ENSEMBLE MACHINE LEARNING APPROACH FOR IDENTIFYING THREATS IN SECURITY OPERATIONS CENTER

FEMI-OYEWOLE, FAVOUR OLASUNBO
(17PCH01659)
B.Sc Computer Science, Olabisi Onabanjo University, Ago-Iwoye
M.Sc Computer Science, University of Nigeria, Nsukka,
M.Sc Information Technology, University of Liverpool, United Kingdom

AUGUST, 2024

ENSEMBLE MACHINE LEARNING APPROACH FOR IDENTIFYING THREATS IN SECURITY OPERATIONS CENTER

BY

FEMI-OYEWOLE, FAVOUR OLASUNBO
(17PCH01659)
B.Sc Computer Science, Olabisi Onabanjo University, Ago-Iwoye
M.Sc Computer Science, University of Nigeria, Nsukka,
M.Sc Information Technology, University of Liverpool, United Kingdom

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY (Ph.D) DEGREE IN COMPUTER SCIENCE, DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA

AUGUST, 2024

ACCEPTANCE

This is to attest that this thesis is accepted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy in Computer Science in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I, FEMI-OYEWOLE, FAVOUR OLASUNBO (17PCH01659), hereby declare that this research was carried out by me under the supervision of Prof. Victor C. Osamor of the Department of Computer and Information Sciences, Covenant University, Ota and Prof. Okunbor Daniel of the Department of Computer and Information Sciences, Covenant University, Ota. I attest that the thesis has not been presented either wholly or partly for the award of any degree elsewhere. All sources of data and scholarly information used in this thesis are duly acknowledged.

FEMI-OYEWOLE, FAVOUR OLASUNBO

Signature and Date

CERTIFICATION

This is to certify that the research work titled "**ENSEMBLE MACHINE LEARNING APPROACH FOR IDENTIFYING THREATS IN SECURITY OPERATIONS CENTER**" is an original research work carried out by **FEMI-OYEWOLE, FAVOUR OLASUNBO (17PCH01659),** in the Department of Computer and Information Sciences, Covenant University, Ota, Ogun State, Nigeria, under the supervision of Prof. Victor C. Osamor and Prof. Okunbor Daniel. We have examined and found the work acceptable for its contribution to knowledge and literary presentation.

Prof. Victor C. Osamor **Signature and Date**
(Supervisor)

Prof. Okunbor Daniel **Signature and Date**
(Co-Supervisor)

Prof. Olufunke Oladipupo **Signature and Date**
(Head of Department)

Prof. Olusegun Folorunso
(External Examiner) **Signature and Date**

Prof. Akan B. Williams **Signature and Date**
(Dean, School of Post-Graduate Studies)

DEDICATION

This work is dedicated, firstly to Almighty God, the one in whom I move, live and have my being. Then, I dedicate this thesis to my family, whose unwavering support and encouragement have been instrumental in my academic journey. Their love and belief in my abilities have inspired me to pursue excellence in the field of Cyber Security. I am deeply grateful for their constant presence and sacrifices, which have made this achievement possible.

ACKNOWLEDGMENTS

I am humbly grateful to my heavenly Father, God, whose blessings, grace, and guidance have been the foundation of my academic journey. I acknowledge His unwavering presence, divine wisdom, and unconditional love, which have sustained me throughout the process. With deep gratitude, I recognize that every accomplishment and opportunity is a result of His divine favour. I offer my heartfelt thanks to my God for His constant support and blessings on this path of knowledge and growth.

I wish to express my deep sense of gratitude and thanks to the Chancellor and Chairman, Board of Regents, Covenant University, Dr. David O. Oyedepo, for the academic and spiritual platform created. I sincerely thank the Vice-Chancellor, Prof. Abiodun H. Adebayo, the Deputy Vice-Chancellor, Prof. Olujide A. Adekeye, and the management team of Covenant University for running the vision. My special appreciation goes to the Dean, School of Postgraduate Studies, Prof. Akan B. Williams, for his dedication to creating an enabling research environment and producing world-class researchers. I also applaud the Sub-Dean, School of Postgraduate Studies, Dr. Hezekiah Falola, for his steadfast counsel throughout my programme.

I express my heartfelt gratitude to the following individuals who have played significant roles in the completion of this project. I am sincerely thankful to Prof. Olufunke Oladipupo, the Head of the Department, and Dr. Itunuoluwa Isewon (The Postgraduate Coordinator), for their guidance and support throughout my research journey. My deepest appreciation goes to my esteemed supervisors, Prof. Victor Osamor and Prof. Daniel Okunbor, whose expertise, valuable insights, and continuous encouragement have been invaluable. Also, I acknowledge the members of the Intelligent System Cluster, for their contributions and fruitful discussions. Lastly, I extend my gratitude to all the senior faculty members in the Department of Computer and Information Sciences for their knowledge-sharing and assistance. Their collective efforts have greatly enriched my academic experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
APTs	Advanced Persistent Threats
API	Application Programming Interface
AUC	Area Under the Curve
AI	Artificial Intelligence
ANN	Artificial Neural Network
BPTT	Backpropagation through Time
BYOD	Bring Your Own Device
CPAs	Certified Public Accountants
CIO	Chief Information Officer
CISO	Chief Information Security Officer
CASB	Cloud Access Security Brokers
CSPM	Cloud Security Posture Management
CWPP	Cloud Workload Protection Platform
CSIRT	Computer Security Incident Response Team
CIA	Confidentiality, Integrity and Availability
CDN	Content Delivery Network
CNNs	Convolutional Neural Networks
CSRF	Cross Site Request Forgery
XSS	Cross Site Scripting
CDC	Cyber Defence Centers
CFC	Cyber Fusion Centers
CSIRT	Cyber Security Incident Response Teams
CSOC	Cyber Security Operation Centers
CICIDS2017	Cybersecurity Intrusion Detection System 2017
CRMRP	Cybersecurity Risk Management Reporting Framework
DLP	Data Leakage Monitoring
DNNs	Deep Neural Networks
DoS	Denial of Service
IDS	Intrusion Detection Systems
DDoS	Distributed Denial of Service
DGAs	Domain Generation Algorithms

DNS	Domain Name System
DHCP	Dynamic Host Configuration Protocol
EDR	Endpoint Detection and Response
EM	Expectation-Maximization
XGBoost	Extreme Gradient Boosting
FN	False Negatives
FP	False Positives
GDPR	General Data Protection Regulation
HMM	Hidden Markov Models
IOCs	Indicators of Compromise
ICMP	Internet Control Message Protocol
IQR	Interquartile Range
IPS	Intrusion Prevention Systems
JOC	Joint Operations Centers
k-NN	k-Nearest Neighbours
KPIs	Key Performance Indicators
LLMs	Large Language Models
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MSSP	Managed Security Service Provider
MDI	Mean Decrease Impurity
MTTD	Mean Time to Detect
MTTR	Mean Time to Recover
NLP	Natural Language Processing
OSINT	Open-Source Intelligence
PCI DSS	Payment Card Industry Data Security Standard
PMCs	Protective Management Controls
QP	Quadratic Programming
RNNs	Recurrent Neural Networks
Rsh	Remote Shell
SEL	Security Event Logs
SEM	Security Event Management
SIEM	Security Information and Event Management

SIM	Security Information Management
SMBs	Small and Medium-Sized Businesses
SMOTE	Synthetic Minority Over-sampling Technique
SOC	System and Organization Controls
SOCs	Security Operations Centers
SOAR	Security Orchestration, Automation and Response
SFM	Select from Model
SQL	Structured Query Language
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TIPs	Threat Intelligence Platforms
TN	True Negatives
TP	True Positives
UEBA	User Entity Behaviour Analytics
WAFs	Web Application Firewalls

ABSTRACT

Cyberattacks can be prevented by identifying threats before they cause damage, requiring robust cybersecurity measures. However, recent years have seen an increase in cyber threats and data breaches, often exploiting infrastructure weaknesses. These attacks lead to significant financial losses and compromised personal information, necessitating proactive defence strategies. Traditionally, detecting threats involves laborious log analysis, but machine learning can automate this process in intrusion detection systems (IDS). This study aims to implement a blended ensemble approach for cyberattack detection in security operation centers, combining predictions from base classifiers like Random Forest, XGBoost, HMM, and LSTM, Feature selection was performed by aggregating importance scores from these classifiers, with selected features used to improve the model's performance. A web application interface was developed using the Python Flask framework. The integration of trained models into the application programming interface (API) facilitated model training and dependency management. The testing and evaluation were performed on both real production network traffic flows and the testing set of the CICIDS2017 Thursday-WorkingHours-Morning.pcap_ISCX.csv dataset, as well as the generated real-time network traffic dataset. Real web attacks were intentionally executed on the server where the API/Intrusion Detection System was implemented, and these unlabelled attack network flows were accurately labelled by the IDS. To implement the ensemble model, the "Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv" was extracted from the renowned CICIDS2017 Thursday Morning Hours Dataset was utilized to train the model. To enhance the diversity of network traffic patterns and potential security incidents, real-time network traffic was generated using Sqlite, Zenmap Nmap, ID2T, and Python. The generated real-time network traffic was also used to train the model to detect unseen attacks. The proposed model performed well on the balanced Thursday Morning Dataset. With precision, recall, and F1-score all at 0.99, the model achieved an overall accuracy of 99% across the binary classification task, highlighting its robustness and effectiveness in handling real-time malicious traffic. These findings validate the model's ability to detect real-time network traffic patterns, particularly in the context of potential security incidents. The proposed model demonstrated high performance on the generated dataset, achieving a precision of 1.00 for detecting malicious threats, thereby correctly identifying all instances without false positives. The recall of 1.00 further underscored its capability to detect all actual instances of malicious activity. An F1-score of 1.00 for legitimate traffic reflected the model's balanced precision and recall, ensuring reliable classification across categories. Additionally, the cross-validation results exhibited consistently high accuracy, with an average accuracy of approximately 0.999 across five folds. This outcome confirms the model's robustness and generalizability across various data subsets, highlighting its potential for reliable real-time threat detection and enhanced cybersecurity in practical applications.

Keywords: Cyberattack, Cybersecurity, Ensemble Model, Machine Learning