

TABLE OF CONTENTS

APPROVAL PAGE	ii
Declaration	iii
PREFACE	iv
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	ix
Chapter I. INTRODUCTION	1
1.1. Background Problem	1
1.2. Problem Statement	3
1.3. Research Goal	4
1.4. Benefits	5
CHAPTER II. RELATED WORKS	6
2.1. Handling Imbalanced Dataset using Data Generation Method	6
2.2. Data Preprocessing in NIDS	7
2.3. Using Deep Belief Network	8
CHAPTER III. THEORETICAL FOUNDATION	9
3.1. Network Intrusion	9
3.2. Deep Belief Network	28
3.2.1. Restricted Boltzmann Machine	29
3.2.2. Deep belief neural network pre-training	31
3.2.3. Supervised fine-tuning based on BP algorithm	32
3.3. Synthetic Data Generation	32
3.3.1. Synthetic Minority Over-sampling Technique (SMOTE)	33
3.3.2. Generative Adversarial Networks (GANs)	35
3.3.3. Conditional Tabular Generative Adversarial Network (CTGAN)	38
3.4. Preprocessing	43
3.4.1. Maximum Minimum Normalization	43
3.4.2. Categorical Encoding	43
3.5. Model Evaluation	44
3.5.1. F1 Score	45
3.5.2. Area Under Curve (AUC)	46
CHAPTER IV. RESEARCH METHODOLOGY	51
4.1. Problem Description	51
4.2. Research Steps	52
4.2.1. Dataset Used	53
4.2.2. Preprocessing	58

4.2.2. Synthesize Data	60
4.2.3. Model Training	60
4.2.4. Model Evaluation using Area Under the Receiver Operating Characteristic Curve (AUC ROC)	62
CHAPTER V. IMPLEMENTATION	63
5.1. Main Program Flow	63
5.2. Data Acquisition	63
5.2. Preprocessing	64
5.3. Synthesize Data	65
5.4. Model Training	67
5.5. Model Evaluation	70
Chapter VI. RESULTS AND DISCUSSION	73
6.1. CTGAN Model	73
6.2. CTGAN-DBN Model	76
Chapter VII. CONCLUSION	82
7.0.1. Future Works	82
REFERENCES	83
Full Implementation	86
Models Performance	106

LIST OF TABLES

Table 4.1. UNSW-NB15 traffic distribution.....	54
Table 4.2. UNSW-NB15 dataset attributes list.....	54
Table 4.3. UNSW-NB15 traffic distribution after CTGAN.....	57
Table 4.4. Snippet of attribute proto.....	58
Table 4.5. Result of one-hot encoding.....	58
Table 4.6 Result of Min-Max Normalization.....	59
Table 4.6. Configuration of CTGAN (Bourou et al. ,2018).....	60
Table 4.7. DBN hyperparameter.....	61
Table 6.1. Synthesized data created by CTGAN.....	75
Table 6.2. Original data.....	76
Table 6.3. CTGAN-DBN model comparison.....	80

LIST OF FIGURES

Figure 3.1. Example of network intrusion data.....	9
Figure 3.2. Example of normal network.....	10
Figure 3.3. Difference in attack and normal attributes.....	28
Figure 3.4. RBM Structure according to Yaqi et al. (2018).....	29
Figure 3.5. RBM Structure (Hinton et al., 2006).....	31
Figure 3.6. Example of it works.....	34
Figure 3.7. Overlapping in SMOTE (Ahsan et al., 2023).....	35
Figure 3.8. GAN Algorithm (Goodfellow, 2016).....	36
Figure 3.9. Generative Adversarial Network Architecture and its Components (Rahul, 2023).....	36
Figure 3.10. Generator and Discriminator Network Structure (Zhang et al., 2020)....	37
Figure 3.11. An example of mode-specific normalization (Xu, 2019).....	39
Figure 3.12. Numerical and One-hot Encoding (Padarian, 2019).....	44
Figure 3.13. Example of Machine's result.....	47
Figure 3.14. Example of Machine's result.....	47
Figure 3.15. Graph of AUC ROC.....	49
Figure 3.16. Comparison of AUC between Two Models.....	49
Figure 4.1. Complete of CTGAN DBN architecture.....	52
Figure 5.1. Combining dataset.....	64
Figure 5.2. Cleaning and applying the column names to the combined dataset.....	64
Figure 5.3. Drop of unnecessary attributes.....	64
Figure 5.4. One-hot encoding and scaling of attributes.....	65
Figure 5.5. Metadata creation.....	66
Figure 5.6. Calculation of data creation.....	67
Figure 5.7. Synthesized data creation.....	67
Figure 5.8. Data split.....	68
Figure 5.9. DBN structure.....	69
Figure 5.10. DBN pre-training and fine-tuning.....	70
Figure 5.11. CTGAN Implementation.....	70
Figure 5.12. Validation and training loss.....	72
Figure 5.13. Testing.....	72
Figure 5.14. AUC PR.....	72
Figure 6.2. Data to be generated.....	74
Figure 6.3. CTGAN score.....	75
Figure 6.4. Result of data synthesizing.....	75
Figure 6.5. Data similarity score.....	75

Figure 6.7. AUC PR of CTGAN-DBN.....	78
Figure 6.8. AUC PR of regular DBN.....	79
Figure 6.10. Prediction of benign class.....	81
Figure 6.11. Prediction of intrusion class.....	81
Figure 1.0.1. Data Acquisition.....	86
Figure 1.0.2. Data Synthesization.....	88
Figure 1.0.3. Data Preprocessing and Splitting.....	93
Figure 1.0.4. Model Creation.....	96
Figure 1.0.5. Model Training and Testing.....	103
Figure 1.0.6. Model Measurement.....	105