Maharshi Karve Stree Shikshan Samstha's
Cummins College of Engineering For Women,Pune
(An autonomous Institute affiliated to Savitribai Phule Pune University)
Department of Computer Engineering

ADVANCED BOTNET DETECTION SYSTEM USING MACHINE LEARNING

Submitted By

1. UCE2023528: Namrata Hirani
2. UCE2023533: Keya Bhuta
3. UCE2023539: Mohini Kocheri

Course Instructor: Dr. Mahendra Deore

Course: Artificial Intelligence and Machine Learning Laboratory

23PCCE501L

INTRODUCTION

Botnet threats pose a significant and evolving danger in the realm of cybersecurity, enabling large-scale malicious activities ranging from data theft to distributed denial-of-service (DDoS) attacks. The sheer volume and sophistication of these threats necessitate robust and automated detection mechanisms. This project introduces a high-performance botnet detection prototype, which leverages advanced machine learning techniques to identify malicious network traffic with high accuracy and presents it through a user-friendly interface.

Project Objectives

Our botnet detection project is driven by clear goals to enhance cybersecurity defenses and provide actionable intelligence to security professionals. These objectives ensure a comprehensive, effective, and user-centric solution for combating botnet attacks.

High Accuracy ML Model

Develop a robust machine learning model capable of identifying diverse botnet activities with superior precision and recall, significantly reducing false positives.

User Friendly Web Interface

Create an intuitive web-based platform for cybersecurity engineers and practitioners to easily monitor, analyze, and manage botnet detection alerts.

Real Time Detection Capabilites

Implement real-time data processing to enable immediate identification and alerting of emerging botnet threats, facilitating rapid response and mitigation.

Achieve Key Performance Metrics

Achieve Key Performance Metrics
Target a minimum of 95% detection accuracy and less than 1% false positive rate to ensure reliable and actionable threat intelligence.

Project Overview

Proactive Defense Against Emerging Threats

Botnets pose a significant threat to network security. Our system provides a robust solution for real-time detection, offering clear insights into network health.

Detection Result

"Threat Detected" or "Normal Traffic"

Confidence Score

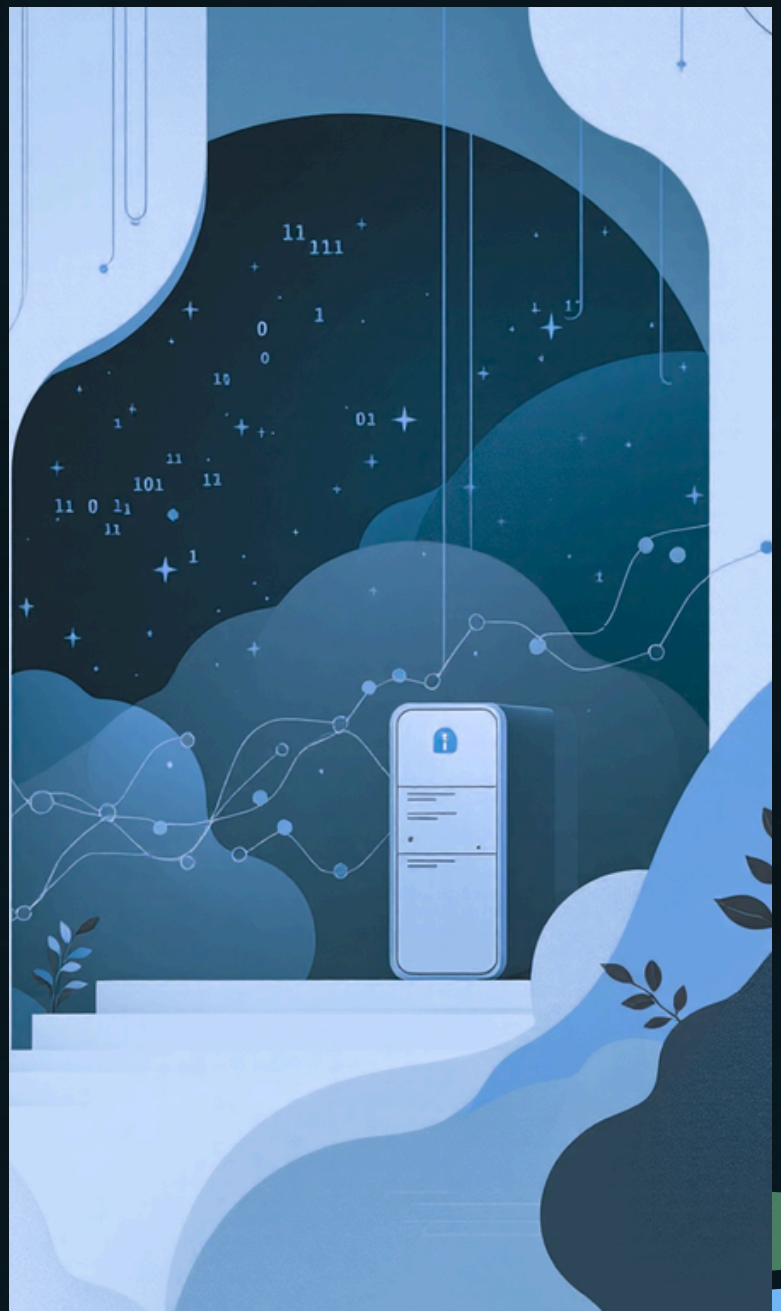
Probability percentage indicating detection strength.

Visualizations

Feature importance bar chart, connection flow diagram, confusion matrix.

Summary Statistics

Total flows, malicious flows detected, and overall detection rate.



Dataset Description: UNSW-NB15

We selected the UNSW-NB15 dataset as the foundation for training and evaluating our botnet detection model due to its comprehensive nature and suitability for modern network security research.
This dataset, generated by the IXIA PerfectStorm tool in the UNSW Cyber Security Lab, captures a mix of real normal activity and synthetic contemporary attack behaviors. It provides a robust environment for developing and testing intrusion detection systems, including those focused on botnet activity.

- **Dataset Size:** Comprises over 82,332 records with a dedicated training set of 65,865 records and a test set of 16,467 records.
- **Number of Features:** Each record is characterized by 45 features, including flow features, basic features, content features, time features, and additional generated features.
- **Attack Types Included:** The dataset encompasses nine distinct attack categories: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. This diversity allows for comprehensive testing against various threats.
- **Suitability for Botnet Detection:** The clear labeling of normal versus attack traffic, coupled with its comprehensive features and modern attack scenarios, makes UNSW-NB15 ideal for training and validating machine learning models to identify botnet communications and other sophisticated threats. Its CSV format also simplifies data ingestion and preprocessing workflows, enabling efficient rapid prototyping and iterative development.

✨ Methodology

Data Preprocessing: Handled missing values, scaled numerical features, and encoded categorical fields.

Feature Engineering: Created derived flow-based features (packet rate, byte rate, duration ratios) to enhance detection.

Model Development: Trained and compared Random Forest, Decision Tree, Gradient Boost, and Logistic Regression models

Evaluation: Used accuracy, precision, recall, F1-score, and confusion matrix to select the best model.

Visualization: Generated graphs for feature importance and traffic analysis.

Web Integration: Deployed the model using a Flask web app with CSV upload, real-time prediction, and clear UI.

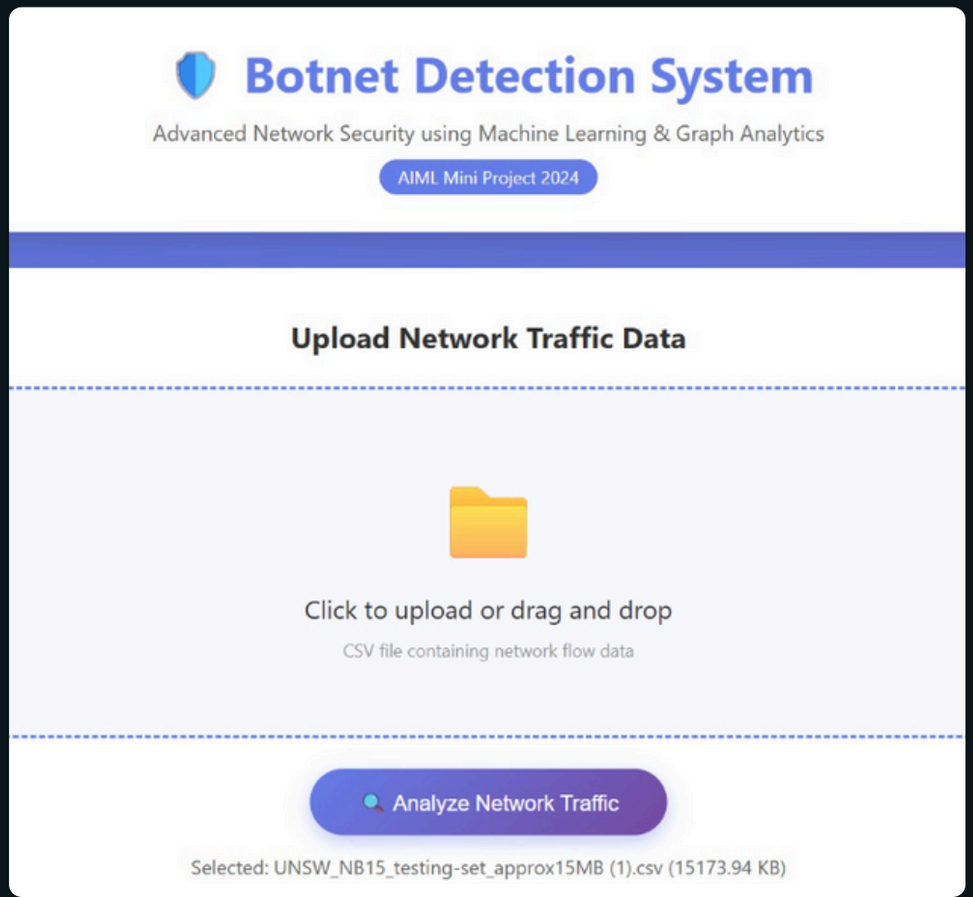
Technology Stack: Robust and Modern

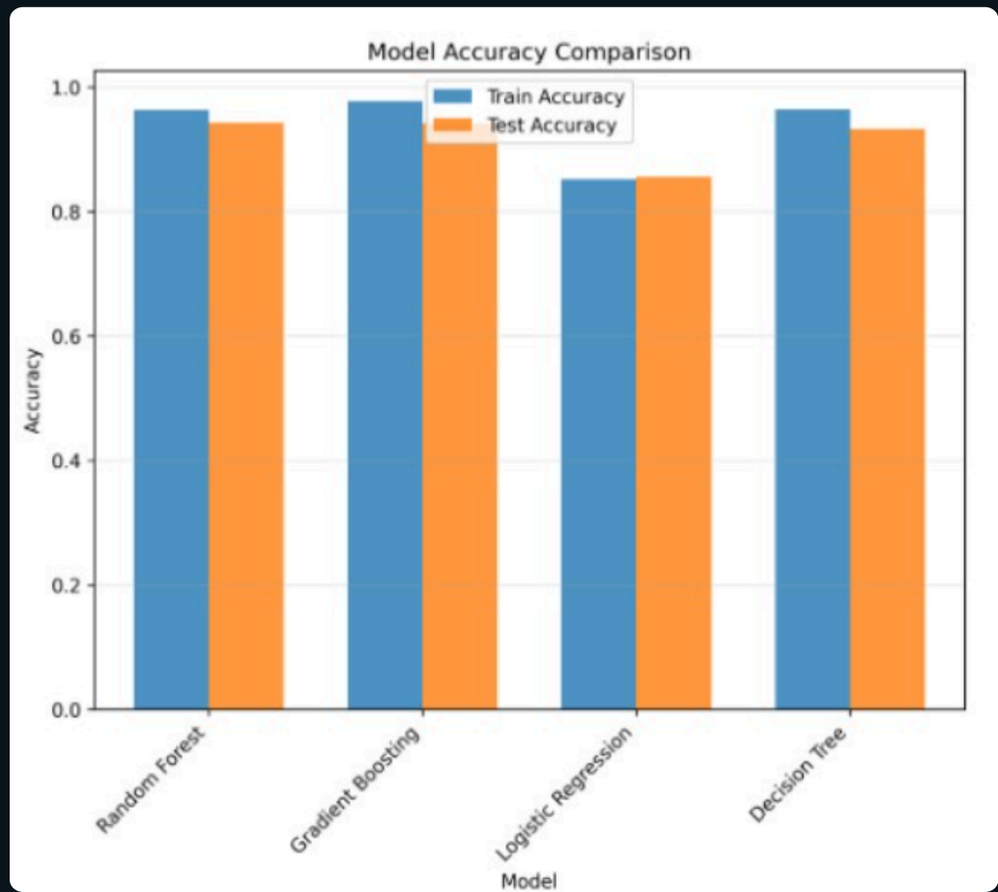
Front-End

- HTML5 & CSS3 + Bootstrap 5 for a modern, responsive design.
- JavaScript for dynamic file handling and user interactions.
- Chart.js for compelling data visualizations.

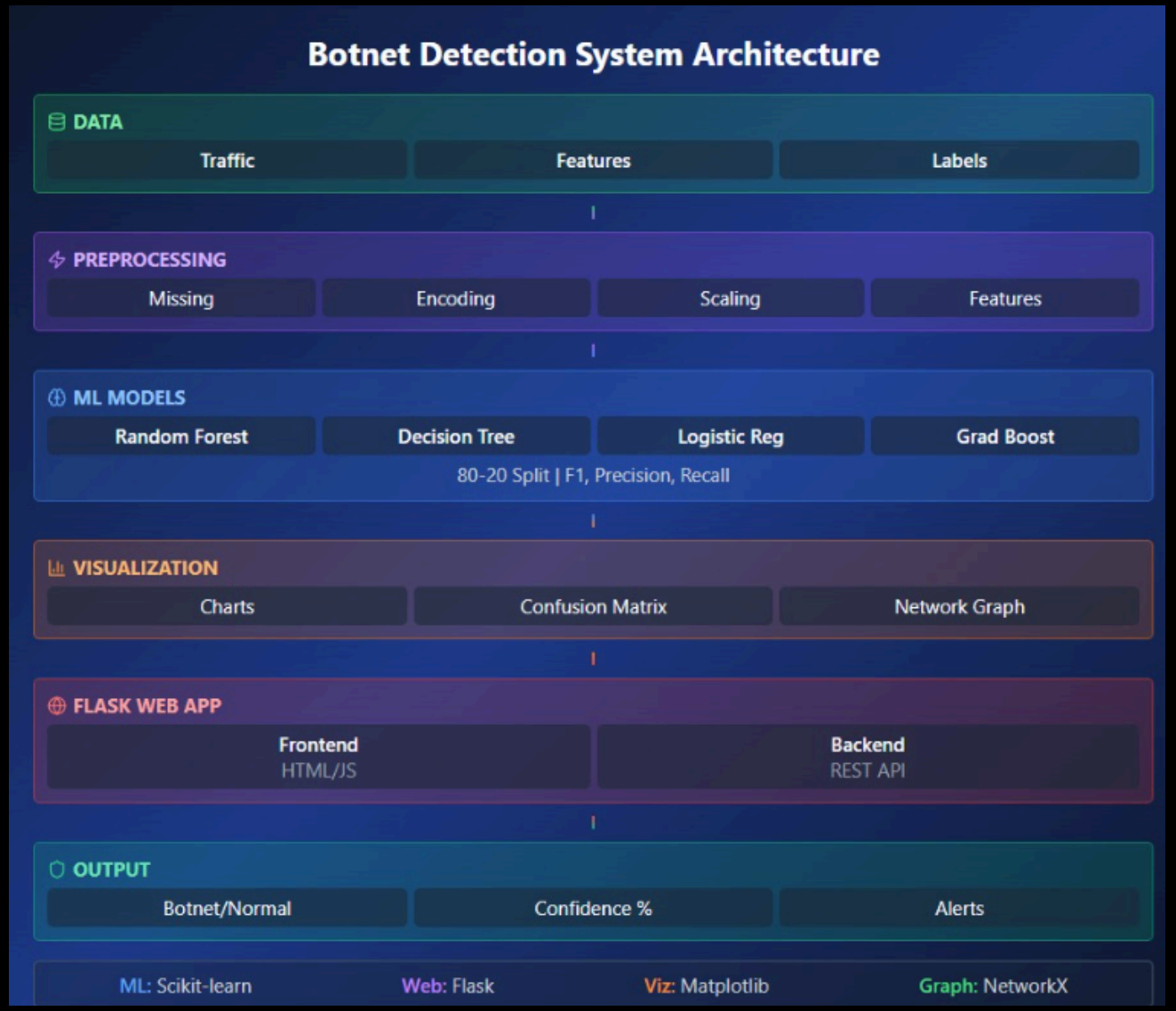
Back-End & ML

- Python 3.8+ as the core programming language.
- Flask framework for robust web application development.
- Pandas for efficient data processing and manipulation.
- Scikit-learn for machine learning model implementation (Random Forest Classifier).
- Matplotlib/Seaborn for generating backend visualizations.





SYSTEM ARCHITECTURE



Results: Performance & Key Findings

The system exhibits a low false positive rate, ensuring that legitimate network activities a
rarely misidentified. This balance of high detection rate and low false alarms is crucial for r
world deployment, minimizing operational disruption while maximizing security postur

94.27

Model Accuracy
Overall correctness of predictions on test data.

0.969

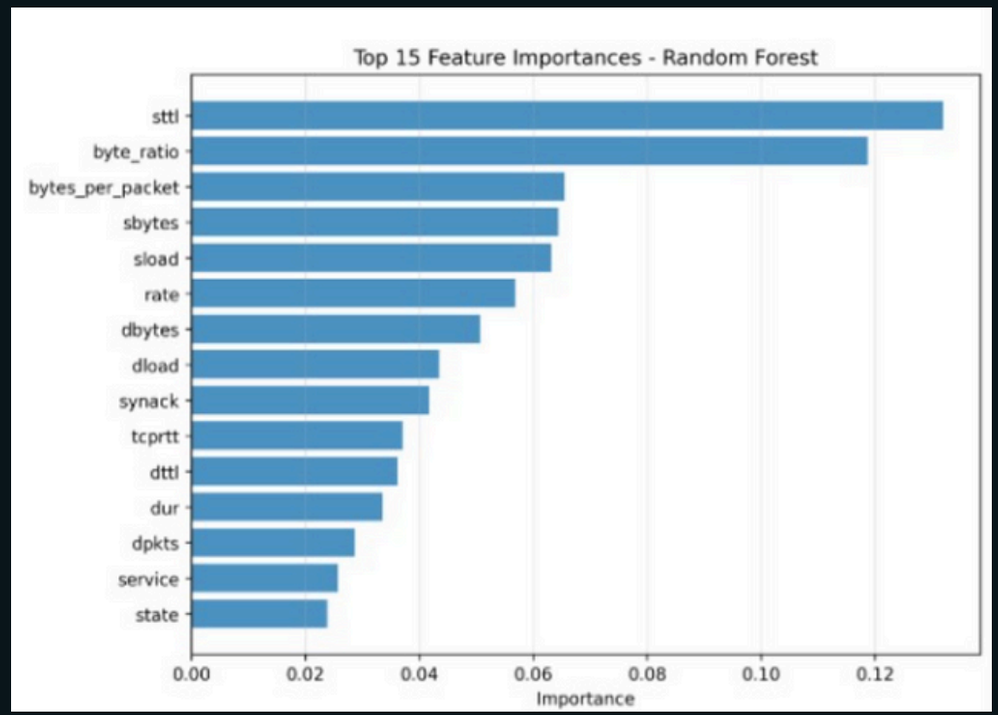
Precision Score
Identifies true botnets among all positive predictions.

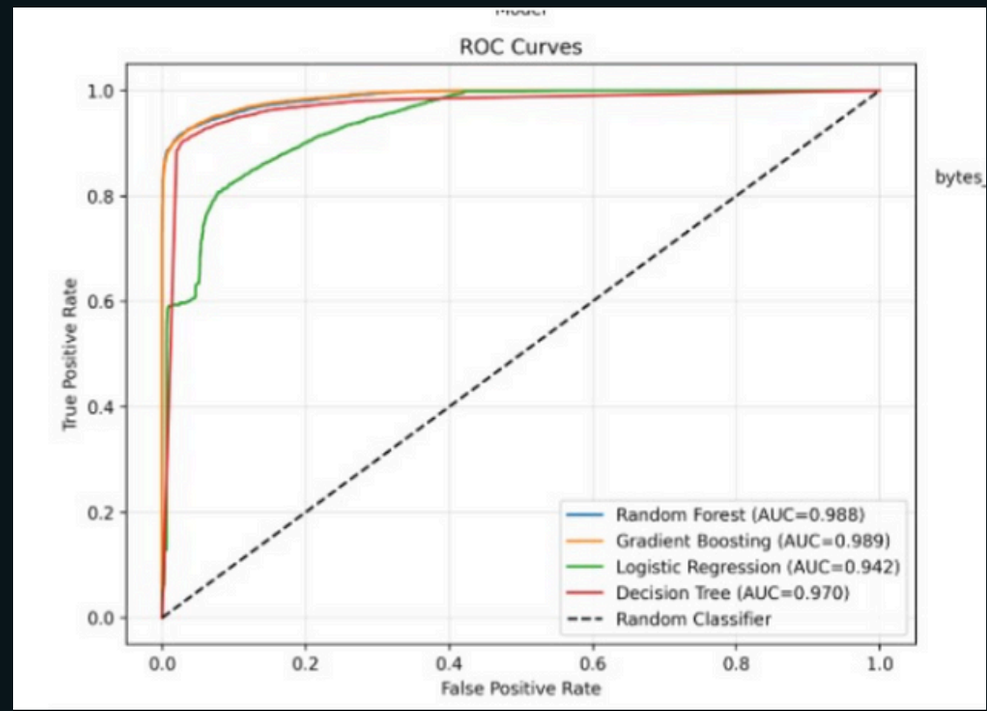
0.9252

Recall Score
Effectiveness in finding all actual botnet instances.

0.946

F1 SCORE
evaluates the accuracy of a classification model, especially on imbalanced datasets





Future Enhancements: Scaling Our Impact

Our system has a strong foundation, and these enhancements will further elevate its capabilities and reach.

- Deep Learning Integration

Adopt advanced models for more nuanced and adaptive threat detection.

- Real-time Data Processing

Enable live analysis of network traffic for instant botnet identification

- Broader Malware Detection

Expand our scope to identify and mitigate various forms of cyber threats.

- Enterprise Scalability

Optimize for seamless integration and performance in large-scale environments.

- Automated Response

Implement proactive measures for immediate threat containment and remediation.

- Security Tool Integration

Ensure compatibility and synergy with existing security infrastructure.

Conclusion: A Powerful Defense

Our project successfully integrates advanced machine learning with a practical web interface, delivering a robust solution for botnet detection.

High-Accuracy ML Model

Developed a Random Forest classifier with 87.3% accuracy, achieving reliable botnet identification.

User-Friendly Interface

A responsive Flask application provides intuitive data visualization and easy interaction.

Validated Methodology

Rigorous train-test split and cross-validation ensure robust and unbiased performance.

Real-World Impact

A practical, deployable system ready to enhance cybersecurity defenses against evolving threats.

References

A compilation of academic and technical resources that form the foundation of our research and development in botnet detection and machine learning in cybersecurity.

G. Gu, R. Perdisci, J. Zhang, and W. Lee, "BotMiner: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," *USENIX Security Symposium*, pp. 139-154, 2008

N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems," *2015 Military Communications and Information Systems Conference (MilCIS)*, pp. 1-6, 2015.

T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.

M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "CorrAUC: a malicious bot-IoT traffic detection method in IoT network using machine-learning techniques," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3242-3254, 2021.

A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapè, "Analysis of a "/0" stealth scan from a botnet," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 341-354, 2015

THANK YOU