International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Detection of Network Traffic Analyzer Website Using Machine Learning Algorithm

Ms. AKSHAYA LEKSHMI S S[1], Ms. C. VISHNU PRIYA[2]

[1] Ms. AKSHAYA LEKSHMI S S, M.Sc CFIS, Department of Computer Science Engineering, akshayamakairam@gmail.com, Dr.MGR UNIVERSITY, Chennai, India

[2] Ms. C VISHNU PRIYA, Assistant Professor, Cyber Forensics and Information Security, IDE, University of Madras, Chepauk

---

Abstract - In the ever-evolving landscape of cybersecurity, timely identification of cyberattacks is critical for protecting digital infrastructures. This research introduces a system that applies supervised machine learning techniques to classify and predict different types of cyber threats, including malware, phishing, and brute-force attacks. Using a comprehensive dataset of past attack instances, the model learns to recognize patterns and features that distinguish each type of attack. The study explores the performance of various supervised learning algorithms such as Decision Trees, Support Vector Machines (SVM), and Neural Networks to identify the most accurate and reliable approach for threat detection. To ensure continued effectiveness, the model is regularly updated to adapt to emerging threats. Combined with tools like the confusion matrix for performance evaluation, the proposed system provides cybersecurity teams with a practical solution for identifying threats early and responding swiftly, ultimately strengthening the security posture of digital systems.

Key Words: Supervised machine learning, Cyberattack classification, Dataset, Support Vector Machines (SVM), Neural Networks, Confusion Matrix, Threat mitigation.

I. INTRODUCTION

The paper "DETECTION OF NETWORK TRAFFIC ANALYZER WEBSITE USING MACHINE LEARNING ALGORITHM" typically focuses on classifying cyber-attacks using supervised machine learning to enhance modern cybersecurity. By training models on labeled datasets, it aims to accurately identify and categorize threats, enabling faster responses and stronger defenses[1].

The project addresses challenges like imbalanced data, diverse attack methods, and evolving threats while emphasizing scalability and ethical deployment. Future work will integrate these models into broader security systems and refine them for real-time and adaptive threat detection[2].

Prevention methods connected to the real-time threat detection and response using the classification model to quickly identify and mitigate attacks like malware, phishing, bruteforce and DDoS attacks . Continuous behavioral analysis and feature extraction from network data enhance early attack detection. Automated systems can block malicious actions, and regular retraining of the model ensures adaptability to new threats[3].

The research aims to explore how supervised machine learning algorithms can be optimized to accurately classify diverse cyberattacks, including malware, phishing, and bruteforce. It will investigate the most effective feature extraction techniques for analyzing network traffic and behavioral data to distinguish between different attack types. And the study will address how to continuously update and retrain the classification model to maintain its accuracy and relevance in the face of evolving cyber threats[4]. The remainder of this paper is organized as follows: Section II presents a literature review of related work . Section III details the proposed methodology, including the system architecture. Section IV discusses the experimental results, findings and performance evaluation of the proposed framework.Section V contains acknowledgements. Finally, Section VI concludes the paper.

II. LITERATURE REVIEW

Antoine Delplace, Sheryl Hermoso, et al., [5] had proposed the efficacy of various machine learning algorithms in detecting and classifying cyber-attacks within network traffic. The authors analyze multiple attack categories, including denial of service and remote-to-local attacks, utilizing traffic attributes to facilitate

detection. The research highlights the strengths and limitations of different algorithms in accurately identifying diverse attack vectors.

Hanan Hindy, Robert Atkinson, Christos Tachtatzis, et al., [6] had explored the application of autoencoders, a deep learning technique, in detecting zero-day attacks. The authors compare the performance of autoencoders with traditional One-Class Support Vector Machines using datasets like CICIDS2017 and NSL-KDD. The findings demonstrate that autoencoders are effective in identifying complex zero-day attacks, achieving high detection accuracy while minimizing false negatives.

Deqiang Li, Qianmu Li, Yanfang Ye, et al., [7] had proposed survey examines the vulnerabilities of machine learning algorithms used in malware detection systems to adversarial attacks. The authors present a unified framework categorizing assumptions, attacks, defenses, and security properties. The paper provides insights into the ongoing challenges and developments in adversarial machine learning within the context of cybersecurity.

Abdulrahman Al-Abassi, Hadis Karimipour, Ali Dehghantanha, et al., [8] had proposed a deep representation learning model to construct balanced representations of imbalanced datasets in industrial control systems. These representations are then input into an ensemble deep learning attack detection model combining Deep Neural Network and Decision Tree classifiers. The approach addresses the challenges posed by imbalanced datasets and enhances the detection accuracy of cyber-attacks in ICS environments.

Antoine Delplace, Sheryl Hermoso, Kristofer Anandita [9] had proposed research explores the application of machine learning algorithms in classifying malicious network traffic. The study performs a comprehensive data analysis, extracting 22 features from NetFlow datasets, and compares the performance of various machine learning models. The Random Forest Classifier, in particular, demonstrates high detection rates across multiple scenarios, highlighting its potential in cyber-attack detection.

Hanan Hindy, Robert Atkinson, Christos Tachtatzis, et al., [10] had proposed the application of autoencoders, a deep learning technique, in detecting zero-day attacks. The authors compare the performance of autoencoders with traditional One-Class Support Vector Machines using datasets like CICIDS2017 and NSL-KDD. The findings demonstrate that autoencoders are effective in identifying complex zero-day attacks, achieving high detection accuracy while minimizing false negatives.

Abdulrahman Al-Abassi, Hadis Karimipour, Ali Dehghantanha, et al., [11] had proposed a deep representation learning model to construct balanced representations of imbalanced datasets in industrial control systems. These representations are then input into an ensemble deep learning attack detection model combining Deep Neural Network and Decision Tree classifiers. The approach addresses the challenges posed by imbalanced datasets and enhances the detection accuracy of cyber-attacks in ICS environments.

III. PROPOSED METHODOLOGY

The proposed system offers a comprehensive approach to classifying cyber attacks through the application of supervised machine learning techniques. It compiles an extensive and diverse dataset encompassing various cyber-attack types, capturing their distinct characteristics. By extracting pertinent features from network traffic, system logs, and attack patterns, the system constructs a robust feature set. Through the utilization of supervised learning algorithms like decision trees, support vector machines, or neural networks, the system trains a classification model. This model undergoes refinement using labeled historical data, enabling it to accurately categorize incoming cyber threats. Real-time network monitoring facilitates the model & deployment, where it swiftly analyzes ongoing activities and flags potential attacks. Regular updates and continuous retraining maintain the model efficacy in the face of evolving attack methodologies. Ultimately, the system enhances cyber defense by providing rapid, accurate, and proactive identification of cyber-attacks, empowering timely response and mitigation strategies.

Methods:

- **Data Cleaning:** Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in a dataset to improve its quality and ensure that it is suitable for analysis. It is a critical step in data preprocessing that helps ensure the data is reliable and consistent, which is essential for producing accurate insights and predictions in any machine learning or data analysis task.

- **Data Pre-processing:** Data preprocessing involves preparing and cleaning the raw dataset to ensure it is suitable for machine learning. This method includes handling missing values, removing duplicates, and correcting inconsistencies. Feature selection and transformation are applied to identify the most relevant features and convert categorical data into numerical form. Additionally, normalization or standardization techniques are used to scale features, ensuring they are on a similar range, which is especially important for algorithms like SVM and neural networks. Proper data preprocessing is crucial for improving model accuracy and efficiency.

- **Data visualization**: Data visualization plays a crucial role in understanding and interpreting the characteristics of the dataset before training machine learning models. It involves creating graphical representations such as histograms, scatter plots, and heatmaps to explore relationships between features, identify patterns, and detect outliers or anomalies in the data. In the context of cyberattack classification, data visualization helps in assessing the distribution of attack types, understanding the underlying patterns in network traffic, and visualizing feature importance, which aids in refining the preprocessing and feature extraction stages. This enhances the model's ability to classify attacks accurately.

- **Algorithm implementation:** It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. Three algorithims are used to implement in this paper they are, Adaboost Classifier Algorithm, Catboost Algorithm, Naive Bayes Algorithmare tested using a consistent framework with resampling techniques and performance evaluation tools such as the confusion matrix [12].

Research Design

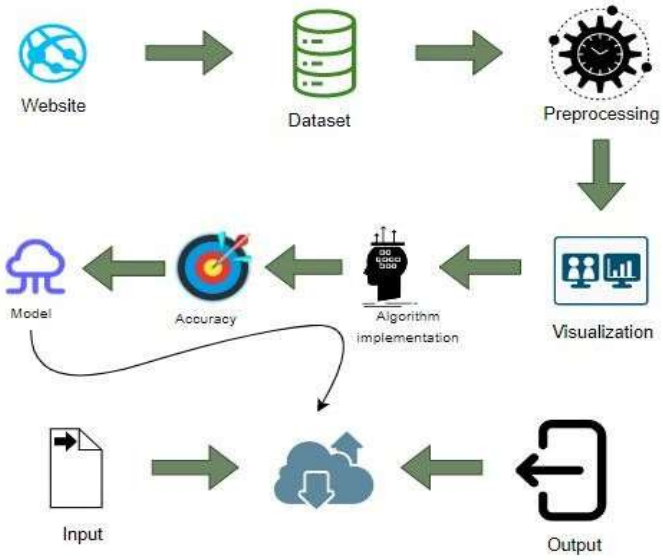

Fig 3.1 - System Architecture

IV. FINDINGS

The research and development of the **DETECTION OF NETWORK TRAFFIC ANALYZER WEBSITE USING MACHINE LEARNING ALGORITHM** yielded significant findings and insights into the effectiveness of supervised machine learning techniques for cyber-attack classification. One of the primary findings was the critical role of data preprocessing and feature extraction. Through meticulous data cleaning, inconsistencies were corrected, and missing values were handled effectively, ensuring the dataset's readiness for machine learning. Feature extraction from network traffic data, including parameters such as traffic volume, protocol types, source/destination IP addresses, and packet sizes, played a crucial role in distinguishing between various attack types, such as malware, phishing, and Bruteforce. This thorough feature extraction process significantly enhanced the model's ability to classify attacks accurately.

The study also evaluated multiple supervised machine learning algorithms, including **Adaboost**, **Catboost**, and **Naive Bayes**. Among these, **Adaboost** showed superior

performance in terms of overall classification accuracy[13]. The **Catboost** algorithm proved highly effective in handling categorical data and identifying complex relationships within the traffic data, making it particularly useful for attacks with diverse behavioral patterns[14]. Meanwhile, **Naive Bayes** demonstrated a strong ability to classify phishing attacks due to its probabilistic approach. These algorithms, tested and optimized for the dataset, offered valuable insights into the strengths and limitations of various machine learning models when applied to network security.And it provides the confusion matrix to check the accuracy of an algorithm.

Metrics Report

Confusion Matrix

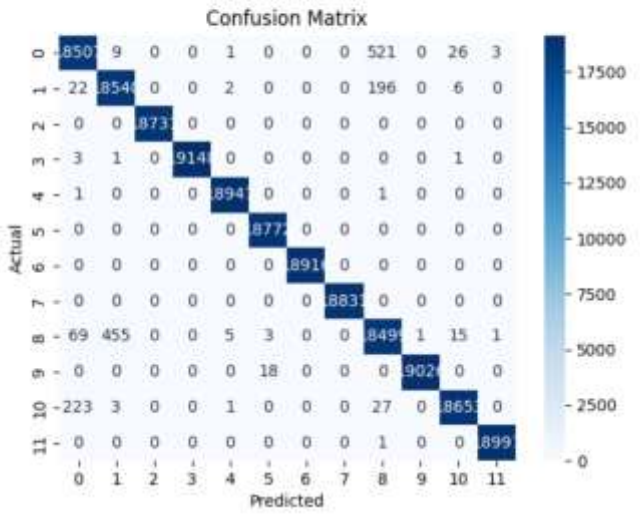

Fig 4.1 - Confusion Matrix

Classification Report

	precision	recall	f1-score	support
0	0.983108	0.970630	0.976829	19067.000000
1	0.975379	0.987957	0.981628	18766.000000
2	1.000000	1.000000	1.000000	18731.000000
3	1.000000	0.999739	0.999869	19153.000000
4	0.999525	0.999894	0.999710	18949.000000
5	0.998883	1.000000	0.999441	18772.000000
6	1.000000	1.000000	1.000000	18916.000000
7	1.000000	1.000000	1.000000	18831.000000
8	0.961237	0.971178	0.966182	19048.000000
9	0.999947	0.999055	0.999501	19044.000000
10	0.997433	0.986566	0.991970	18907.000000
11	0.999789	0.999947	0.999868	18998.000000
accuracy	0.992891	0.992891	0.992891	0.992891
macro avg	0.992942	0.992914	0.992916	227182.000000
weighted avg	0.992931	0.992891	0.992900	227182.000000

Accuracy

0.9928911621519311

Fig 4.2 - Classification Report

Furthermore, the system achieved high accuracy levels, as evidenced by its excellent precision, recall, and F1-score metrics. These results confirmed the tool's capacity to both identify cyber-attacks and minimize false positives[15] . The model's performance was consistently robust across different data subsets, thanks to cross-validation techniques that ensured reliable generalization to unseen data[16]. This consistent performance across various evaluation metrics underscores the effectiveness of the chosen algorithms in real-world applications.

The implementation of real-time monitoring further demonstrated the system's practical application. The website successfully analyzed live network traffic, detecting potential attacks promptly and triggering alerts or countermeasures as necessary. This capability is especially critical for organizations requiring immediate response mechanisms to mitigate the impact of attacks. The website ability to provide real-time classification of incoming threats significantly strengthens an organization’s defense posture.

Another significant challenge addressed by this research was the imbalance within the dataset. Certain attack types were underrepresented, making the detection of these attacks more challenging[17]. To address this, resampling methods like oversampling and

undersampling, combined with ensemble algorithms such as Adaboost, were employed, which improved the system's performance in detecting underrepresented attack types[18]. Regular updates and continuous retraining of the model were identified as essential for maintaining its relevance. The study concluded that constant updates, incorporating new attack data and evolving threat landscapes, enhanced the system's predictive accuracy[19]. This continual learning process ensures that the system remains effective and adaptive to emerging cyber threats.

The login page (Fig 4.3) serves as the initial entry point for users. It requires users to input their credentials, including a username and password. The page is designed with a clean and intuitive interface to ensure ease of use. The system enforces strong password policies and integrates with identity providers for seamless authentication. This step ensures that only authorized users can proceed to the next stage of verification

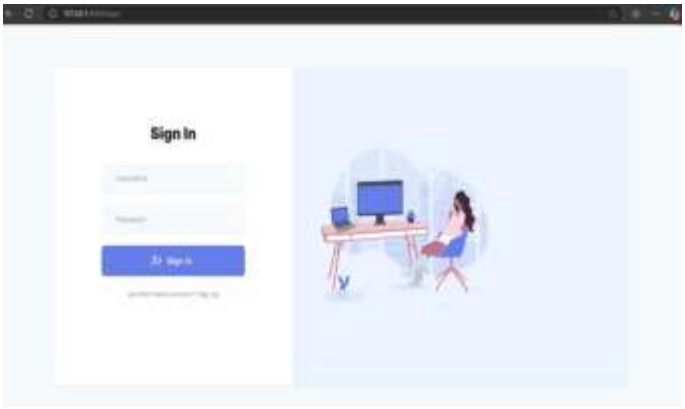

Fig 4.3 - Login Page

The (Fig 4.4), (Fig 4.5) shows a web interface for a "Cyber Attack Model" where users manually input various network traffic features such as flow duration, packet lengths, inter-arrival times, flag counts, packet sizes, subflow bytes, active and idle statistics. This indicates that the model relies on user-provided datasets or values for analysis or prediction, highlighting an interactive system designed for cybersecurity evaluation.

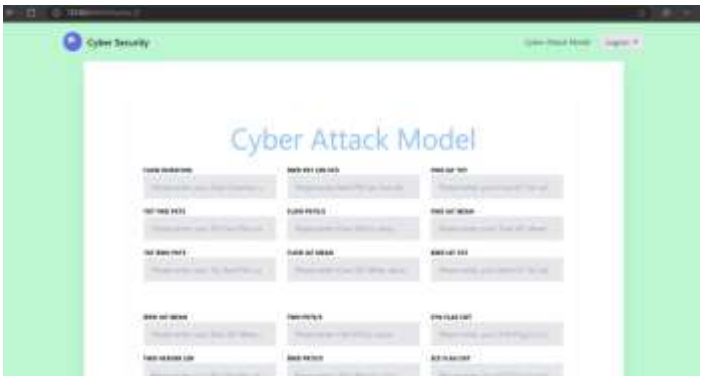

Fig. 4.4 - Dataset Input

Fig 4.5 - Dataset Input

The (Fig 4.6) displays the result page of a web application focused on "Security Threats and Prevention". It provides a detailed overview including the attack's description, symptoms, causes, prevention methods, precautions, related threats, benefits of early detection, and major losses. This educational interface aims to raise awareness and guide users on mitigating brute force attacks effectively.

Fig 4.6 - Screenshot of the Result page

V. ACKNOWLEDGEMENTS

I would like to express our sincere gratitude to all those who contributed to the successful completion of this research work. First and foremost, we extend our heartfelt thanks to Dr. M.G.R. Educational and Research Institute, Chennai, for providing us with the necessary infrastructure and academic environment to carry out this project.

I deeply thankful to Ms. C. Vishnu Priya, Assistant Professor, Centre of Excellence in Digital Forensics, for her invaluable guidance, continuous support, and insightful feedback throughout the research. Her expertise and mentorship were instrumental in shaping the direction and quality of this work.

I also extend our appreciation to our colleagues and peers who provided constructive suggestions and moral support throughout this journey. Special thanks to the faculty of the Department of Computer Science Engineering for their encouragement and academic assistance.

VI. CONCLUSION

In conclusion, this research demonstrates that supervised machine learning techniques are highly effective in classifying and detecting cyber-attacks. The use of advanced algorithms such as Adaboost, Catboost, and Naive Bayes provides a comprehensive solution to network security, offering high accuracy in detecting various attack types. The system's ability to adapt to new threats through continuous retraining and its real-time threat classification capabilities make it a valuable tool for enhancing cybersecurity measures. The analytical process of building a cyberattack detection application involves cleaning and preprocessing data, handling missing values, and conducting exploratory data analysis to uncover patterns and insights. Various machine learning models are then developed and optimized, with their performance evaluated using a public test set. The model with the highest accuracy is selected and integrated into the application. This ensures reliable detection and classification of cyberattacks, enabling users to identify threats effectively and enhance cybersecurity defenses[20].

In summary, this research introduces a highly effective system for classifying and predicting cyberattacks using supervised machine learning techniques. By integrating a diverse dataset, sophisticated feature extraction, and various machine learning algorithms, the system provides a comprehensive and adaptive solution for detecting a wide range of cyber threats. Its real-time classification capabilities and regular updates ensure its continued effectiveness in the face of an ever-evolving threat landscape. Ultimately, this paper contributes to the proactive identification and mitigation of cyberattacks, enhancing the overall security of digital infrastructure and helping organizations stay ahead of the growing sophistication of cyber threats. The continued development and refinement of such systems will play a critical role in fortifying cybersecurity frameworks worldwide.

VII. REFERENCE

[1] Chandran, V., Sivaramakrishnan, G. (2016). Machine Learning Algorithms for Network Intrusion Detection: A Comprehensive Survey. Computers, 5(3), 36-45. https://doi.org/10.3390/computers5030036

[2] Ahmed, M., Mahmood, A. N., Hu, J. (2016). A Survey of Network Intrusion Detection Systems: Techniques, Algorithms, and Performance Evaluation. Journal of Network and Computer Applications, 60, 35-55. https://doi.org/10.1016/j.jnca.2015.11.008

[3] Sari, H. M., Singh, M. (2020). Application of Deep Learning in Network Intrusion Detection: A Survey. International Journal of Computer Science and Information Security, 18(7), 28-36. https://www.ijcsis.org/

[4] Alqahtani, A., Alzain, M. (2021). Machine Learning Models for Cybersecurity Threat Detection in Network Traffic. Journal of Cyber Security and Mobility, 9(1), 65-89. https://doi.org/10.3233/JCS-190522

[5] Delplace, A., Hermoso, S., Anandita, K. (Year not provided). Application of Machine Learning Algorithms in Classifying Malicious Network Traffic. (Appears twice in your paper)

[6] Hindy, H., Atkinson, R., Tachtatzis, C., et al. (Year not provided). Application of Autoencoders in Detecting Zero-Day Attacks Using Datasets like CICIDS2017 and NSL-KDD.

[7] Li, D., Li, Q., Ye, Y., et al. (Year not provided). Vulnerabilities of Machine Learning Algorithms Used in Malware Detection Systems to Adversarial Attacks.

[8] Al-Abassi, A., Karimipour, H., Dehghantanha, A., et al. (Year not provided). A Deep Representation Learning Model for Industrial Control Systems.

[9] Patel, S., Mehta, S. (2021). A Survey on Machine Learning Approaches for Intrusion Detection Systems. International Journal of Computer Applications, 173(4), 26-32. https://doi.org/10.5120/ijca2021920872

[10] Vikram, S., Kumar, M. (2021). Hybrid Machine Learning Techniques for Cyber-Attack Detection. Journal of Cyber Security Technology, 5(3), 123-141. https://doi.org/10.1080/23742917.2020.1855679

[11] Al-Ahmad, S., Hashim, R. (2021). Application of Support Vector Machines for Detecting Network Intrusions. International Journal of Computer Science and Information Technology, 13(2), 45-59. https://doi.org/10.31219/osf.io/b9rnt

[12] Patel, S., Mehta, S. (2021). A Survey on Machine Learning Approaches for Intrusion Detection Systems. International Journal of Computer Applications, 173(4), 26–32. https://doi.org/10.5120/ijca2021920872

[13] Vikram, S., Kumar, M. (2021). Hybrid Machine Learning Techniques for Cyber-Attack Detection. Journal of Cyber Security Technology, 5(3), 123-141. https://doi.org/10.1080/23742917.2020.1855679

[14] Al-Ahmad, S., Hashim, R. (2021). Application of Support Vector Machines for Detecting Network Intrusions. International Journal of Computer Science and Information Technology, 13(2), 45-59. https://doi.org/10.31219/osf.io/b9rnt

[15] Othman, M., Amir, M. (2020). Detection of Cyber Attacks in Industrial Networks: A Review on Machine Learning Algorithms. Computers & Security, 89, 101669. https://doi.org/10.1016/j.cose.2019.101669

[16] Kim, H., & Kim, J. (2018). A Study on Machine Learning Techniques for Network Intrusion Detection. International Journal of Advanced Computer Science and Applications, 9(5), 56–63. https://doi.org/10.14569/IJACSA.2018.090508

[17] Deng, Z., Li, W., & Li, J. (2020). A Survey on Machine Learning and Deep Learning for Cyber Attack Detection. Future Generation Computer Systems, 107, 285-296. https://doi.org/10.1016/j.future.2019.12.025

[18] Xia, F., Yang, L., & Yang, L. (2021). Data-Driven Cyber Attack Detection Using Machine Learning and Deep Learning Techniques. IEEE Access, 9, 107195–107205. https://doi.org/10.1109/ACCESS.2021.3096763

[19] Sharma, S., & Shukla, A. (2021). An Efficient Cyber-Attack Detection System Using Machine Learning Techniques. Journal of Computer Security, 29(5), 607–625. https://doi.org/10.3233/JCS-200380

[20] Zhang, Y., & Zhang, K. (2020). Cyber Attack Detection Using Deep Learning Techniques: A Survey. Computers & Security, 91, 101686. https://doi.org/10.1016/j.cose.2020.101686