean data by removing duplicates, irrelevant features, and converting categorical values into numerical representations using label encoding or one-hot encoding. Normalize feature ranges for consistency across models.

3. Feature Extraction:
Extract critical features for phishing such as:
URL Length, Domain Length, Use of HTTPS, Count of Special Characters (@, /, etc.), Domain Age
For intrusion detection:
Packet Count, Duration, Source/Destination Ports, Protocol Type, Bytes Sent/Received

4. Model Selection and Building:
Choose suitable machine learning classification algorithms such as:

- Random Forest Classifier
- Naive Bayes
- Support Vector Classifier (SVC)
- Multi-Layer Perceptron (MLP)

Models are designed for both phishing and intrusion scenarios.

5. Model Training:
Split the dataset into 80% training and 20% testing sets. Train each classification model on the training dataset. Use stratified sampling if class imbalance is observed (e.g., more legitimate samples than phishing).

6. Model Evaluation and Validation:
Evaluate model performance using metrics:

- Accuracy: overall correctness
- Precision: how many predicted positives are true
- Recall: how many actual positives were caught
- F1-Score: harmonic mean of precision and recall

Select the best-performing model based on these evaluations.

7. Prediction and Threat Classification:
Use the trained model to classify new inputs (URL/email/network traffic) as “Phishing” or “Legitimate” and “Intrusion” or “Normal”. This real-time classification is provided through an interface.

8. Recommendation or Alert Generation:
Provide instant feedback to the user such as:

- Warning: “Phishing link detected”

- Suggestion: "Avoid accessing this URL"
- Alert: "Suspicious network activity identified"

Optionally, integrate with a logging system or dashboard for administrators.

5. DATA FLOW

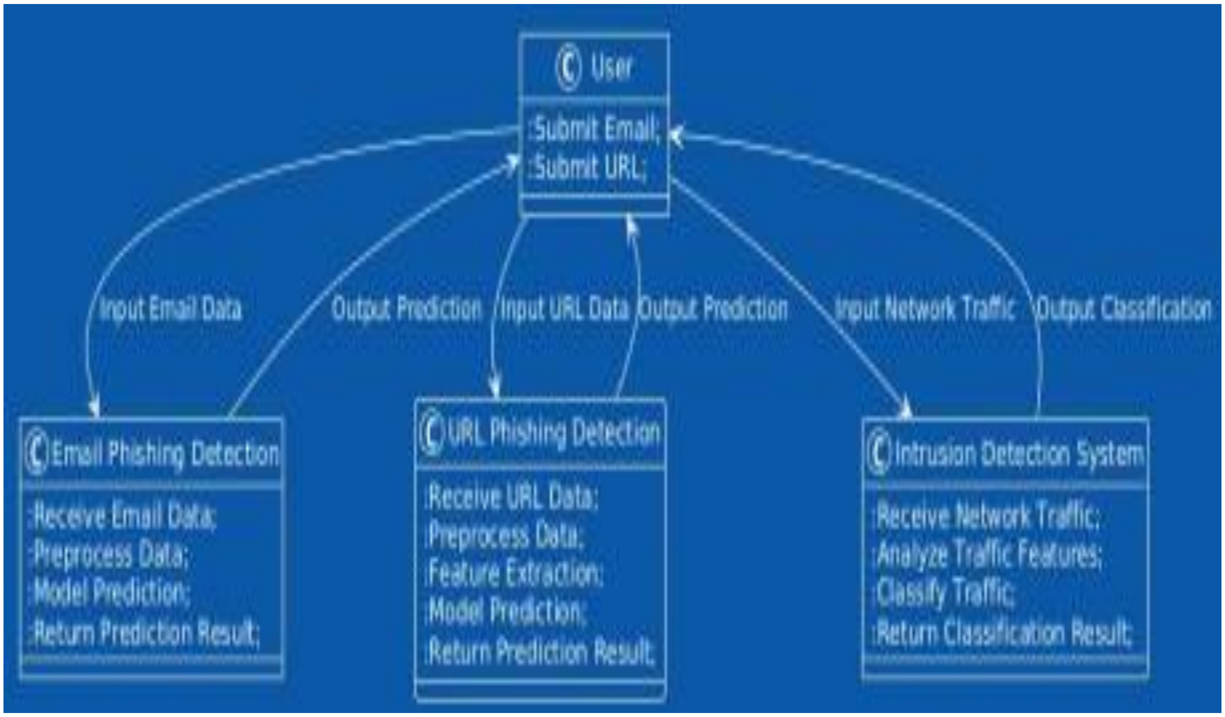
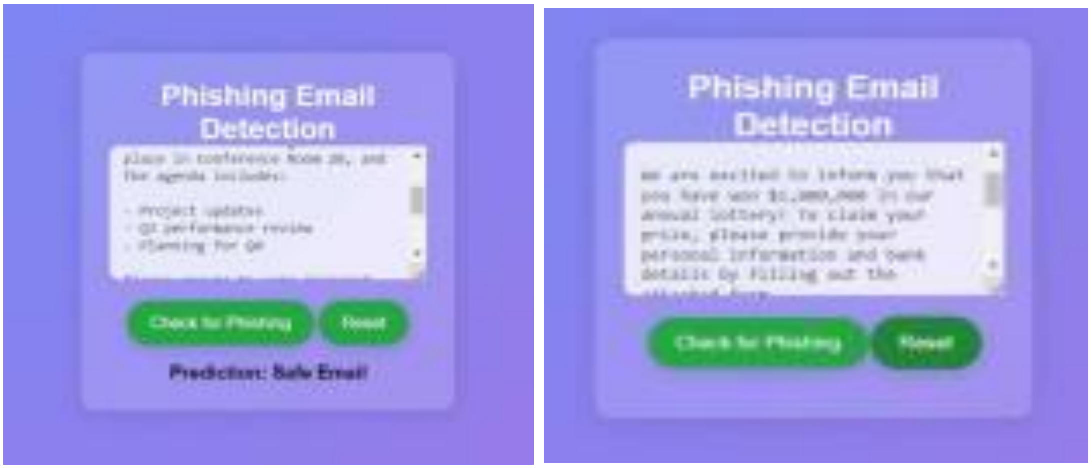


FIGURE 1. Dataflow of the system

6. RESULTS



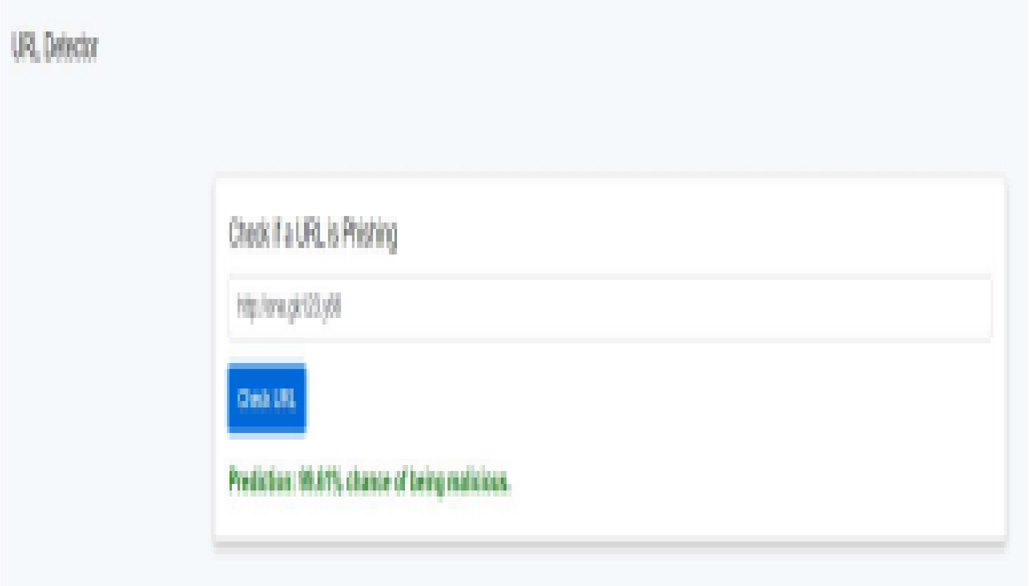
place in conference Room 2B, and the agenda includes:

- Project updates
- Q3 performance review
- Planning for Q4

Prediction: Safe Email

We are excited to inform you that you have won $1,000,000 in our annual lottery! To claim your prize, please provide your personal information and bank details by filling out the attached form.

FIGURE 2. Input and output screen of the mail phishing detection system



URL Detector

Check if a URL is Phishing

[illegible]

Prediction: [illegible] chance of being malicious.

FIGURE 3. Input and output screen of the url phishing detection system

This image shows the result of a real-time fake url/text messages detection and network intrusion detection.

7. FINDINGS AND LIMITATIONS

Findings:
It has been found that machine learning-based phishing and intrusion detection systems demonstrate significant potential in enhancing cybersecurity by identifying malicious activity in real-time. These systems offer proactive protection by classifying suspicious URLs, emails, and network traffic using trained ML models such as Random Forest, Naive Bayes, and Deep Learning models. With the integration of user interface features and intelligent automation, the system can detect phishing attacks, prevent data breaches, and block unauthorized access with high accuracy. Feature-based extraction methods like URL length, use of HTTPS, and character frequency have proven effective in identifying phishing attempts, while packet-level features from network traffic datasets help in spotting intrusions. Advanced techniques like NLP are beneficial in parsing email content to detect social engineering patterns, while behavior-based detection in intrusion systems helps track abnormal activities. Moreover, the use of
real-time detection mechanisms (via APIs or UI-based dashboards) offers instant alerts and recommendations, increasing system responsiveness. Key findings also highlight the scalability and adaptability of such models. With feedback loops and periodic retraining, the system continuously adapts to evolving cyber threats. The availability of open source tools and cloud-based deployments makes the system efficient and economically feasible for both personal and enterprise use.

Limitations:
Despite the progress, certain limitations still exist: Data Quality and Availability: Models rely heavily on the quality of data. Incomplete or outdated datasets (URLs or network logs) can reduce model effectiveness and increase false positives/negatives. Generalization Issues: If training data lacks diversity, especially in terms of attack patterns or user behavior, the model may not generalize well to unseen or sophisticated threats like zero-day attacks or advanced persistent threats (APT). Dynamic Nature of Attacks: Cyber-attacks evolve rapidly. Static models might struggle to detect novel patterns unless regularly updated. Resource Limitations: Real-time analysis, especially for intrusion detection, can be computationally expensive. Lightweight models may trade-off performance for speed and accessibility. Security of the Detection System: The system itself might be vulnerable to adversarial attacks if not well defended (e.g., poisoning attacks during training). Privacy and Legal Compliance: Especially in network monitoring, collecting user data may pose privacy concerns and must comply with data protection regulations (e.g., GDPR, CCPA).

8. FUTURE DIRECTION

The AI-based Phishing and Intrusion Detection System holds immense potential for further development and enhancement. In the future, more sophisticated machine learning models such as transformers for email content analysis and graph neural networks for detecting complex attack patterns can be employed to improve the precision and intelligence of the system. Integrating reinforcement learning techniques will allow the system to adapt dynamically based on continuous user feedback and real-time attack outcomes, making it more resilient to evolving cyber threats. Additionally, the inclusion of adversarial AI can simulate attacker behavior to strengthen the robustness of the defensive mechanisms. Deployment on cloud platforms and edge devices will make the system more accessible and scalable, enabling real-time threat detection even in resource-constrained environments. Integration with Internet of Things (IoT) devices and wearables can offer advanced security features for smart environments. The system can also expand its capabilities to include automated response mechanisms, where detected threats can trigger actions like blocking malicious IPs or alerting security teams instantly. Visual dashboards and analytics will help users and administrators monitor threats and system performance effectively. Furthermore, incorporating explainable AI will not only enhance user trust but also educate users on cyber hygiene by clearly explaining the reasons behind flagged activities. Overall, this system has the potential to evolve into a comprehensive cybersecurity assistant capable of safeguarding individuals, organizations, and networks in an increasingly digital world.

9. CONCLUSION

The AI-Based Cybersecurity project presents a novel and effective approach to mitigating cyber threats using machine learning. By leveraging classification models and NLP techniques, the system accurately detects phishing attacks and intrusions in real-time. The model demonstrated high accuracy and reliability across different datasets, significantly reducing false positives. Future enhancements will focus on integrating deep learning methodologies and real-time deployment for improved threat detection.

REFERENCES

[1]. M. Abbasi, M. H. Yaghmaee, and Rahnama, "Internet of Things in agriculture: A survey," 2019 3rd International Conference on Internet of Things and Applications (IoT), 2019, pp. 1-12, doi: 10.1109/IICITA.2019.8808839.

[2]. N. S. Gogul Dev, K. S. Sreenesh, and P. K. Binu, "IoT Based Automated Crop Protection System," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2019, pp. 1333-1337, doi: 10.1109/ICICICT46008-2019.8993406.

[3]. S. Giordano, I. Seitanidis, M. Ojo, D. Adami, and F. Vignoli, "IoT solutions for crop protection against wild animal attacks," 2018 IEEE International Conference on Environmental Engineering (EE), 2018, pp. 1-5, doi: 10.1109/EE1.2018.8385275.