

TABLE OF CONTENTS

PREFACE	ix
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	viii
LIST OF EQUATIONS	ix
ABSTRACT	x
CHAPTER I	
INTRODUCTION	1
1.1. Research Background.....	1
1.2. Research Problem.....	4
1.3. Research Objectives.....	4
1.4. Research Scope.....	4
1.5. Research Advantage.....	5
CHAPTER II	6
LITERATURE REVIEW	6
CHAPTER III	
BASIC THEORY	15
3.1. Credit Scoring.....	15
3.2. Telecommunications Data in Credit Scoring.....	15
3.3. Synthetic Minority Oversampling Method (SMOTE).....	16
3.4. Information Value (IV).....	17
3.5. Logistic Regression.....	18
3.6. Multi-Layer Perceptron.....	19
3.7. Grid Search and Cross-Validation.....	20
3.8. Shapley Additive Explanations (SHAP).....	21
3.9. Kolmogorov-Smirnov (KS).....	21
3.10. Area Under the Curve (AUC).....	22
3.11. Confusion Matrix.....	23
CHAPTER IV	25
4.1. Research Description.....	25
4.2. Research Procedure.....	26
4.2.1. Data Acquisition.....	28
4.2.2. Data Preprocessing and Feature Engineering.....	29
4.2.3. Feature Selection.....	29
4.2.4. Train Test Evaluation Split.....	30
4.2.5. Model Fitting.....	30
4.2.6. Model Evaluation.....	32

CHAPTER V.....	34
5.1. Tools and Materials.....	34
5.2. Data Acquisition.....	34
5.3. Data Preprocessing.....	35
5.4. Feature Engineering.....	37
5.5. Feature Selection.....	39
5.6. Train, Test, Evaluation Split.....	40
5.7. Model Development.....	41
5.7.1. Logistic Regression + GridSearchCV.....	41
5.7.2. SMOTE + Logistic Regression + GridSearchCV.....	43
5.7.3. Multilayer Perceptron.....	43
5.7.4. SMOTE + Multilayer Perceptron.....	45
5.8. SHAPLeY Values for Explainability.....	45
CHAPTER VI.....	47
RESULTS AND DISCUSSION.....	47
6.1. Class Imbalance.....	47
6.2. Result of Feature Selection.....	47
6.3. Model Summary.....	49
6.4. Model Evaluation.....	51
6.4.1. Confusion Matrix.....	52
6.4.2. Accuracy.....	53
6.4.3. AUC and K-S Score.....	53
6.4.4. SHAPLeY Additive Explanation Violin Plot.....	55
CHAPTER VII.....	58
CONCLUSION.....	58
7.1. Conclusion.....	58
7.2. Future Works.....	59
REFERENCES.....	60

LIST OF FIGURES

Figure 3.1. Transaction value of digital payments in Southeast Asia in 2023, by country and segment (in billion U.S. dollars) (Statista, 2023).....	16
Figure 3.2. Visual Representation of SMOTE (Schubach et al., 2017).....	16
Figure 3.3. Logistic Curve Model for a Dichotomous Dependent Variable (Menard, 2010).....	18
Figure 3.4. A single neuron (or perceptron) in a Multilayer perceptron model (Afan et al., 2021).....	19
Figure 3.5. Basic Multilayer Perceptron Model (Afan et al., 2021).....	20
Figure 3.6. Visualization of GridSearchCV Concept.....	20
Figure 3.7. The Kolmogorov-Smirnov Statistic Graph (Adeodato and Melo, 2016).....	22
Figure 3.8. The ROC graph (Adeodato and Melo, 2016).....	23
Figure 3.9. Confusion Matrix (de Giorgio et al., 2023).....	23
Figure 4.1. Research Procedure Flowchart.....	26
Figure 4.2. Feature Selection Flowchart.....	29
Figure 4.3. Model Fitting Flowchart.....	30
Figure 4.4. Logistic Regression Architecture w/o SMOTE.....	31
Figure 4.5. Logistic Regression Architecture w/ SMOTE.....	31
Figure 4.6. Resulting Model Architecture After Manual Fine-Tuning.....	32
Figure 5.1. Data Preprocessing.....	36
Figure 5.2. Function to calculate statistics of transactions and create new columns.....	38
Figure 5.3. Feature Selection by Information Value.....	39
Figure 5.4. Feature Selection by Correlation Matrix.....	40
Figure 5.5. Train, test, evaluation split.....	41
Figure 5.6. Model Pipeline Logistic Regression + GridSearchCV.....	42
Figure 5.7. SMOTENC Process.....	43
Figure 5.8. Multilayer Perceptron Model Development.....	44
Figure 5.9. Calling SHAPLeY Additive Explainer Library for Logistic Regression.....	45
Figure 5.10. Calling SHAPLeY Additive Explainer Library for Multilayer Perceptron.....	46
Figure 6.1. Class Imbalance Before and After SMOTENC.....	47
Figure 6.3. MLP Accuracy and Loss Graph (Non-SMOTE).....	51
Figure 6.4. MLP Accuracy and Loss Graph (SMOTE).....	51
Figure 6.5. Confusion Matrices of LR and MLP (Non-SMOTE).....	52
Figure 6.6. Confusion Matrices of LR and MLP.....	52
Figure 6.7. ROC Curve and K-S Graph of LR w/o SMOTE.....	54

Figure 6.8. ROC Curve and K-S Graph of MLP w/o SMOTE.....	54
Figure 6.9. ROC Curve and K-S Graph of LR w/ SMOTE.....	55
Figure 6.10. ROC Curve and K-S Graph of MLP w/ SMOTE.....	55
Figure 6.11. SHAP Violin Plot LR(Left) and MLP(Right) without SMOTE..	56
Figure 6.12. SHAP Violin Plot LR(Left) and MLP(Right) with SMOTE.....	57

LIST OF TABLES

Table 2.1. Table of Literature Review.....	11
Table 4.1. Dry-run.....	26
Table 5.1. Dataset Dictionary.....	34
Table 5.2. Feature Engineering Table.....	36
Table 5.3. Feature Engineering Table.....	38
Table 6.1. Table of Intersecting Variables.....	49
Table 6.2. Multilayer Perceptron Model Summary.....	50
Table 6.3. Consolidated Accuracy from All Models.....	53
Table 6.4. Consolidated AUC and K-S Score from All Models.....	55

LIST OF EQUATIONS

Equation 3.1. Information Value.....	17
Equation 3.2. Logistic Regression Function.....	18
Equation 3.3. SHAP Value Function.....	21
Equation 3.4. Kolmogorov–Smirnov (KS) test statistic.....	21