IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

"Subterfuge Sentry: Guarding Against Sneaky Malware Tactics"

Ms. Samruddhi Babasaheb Shinde
Dr. S.P. Pawar
Department of Computer Science Engineering
SVERI's College of Engineering, Pandharpur, Maharashtra, India.

Abstract- The goal of this project is to create a reliable system that can instantly determine the security status of web links in an era where internet safety is crucial. Support Vector Machines (SVM), Random Forest, Boosting, and Multilayer Perceptron (MLP) are the four different machine learning algorithms that are incorporated into the suggested solution, which makes use of a browser plugin. Every algorithm brings something special to the analysis of several aspects related to connections on the web, offering an all-encompassing assessment of their security.

The first step of the project is to gather a variety of datasets with annotated examples of secure and insecure links. These datasets include a wide range of attributes, including URL structure, SSL/TLS certificate details, and historical threat intelligence data. Each algorithm is then trained and optimized using the dataset, maximizing its capacity to correctly categorize links into secure and insecure groups.

After the algorithms are trained individually, a unique technique called ensemble learning is presented to improve overall accuracy. A voting technique that incorporates the predictions of SVM, Random Forest, Boosting, and MLP is used by the ensemble model. By combining the advantages of each method, this cooperative decision-making process minimizes potential flaws and produces a more robust and trustworthy link security detection system. Comprehensive testing and validation utilizing a variety of datasets and real-world scenarios are used to assess the project's success. The system's total accuracy is tested to show how well the ensemble learning approach improves link security detection. The project's outputs support further initiatives to improve internet security and provide users with the means to safely traverse the digital environment.

***Key Words*:** Link Security, Ensemble Learning, Machine Learning, Support Vector Machines (SVM), Random Forest, Boosting, Multilayer Perceptron (MLP), Voting Algorithm, Browser Extension, Web Security, Online Threat Detection, Real-time Classification, Feature Extraction, Accuracy Assessment, Cybersecurity, ThreatIntelligence, Online Safety.

INTRODUCTION

Given the widespread prevalence of digital connectivity in today's world, online activity security is critical. The identification of potentially dangerous web connections that could direct users to security risks like malware, phishing, or other cyberattacks is a crucial component of today's cybersecurity environment. With the use of a browser extension that is outfitted with cutting-edge machine learning algorithms for real-time link security detection, our research presents a fresh solution to this issue.

The system incorporates four independent algorithms, namely Support Vector Machines (SVM), Random Forest, Boosting, and Multilayer Perceptron (MLP), each of which brings a unique set of strengths to the analysis of different aspects related to online links. These attributes include things like the SSL/TLS certificate information, the URL structure, and threat intelligence data from the past. Through utilizing these algorithms' advantages, our goal is to develop a thorough and efficient link security detection system.

We present an ensemble learning strategy to further improve our system's accuracy. An effective decision-making process is produced by the ensemble model's Voting algorithm, which aggregates the predictions of several algorithms. By working together, the shortcomings of separate algorithms are lessened, and the total accuracy of link security categorization is raised.

The project prioritizes user accessibility and convenience in addition to the technical aspects of algorithmic implementation. A browser extension that is easy to use and integrates effortlessly into various online browsers is designed. This addon informs users in real time on the

security state of URLs they come across while browsing. By providing customers with this knowledge, they can make educated make choices and safely traverse the digital terrain. Our goal is to show the effectiveness of our system through extensive testing, validation, and assessment utilizing a variety of datasets and real-world scenarios. The accomplishment of this project provides consumers with a trustworthy tool to browse the web with confidence in the face of changing cyberthreats, thereby supporting continued efforts to strengthen online security.

LITERATURE SURVEY

1. A Cascade Approach" presents a comprehensive exploration of malware detection methodologies, emphasizing the novelty of their proposed cascade approach. The main findings reveal that the cascade one- sided perceptron (COS-P) algorithm, including its mapped and kernelized versions, exhibits promising accuracy and sensitivity in distinguishing malware from benign files. The study showcases the limitations of conventional signature-based techniques in handling diverse malware behaviors and highlights the need for advanced detection mechanisms. By organizing sources into themes, it becomes evident that the progression of research in this field revolves around enhancing the effectiveness of machine learning-based detection methods, particularly focusing on ensemble techniques and feature engineering to improve classification performance. The key strengths of the paper lie in its systematic comparison of various COS-P adaptations, comprehensive experimentation, and the introduction of algorithmic optimizations to enhance scalability. However, some weaknesses include the absence of real-time testing and an exhaustive analysis of potential false positives. In conclusion, the cascade approach demonstrates its potential as an advanced malware detection solution, showcasing significant advancements over traditional methods and underscoring the value of ensemble learning and algorithmic cascades in cybersecurity.[1]

2. "Malware Detection & Classification using Machine Learning" addresses the crucial issue of malware detection in the current digital landscape. In light of the escalating risk posed by constantly evolving and polymorphic malware, the paper emphasizes the need for effective detection techniques beyond traditional signature-based methods. Recognizing the limitations of conventional tools in tackling dynamic malware behaviors, the authors propose leveraging Machine Learning (ML) techniques for detection. The paper outlines a methodology involving the extraction of behavioral patterns through static or dynamic analysis, followed by the application of diverse ML algorithms to determine whether a given file is malware or not. The study examines both behavioral- based detection methods and the potential of ML algorithms to create a social-based malware recognition and classification model. The authors highlight the significance of ML-driven detection due to the growing volume of novel malware, presenting an urgent need for improved cybersecurity measures. They categorize various types of malwares, such as adware, spyware, viruses, worms, Trojans, rootkits, ransomware, and more, underlining the diverse threat landscape. The paper also discusses malware discovery techniques, dividing them into signature-based and behavior-based approaches. It details the features used for analysis and outlines several machines learning algorithms, including Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Decision Trees, applied in different studies for malware detection and
classification. Experimental results demonstrate the superiority of ML-driven techniques over traditional signature-based approaches, with enhanced accuracy and efficiency in malware detection. Overall, the paper underscores the potential of Machine Learning to revolutionize malware detection and offers valuable insights into various algorithms that have shown promise in this domain.[2]

3. The paper presents a comprehensive study on the effectiveness of deep neural networks, specifically DenseNet, for detecting malware through a visual feature approach. The authors investigate the vulnerability of deep learning models to adversarial attacks, focusing on Gaussian noise and the Fast Gradient Sign Method (FGSM). Using benchmark datasets, the proposed DenseNet model achieves high accuracy and F1-scores for malware detection. The study evaluates the model's resilience to adversarial attacks and showcases its robustness against poisoning and evasion attacks. The research sheds light on the importance of developing malware detection systems that can withstand adversarial attempts, contributing to the advancement of secure computing environments. limitation is the focus solely on visual features while not considering other essential malware detection features, potentially affecting the real- world applicability of the proposed DenseNet approach.[3]

4. Incorporating a hybrid approach, the study employs a Voting Classifier to effectively amalgamate numerous machine learning models for classification through majority voting, thus creating a singular robust classifier. This approach demonstrates manageable execution times, rendering it potentially suitable for extensive malware analysis, particularly in large-scale applications.[4]

5. A static malware detection system that leverages data mining techniques. The study evaluates the effectiveness of SVM, J48, and Naïve Bayes classifiers in detecting malware. Interestingly, results indicate that classifiers based on the DLL name feature exhibit notably poor

detection rates. Furthermore, the Naïve Bayes classifier consistently demonstrates subpar detection accuracy across various scenarios. This study sheds light on the intricate interplay between data mining methods and their application to static malware detection, highlighting the nuanced performance variations among different classifiers and features.[5]

AIM & OBJECTIVES

- Developing Machine Learning Models.
- Feature Extraction and Dataset Creation.
- Ensemble learning integration.
- Browser Extension Development.
- Real time link security detection.

MOTIVATION

This project's driving force is the growing importance of cybersecurity in the digital era. The hazards of dangerous activities like malware, phishing, and cyberattacks have increased along with our dependence on online platforms and increased interconnection. the driving force behind creating a reliable machine learning and ensemble learning detection solution for link security.

SCOPE

This project has a broad scope that includes several facets of user interface, cybersecurity, and machine learning. The project's goal is to provide a thorough link security detection system with an emphasis on user empowerment and real-time security.

PROBLEM DEFINITION

In the digital age, users face an escalating risk of encountering malicious web links that can lead to a range of security threats, including phishing attacks, malware infections, and other cyber-attacks. Traditional methods of link analysis often struggle to keep pace with the evolving tactics employed by cybercriminals, leaving users vulnerable to online threats. The problem addressed by this project is the lack of an efficient and real-time link security detection system that can accurately classify the security status of web links and empower users to make informed decisions during their online activities.

SYSTEM ARCHITECTURE

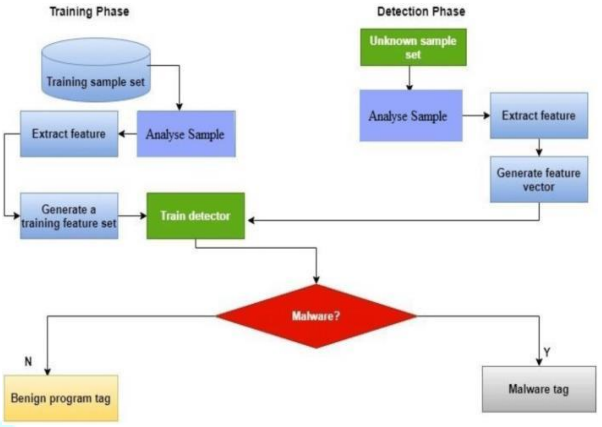


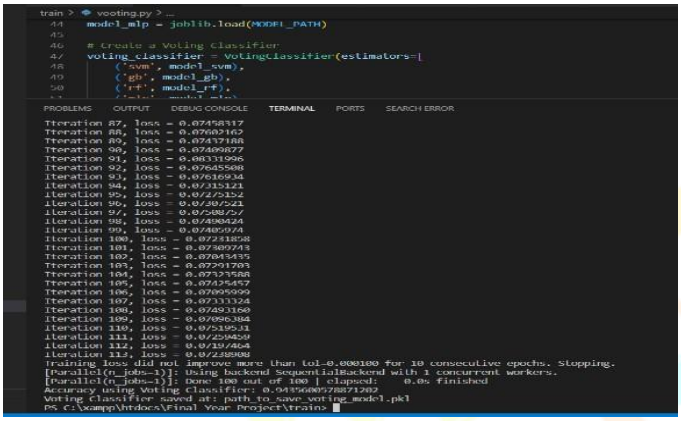
Fig -1: System Architecture Diagram

- The user interacts with the system through a browser extension interface. The extension seamlessly integrates into popular web browsers, ensuring a user-friendly experience.
- This module is responsible for conducting real-time link security assessments. It interfaces with the machine learning models and utilizes ensemble learning techniques to classify links as either secure orinsecure.
- The system incorporates four machine learning algorithms: Support Vector Machines (SVM), Random Forest, Boosting, and Multilayer Perceptron (MLP). Each model is trained on a diverse dataset containing labeled instances of secure and insecure links.
- The ensemble learning module combines the predictions of individual machine learning models (SVM, Random Forest, Boosting, MLP) using a Voting algorithm. This collaborative decision-making process aims to improve overall link security detection accuracy.
- The system extracts relevant features from web links,including URL structure, SSL/TLS certificate details,and historical threat intelligence data. These features serve as inputs to the machine learning models, contributing to the accuracy of link security assessments
- The link security detection module communicates in real-time with the browser extension interface, providing instantaneous feedback to the user as they navigate the web. This ensures timely alerts and empowers users to make informed decisions.

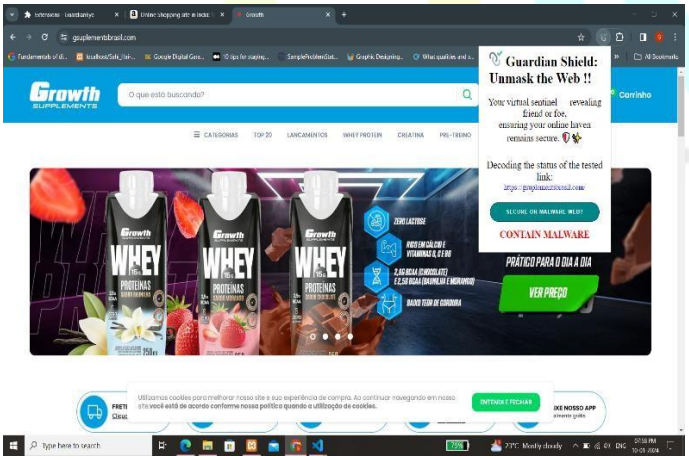
- In the case of a potential security threat, the system triggers an alert mechanism within the browser extension interface. Users receive clear and actionable alerts, informing them of the security status of the linkthey are about to access.

RESULT

Final Voting Classifier Accuracy:



Final Output:



CONCLUSION

In summary, a major step toward improving online security has been made with the creation of a user-friendly browser extension and a link security detection system that makes use of ensemble machine learning techniques. This system's complete approach tackles the growing problems caused by harmful online links by fusing cutting-edge machine learning algorithms with real-time user engagement.

REFERENCES

[1] N. Milosevic, “History of malware,” 02 2013.

[2] Internet Crime Report, 2021, https://www.ic3.gov/

[3] Mouhammd Alkasassbeh, Mohammad A. Abbadi, Ahmed M. AlBustanji. " LightGBM Algorithm for Malware Detection." Applied Sciences volume 1230 (2022)

[4] 14. Or-Meir, O.; Nissim, N.; Elovici, Y.; Rokach, L. Dynamic malware analysis in the modern era—A state of the art survey. ACM Comput. Surv. 2019, 52, 1–48. [CrossRef] 15. Albulayhi, K.; Abu Al-Haija, Q.; Alsuhibany, S.A.; Jillepalli, A.A.; Ashrafuzzaman, M.; Sheldon, F.T. IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method. Appl. Sci. 2022,12, 5015.

[5] Document management – portable document format – part 1: Pdf 1.7. Standard, International Organization for Standardization, Geneva, CH, Mar. 2008.

[6]PDF properties and metadata, Adobe Acrobat Accessed 6,Dec 2022

[7] Aslan, Ömer & Samet, Refik. (2020). A Comprehensive Review on Malware Detection Approaches. IEEE Access. 8. 1-1. 10.1109/ACCESS.2019.2963724.

[8] Elingiusti, Michele & Aniello, Leonardo & Querzoni, Leonardo. (2018). PDF-Malware Detection: A Survey and Taxonomy of Current Techniques. 10.1007/978-3-319- 73951-9_9.

[9] Albahar, Marwan & Thanoon, Mohammed & Alzilai, Monaj & Alrehily, Alla & Alfaar, Munirah & Al-Ghamdi,

Maimoona & Alassaf, Norah. (2021). Toward Robust Classifiers for PDF Malware Detection. Computers, Materials and Continua. 69. 2181-2202. 10.32604/cmc.2021.018260.

[10] VirusTotal https://virustotal.com/.

[11] Contagio Malware Dump, "External data source," [Online]. Available: http://contagiodump.blogspot.com.au

[12] Falah, Ahmed & Pan, Lei & Huda, Shamsul & Pokhrel, Shiva & Anwar, Adnan. (2021). Improving malicious PDF classifier with feature engineering: A data-driven approach. Future Generation Computer Systems. 115. 314-326. 10.1016/j.future.2020.09.015.

[13] CIC-Evasive-PDFMal2022 Dataset CIC-Evasive-PDFMal2022 | Datasets | Canadian Institute for Cybersecurity | UNB

[14] Abu Al-Haija, Q.; Odeh, A.; Qattous, H. PDF Malware Detection Based on Optimizable Decision Trees.Electronics 2022, 11, 3142.

[15] Chandran, P. & Hema, Rajini & Jeyakarthic, M.. (2022). Invasive weed optimization with stacked long short term memory for PDF malware detection and classification. International journal of health sciences. 4187- 4204. 10.53730/ijhs.v6nS5.9540. Kumar, Akshi. (2018). Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series. 1142. 012012. 10.1088/1742-6596/1142/1/012012.