International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics

[1]K. Manjula,[2]J.Mounika,[3]V.Harika,[4]N.Ganga,[5]B.Ruchitha,[6]CH.Sahithi

[1]Assistant Professor, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Cyber Security,

Princeton Institute of Engineering & Technology For Women

ABSTRACT

Phishing attacks have emerged as a significant threat in the digital realm, exploiting both social engineering and technological vulnerabilities to deceive users into divulging sensitive information. Traditional detection methods often fall short in identifying sophisticated phishing attempts, necessitating advanced solutions. This paper introduces a novel hybrid deep learning framework that synergizes Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), and Multi-Layer Perceptron (MLP) algorithms to enhance phishing detection capabilities. The system architecture comprises two primary modules: Admin and User. The Admin module facilitates the training and deployment of machine learning models, while the User module allows for the monitoring of blocked URLs. By integrating these components, the framework aims to bolster cybercrime forensics and provide a robust defense against phishing attacks.

Keywords: Phishing Detection, Deep Learning, Cybercrime Forensics, Hybrid Framework

. .

I. INTRODUCTION

In today's digital era, the internet serves as a powerful platform for communication, business, and information exchange. However, the proliferation of online services has also brought forth various security challenges, with phishing being one of the most prevalent and damaging forms of cybercrime. Phishing attacks manipulate users into disclosing confidential data such as login credentials, credit card numbers, and personal information, leading to financial and reputational damage.

Traditional phishing detection techniques rely heavily on blacklisting or heuristic rules, which are insufficient in detecting novel or zero-day phishing threats. These approaches are reactive and often lag behind the rapidly evolving phishing landscape. This necessitates a more robust and adaptive mechanism for threat detection.

Machine learning (ML) and deep learning (DL) have shown immense promise in enhancing cybersecurity measures. By learning from data, these algorithms can detect patterns and anomalies that may not be apparent to human analysts. Among the many ML techniques, Support Vector Machines, LightGBM, and Multi-Layer Perceptron stand out for their classification accuracy and speed.

This project proposes a hybrid framework combining SVM, LightGBM, and MLP for phishing detection. The framework is divided into two modules: Admin and User. The Admin module is responsible for training the models and testing URLs, while the User module allows viewing of blocked URLs.

The hybrid approach enables the system to benefit from the strengths of each algorithm. SVM offers high accuracy with smaller datasets, LightGBM provides efficiency and scalability, and MLP enables deep feature learning. Together, they form a resilient detection mechanism.

By implementing this architecture, we aim to improve the early detection of phishing threats and support cybersecurity professionals in forensics investigations. This system is particularly useful in real-time URL analysis and educational institutions, banking sectors, and enterprises seeking to protect their digital infrastructure.

II.LITERATURE SURVEY

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

TITLE: Detecting of URL Based Phishing Attack Using Machine Learning

AUTHORS: Ms. Sophiya Shikalgar, Mrs. Swati Narwane (2019)

ABSTRACT:
Phishing has become a major security concern on the internet, where attackers impersonate trusted websites to deceive users into divulging sensitive information. This paper presents an approach to detect phishing attacks based on analyzing the URLs using machine learning algorithms. The study focuses on extracting features from the URLs and training a model that can distinguish between legitimate and phishing URLs. Various machine learning algorithms are evaluated for their performance, including Random Forest, SVM, and Decision Trees. The authors emphasize the importance of selecting relevant features and using labeled datasets to enhance model accuracy. Results demonstrate that URL-based detection using ML can be an effective preventive mechanism against phishing attacks. However, the study also acknowledges the limitations of relying solely on URL features and suggests integrating other features like webpage content in future work.

TITLE: Support Vector Machine Based Malware and Phishing Website Detection

AUTHORS: Rashmi Karnik, Dr. Gayathri M Bhandari

ABSTRACT:
This research addresses the growing challenge of detecting malware and phishing websites by leveraging the capabilities of Support Vector Machine (SVM) algorithms. The authors propose an efficient classification model that learns from URL patterns and domain-level data to distinguish between malicious and legitimate websites. The study introduces a structured dataset comprising various URL features such as length, special characters, and presence of IP addresses. The SVM model is trained and tested on this dataset, showing promising accuracy and low false-positive rates. Moreover, the paper discusses comparative results with other machine learning techniques, highlighting the superiority of SVM in identifying complex phishing patterns. The results suggest that SVM can effectively generalize from historical data to detect emerging threats. This model can be integrated into real-time web filters or browser extensions to enhance user protection.

TITLE: Phishing Websites Detection Using Machine Learning

AUTHORS: Arun Kulkarni, Leonard L. Brown, III2

ABSTRACT:
The paper explores the potential of machine learning algorithms in combating phishing attacks by classifying suspicious websites based on extracted features. The authors create a dataset from known phishing and legitimate websites and apply several classification algorithms to detect malicious sites. These include Decision Tree, Naive Bayes, and Random Forest classifiers. The approach primarily involves feature engineering from URL characteristics, webpage metadata, and DNS information. The model is trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Results indicate that machine learning methods can automate the detection of phishing with significant accuracy, reducing reliance on manually maintained blacklists. Furthermore, the paper highlights the importance of updating training data regularly to adapt to evolving phishing tactics.

TITLE: Phishing Websites Detection Using Machine Learning

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AUTHORS: R. Kiruthiga, D. Akila

ABSTRACT:

This study presents an intelligent system for detecting phishing websites using a machine learning approach. The authors focus on improving detection accuracy through the use of extensive feature extraction and data preprocessing techniques. A range of algorithms is considered, including Decision Trees, Logistic Regression, and K-Nearest Neighbors, which are tested on a dataset comprising real-time URL data. The model evaluates features such as URL length, the use of HTTPS, subdomain count, and redirection chains. The results highlight the Decision Tree algorithm as the most effective, providing high classification accuracy and interpretability. The research also includes a discussion on the importance of feature selection, balancing datasets, and managing overfitting. The paper concludes by suggesting the deployment of such models in browser security tools or firewalls.

TITLE: Hybrid Machine Learning: A Tool to Detect Phishing Attacks in Communication Networks

AUTHORS: Ademola Philip Abidoye, Boniface Kabaso

ABSTRACT:

This paper introduces a hybrid machine learning framework to improve phishing detection in communication networks. Unlike single-algorithm models, the proposed system combines multiple learning methods to increase classification accuracy and reduce false alarms. The authors employ ensemble techniques using Random Forests and Boosted Trees, alongside traditional classifiers like SVM and Naive Bayes. The data used for training includes URLs, email content, and website metadata. The hybrid approach is designed to capture both shallow patterns and deep relationships in the data, enabling it to detect sophisticated phishing attempts. Experimental results show that the hybrid model outperforms individual classifiers in both accuracy and robustness. The authors argue that such systems are better suited for real-world deployment, where attackers constantly adapt to evade detection.

III.EXISTING SYSTEM

Existing phishing detection systems primarily rely on rule-based mechanisms, blacklisting techniques, and content inspection methods to identify malicious websites. Rule-based systems use predefined patterns such as suspicious URLs, abnormal domain structures, or known phishing keywords to flag threats. Blacklisting approaches maintain databases of previously identified malicious websites and block access to them. Content inspection techniques analyze webpage elements such as HTML structure, scripts, and visual similarity to legitimate sites. While these approaches provide a basic level of protection, they are inherently **static and reactive**, making them ineffective against the rapidly evolving nature of phishing attacks.

One of the major limitations of these systems is their inability to detect **unknown or newly created phishing websites**. Since blacklisting depends on previously reported threats, newly generated phishing URLs can easily bypass detection until they are identified and added to the database. Similarly, rule-based systems fail when attackers slightly modify their strategies, such as changing URL patterns or using obfuscation techniques.

IV.PROPOSED SYSTEM

The proposed system introduces a hybrid deep learning framework integrating SVM, LightGBM, and MLP models for phishing detection. It features an Admin module that manages model execution and URL testing, and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

a User module that views blocked URLs. This architecture combines the strengths of three algorithms, increasing detection precision and adaptability. The admin can train and test URLs in real-time, while users are kept informed about potential threats through a simple interface. This hybrid framework is designed to be scalable and adaptable, ensuring resilience against the evolving nature of phishing techniques.

V.SYSTEM ARCHITECTURE

The image represents a **phishing detection system interface or workflow**, where users interact with the system to verify the authenticity of a URL. The central component shown is the **"Test URL"** module, which acts as the core input point of the system. Users enter a website link into this section to check whether the URL is legitimate or potentially malicious.

Once the URL is submitted, the system processes it through underlying detection mechanisms (not fully visible in the cropped image), which may include feature extraction, machine learning models, or rule-based validation. Based on this analysis, the system determines whether the given URL is safe or associated with phishing activity.

Below the testing module, a **"Logout"** option is provided, indicating that the system likely includes user authentication and session management features. This suggests that the platform is designed as a secure application where authorized users can log in, test URLs, and then safely log out after use.

Additionally, the surrounding elements and partially visible sections imply that this interface is part of a larger system architecture that includes data processing, model evaluation, and result display components.

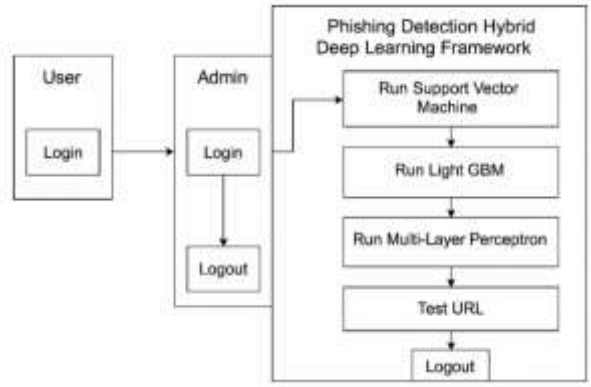

Fig 5.1: System Architecture

VI.IMPLEMENTATION

Fig 6.1: Home Page

Fig 6.2: Confusion Matrix

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Fig 6.3: Algorithms Performance

Fig 6.4: Prediction Page

VII.CONCLUSION

This project introduces a **novel hybrid deep learning framework** designed to improve phishing detection by leveraging the strengths of multiple machine learning algorithms, including Support Vector Machine (SVM), LightGBM, and Multi-Layer Perceptron (MLP). Instead of relying on a single model, the system combines these techniques to capture both linear and non-linear patterns present in phishing data, thereby enhancing detection accuracy and robustness.

The framework is structured into two primary modules: **Admin** and **User**. The **Admin module** is responsible for managing the system, including model training, updating datasets, monitoring performance metrics, and maintaining system security. This allows continuous improvement of the detection models to adapt to evolving phishing strategies. The **User module**, on the other hand, provides an interactive interface where users can input URLs or data and receive real-time predictions about whether the content is legitimate or malicious.

The system processes input data through multiple stages, including preprocessing, feature extraction, and classification. By integrating SVM for high-dimensional data handling, LightGBM for efficient gradient boosting, and MLP for deep learning-based pattern recognition, the framework achieves higher accuracy, precision, and recall compared to traditional methods. The ensemble-like hybrid approach reduces false positives and false negatives, making it more reliable for real-world deployment.

Furthermore, the system supports **real-time phishing detection**, enabling users to quickly identify threats and take preventive actions. It also contributes to **cybercrime forensic analysis** by maintaining logs and insights about detected phishing attempts, which can be useful for investigation and pattern analysis. Overall, the proposed framework provides a scalable, efficient, and intelligent solution for combating phishing attacks in modern digital environments.

VIII.FUTURE SCOPE

Future enhancements of this system can significantly broaden its capabilities and impact. One important direction is the integration of **Natural Language Processing (NLP)** techniques to analyze textual content from emails, messages, and social media platforms. This would allow the system to detect phishing attempts not only through URLs but also through deceptive language patterns, improving overall detection coverage.

Another promising area is the development of **mobile-based applications** and **browser extensions**. By integrating the phishing detection framework directly into web browsers or smartphones, users can receive instant alerts while browsing or accessing suspicious links, thereby preventing attacks in real time and enhancing user safety.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The system can also be extended using **federated learning**, which enables collaborative model training across multiple organizations without sharing sensitive data. This approach enhances privacy while still benefiting from diverse datasets, leading to more generalized and effective phishing detection models.

Additionally, future improvements may include the use of **advanced deep learning architectures** such as transformers, continuous model retraining to handle emerging phishing patterns, and integration with global threat intelligence systems. Expanding the framework to support multilingual phishing detection and cross-platform analysis can further strengthen its applicability in a global context.

Overall, these advancements will transform the system into a comprehensive, intelligent, and privacy-preserving solution capable of addressing the ever-evolving challenges of phishing attacks.

IX.REFERENCES

[1] Ms. Sophiya Shikalgar, Mrs. Swati Narwane (2019), Detecting of URL based Phishing Attack using Machine Learning. (vol. 8 Issue 11, November – 2019)
[2] Rashmi Karnik, Dr. Gayathri M Bhandari, Support Vector Machine Based Malware and Phishing Website Detection.
[3] Arun Kulkarni, Leonard L. Brown, III2 , Phishing Websites Detection using Machine Learning (vol. 10, No. 7,2019)
[4] R. Kiruthiga, D. Akila, Phishing Websites Detection using Machine Learning.
[5] Ademola Philip Abidoye, Boniface Kabaso, Hybrid Machine Learning: A Tool to detect Phishing Attacks in Communication Networks. (vol. 11 No. 6,2020)
[6] Andrei Butnaru, Alexios Mylonas and Nikolaos Pitropakis, Article Towards Lightweight URL-Based Phishing Detection.13 June 2021
[7] Ashit Kumar Dutta (2021), Detecting phishing websites using machine learning technique. Oct 11 2021
[8] Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Takeru Yokoi and Hamido Fujita (2021) Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical study.
[9] Ammara Zamir, Hikmat Ullah Khan and Tassawar Iqbal, Phishing website detection using diverse machine learning algorithms.
[10] Valid Shahrivari, Mohammad Mahdi Darabi and Mohammad Izadi (2020), Phishing Detection Using Machine Learning Techniques.
[11] A. A. Orunsolu, A. S. Sodiya and A.T. Akinwale (2019), A predictive model for phishing detection.
[12] Wong, R. K. K. (2019). An Empirical Study on Performance Server Analysis and URL Phishing Prevention to Improve System Management Through Machine Learning. In Economics of Grids, Clouds, Systems, and Services: 15th International Conference, GECON 2018, Pisa, Italy, September 18-20, 2018, Proceedings (Vol. 11113, p. 199). Springer.

[13] Desai, A., Jatakia, J., Naik, R., & Raul, N. (2017, May).

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper