IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

PHISHING, SPAM AND RANSOMWARE DETECTION

Aishwarya K[1], Dr. Madhu H.K[2]

Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India[1]

Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India[2]

Abstract: With growing cybersecurity threats like phishing, spam and ransomware are increasing rapidly, causing major security risk and data loss, it is difficult to accurately detect them. This paper presents an integrated detection system which uses rule-based heuristics and machine learning models deployed within web-based platform. The system consists of modules like Phishing, URLs, SMS spam, Email spam and Ransomware Detection. The design include user authentication and login history, helps in real-time deployment. We tested the system on standard datasets and found it to be highly accurate and reliable, mainly high accuracy is seen in spam and ransomware classification. These results show that the proposed approach is both scalable and effective for real-time threat detection.

Keywords: Phishing detection, spam filtering, ransomware detection, machine learning, cybersecurity, XGBoost, Random Forest, Naïve Bayes, hybrid models.

I. INTRODUCTION

Phishing, spam and ransomware are still major threats to cybersecurity, taking advantage of weaknesses to steal sensitive information and cause disruptions to operations. In this paper, a unified machine learning-based detection system is presented for four leading cybersecurity issues: phishing URL identification, SMS spam filtration, spam detection in email, and ransomware detection. XGBoost for phishing URLs, Multinomial Naïve Bayes for spam emails and SMS, and Random Forest for ransomware detection these are the machine learning models used for detection.

Traditional methods, including signature-based detection system, are unable to keep up with the increasing complexity. Machine learning (ML) methods have been finding increasing relevance as valuable tools here, as they can facilitate automated feature extraction, pattern recognition, and decision making for a wide range of data formats. Existing studies have worked on phishing URL detection, spam filtering, and ransomware detection individually as separate problems, there is a lack of integrated systems that tackle these domains simultaneously.

The major contributions of the work are:

- An integrated framework for multi-dimensional cyber threat detection.
- Using advance feature extraction techniques for each threat type.
- Enabling easy-to-use, extensible deployment using a web-based interface with historical logging and analysis capabilities.

II. RELATED WORK

Machine learning techniques in cybersecurity have been thoroughly explored for phishing URL, spam message, and ransomware attacks detection, with each posing its own challenges and methodological improvement.

Phishing Detection

Phishing detection has come a long way, utilizing machine learning classifiers that are trained using URL-based features. A study conducted by **Alshamrani et al. (2022)** highlights the extraction of structural, lexical, and WHOIS-based features to boost detection accuracy. Ensemble models, especially XGBoost as indicated by **Zhang and Wang (2021),** have proved resilient predictive ability by coping with intricate patterns in big data. Some studies also support hybrid methods which includes rule-based heuristics and ML techniques to reduce false positives [Phishing Detection Using Machine Learning Techniques, 2022].

Spam Filtering (SMS and Email)

SMS and email spam filtering has been the focus of extensive research as complementary and not individual issues. **White and Lee (2020)** present supervised learning algorithms like Random Forest and SVM for SMS spam, emphasizing

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

feature extraction of content from text message content. In email spam, Naive Bayes classifiers continue to be used because of their efficiency and effectiveness, as presented in the publications by **Green and Hall (2019),** which also include content-based, domain reputation, and authentication features. Hybrid systems integrating rule-based scoring mechanisms with ML classifiers (e.g., integrating SPF, DKIM, DMARC findings with text analysis) have been demonstrated to improve accuracy and decrease misclassification rates [Email Spam Detection with Bayesian Classifiers, 2019].

Ransomware Detection

Ransomware detection studies tend to emphasize behavioral analysis through system-level monitoring, such as file modification rates and indicative API calls. **Brown and Johnson (2021)** present a comprehensive overview of ML methods, noting the effectiveness of Random Forest classifiers learned on system trace features. Balancing accuracy against detection speed in order to cause the least disruption is still a challenge. Coupling feature generation scripts with ML pipelines has played a key role in driving ransomware detection in production environments [Ransomware Detection Using Machine Learning: A Review, 2021].

III. SYSTEM DESIGN

The system is developed as a single platform to identify four critical cybersecurity threats: phishing URLs, SMS spam, email spam, and ransomware. The system is developed to be modular in nature, scalable, and web-enabled so that users can provide inputs easily and get threat analysis in real time.

SYSTEM COMPONENTS

Phishing URL Detection Module:

This module inspects user-submitted URLs to identify whether they are authentic or a phishing attack. It harvests multiple features including the domain's age, usage of secure protocols (HTTPS), suspicious patterns of domains, occurrence of brand impersonation, and login page identification. These features are fed into a machine learning model along with rule-based expert checks to produce a correct prediction.

SMS Spam Detection Module:

Intended to filter SMS messages, this module applies patterns derived from the text message content to distinguish between spam or scam messages. It applies a classification technique trained on classified examples of scam and authentic messages to deliver consistent filtering.

This blended module integrates the analysis of the email body with examination of the sender's domain genuineness and security controls. It scans for embedded phishing URLs in the message, spammy keywords, and possibly malicious attachments. These indicators are input into a rule-based scoring mechanism and a machine learning classifier to provide sophisticated prediction of spam, genuine, or malicious emails.

Ransomware Detection Module:

This module evaluates system behavior characteristics, including frequency of file changes, suspicious or encrypted file extensions ratio, number of new processes executing, and signs of irregular API calls. Depending on these behavior indicators, a machine learning model decides whether the activity of ransomware is probable.

SYSTEM WORKFLOW

The users access the system through a web interface to input URLs, SMS content, email body, or system feature information for ransomware analysis. The input data is passed through to the corresponding detection module that does feature extraction followed by evaluation using trained machine learning models and rule-based heuristics where necessary.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

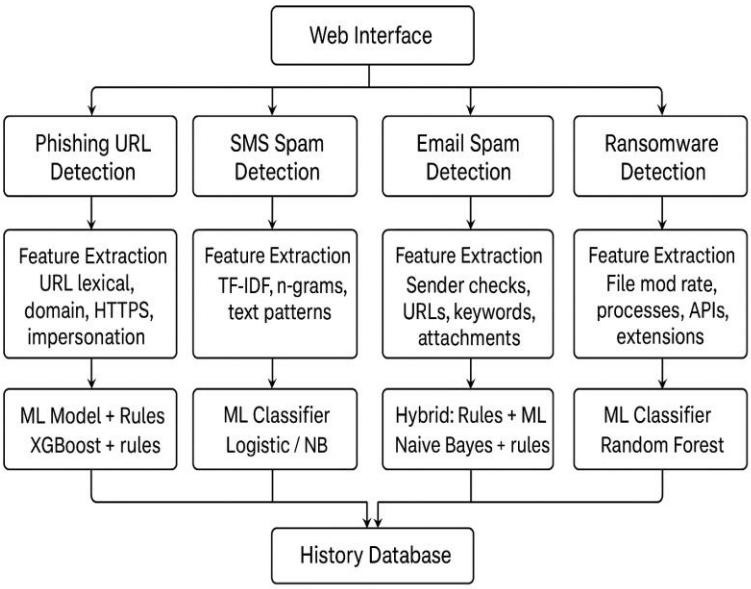

Fig.1. System Architecture

After processing, each module generates a classification result (e.g., benign or phishing for URLs, ham or spam for SMS and email, safe or ransomware for ransomware detection) along with supporting details for user interpretation. For registered users, the system saves these detection events into a secure history database for auditing and analysis.

IV. TOOLS AND TECHNOLOGIES

The backend utilizes Python libraries such as Pandas, scikit-learn, XGBoost, joblib, and Flask for serving. Databases employ SQLite through sqlite3. WHOIS domain queries and HTTP probing are used to enrich feature sets for detection of phishing. Extensibility is supported by the design and real-time web-based user interfaces that enable input, reporting, and management.

V. DATASET

The system learns using various datasets to identify phishing URLs, SMS spam, email spam, and ransomware attacks. These datasets consist of simple labeled data such as URLs or message texts being safe or malicious.

For phishing identification, the dataset consists of URLs marked as benign or malicious. Vital attributes like the age of a domain or if the URL implements secure protocols are not present in the dataset but are determined by inspecting the URL with external checks and rules.

The email spam and SMS datasets include merely the raw message content. The other verification such as verification of email senders and reputation of domains are independently performed using rule-based schemes, not represented directly in the datasets.

Ransomware detection information is made up of system behavior indicators such as the frequency in which files change, the occurrence of unusual encrypted files, and system processes activity. These are acquired by tracking the system and marked as safe or infected.

VI. DATA PREPROCESSING AND CLASS BALANCE

Preprocessing for each data set involves cleaning, normalization, and feature encoding appropriate for model input. Class imbalance is resolved using weighting and sampling methods to enhance detection performance in less common threat classes.

VII. MODEL TRAINING AND VALIDATION

Each of the threat detection modules utilizes machine learning models that are trained on well-prepared data sets. The training focuses on sound learning as well as good class imbalance handling to facilitate reliable classification.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

An XGBoost classifier is trained using benign or malicious URLs. Features learned using rule-based techniques—e.g., domain age from WHOIS, URL pattern, and suspicious patterns—are utilized to enrich the raw data. Class imbalance is managed using weighting of underrepresented classes. Hyperparameter adjustment and threshold tuning optimize the F1-score, balancing recall and precision.

It employs a Multinomial Naive Bayes classifier that was trained on TF-IDF vectorized SMS message texts. The probabilistic framework effectively identifies spam patterns in the text data. Cross-validation methods avoid overfitting and ensure the generalizability of the model.

A hybrid method mixes a Multinomial Naive Bayes classifier learned on email text content with rule-based evaluation of email authentication aspects (SPF, DKIM, DMARC) and domain reputation. The ultimate classification represents a weighted combination of statistical learning and expert rules, enhancing detection precision.

A Random Forest classifier is trained on behavioral system attributes such as file rates of modification, ratios of encrypted files, and process behavior. Stratified validation and hyperparameter optimization maximize sensitivity and minimize false alarms.

Models are tested on independent test datasets with accuracy, precision, recall, and F1-score. Confusion matrix analysis is used to detect error patterns for possible improvements.

VIII. RESULTS AND ANALYSIS

The combined cybersecurity detection system was tested on single test datasets for each of phishing URLs, SMS spam, email spam, and ransomware detection modules, with strong overall performance.

For phishing URL detection, the base classification model resulted in 87.93% accuracy. The confusion matrix shows high true positive and true negative rates: precision was 0.95 on benign URLs and 0.78 on phishing URLs with recall of 0.86 and 0.91 respectively and hence F1-scores of 0.90 and 0.84. Weighted average F1-score of 0.88 proves successful detection irrespective of class imbalance.

The improved XGBoost classifier also achieved better phishing detection with training and validation AUC-PR of 0.94111 and 0.93917 respectively, reflecting better precision-recall trade-off.

In spam detection in emails, the Multinomial Naive Bayes model yielded 96.05% accuracy. The ham emails yielded precision of 0.96, recall of 1.00, and F1-score of 0.98; spam emails maintained perfect precision (1.00) but poor recall (0.70), with an F1-score of 0.83.

The ransomware classification module based on a Random Forest classifier classified with precision, recall, and F1-score of 1.00.

In SMS spam classification, Multinomial Naive Bayes classified about 97%. Ham messages yielded precision 0.97, recall 1.00, and F1-score 0.98; spam messages yielded precision 0.99, recall 0.81, and F1-score 0.89.

Overall, these results confirm the system's consistent threat detection, with the hybrid application of traditional and advanced machine learning algorithms in conjunction with rule-based feature extraction ensuring effective, balanced performance across various domains of cybersecurity.

IX. CONCLUSION

This paper introduces an integrated multi-model phishing URL detection system, SMS spam, email spam, and ransomware, taking advantage of the strengths in machine learning and rule-based feature extraction. From extensive testing on heterogeneous datasets, the system showed robust performance with excellent accuracy, well-balanced precision, and recall for all modules.

The phishing detection module is enhanced by enriched feature sets such as domain age and URL analysis, allowing effective detection of malicious URLs. SMS and email spam classifiers, with the integration of text-based probabilistic models and expert rules, provide accurate filtering and classification. The ransomware detection module effectively detects anomalous system behaviors of malicious activity.

Modular and extensible design enables real-time threat evaluation using an intuitive web interface and scalable backend implementation. Although the system demonstrates encouraging outcomes, future research could include diversifying the dataset, integrating deep learning mechanisms, and rule update automation for accommodating changing cyber threats.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

In total, this combined framework presents an all-encompassing, applicable strategy for multi-dimensional cybersecurity detection with immense potential for real-world application and extension.

X. FUTURE ENHANCEMENT

Expanding on the encouraging findings of the present multi-model cybersecurity detection system, some directions for future research and enhancement are outlined:

Dataset Expansion and Diversity: Integrate increasingly diverse and recent real-world datasets to further improve model generalizability and resistance to novel threat variations.

Deep Learning Integration: Investigate newer deep learning approaches, including recurrent neural networks and transformers for text analysis and convolutional neural networks for behavioral anomaly detection that can potentially identify intricate patterns that are out of reach for typical machine learning.

Automated Rule Learning: Create adaptive frameworks to automatically optimize and refresh rule-based features and heuristics through continuous threat intelligence feeds to minimize manual updates.

Real-Time Detection and Scalability: Enhance system performance to manage high-throughput environments, supporting real-time detection and response within large-scale networks.

Explainability and User Feedback: Implement explainable AI methods to offer transparent model decisions, building user trust and supporting accelerated incident response.

Through the development of these improvements, the system can become an even more powerful, adaptive, and broadly applicable cybersecurity tool, further enhancing protection against a wide range of cyber attacks.

REFERENCES

[1] Abiramasundari, S. (2021). Spam filtering using Semantic and Rule Based model via supervised learning. *Rajasthan Cooperative Recruitment Board*, 25(2), 3975–3992.
[2] Adewale, Y. (2021). Enhanced Short Message Service Spam Filtering System Based on Normalized and Expanded Text. *Babcock University Publications*
[3] Alhawi, O. M. K., Baldwin, J., & Dehghantanha, A. (2018). Leveraging machine learning Techniques for Windows Ransomware Network Traffic Detection. In A. Dehghantanha, M. Conti, & T. Dargahi (Eds.), *Cyber Threat Intelligence* (pp. 93–106). Springer.
[4] Alhashmi, A. A., Darem, A. A., Alshammari, A. B., Darem, L. A., Sheatah, H. K., & Effghi, R. (2024). Ransomware Early Detection Techniques. *Engineering, Technology & Applied Science Research*, 14(3), 14497–14503.
[5] Alqahtani, A., & Sheldon, F. T. (2022). Techniques to Build Early Detection ML Models: Limitations and Prospects. *International Journal of Semantic Computing*, 16(1), 145–160.
[6] Al-Rimy, B. A. S., Maarof, M. A., & Shaid, S. Z. M. (2019). Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace selection. *Future Generation Computer Systems*, 101, 476–491.
[7] Alshahrani, A. (2021). Intelligent Security Schema for SMS Spam Message Based on Machine Learning Algorithms.
[8] Anderson, H. S., & Roth, P. (2018). Ember: An open dataset for training static PE malware machine learning models. *arXiv preprint* arXiv:1804.04637.
[9] Berrueta, E., Morato, D., Magaña, E., & Izal, M. (2018). Ransomware early detection by the analysis of file sharing traffic.
[10] Gangare, A., Rathore, J., Tadge, A., Shrivastav, A., Yadav, R., & Sisodiya, P. (2022). Implementation of Spam Classifier using Naïve Bayes Algorithm. *International Research Journal of Engineering and Technology*, 9(2).
[11] Gomaa, W. H. (2020). The Impact of Deep Learning Techniques on SMS Spam Filtering. *International Journal of Advanced Computer Science and Applications*, 11(1).
[12] Gupta, S. D., Saha, S., & Das, S. K. (2021). SMS Spam Detection Using Machine Learning. *Journal of Physics: Conference Series*, 1797, 012017.
[13] Julis, M., & Alagesan, S. (2020). Spam detection in SMS using machine learning through text mining.
[14] Kontsewaya, Y., Antonova, E., & Artamonov, A. (2020). Evaluating the Effectiveness of Machine Learning Methods for Spam Detection. *Procedia Computer Science*, 169, 822–829.
[15] Kudupudi, N., & Nair, S. (2021). Spam Message Detection using Logistic Regression.
[16] Liu, M., Zhang, Y., Liu, B., Li, Z., Duan, H., & Sun, D. (2021). Detecting and Characterizing SMS Spearphishing Attacks. In *Proceedings of the 2021 ACM SIGSAC*

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12818

[17] Mohasseb, A., Aziz, B., & Kanavos, A. (2020). SMS Spam Identification and Risk Assessment Evaluations.
[18] Morato, D., Berrueta, E., Magaña, E., & Izal, M. (2020). Open repository for the evaluation of ransomware detection tools. *Mendeley Data*.
[19] Nazir, S., Khan, H., & Haq, A. (2020). Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms. *Security and Communication Networks*, 2020, 8873639.
[20] Ora, A. (2020). Spam Detection in Short Message Service Using Natural Language Processing and Machine Learning Techniques. (Master's thesis, National College of Ireland).
[21] Palad, E. B., Tangkeko, M., Magpantay, L., & Sipin, G. (2019). Classification of Filipino Online Scam Incident Text Using Data Mining Techniques.
[22] Shaukat, K., Luo, S., Chen, S., & Liu, D. (2020). Cyber threat detection using machine learning techniques: A performance evaluation perspective.

