

**NETWORK INTRUSION DETECTION MODEL USING
ENSEMBLE-BASED LEARNING**

**ONIETAN, IYANU-OLUWA CHRISTOPHER
(20PCG02292)**

BSc. Computer Science, Samuel Adegboyega University, Ogwa

AUGUST, 2023

**NETWORK INTRUSION DETECTION MODEL USING
ENSEMBLE-BASED LEARNING**

BY

**ONIETAN, IYANU-OLUWA CHRISTOPHER
(20PCG02292)**

BSc. Computer Science, Samuel Adegboyega University, Ogwa

**A DISSERTATION PRESENTED TO THE SCHOOL OF
POSTGRADUATE STUDIES AS A PARTIAL FULFILLMENT OF
THE PREREQUISITES FOR OBTAINING A MASTER OF SCIENCE
(M.SC) DEGREE IN THE DEPARTMENT OF COMPUTER SCIENCE
WITHIN THE COLLEGE OF SCIENCE AND TECHNOLOGY AT
COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA**

AUGUST, 2023.

ACCEPTANCE

This is to confirm that this dissertation is deemed worthy for the purpose of meeting partial requirements towards the attainment of the Master of Sciences in Computer Science degree from the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, 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, **ONIETAN IYANU-OLUWA CHRISTOPHER** (20PCG02292), assert that I conducted this project at Covenant University in Ota, Ogun State, Nigeria, under the guidance of Dr. Jonathan Oluranti from the Department of Computer and Information Sciences in the College of Science and Technology. I confirm that this research has not been previously submitted, either wholly or in part, for any other academic degree. This dissertation appropriately acknowledges all sources of data and scholarly information.

ONIETAN, IYANU-OLUWA CHRISTOPHER

Signature and Date

CERTIFICATION

We certify that this dissertation titled “**NETWORK INTRUSION DETECTION MODEL USING ENSEMBLE-BASED LEARNING**” is an original research carried out by **ONIETAN, IYANU-OLUWA CHRISTOPHER (20PCG02292)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. Oluranti Jonathan. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Computer Science.

Dr. Oluranti Jonathan
(Supervisor)

Signature and Date

Prof. Olufunke O. Oladipupo
(Head of Department)

Signature and Date

Prof. Adio T. Akinwale
(External Examiner)

Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

Signature and Date

DEDICATION

With deep gratitude, I dedicate this project to You. Your guidance has been my constant light. May this work reflect my faith and gratitude. Thank you for your presence in my life. To my family, friends, mentors, and the countless individuals who have contributed to my journey, I dedicate this project. Your unwavering support and guidance have been invaluable, and this work is a reflection of our collective efforts. Thank you for being a constant source of inspiration and encouragement.

ACKNOWLEDGEMENTS

I want to express my gratitude to everyone who helped me out and gave me advice when I was working on this thesis. All the support, advice, and ideas they offered were priceless.

I would want to start by expressing my deepest gratitude to Dr. Oluranti Jonathan, my supervisor, for all of the help, advice, and encouragement they have given me. His perceptive criticism and guidance were crucial in determining the course of this study.

I appreciate my family's undying support and affection so much. Their tolerance, faith, and confidence in me encouraged me to push past obstacles and complete the task at hand.

I would like to thank everyone in my community who has helped me grow via shared experiences, conversations, and friendship. Their unique insights deepened my comprehension of the material and made the trip more fun.

I would like to express my gratitude to everyone who helps make the Computer and Information Sciences department such a great place to study and conduct research. All of the offered tools, space, and chances have been crucial to the accomplishment of this mission.

TABLE OF CONTENTS

CONTENTS	PAGES
COVER PAGE	i
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xii
ABSTRACT	xiii
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Study	1
1.2 Statement of the Problem	3
1.3 Aim and Objectives of Study	4
1.4 Significance of Study	5
1.5 Project Organization	5
CHAPTER TWO: LITERATURE REVIEW	6
2.1 Preamble	6
2.2 Fundamental Concept	6
2.3 Detection Techniques for Intrusion Detection System	8
2.3.1 Signature-Based Detection	8
2.3.2 Anomaly-Based Detection	9
2.3.3 Stateful Protocol Analysis	13
2.4 Challenges of Anomaly Detection	14
2.5 Machine Learning	15
2.5.1 Supervised Machine Learning Techniques	17
2.5.2 Unsupervised Machine Learning	18
2.6 Related Works	20

CHAPTER THREE: RESEARCH METHODOLOGY	23
3.1 Preamble	23
3.2 Proposed Model	25
3.2.1 Data Collection	26
3.2.2 Data Preprocessing	31
3.2.3 Feature Engineering	36
3.2.4 Feature Selection	37
3.2.5 Explorative Data Analysis	38
3.2.6 Data Sampling	39
3.2.7 Handling Imbalanced Dataset	39
3.2.8 Base Classifiers	41
3.2.9 Meta Classifier	44
3.2.10 Hyperparameter Tuning	47
3.2.11 Dataset Description	48
3.3 Evaluation Metrics	53
CHAPTER FOUR: RESULTS AND DISCUSSION	56
4.1 Result of the Proposed Model	56
4.2 Approach 1: ID2T Dataset	56
4.2.1 Binary Class Classification Report of Base-Classifiers Using the SMUTE Balancing Technique	56
4.2.2 Binary Class Classification Report of Majority Voting Ensemble Classifier Using the SMUTE Balancing Technique	57
4.2.3 Binary Class Classification Report of the Stacking Model Using the SMUTE Balancing Technique	57
4.2.4 Binary Class Classification Report of the Bagging Model Using the SMUTE Balancing Technique	59
4.3 Approach 2: ID2T Dataset	60
4.3.1 Binary Class Classification Report of Base-Classifiers using the SMOTE Balancing Technique.	60
4.3.2 Binary Class Classification Report of the Majority Voting Model Using the SMOTE Balancing Technique	62
4.3.3 Binary Class Classification Report of the Stacking Model using the SMOTE Balancing Technique.	64
4.3.4 Binary Class Classification Report of the Bagging Model using the SMOTE Balancing Technique	66

4.4.	Approach 3: CIC-IDS2017	68
4.4.1	Multi Class Classification Report of the Base Classifier Models Using the SMUTE Balancing Technique	68
4.4.2	Multi Class Classification Report of the Majority Voting models Using the SMUTE Balancing Technique	71
4.4.3	Multi Class Classification Report of the Stacking model using the SMUTE Balancing Technique.	72
4.4.4	Multi Class Classification Report of the Bagging model using the SMUTE Balancing Technique	73
CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS		76
5.1	Summary	76
5.2	Contribution to Knowledge	76
5.3	Conclusion	76
5.3	Recommendations	77
REFERENCES		78

LIST OF TABLES

TABLES	TITLES OF TABLES	PAGES
1.1	Summary of Objectives	4
3.1	30 Features Were Outputted	26
3.2	Datasets Features Description	50
3.3	Classes of Attacks for the ID2T Dataset	52
3.4	Classes of Attacks for the CICIDS2017 Dataset	52
3.5	Datasets Files and Labels	53
4.1	Binary Class Classification Report of Base-Classifiers Using the SMUTE Balancing Technique	57
4.2	Binary Class Classification Report of the Majority Voting Model Using the SMUTE Balancing Technique	58
4.3	Binary Class Classification Report of the Stacking Model Using the SMUTE Balancing Technique	59
4.4	Binary Class Classification Report of the Bagging Model Using the SMUTE Balancing Technique	59
4.5	Binary Class Classification Report of Base-Classifiers Using the SMOTE Balancing Technique.	61
4.6	Binary Class Classification Report of the Majority Voting Model Using the SMOTE Balancing Technique	63
4.7	Binary Class Classification Report of the Stacking Model Using the SMOTE Balancing Technique	65
4.8	Binary Class Classification Report of the Bagging Model Using the SMOTE Balancing Technique	67
4.9	CIC-IDS2017 Encoded Class Labels for Multi Class Classification	69
4.10	Multi Class Classification Report of the Base Classifier Model Using the SMUTE Balancing Technique	69
4.11	Multi Class Classification Report of the Majority Voting Models Using the SMUTE Balancing Technique	71

4.12	Classification Report of the Stacking Classifier on the Undersampled CICIDS2017 Dataset	73
4.13	Classification Report of the Bagging Classifier on the Undersampled CICIDS2017 Dataset	74

LIST OF FIGURES

FIGURES	TITLES OF FIGURES	PAGES
2.1	Data Distribution to Detect Potential Outliers	12
2.2	Log Transformed Data to Identify Potential Outliers	13
2.3	Classification of Particular Difficulties Posed by Anomaly Detection in High-Dimensional Big Data	15
2.4	Machine Learning Techniques	16
3.1	Proposed Majority Voting Intrusion Detection Model	25
3.2	Installing the ID2T Library and Dependencies	28
3.3	ID2T Setup	28
3.4	Figures Above Define Nping Connection Parameters by Saving all Connection Strings	29
3.5	Customization of Record Generation Process	29
3.6	Attack Types	30
3.7	Generate Dataset in CSV Format	30
3.8	Data Processing Phase	31
3.9	Checking for Missing Data	32
3.10	Checking for Duplicate Value	32
3.11	Data Integration	33
3.12	Data Transformation	34
3.13	Feature Scaling	35
3.14a	Assignment of Countries Without Country Codes to a Related Country	36
3.14b	Getting Continent Names from Generated Country Codes	37
3.15a	Code to Show a Class Distribution of Benign and Malicious Attacks	38

3.15b	Show Attack Type Distribution	39
3.16	Undersampling Dataset	40
3.17	Oversampling Dataset	41
4.15a	Logistic Regression with Undersampling	60
4.1	Confusion Matrix of the Majority Voting Model	63
4.2	Confusion Matrix of the Stacking Model	64
4.3	Confusion Matrix of the Bagging Model	66
4.4	Performance of the Proposed Ensemble Models	67
4.5	Performance of Comparison of Ensemble Models Multiclass Classification	73

LIST OF ABBREVIATIONS

AIDS	Anomaly-based IDS
ANFIS	Adaptive Network-based Fuzzy Inference System
AUC	Area Under the Curve
CIA	Confidentiality, Integrity, and Availability
CIC-IDS2017	Canadian Institute for Cybersecurity Intrusion Detection Systems 2017
CSV	Comma-Separated Value
DARPA	Defense Advanced Research Projects Agency
DDoS	Distributed Denial-of-Service
DoS	Denial of Service
DT	Decision Tree
EDA	Exploratory Data Analysis
ET	Extra Tree
FP	False Positive
FN	False Negative
FTP	File Transfer Protocol
HIDS	Host-based Intrusion Detection Systems
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
ID2T	Intrusion Detection Dataset Toolkit
IDS	Intrusion Detection Systems
IETF RFC	Internet Engineering Task Force Request for Comments
IMAP	Internet Message Access Protocol
IoT	Internet of Things

IPS	Intrusion Prevention System
IPv6	Internet Protocol version 6
IRC	Internet Relay Chat
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
kNN	K-Nearest Neighbour
LSTM	Long Short-Term Memory
ML	Machine Learning
MANETs	Mobile Ad Hoc Networks
NGIPS	Next-Generation IPS
NMAP	Network Mapper
NIDS	Network Intrusion Detection System
NIST	National Institute of Standards and Technology
NSL-KDD	Network Security Laboratory Knowledge Discovery in Databases
PCAP	Packet Capture
POD	Ping of Death
POP3	Post Office Protocol version 3
PPP	Point-to-Point Protocol
QoS	Quality of Service
TN	True Negative
ToS	Type of Service
TP	True Positive
TCP	Transmission Control Protocol
TTL	Time to Live
SMOTE	Synthetic Minority Oversampling Technique

SMUTE	Synthetic Majority Undersampling Technique
SNMP	Simple Network Management Protocol
SSH	Secure Shell
UDP	User Datagram Protocol
SVM	Support Vector Machine
WASC	Web Application Security Consortium
WSN	Wireless Service Networks

ABSTRACT

The interconnectedness of devices, technologies, networks and the services they provide has continued to increase. This has also resulted in increased cases of cyber threats and intrusions. detection has become a major concern for organizations. Intrusion detection system is one way to address the issue of intrusions and anomaly network traffic. Existing machine learning algorithms has performed well on intrusion detection however, the issues of high false positive rates as well as low accuracy still persists. This is largely due to fact that individual models are not able to efficiently detect previously unknown intrusions on their own. Other the hand, ensemble models have proven to be more efficient in identifying intrusions and anomalies in networks since they combine the predictive powers of several base models. However, the efficiency of ensemble models has not been sufficiently considered where imbalanced datasets are involved. This study therefore proposes and investigates the performance of various ensemble models when applied to conspicuously imbalanced datasets. Two largely imbalanced datasets were acquired namely IDT2 and CICIDS2017. Additional datasets were generated from each of the acquired datasets using SMOTE and SMUTE, oversampling and under-sampling techniques respectively. In order to investigate the performance of ensemble models, three ensemble models were constructed namely Bagging, Majority voting and Stacking. The performance of each model was effectively determined and compared.

Keywords: Network; Intrusion Detection Systems, Intrusion Prevention Systems, Machine Learning; Cybersecurity; Ensemble Models.