UNIVERSITY OF BAHRAIN
College of Information Technology



Towards Electronic Fund Transfer Fraud Detection Using Machine Learning

A Thesis Submitted in Partial Fulfilment of the Requirements for the Master of Science Degree in Cyber Security

Submitted by
Hussain Ismaeel
202100048

Supervised by
Dr. Wael Elmedany
Associate Professor
University of Bahrain

Kingdom of Bahrain
May 2023

University of Bahrain
Deanship of Graduate Studies
And scientific Research

جامعة البحرين
عمادة الدراسات العليا والبحث العلمي

Dissertation Deposit Form

Author Name:	Hussain Abdulaziz Saeed Ahmed Ismaeel		
Email:	202100048@stu.uob.edu.bh	Contact No:	38888924
College:	**Information Technology**		
Department:	**Computer Science**		
Thesis Title:	**Towards Electronic Fund Transfer Fraud Detection Using Machine Learning**		
Year:	2023		

Author's Declaration: I agree to the following conditions:

The above thesis will be available (in the University of Bahrain library and externally) and reproduced as necessary at the discretion of the University of Bahrain. It may also be digitized by the University of Bahrain and made available via the university's repository or via other parties on the Internet.

This condition shall apply to ALL copies including electronic copies.

The above thesis has been provided on the understanding that it is copyright protected material and that no quotation from this thesis may be published without proper acknowledgement.

Signature of the Author:	Hussain Abdulaziz Saeed Ahmed Ismaeel	Date:	22-Jun-23

10 Approved at the University Council 6th meeting, 2013
In accordance with Decision No. 736/2013 dated 14/3/2013
Modified by University Council Decision No. 2043/2013 dated 20/11/2013.

DECLARATION

I declare that this work is the result of my own investigation and that it has not already been accepted in substance for any degree, nor is it currently submitted for any degree.

Signed:  Date: 22/06/2023

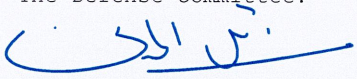
Husain Abdulaziz Saeed Ahmed Ismaeel

The Thesis defense committee considers the thesis titled:

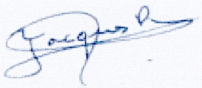
Towards Electronic Fund Transfer Fraud Detection Using Machine Learning

Submitted by Husain Abdulaziz Saeed Ahmed Ismaeel to the College of Information Technology is satisfactory and acceptable for the Master of Science in Cyber Security.

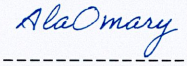
The Defense Committee:



Dr. Wael Elmedany (Supervisor) — Department of Computer Engineering, University of Bahrain



Prof. Jacques Demerjian (External Examiner) — Lebanese University



Dr. Alauddin Al-Omary (Internal Examiner) — Department of Computer Engineering, University of Bahrain

ABSTRACT

Electronic payment methods have significantly evolved and became dominant in the contemporary financial system, offering a variety of seamless payment channels of electronic transactions. The number of electronic payments transactions has taken a drastic leap in the last decade, processing millions of transactions on a daily basis. However, the dramatic increase in digital payment transactions has been accompanied with dreadful number of fraudulent activities which result in substantial financial losses to financial institutions and their consumers. Further, the rapid technological advancement plays a key role in the increase of fraud transactions and are projected to ever grow which posture vital risk. Thus, fraud prevention and detection has become a necessary requirement to ensure reliability and credibility of financial institutions. Data mining and machine learning techniques have been prominent strategies by researchers to develop methods to detect frauds and reduce the risk. In that context, designing efficient and practical algorithms have encountered several challenges including the lack of interpretability of fraud definition, the unavailability of fraud transactional data due to privacy regulation, and the sophisticated indistinguishable characteristics of fraud transactions. In this thesis, we propose a machine learning model to detect fraud Electronic Funds Transfer (EFT) transactions. To assess the model's efficacy, experiments are carried out on a vast simulated transactions dataset spanning two years which incorporated Synthetic Minority Oversampling Technique (SMOTE) resampling technique along with cross validation using stratified K-fold. The evaluation involved the utilization of multiple metrics to assess the model comprehensively, which comprised the confusion matrix, Receiver Operating Characteristic Area Under the Curve (ROC AUC), Precision-Recall Area Under the Curve (PR AUC), precision, recall, F1-score, and time efficiency. The performance results show employing SMOTE technique improve the model detection rate, and the proposed XGBoost model proved to be the most efficient and effective, achieving a perfect ROC AUC and a PR AUC of 0.994.

Contents

APPROVAL . . . . . A
DECLARATION . . . . . B
ABSTRACT C
TABLE OF CONTENTS . . . . . D
LIST OF FIGURES . . . . . G
LIST OF TABLES . . . . . H
ACKNOWLEDGMENTS I
DEDICATION . . . . . J
LIST OF ABBREVIATIONS . . . . . K

1 Introduction **1**

1.1 Overview 1
1.2 The Battle Against Financial Fraud 3
1.3 Credit Card Fraud Types 5
1.3.1 Physical credit card fraud 6
1.3.2 Credit card Information fraud 6
1.4 The Role of Electronic Fund Transfer System 7
1.5 Research Problem 10
1.6 Research Objectives 11
1.7 Research Significance 12
1.8 Thesis Outline 12

2 Literature Review **13**

2.1 Literature Review 13

2.1.1 Supervised Machine Learning Techniques 14

2.1.2 Unsupervised Machine Learning Techniques 17

2.2 Data Augmentation Techniques 21

2.3 Transaction Fraud Detection Datasets 23

2.4 Summary 25

3 Methodology **26**

3.1 Introduction 26

3.2 Proposed Model 26

3.3 Dataset Description 28

3.4 Data Preprocessing 31

3.4.1 Categorical features 32

3.4.2 Standardization 32

3.5 Feature Selection 33

3.6 Class Imbalance 34

3.6.1 Re-sampling Methods 34

3.6.2 Cost-sensitive Methods 35

3.7 Evaluation Metrics 36

3.7.1 Confusion Matrix 37

3.7.2 Receiver Operating Characteristic 37

3.7.3 Precision-Recall Curve 38

3.8 Machine Learning Algorithms 38

3.8.1 Random Forest 39

3.8.2 Extreme Gradient Boosting (XGBoost) 40

3.8.3 Support Vector Machine 40

3.9 Hyper-Parameter Tuning 41

3.9.1 Random Forest Hyper-Parameters Selection 42

3.9.2 XGBoost Hyper-Parameters Selection 42

3.9.3 Support Vector Machine Hyper-Parameters Selection 43

4 Findings and Results **44**

4.1 Introduction 44

4.2 Random Forest Results 44

4.2.1 Original Dataset 45

4.2.2 Balanced Dataset using SMOTE 46

4.2.3 Cross Validation 47

4.3 Extreme Gradient Boosting Results 48

4.3.1 Original Dataset 48

4.3.2 Balanced Dataset using SMOTE 49

4.3.3 Cross Validation 50

4.4 Support Vector Machine Results 51

4.4.1 Original Dataset 52

4.4.2 Balanced Dataset using SMOTE 53

4.4.3 Cross Validation 54

4.5 Comparative Analysis 55

4.5.1 Time Efficiency 60

4.5.2 Comparison with Previous Works 60

5 Conclusion, Recommendations and Future Work **63**

5.1 Conclusion 63

5.2 Recommendations and Future Work 64

References **66**

A Interview Questions **78**

Abstract (in Arabic) . . . . .

List of Figures

1.1 Electronic Funds Transfer System Overview. 8
1.2 One Leg Transaction. 9
1.3 Two Leg Transaction. 10

2.1 Data Augmentation Techniques. 22

3.1 Real-Time EFT Fraud Detection System architecture. . . . 27
3.2 Proposed Fraud Detection System. 28
3.3 Distribution of dataset classes. 29
3.4 Dataset Correlation Matrix Heat map. 33
3.5 Class Distribution for Imbalanced Dataset and Oversampled Dataset. 35
3.6 Random Forest Prediction Overview. 39

4.1 Random Forest Original Dataset Results. 45
4.2 Random Forest Resampled Dataset Results. 47
4.3 XGBoost Original Dataset Results. 49
4.4 XGBoost Resampled Dataset Results. 50
4.5 SVM Original Dataset Results. 52
4.6 SVM Resampled Dataset Results. 54
4.7 Original Dataset Results Comparison. 57
4.8 Resampled Dataset Results Comparison. 57
4.9 ROC AUC Results Comparison. 59
4.10 PR AUC Results Comparison. 59

List of Tables

1.1 Comparison of Conventional Fraud Detection and ML Fraud Detection 5

2.1 Comparison of Fraud Detection Systems 20
2.2 Transactions Fraud Datasets Summary 24

3.1 Dataset Feature Description 30

4.1 Random Forest Original Dataset Results 46
4.2 Random Forest Resampled Dataset Results 47
4.3 RF Model Performance 48
4.4 XGBoost Original Dataset Results 49
4.5 XGBoost Resampled Dataset Results 51
4.6 XGBoost Model Performance 51
4.7 SVM Original Dataset Results 53
4.8 SVM Resampled Dataset Results 54
4.9 SVM Model Performance 55
4.10 All Experiment Results 56
4.11 All Experiment Results 60
4.12 Comparison with State of the Art 61

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. Wael Elmedany for his guidance, support, and encouragement throughout the research and writing process. His insightful feedback and constructive criticism have been invaluable in shaping this thesis. I would also like to thank my family for their unwavering support and encouragement. Their understanding and patience have been crucial in enabling me to complete this thesis. Additionally, I would like to acknowledge the assistance and resources provided by my colleagues and other individuals who have generously shared their time and expertise. Finally, I extend my thanks to the university for providing me with the opportunity to pursue my academic goals and for the resources and facilities that have made this research possible.

DEDICATION

I dedicate this thesis to my parents, whose unwavering love and support have been the driving force behind my academic achievements. I also dedicate this thesis to my family, whose love and support have been a source of motivation and inspiration. Finally, I dedicate this thesis to all the individuals and organizations that have supported me throughout my academic journey, including my manager, mentors, and colleagues.

LIST OF ABBREVIATIONS

ACC	Accuracy
ADASYN	Adaptive Synthetic Sampling
AE	Auto Encoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
ATM	Automated Teller Machine
AVS	Address Verification Service
CBB	Central Bank of Bahrain
CCF	Credit Card Fraud
CNP	Card Not Present
CS	Cost Sensitive
CVM	Card Verification Method
DL	Deep Learning
DT	Decision Tree
EDA	Exploratory Data Analysis
EFT	Electronic Funds Transfer
EMV	Europay, Mastercard, and Visa
FDS	Fraud Detection System

FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GA	Genetic Algorithm
GAN	Generative Adversarial Networks
GBM	Gradient Boosting Machine
IF	Isolation Forest
KNN	K-Nearest Neighbour
LGBM	Light Gradient Boosting Machine
LOF	Local Outlier Factor
LR	Logistic Regression
MCC	Merchant Category Code
ML	Machine Learning
NB	Naïve Bayes
NN	Neural Network
OCSVM	One-Class Support Vectore Machine
PCA	Principal Component Analysis
PII	Personal Identifiable Information

POS Point-of-Sale

PR Precision-Recall

PR AUC Precision-Recall Area Under the Curve

RBF Radial Basis Function

RBM Restricted Boltzmann Machine

RF Random Forest

ROC Receiver Operating Characteristic

ROC AUC Receiver Operating Characteristic Area Under the Curve

RUS Random Under Sampler

SMOTE Synthetic Minority Oversampling Technique

SOM Self Organizing Maps

SVM Support Vector Machine

TN True Negative

TNR True Negative Rate

TP True Positive

TPR True Positive Rate

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

1.1 Overview

Digital evolution is dramatically transforming the contemporary financial systems towards the adoption of more simple and convenient financial services. As technology advances, electronic payment methods are increasing to serve an accelerating online marketplace demand, and outperforming traditional financial instruments in offering services that are more efficient and easily accessible (Tasca, Aste, Pelizzon, & Perony, 2016). The wide adoption of variety forms of digital payments including mobile wallets and online banking has been crucial in raising the volume of Electronic Fund Transfer (EFT) transactions. According to Central Bank of Bahrain (CBB) published report for 2019, transactions processed by digital wallets only have surged by 566.7% to 8.0 million in 2019 from 1.2 million reported in 2018 (CBB, 2020).

This great conversion in payment methods has exacerbated the security and privacy concerns of the emerged technologies flaws and vulnerabilities. Cybercrimes have taken the opportunity to capitalize from the increased load of digital transactions to target a lucrative avenue (Hasan & Rizvi, 2022). One of the constant and costly reality that accompanied the extensive increase of online payment transactions is fraud. The prolifera-

tion of financial fraud has increased tremendously in recent years which has led to substantial losses to card issuers and financial institutions (Tiwari, Mehta, Sakhuja, Kumar, & Singh, 2021). The dreadful financial losses do not only impact financial institutions but also affect individual customers. In addition, successful financial frauds consequently can cause significant non-financial reputational damage. According to Nilson report, in 2020 the gross card fraud losses have incurred of $28.58 Billion and is projected to increase to $49.32 Billion by the year 2030 (Report, 2021).

Subsequently, fraud prevention and detection have been an ongoing battle along with the rise of electronic payment methods. Financial institutes and regulators have introduced several regulations and compulsory requirements for processing EFT transactions in an attempt to decline the rate of fraud payments and prevent fraudulent attempts (Sanad & Al-Sartawi, 2021). Additionally, rule-based statistical fraud detection techniques were widely employed by various financial institutes. While these strategies have declined financial frauds to a certain level, fraudsters are continuously investigating further sophisticated strategies to bypass these prevention methods (Raghavan & El Gayar, 2019). Moreover, since rule-based statistical techniques are manually crafted by fraud experts, fraudsters have always managed to overcome the process and making it obsolete (Hsin, Dai, Ti, & Huang, 2021).

Data mining techniques and Machine Learning (ML) models are promising research area to detect anomalies in various applications such as fraud detection (Adepoju, Wosowei, Jaiman, et al., 2019). Data mining techniques can be used to finding insightful and unique patterns from huge datasets taking advantage of its ability to extract descriptive information and key characteristics by analyzing various categorical attributes of the dataset individually as well as jointly (Zaki, Meira Jr, & Meira, 2014; Yee, Sagadevan, & Malim, 2018). On the other hand, ML Models can be used to quickly

identify and detect hidden patterns and interpretation from massive amount of data to detect fraudulent transaction operations (Varmedja, Karanovic, Sladojevic, Arsenovic, & Anderla, 2019; Khatri, Arora, & Agrawal, 2020). Accordingly, many experts endorse that data analytics is to be the core of future fraud detection systems and fraud detection techniques shall be driven and centered on data (M. A. Ali, Azad, Centeno, Hao, & van Moorsel, 2019; Nassif, Talib, Nasir, & Dakalbab, 2021).

1.2 The Battle Against Financial Fraud

With the rise of technological advancement, the advent of a plethora of digital financial services has played vital role in increasing the financial cybercrimes including financial fraud (Kaur et al., 2021). The rampant of financial fraud in its variant forms has caused the financial industry serious repercussions for both individuals and organizations which cannot be overlooked anymore due to the large-scale occurrence of fraudulent activities. While financial fraud crimes are not new to the financial industry. Yet, the pace and the magnitude of financial frauds have been at constant increase over the past few decades (Reurink, 2018). Financial fraud involves the illegal use of financial assets or financial transactions to obtain goods or services, or to gain access to sensitive financial information. Financial statement fraud, Credit Card Fraud (CCF), and transaction fraud are among the most frequently occurring forms of financial fraud. Financial fraud crimes can be classified based on the perpetrator of the fraud activity as well as according to the organization involved (Zhu et al., 2021).

Nevertheless, among the variant types of financial fraud, credit card transactions fraud is one of the most prevalent and pervasive form of financial fraud due to several factors including the emergence and proliferation of online payments methods, the evolving tactics and techniques of the modus

operandi of fraudulent activities, and the growing numbers of Card-Not-Present (CNP) payments (i.e., payment executed over the phone or online) (Mekterović, Karan, Pintar, & Brkić, 2021). CCF is a type of financial fraud that involves the unauthorized use of a credit card or card information to make fraudulent purchases or fraudulent fund transfer transactions.

To combat CCF, financial institutions and banks have enforced a range of preventive measures and techniques to reduce the occurrence of fraud including security protocols, data encryption or tokenization, Europay, Mastercard and Visa (EMV) chip, and multi-factor authentication (Singh & Jain, 2020). These measures are designed to prevent fraudsters from attempting CCF transactions and to prevent financial losses. In addition, financial institutions have employed variety of detection techniques such as rule-based systems, scoring system, hot cards, and real-time monitoring (Pourhabibi, Ong, Kam, & Boo, 2020). These measures are meant to detect transactions requests based on expert set rules of potential fraud behaviour or features.

However, fraudsters are constantly developing new tactics and techniques to pass existing preventive and detection strategies. Furthermore, the existing techniques are only effective to known forms of frauds and require continuous costly updates and investigations. The need for competent and vigilant detection techniques are inevitable to keep up with the new evolving form of fraud techniques (Priya & Saradha, 2023). Thus, data driven automated detection techniques have become necessary for financial institutions to combat and detect fraudulent activities as soon as they occur. The fraud defence mechanisms can be categorized into prevention and detection layers (Rodrigues et al., 2022). Some research extends the layers of fraud prevention to three layers and four layers model including containment and enforcement (Dastidar, Granitzer, Siblini, et al., 2022). Table 1.1 present a comparison between the conventional fraud detection approach and ML fraud detection in term of approach, scalability, accuracy, and efficiency.

Table 1.1: Comparison of Conventional Fraud Detection and ML Fraud Detection

Aspect	**Conventional Fraud Detection**	**ML Fraud Detection**
Approach	Rule-based	Adaptive and data-driven
Analysis	Manual analysis by fraud analysts	Automated analysis
Scalability	Limited scalability for large volumes of data and real-time transactions	Scalable for handling big data and real-time processing
Identifying Complex Patterns	Difficulty in identifying complex fraud patterns	Ability to detect complex and evolving fraud patterns
Accuracy	Moderate accuracy and precision	Improved accuracy and precision
Real-Time Detection	Limited real-time detection capabilities	Real-time detection and proactive prevention
Efficiency	May be time-consuming and prone to human error	Efficient processing and analysis

1.3 Credit Card Fraud Types

CCF refer to the use of physical card or card information to perform payment or transfer funds between accounts illegally. Fraudsters can obtain physical custody or card information using various ways (Arora, 2022). Due to the varied and mounting types of CCF, no consensus has been achieved on how to categorize the types of CCF. According to the fraudulent cases covered in the literature, fraud types can be categorized based on the possession of card information (Mittal & Tyagi, 2020).

1.3.1 Physical credit card fraud

1. Stolen/lost card: the use of stolen or lost credit cards to make purchases or payment. While this also require obtaining the PIN code. Fraudsters can use the contact-less features of cards to make purchases.
2. Application card: the use of false or fake information to apply or replace a credit card under someone else identity.
3. Mail intercept card: the interception of mails of credit cards by fraudsters and obtaining cards intended to go to its authorized users and using it.
4. Counterfeit card: the use of a device to steal credit card information, and clone encoded information into a new created forged card.

1.3.2 Credit card Information fraud

1. Data breach: credit card information can be leaked due to a data breach of ecommerce websites or due to insider data theft which fraudsters can use it to perpetrate fraudulent activities.
2. Credit card imprint: obtaining credit card information by skimming the card or through a service provider and later use it to perform fraudulent transactions.
3. Account takeover: with the advent of various digital wallet, fraudsters can takeover account information to access all saved credit cards and make fraudulent activities.
4. Phishing/spoofing: fraudsters can trick users to provide their credit card information through malicious emails or commit payments through malicious/forged websites.

1.4 The Role of Electronic Fund Transfer System

Electronic Funds Transfer (EFT) refers to the process of transferring funds electronically from one bank account to another. EFT can be used for a variety of financial transactions, such as paying bills, making purchases, and transferring funds between accounts. EFT system support various electronic channels, such as online banking, mobile banking, Automated Teller Machine (ATM), and Point-Of-Sale (POS) terminals. EFTs can be processed in real-time, allowing for fast and secure transactions, and they are often less expensive than traditional paper-based transactions. Figure 1.1 illustrates the multiple components which makes the EFT System. Like any financial system, EFT systems can be susceptible to fraud. Fraudulent EFT transactions can occur through a variety of methods including phishing, identity theft, and account takeover. To prevent EFT fraud, it is essential to understand how transactions flow within the payment ecosystem to effectively examine potential fraud crimes and attacks.

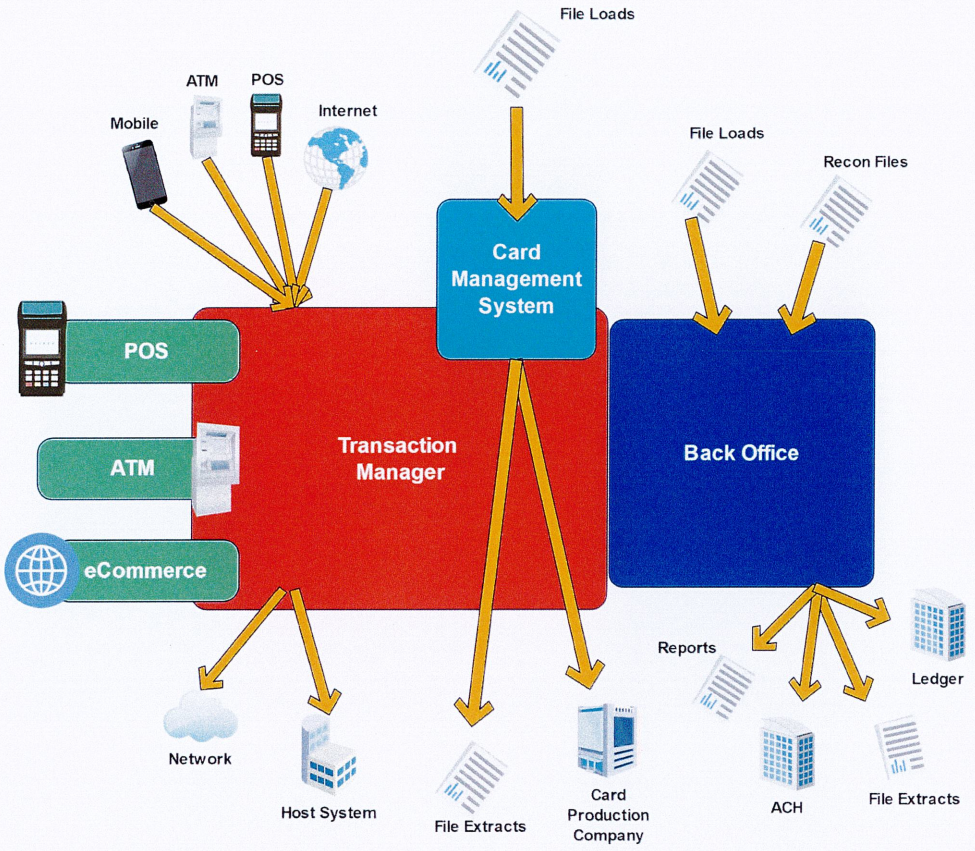


Figure 1.1: Electronic Funds Transfer System Overview.

In the context of EFT, the workflow for processing transactions comprises several stages and includes multiple financial entities, such as acquirers and issuing banks. Moreover, the processing workflow varies depending on the type of transaction and the card used in the transaction. Generally, financial entities involved in the process of EFT transactions are acquiring banks and issuing banks. The financial organization that provides credit or debit cards to its clients is known as the issuing bank. The issuing bank is responsible for authorizing transactions and ensuring the customer's account has the funds to be transferred to the recipient's account. On the other hand, the financial institution that handles transactions on behalf of the recipients is referred to as the acquiring bank. The acquiring bank is responsible for verifying the transaction details, settling the transaction, and funds from

one account to another. Therefore, the issuing bank and acquiring bank play critical roles in the functioning of EFT systems as well as the processing of EFT transactions, and their relationship is key to ensuring that transactions are processed quickly, securely, and accurately.

Another important aspect in EFT transactions fraud detection, is the message pair. A message pair refer to the two messages exchanged between the parties involved in the transaction. There are two types of message pairs in EFT transactions: single message pair and dual message pair.

In a single message pair transaction, the authorization and clearing occur simultaneously, and the funds are transferred immediately. This means that there is little time for fraud detection and prevention measures to be implemented before the funds are transferred. Therefore, single message pair transactions are generally considered to have a higher risk of fraud. Figure 1.2 demonstrate the workflow of a single message pair transactions such as cash withdrawals.

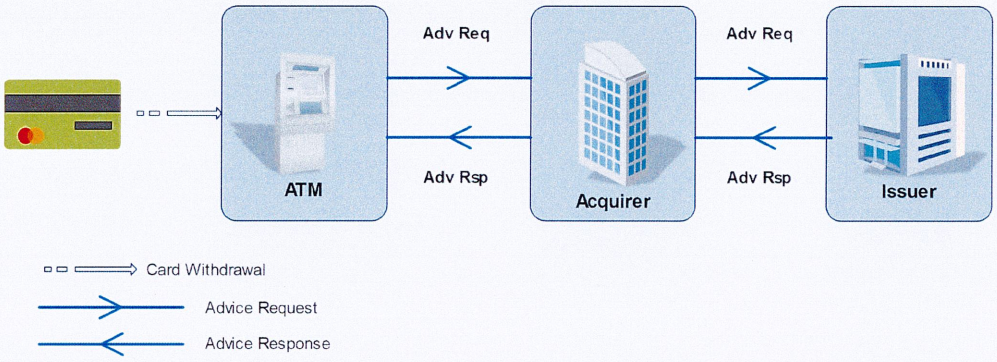


Figure 1.2: One Leg Transaction.

In contrast, a dual message pair transaction involves a delay between the authorization and clearing messages, which provides more time for fraud detection and prevention measures to be implemented. For example, the issuing bank can perform additional verification checks during the time delay to ensure that the transaction is legitimate. As a result, dual message pair

transactions are generally considered to have a lower risk of fraud. Figure 1.3 demonstrate the workflow of a dual message pair transactions such as POS and online purchases.

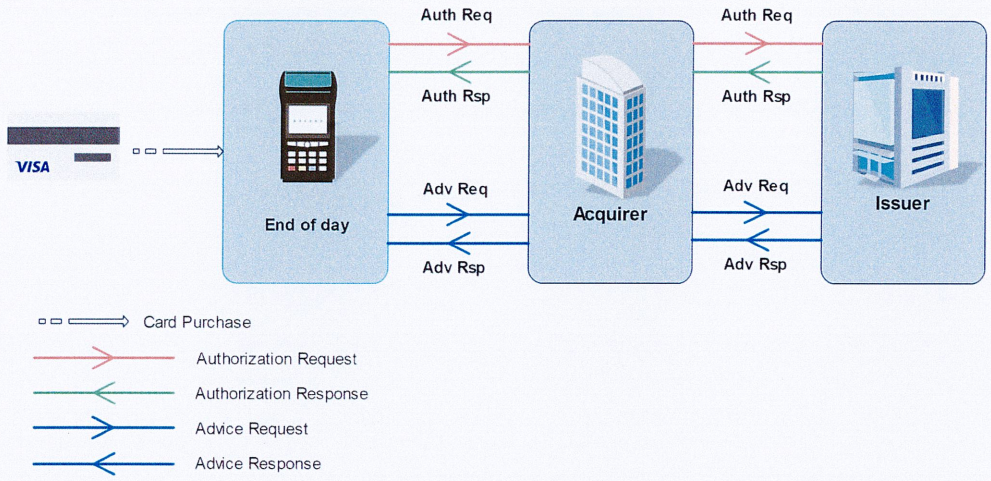


Figure 1.3: Two Leg Transaction.

Overall, the choice between single message pair and dual message pair depends on the transaction value, risk level, and the requirements of the parties involved in the transaction. Therefore, understanding the type of message pair used in an EFT transaction is important for assessing the level of risk involved and for implementing appropriate fraud detection and prevention measures.

1.5 Research Problem

Financial frauds in its variant ways and its different intent have gained a considerable amount of traction by fraudsters and cyber criminals due to the potential financial benefits. Fraud activities have increased dramatically and have evolved significantly to form an alarming rate to service providers globally. Financial institutions suffer great financial and non-financial losses

due to the accelerating fraud transactions. The conventional and legacy prevention and detection techniques are becoming inapplicable and inefficient to detect fraud attempts with high accuracy. As such, fraud prevention and detection have become necessary requirements to ensure the reliability and credibility of financial institutions. Data mining and machine learning techniques have emerged as prominent strategies by researchers to develop methods for detecting fraud and reducing risk. However, the design of efficient and practical algorithms has encountered several challenges, including the unavailability of fraud transactional data due to confidentiality, the lack of domain specific metrics to appraise the model's effectiveness, and the class distribution imbalance nature of fraudulent and geniune transactions.

1.6 Research Objectives

This research aims to address the increasing problem of fraudulent activities in electronic payment transactions, which pose a significant risk to the reliability and credibility of financial institutions. The objectives of this research can be summarized as follows:

1. Investigate the role of machine learning techniques in developing effective fraud detection methods.
2. Develop efficient machine learning model to detect EFT fraudulent transactions in real-time.
3. Evaluate the model's performance using a large simulated dataset spanning two years, incorporating resmapling techniques to address the class imbalance of the dataset.
4. Utilize multiple evaluation metrics, including time efficiency, to comprehensively assess the model's effectiveness and actual outcome.

1.7 Research Significance

The rising number of fraudulent transactions could have serious effects on the electronic payment ecosystem such as huge financial devastation and destroyed reputation. Fraud prevention and detection are critical activities for organizations that aim to protect themselves from financial losses, reputational damage, and legal liabilities. This research provides financial expertise insights of using machine learning for fraud detection. This research aims to help financial institutions including banks and merchants to identify and detect fraudulent transactions from legitimate transactions in real-time with high accuracy to reduce financial losses.

1.8 Thesis Outline

The organization of this thesis is as follows: chapter 2 provide brief review of the literature and the various ML methods employed to detect fraud; chapter 3 describes the methodology and data preprocessing techniques utilized for this research. Chapter 4 illustrates the results and findings of the ML model implementation. In chapter 5, the thesis is concluded and presents the recommendation and future work.

Chapter 2

Literature Review

2.1 Literature Review

Since the advent of digital payments, combating financial frauds have been on top of financial institutions endeavor due to the growth of fraud activities and the potential associated financial loss. With the increased load of online transactions, traditional rule-based fraud detection system became impractical due to the costly manual intervention required by fraud investigators to update and set new rules to cope with the constantly changing strategies by fraudsters (Hilal, Gadsden, & Yawney, 2022). Thus, fraud detection have acquired increasing attention from research community over the past decades (Mansour, Ahmi, Popoola, & Znaimat, 2022). Several research studies exploited machine learning and data analytic techniques to examine financial fraud detection. Fraud detection ML techniques can be categorized into two categories namely supervised ML models and unsupervised ML models (Şahin & Duman, 2011; Dhankhad, Mohammed, & Far, 2018).

2.1.1 Supervised Machine Learning Techniques

To detect fraud transactions, numerous supervised ML methods were proposed to the literature over the past decade including Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbour (KNN) (Srivastava, Kundu, Sural, & Majumdar, 2008; Chaudhary, Yadav, & Mallick, 2012; Shirgave, Awati, More, & Patil, 2019; Jain, Tiwari, Dubey, & Jain, 2019; M. S. Kumar, Soundarya, Kavitha, Keerthika, & Aswini, 2019). Given the nature of the transaction fraud problem being a classification problem, many researchers employed supervised classifiers to classify transaction data into legitimate and fraudulent transactions (Al-Shabi, 2019).

(Thennakoon, Bhagyani, Premadasa, Mihiranga, & Kuruwitaarachchi, 2019) compared four ML algorithms including SVM, KNN, NB, and Logistic Regression (LR) to predict four different fraud scenarios. The proposed system focused on four transactions features including amount, Merchant Category Code (MCC), response code, and address. The dataset used was obtained from a financial institute and subjected to confidential disclosure agreement. The evaluation comprised using the dataset into two forms as raw data form and as transformed numerical form. The result shows that different algorithm outperform the others depending on the fraud pattern with SVM achieving the highest accuracy of 91% (Thennakoon et al., 2019).

To identify credit card fraudulent transactions efficiently, (S. Kumar, Gunjan, Ansari, & Pathak, 2022) proposed the use of feedback system along with SVM algorithm to improve the classification and the detection rate. To train the model the authors used a publicly offline dataset with few transactions and contains several credit card transaction attributes such as merchant id, transactions amount, decline code, count of declines per card and country risk rate. To assess the efficacy of the proposed system, it was compared to the state of the art classification methods including Random Forest (RF),

KNN, NB, and LR. The results of the evaluation demonstrate that SVM achieved the highest accuracy among the other classifiers as well as other performance metrics such as specificity and sensitivity (S. Kumar et al., 2022).

Gradient Boosting Machine (GBM) techniques including Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM) have been used in several studies to detect CCF (Taha & Malebary, 2020). In (Meng, Zhou, & Liu, 2020), a fraud detection system proposed using XGBoost algorithm. The proposed model used the European Cardholder dataset which contains real-life online transactional data. In the study, various techniques, such as undersampling, were utilized in the study to tackle the problem of imbalanced dataset. Further, Synthetic Minority Oversampling Technique (SMOTE) was used in order to balance the legitimate and fraudulent transactions classes. The evaluation comprised testing the model with the three forms of the dataset including the original form of the dataset, the undersampled form and the balanced form. The evaluation emphasized on the Receiver Operating Characteristic (ROC) curve and recall as performance metrics. The result shows that the Area Under Curve (AUC) increases when using SMOTE technique.

(Hajek, Abedin, & Sivarajah, 2022) proposed a fraud detection model based on XGBoost trained with a large dataset containing over 6 million transactions. To evaluate the model, two set of experiments were performed. The first to capture the baseline performance of the model and the second is performed on balanced training sample. The baseline results show that XGBoost outperform other classifiers in the AUC and F1-score metrics. The authors used Random Under Sampler (RUS) undersampling procedure to tackle the imbalance within the dataset. The combination of RUS and XGBoost improved the AUC result but caused a deterioration to the F1-score.

In another study, (Alarfaj et al., 2022) evaluated the accuracy of six dif-

ferent ML algorithms using the European cardholder dataset. The algorithms includes DT, KNN, LR, SVM, RF, and XGBoost. The result achieved accuracy of 97.08%, 93.27%, 97.18%, 97.43%, 97.55%, 94.6% respectively. The study also researched the use of Deep Learning (DL) techniques to detect CCF. While the characteristics of deep learning reveal a promising area of improvement for fraud detection, the dependency of large-scale datasets hinders the implementation of DL methods.

One of the most critical challenges which impedes the development of efficient and effective fraud detection solution is the class imbalance of the available dataset. Due to the vast increase in numbers of online transactions, the number of fraudulent transactions remains considerably low compared to the authentic transactions. (Makki et al., 2019) conducted a rigorous study to address the imbalanced available datasets. The study compared eight distinct ML models that are proposed in the literature. The comparison used accuracy along with True Positive Rate (TPR) and Receiver Operating Characteristic (ROC) curve to best asses the competitive quality of the ML algorithm. The evaluation results concluded that C5.0, SVM, and ANN are the most efficient algorithms. The authors then employed classification balancing techniques including oversampling and Cost Sensitive (CS) to further assess the selected algorithms. The study shows that applying imbalance classification techniques impact the performance and reduce the accuracy of the model. Moreover, imbalance classification approaches encompass several deficiencies and do not adequately address the problem.

The use of Genetic Algorithm (GA) combined with different ML classifiers such as RF, LR, ANN, and NB proposed for CCF detection in (Ileberi, Sun, & Wang, 2022). The proposed framework used two datasets, the European Cardholder dataset and a synthetic dataset generated by (Altman, 2021). The SMOTE method was employed in order to counter the class imbalance of the dataset by generating fresh instances of fraudulent trans-

actions. The evaluation of the model's performance relied on a set of performance metrics, including accuracy, recall, precision, and F1-score. The results of the assessment indicated that the employment of GA feature selection led to a notable improvement in the accuracy of the machine learning classifiers.

Several surveys were carried out to evaluate the effectiveness of the cutting-edge implementation of machine learning models utilized for identifying financial fraud in various domains, such as CCF, counterfeit fraud, bankruptcy fraud, and transactions fraud (Phua, Lee, Smith, & Gayler, 2010; Ngai, Hu, Wong, Chen, & Sun, 2011; Sadgali, Sael, & Benabbou, 2019). In a recent study, a survey was conducted on the research in the same field for the period from 2009 till the 2019 and covered 75 financial fraud relevant studies. The survey showed that the SVM is the highest technique used to investigate the domain followed by NB and RF approaches. Bank Fraud and insurance fraud consumes the majority of the literature due to their importance (Al-Hashedi & Magalingam, 2021). In addition, great proportion of the literature states that the efficiency of ML techniques for detecting fraudulent activities extensively rely on the aggregation of the datasets.

2.1.2 Unsupervised Machine Learning Techniques

The heterogeneous patterns of fraudulent transactions as well as the sparse rate of fraudulent activities posture challenges to problems such as fraud detection. Thus, unsupervised ML algorithms were explored in many studies due to its great abilities in identifying outliers by clustering similar features of raw data into the same class (Niu, Wang, & Yang, 2019). Thus, investigating fraud detection surged the interest in applying unsupervised techniques to class fraudulent transactions as outliers. Several Fraud Detection Systems (FDS) proposed deriving from anomalous detection irrespective of annotations and based on distribution of normal transaction activities (Arfeen &

Khan, 2022). Several algorithms, such as K- means, Isolation Forest (IF), Local Outlier Factor (LOF), and Neural Network (NN) and numerous other techniques were proposed in the literature (Dornadula & Geetha, 2019; Carcillo et al., 2021).

Using the European Cardholder Dataset comprising 284,807 transactions of which 492 are fraudulent transactions, (Mittal & Tyagi, 2019) compared the performance of four unsupervised ML algorithms including Self Organizing Maps (SOM), IF, LOF, and K-means. To evaluate the actual identification rate of fraud, the evaluation metrics involved specificity, sensitivity, and prevalence along with accuracy and confusion matrix metrics. The evaluation also included a comparison of unsupervised algorithms with classical ML algorithms such as SVM, K-NN, RF, and LR. The results shows that IF and LOF outperform other algorithms in the overall effectiveness and superior in dealing with the imbalance of the dataset.

In another study, LOF and IF algorithms were proposed for CCF detection in (Maniraj, Saini, Ahmed, & Sarkar, 2019). The authors employed various pre-processing techniques and standardization techniques to normalize the columns of the dataset. To reduce the time of the evaluation, the testing was also performed using only 10% of the dataset. The metrics used in the evaluation are the overall accuracy score and the model's precision. Although the results managed to achieve an accuracy of over 99.6%, the precision of both LOF and IF remained considerably low and only achieved 33% when using the complete dataset.

(Rai & Dwivedi, 2020) proposed to use unsupervised based NN to efficiently detect frauds in credit cards. The European Cardholder dataset was employed to train the model, which was then evaluated based on certain metrics, including Accuracy, F1-score, precision, Recall, specificity, and False Positive Rate (FPR). To visualize the prediction outcomes, confusion matrix is used for evaluating the accurate performance. Compared to other

unsupervised ML schemes, the proposed scheme outperformed the other algorithms such as Auto Encoder (AE), IF, and LOF in the achieved accuracy, precision and recall. The NN also achieved the lowest FPR and received the highest percentage of balance accuracy.

On a wider scale, a comprehensive comparison performed by (Niu et al., 2019) which incorporated six supervised techniques and four unsupervised techniques. The unsupervised algorithms used are as follows One-Class SVM (OCSVM), Restricted Boltzmann Machine (RBM), Generative Adversarial Networks (GAN), and AE. Due to the highly imbalanced studied data, the comparison used the Receiver Operating Characteristic Area Under the Curve (ROC AUC) as the main performance metric. The observed results demonstrated that supervised techniques have better results on the used dataset. XGboost and RF achieved the highest ROC AUC of 0.9894 and 0.9881 respectively. On the other hand, AE, RBM, and GAN algorithms achieved AUC-ROC score above 0.95.

An accuracy of 99.7% was achieved by implementing LOF in (Cynthia & Thomas George, 2021). The study examined 4 different unsupervised and supervised classifiers namely LOF, IF, SVM and LR. The effectiveness of each model was evaluated using variety of performance metrics. The performance metrics combined the use of three different metrics such as micro average, macro average and weighted average. IF and LR achieved similarly good accuracy of 99.6% and 99.8% respectively. The supervised SVM model on the other hand scored the least accuracy of 97.2%. The analysis of the experiment shows that using unsupervised ML models to detect fraud transactions acquire higher accuracy than those of supervised learning models.

Table 2.1 summarizes the reviewed ML architectures for fraud detection applications highlighting the ML Method, dataset, the advantage of the study and the study limitation.

Table 2.1: Comparison of Fraud Detection Systems

Reference	Method	Dataset	Pros	Cons
(Thennakoon et al., 2019)	SVM, LR, NB, KNN	Private dataset	Analysis of different fraud patterns	Evaluation is measured with accuracy only
(S. Kumar et al., 2022)	SVM, RF, LR, KNN, GBM, DT, NB	Public dataset	Feedback process to enhance the detection rate	Hyper-Parameters were not implemented
(Meng et al., 2020)	XGBoost	European Cardholder dataset	Resampling techniques to balance the data	No data preprocessing performed
(Hajek et al., 2022)	XGBoost	Paysim dataset	Analysis of outlier detection and class imbalance	Evaluatuin is limited by cost savings measure
(Alarfaj et al., 2022)	DT, RF, SVM, LR, KNN, XGBoost, CNN	European Cardholder dataset	Comparison with DL techniques	The proposed model requires larger dataset to improve detection rate
(Rai & Dwivedi, 2020)	Unsupervised NN	European Cardholder dataset	Outperformed other unsupervised methods	Low precision and recall scores
(Cynthia & Thomas George, 2021)	LR, SVM, IF, LOF	European Cardholder dataset	Higher accuracy than supervised methods	Imbalanced dataset is not suited for accuracy evaluation

2.2 Data Augmentation Techniques

Since the performance of ML methods is heavily influenced by the quality of the training data, the class imbalance within the available datasets raises a significant concern related to the accuracy and robustness of the machine learning algorithms (Hsin et al., 2022). Additionally, improper usage of data generation leads to poor performance as it falsely eliminates some properties which results in inaccurate evaluation of the model. various data augmentation strategies have been introduced to tackle the issue of class imbalance within the available datasets for CCF transactions (Strelcenia & Prakoonwit, 2023).

RUS technique is among the most used techniques to under sample the instances of legitimate transactions (Alarfaj et al., 2022; Niu et al., 2019; Hajek et al., 2022). Using RUS approach shows major impact on the result of the classification and decreases the f1-score (Thennakoon et al., 2019). Alternatively, random oversampling has been employed to augment the frequency of fraudulent transactions to achieve balanced classes (Makki et al., 2019; Mittal & Tyagi, 2019). Similarly, random oversampling reduces the accuracy as the new duplicates of incorrectly classified impact the results accordingly.

Synthetic Minority Over-sampling Technique (SMOTE) is another popular technique for re-sampling datasets with unbalanced classes. SMOTE is a type of oversampling technique that synthesize samples for the minority class by interpolating the existing samples within the same class (Mansourifar & Shi, 2020). Implementation of SMOTE helps in improving the performance of the algorithm used and achieving higher AUC score (Meng et al., 2020; Altman, 2021; Esenogho, Mienye, Swart, Aruleba, & Obaido, 2022).

To increase the minority class instances, several studies have used Adaptive Synthetic (ADASYN) sampling for generating synthetic samples of the

fraudulent transactions (Tran & Dang, 2021). Unlike SMOTE, ADASYN is an adaptive technique that focuses on generating more synthetic samples for the minority class instances that are more difficult to learn. While the results shows that ADASYN significantly improves the performance of the ML models, when compared to SMOTE, ML model based on ADASYN achieves lower results than the one based on SMOTE due to the adaptable nature of ADASYN technique (Dang, Tran, Tuan, & Tiep, 2021).

In a hybrid approach, [83] uses both the oversampling SMOTE along with the undersampling Tomek links techniques to address the class imbalance. Tomek links is an undersampling techniques that remove instances from the majority class that are in close proximity to instances from the minority class. The evaluation of the approach showed an increased model performance. However, the model f1-score result dropped significantly. Figure 2.1 present the data augmentation techniques used to balance class distributions in imbalanced datasets.

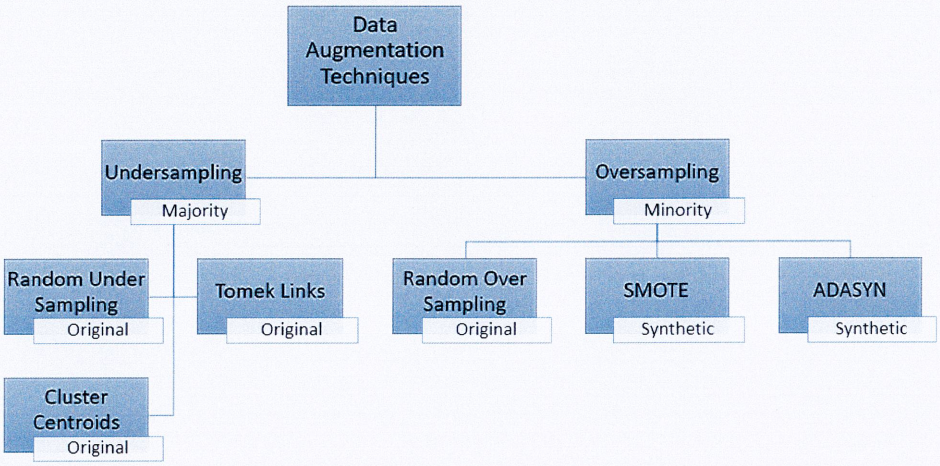


Figure 2.1: Data Augmentation Techniques.

2.3 Transaction Fraud Detection Datasets

The limited availability of fraud transaction datasets has played a crucial role in the development and advancement of fraud detection models. The available public datasets pose several issues towards developing efficient and practical detection models such as high imbalance distributions of fraud and legitimate transactions classes, lack of actual features representation due to confidentiality, and lack of recent datasets which contain up-to-date transaction parameters (Jovanovic et al., 2022). Furthermore, a large proportion of the datasets used in fraud transaction research have been acquired by researchers through partnerships with industry organizations, and have not been made publicly available due to agreements of confidentiality, which obstruct further development and validation of the existing results (Mittal & Tyagi, 2020).

Several datasets have been used to train and test ML models, including the European cardholders CCF Detection dataset, which contains over 284,807 credit card transactions with a binary classification of fraudulent or not fraudulent (GROUP, 2020). A large proportion of the research has been committed using the European cardholder dataset (Dal Pozzolo, Boracchi, Caelen, Alippi, & Bontempi, 2017; Zheng, Liu, Yan, & Jiang, 2018; Ileberi, Sun, & Wang, 2021). Another widely used dataset is the synthetic PaySim dataset which has been generated using a private real-life transactions dataset for mobile users (Lopez-Rojas, Elmir, & Axelsson, 2016). One of the most recent datasets is the IEEE-CIS fraud detection dataset which is provided through a Kaggle competition. The IEEE-CIS dataset presents real-world transactions from Vesta's e-commerce and include variety of features including transaction features and identity feature (Addison Howard, 2019). Table 2.2 present some of the datasets used in the literature and the cardinality of the transactions.

Table 2.2: Transactions Fraud Datasets Summary

Dataset	Cardinality	Reference
European Cardholder CCF Detection dataset	European Cardholder credit cards transactions from 2013 containing 284807 transactions of which 492 are frauds.	(Niu et al., 2019; Ileberi et al., 2022; Alfaiz & Fati, 2022; Strelcenia & Prakoonwit, 2023; Esenogho et al., 2022)
IEEE-CIS fraud detection dataset	Vesta Corporation dataset from 2019 containing 590540 transactions of which 20663 are frauds.	(Society, 2019)
PaySim dataset	Simulated dataset based on real transactions created in 2016 containing over 6 million transactions and around 10 thousand frauds.	(Hajek et al., 2022)
Real transaction data provided by a Bank	Transaction data collected between April 2018 and September 2019 along with fraudulent accounts records supplied by the National Police Agency.	(Hsin et al., 2021, 2022)
PagSeguro payment services	A dataset containing 903801 transactions of which 16639 transactions are fraudulent.	(de Sá et al., 2018)
Worldline financial institution dataset	A dataset from Feb-Dec 2016 and contains 76 million transactions of which 0.36% are fraudulent.	(Carcillo et al., 2021)
UCSD-FICO dataset	A real e-commerce transactions dataset containing 94683 transactions of which 2094 are fraud.	(Taha & Malebary, 2020)

2.4 Summary

CCF is a increasingly prevalent problem that affects both consumers and financial organization. To combat this issue, various techniques and models have been proposed to employ the use of ML to detect CCF. The literature review on CCF detection reveals that there are several approaches to detect fraudulent transactions using various ML methods. Several pre-processing techniques have been suggested to mitigate the issue of dataset imbalance and select the most relevant features or variables that contribute to identifying fraud, with the goal of enhancing the accuracy of fraud detection models.

However, the complexity of the problem presents numerous limitations to the proposed solutions including the limited real-life card fraud dataset, the use of unsuitable evaluation metrics, and the inadequate understanding of the requirements of fraud detection systems in term of computations and time limitation.

Chapter 3

Methodology

3.1 Introduction

In this chapter, the methodology utilized in this study is outlined, encompassing the proposed model, the dataset description, pre-processing methods used, and model fitting, as well as the evaluation metrics utilized to measure the performance of the fraud detection model.

3.2 Proposed Model

There has been a noticeable increase in EFT fraud transactions in recent years which is causing significant losses to both individuals and businesses. Detection of fraud attempts effectively do not only prevent financial losses, but also can alleviate and stop possible future attempts of similar fraud pattern. In this research, we propose an FDS model for combating fraud transactions in real-time, aiming at detecting fraud during the process of a trans-

actions to prevent the recurrence of the detected fraudulent transactions and mitigate any increase load on the EFT system. Figure 3.1 demonstrate the how the FDS will fit within the transaction workflow followed by each step description of the workflow.

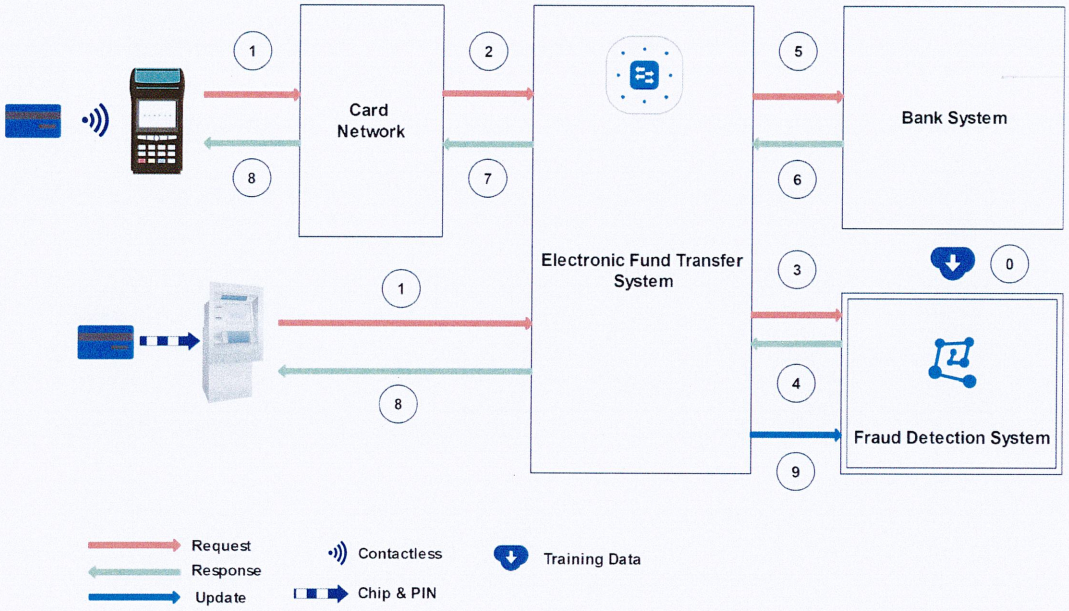


Figure 3.1: Real-Time EFT Fraud Detection System architecture.

Step 1: Transaction is sent to the card network (e.g., MasterCard or Visa).

Step 2: Transaction is received by the EFT system for authorization.

Step 3: EFT system sends the transactions to the FDS for analysis.

Step 4: FDS sends the transaction result to the EFT system.

Step 5: EFT system sends the transaction including the fraud result to the bank system.

Step 6: The bank responds to the EFT system with approval or decline.

Step 7: The EFT system sends the response back to the card network.

Step 8: The response is received by the terminal (e.g., Point-of-sale, Automatic Teller Machine (ATM), or online)

Step 9: The FDS is updated with the transaction response received from the bank.

By leveraging pattern recognition techniques, ML has emerged as a leading field for identifying instances of known fraud activities or transactions that raise suspicion. This can be achieved by utilizing adequate transactional data and applying suitable pre-processing methods. Figure 3.2 illustrates the proposed method for building the FDS.

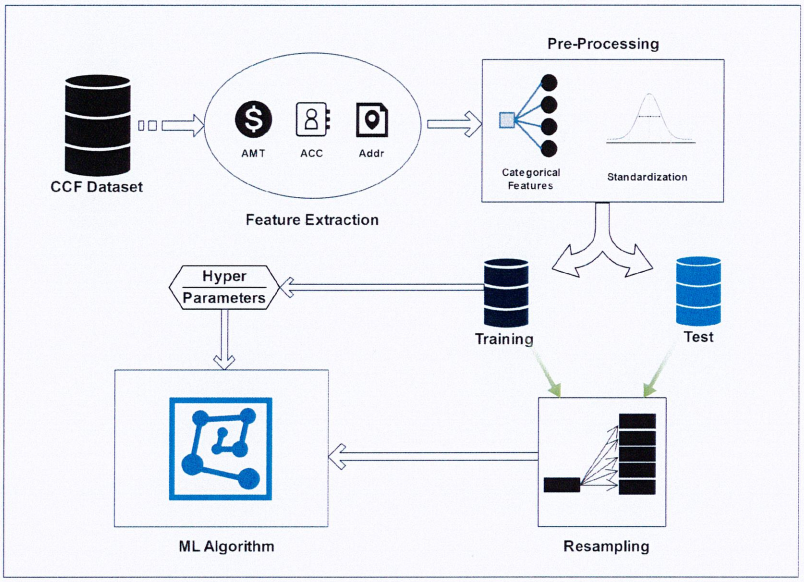


Figure 3.2: Proposed Fraud Detection System.

3.3 Dataset Description

The advancement in the fraud detection research field is confronted by the limited available real life transactions data due to various compliance acts which prohibit the disclosure of Personal Identifiable Information (PII) including financial accounts and credit card information (K. Gupta et al., 2022). In addition, the available datasets are summarized using various techniques including Principal Component Analysis (PCA) to preserve the privacy of the financial details which makes it very difficult to interpret (Grover

et al., 2022). Therefore, in this research study, the dataset used is a simulated dataset which contains a set of synthetic legitimate and fraudulent financial transactions. The dataset is published on Kaggle and was created using Sparkov Data Generation simulator (Shenoy, 2020; Harris, 2022).

The dataset consists of two separated files a training file which include over 1.2 million transactions and a testing file which contains around 500K transactions. It contains genuine and fraudulent transactions for a period of 2 years from 1 Jan 2019 till 31 Dec 2020. The transactions cover credit cards of variant types used across variety of merchant categories. The dataset is highly unbalanced in term of class distribution where fraudulent transactions makes only 0.5% of the total transactions volume combined. The distribution of classes is shown in Figure 3.3. Various features are provided about each transaction, such as the transaction amount, location, and date/time, merchant, and category (e.g., food, shopping, entertainment). The feature details are presented in table 3.1.

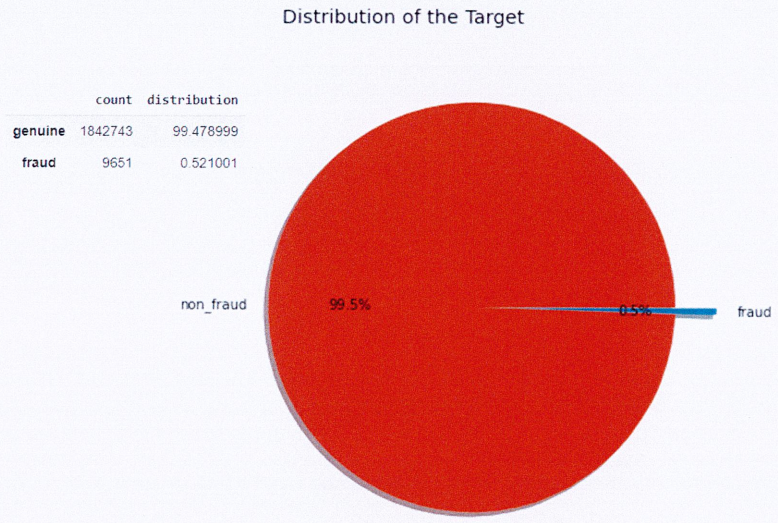


Figure 3.3: Distribution of dataset classes.

Table 3.1: Dataset Feature Description

Feature	**DType**	**Description**
trans_date_trans_time	object	Transaction Date/Time
card_num	int64	Customer Card Number
merchant_name	object	Name of the merchant
merchant_category	object	Merchant type
Amount	float64	Transaction's Amount
cus_first_name	object	Customer first name
cus_last_name	object	Customer last name
cus_gender	object	Customer gender
cus_Address	object	Customer registered address
cus_city	object	Customer registered city
cus_state	object	Customer state
cus_zip	int64	Customer ZIP code
Cus_latitude	float64	Customer Latitude Location
Cus_longitude	float64	Customer Longitude Location
Cus_job	object	Customer registered job
Cus_dob	object	Customer Date of Birth
transaction_no	object	Transaction ID Number
unix_time	int64	Transaction UNIX Time
merchant_latitude	float64	Merchant Latitude Location
merchant_longitude	float64	Merchant Longitude Location
isfraud	int64	Transaction Class (Fraud/non_Fraud)

3.4 Data Preprocessing

Data preprocessing is the process of preparing and cleaning raw data to make it ready for data mining and data analytic (Maharana, Mondal, & Nemade, 2022). It encompasses several steps including data cleaning, transformation of data, and data dimensionality reduction to help reduce inconsistencies and missing values in the dataset (Roy et al., 2021). Data preprocessing is one of the crucial steps in predictive analysis which can significantly impact the efficiency and performance of the resulting model (Khosla & Saini, 2020). While data preprocessing is not always required depending on the dataset being dealt with, it is necessary to perform basic level of data preprocessing such as handling missing values and outliers to improve the data quality and the accuracy of the analysis.

Some of the common techniques used in data preprocessing include:

- Data cleaning: This involves removing or correcting any errors, inconsistencies, or missing values in the data (Alexandropoulos, Kotsiantis, & Vrahatis, 2019).

- Data transformation: This involves converting data into a more appropriate format, such as normalizing or standardizing numerical data, converting categorical data into numerical data, or transforming data to reduce skewness or outliers (Luengo, García-Gil, Ramírez-Gallego, García, & Herrera, 2020b).

- Data reduction: This involves reducing the dimensionality of data by removing irrelevant features or combining similar features (Luengo, García-Gil, Ramírez-Gallego, García, & Herrera, 2020a).

- Data integration: This involves combining multiple datasets into a single dataset to create a more comprehensive view of the data.

In this research work, data cleaning is not required since the dataset used do not contains null or missing values. However, the dataset contains categorical features and variety of data types which required preliminary preprocessing. Therefore, some data preprocessing steps were performed to prepare the training and validation dataset including encoding categorical features and feature standardization.

3.4.1 Categorical features

Encoding categorical features is an important step in preparing data for ML models, as many ML algorithms require numerical data as input. There are several ways to encode categorical features, including one-hot encoding, label encoding, and ordinal encoding (Megdad, Abu-Naser, & Abu-Nasser, 2022). Label encoding were performed in this work to ensure dimensionality reduction of the dataset. Label encoding is a technique used to assign a unique integer value for each category. To ensure consistency between the training and testing dataset for categorical features, the two datasets were concatenated prior to the fitting and transforming of the labels.

3.4.2 Standardization

To enhance the ML algorithm's performance and attain maximum efficiency, feature standardization was performed on the training and testing dataset. Feature standardization is a technique used in ML to transform the features or variables in a dataset so that they have a mean of 0 and a standard deviation of 1 (Salekshahrezaee, Leevy, & Khoshgoftaar, 2023). The aim of feature standardization is to transform the input data so that it has a similar scale and distribution which can make the optimization process of ML algorithms faster and more stable. Feature standardization also preserve the shape of the original distribution of the data, which is important for ML

algorithms. To perform standardization for the numerical features, StandardScaler class from the Scikit-learn library is used.

3.5 Feature Selection

The process of feature selection involves identifying the relevant features from a dataset with respect to the target. The goal of feature selection is to enhance the accuracy and performance of the model avoiding the use of irrelevant or redundant features (Anirudh & Talukdar, 2022). Initially, Exploratory Data Analysis (EDA) is performed to analyze the characteristics of the features within the dataset. In addition, Feature correlation is utilized to eliminate features with high correlation and decrease redundancy in order to enhance the model's performance. Figure 3.4 illustrate the correlation matrix of the original dataset.

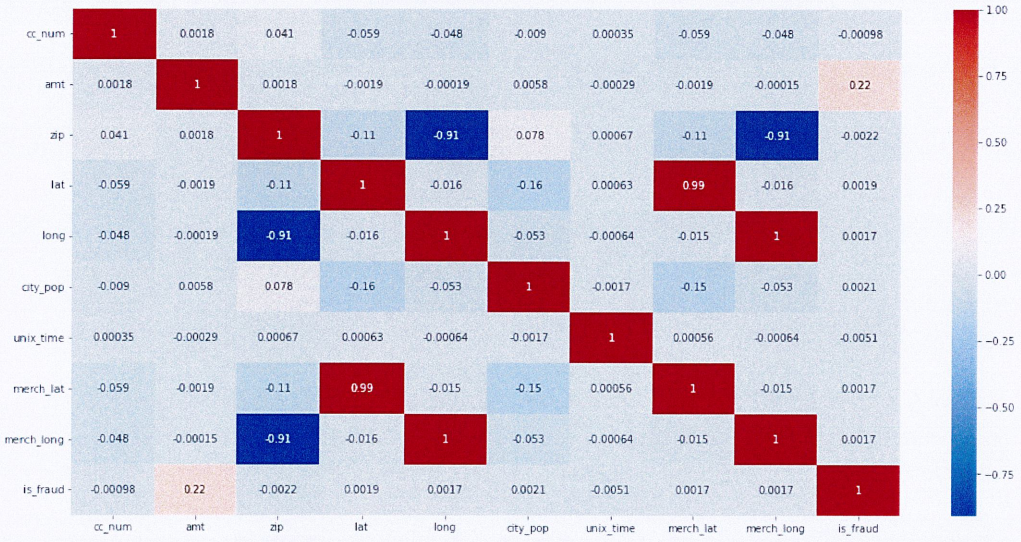


Figure 3.4: Dataset Correlation Matrix Heat map.

Variant feature selection techniques were performed including variance threshold, Pearson correlation coefficient, information gain (Mutual Information), and Lasso technique.

3.6 Class Imbalance

Class imbalance in credit card transactions fraud detection is a common issue, since only a small proportion of the volume of transactions are fraudulent transactions (Dal Pozzolo et al., 2017). The majority of available datasets hold less than 1% of fraudulent transactions which makes ML classifiers produce inaccurate performance on the fraud transactions class. Since the essence of the problem necessitate prioritizing the minority class, handling class imbalance acquired an increased attention and numerous methods to address the class unbalance have been proposed in the literature (Feng et al., 2022; Zhang et al., 2023).

3.6.1 Re-sampling Methods

One way to address the class imbalance is through resampling techniques including undersampling and oversampling. Resampling techniques comprises modifying the dataset to even out the numbers in each class. Undersampling involves removing samples from the higher class to produce an even class distribution within the dataset. Conversely, oversampling comprises adding additional similar samples to the minority class to level the distribution. The choice of which approach to use depends on the size, characteristics of the dataset, and the analysis objective (Tran & Dang, 2021).

Both techniques oversampling and undersampling have their pros and cons and can be employed using different techniques. Oversampling can result in overfitting the ML model to the minority class. On the other side undersampling can result in losing information and useful interpretation from the dropped instances. In term of the techniques used, undersampling and oversampling can be used in a random fashion (Mrozek, Panneerselvam, & Bagdasar, 2020). However, in scenarios where randomness is not preferred, some more sophisticated techniques can be used such as Tomek links, Clus-

ter Centroids, and Synthetic Minority Over-sampling Technique (SMOTE). In addition, in some investigations a combination of both undersampling and oversampling are used to preserve the information of the dataset (Udeze, Eteng, & Ibor, 2022). Figure 3.5 demonstrates the class distribution of both datasets that were used in this work.

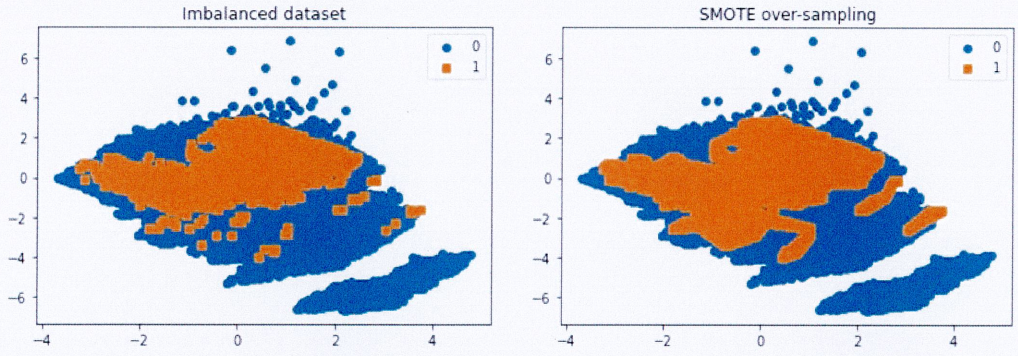


Figure 3.5: Class Distribution for Imbalanced Dataset and Oversampled Dataset.

3.6.2 Cost-sensitive Methods

Another method to address class imbalance in a dataset is to employe cost-sensitive technique on the used ML model. Cost-sensitive learning methods do not require balancing the dataset. It involves assigning different misclassification costs to different classes during model training (Padurariu & Breaban, 2019). When dealing with CCF detection, if classifying a fraudulent instance as normal is more costly than classifying a legitimate transaction as fraudulent. Therefore, a cost-sensitive learning approach would assign a greater cost for misclassifying to the fraudulent class compared to the legitimate class.

3.7 Evaluation Metrics

Choosing the right evaluation metrics to assess the performance of a ML model require cautious consideration to best determine the best fitted evaluation metrics for the investigated problem (Banerjee, Bourla, Chen, Kashyap, & Purohit, 2018). When dealing with imbalanced dataset, the accuracy by itself can be very misleading and just give an illusion of the actual performance. In addition, any classifier which always give a non-fraudulent result will achieve accuracy of 99.5% on the given dataset without the need to perform feature analysis (Itoo & Singh, 2021). The later is called the metric trap which refers to the potential pitfall of relying solely on a single evaluation metric to assess the ML model's performance. Evaluation metrics including accuracy, recall and precision are integral measures for assessing the efficiency of a ML model. Yet, these metrics might not provide a comprehensive analysis of how a model has performed (Alfaiz & Fati, 2022).

In respect to fraud detection problem, many suggest that false negatives (categorizing fraudulent transactions as non-fraudulent) is the key to modelling an algorithm due to the cost of fraud transactions and the need to prevent fraudulent activities from passing the detection system. However, financial loss due to fraud transactions are not less risk and cost than false positives (categorizing non-fraudulent transactions as fraudulent). Depending on the cost of loss to organizations, one can prefer a model with high precision where false positives occurrence is more costly than the cost of fraud transactions. On the other hand, if the cost of false negatives exceeds that of false positive, a model with high recall score is more preferred.

In this work, to best evaluate the efficiency of the models used for fraud detection, three evaluation metrics were selected including confusion matrix, ROC and the Precision-Recall Curve. These metrics are particularly useful when the dataset is imbalanced.

3.7.1 Confusion Matrix

A confusion matrix is a chart utilized to appraise the effectiveness of a ML model's performance in a binary classification task. It provides a summary of the predicted and actual classifications of the model, which can be used to calculate various evaluation metrics. The confusion matrix has four cells, which represent the four possible outcomes of a binary classification:

- True Positive (TP): the model predicted a positive class and the actual class was also positive.
- False Positive (FP): the model predicted a positive class, but the actual class was negative.
- True Negative (TN): the model predicted a negative class and the actual class was also negative.
- False Negative (FN): the model predicted a negative class, but the actual class was positive.

3.7.2 Receiver Operating Characteristic

Receiver Operating Characteristic (ROC) curve is a visual display of the efficacy of a binary classification model, which demonstrates the balance between the true positive rate (sensitivity) and false positive rate (1-specificity) for various thresholds. The true positive rate indicates the fraction of real positive samples that the model correctly identifies as positive, while the false positive rate denotes the fraction of actual negative samples that the model erroneously categorizes as positive. An ideal classifier would attain a true positive rate of 100% and a false positive rate of 0%, manifested by a point at the upper-left edge of the ROC curve. In practice, the ROC curve generally illustrates a balance between the true positive rate and false positive rate, whereby increasing one comes at the expense of the other.

3.7.3 Precision-Recall Curve

Precision-Recall (PR) curves are a visual representation of the trade-off between precision and recall for a machine learning model. They are commonly used to evaluate binary classification models, particularly in cases where the dataset is imbalanced or the cost of false positives and false negatives is not equal. In a PR curve, precision is plotted against recall for different thresholds of the model's output score. Precision is the ratio of true positives to all predicted positives, while recall is the ratio of true positives to all actual positives. A perfect classifier would achieve both 100% precision and 100% recall, which would be represented by a point at the top right corner of the PR curve. Typically, there is a trade-off between precision and recall in practical scenarios, such that increasing one comes at the expense of the other.

3.8 Machine Learning Algorithms

Selecting the right ML model is crucial to ensure the effectiveness of a fraud detection system that can accurately and efficiently classify fraudulent transactions from genuine ones. In this work, we evaluate and compare three popular algorithm that are best suited for the underlying problem which are Random Forest (RF), Extreme Gradient Boost (XGBoost), Support Vectore Machine (SVM). The selected algorithms were initially evaluated on the original dataset. The evaluation also encompasses testing the model on the resampled balanced dataset. In the end, Cross-validation techniques is performed to assess how well the model will perform on new data it has not been trained on, also known as generalization performance.

3.8.1 Random Forest

As a baseline model, Random Forest (RF) is selected for the fraud detection model. RF is a popular ensemble supervised learning algorithm that is applicable to both classification and regression problems. It's core based on the traditional decision trees which are combined to create more powerful model, where each tree is trained on a fraction of the data and a random subset of features, and then the results are aggregated to make the final prediction (Tamtama & Arifudin, 2022).

During the training process, each decision tree predict the target variable by evaluating the chosen features. The ultimate prediction is performed by aggregating the outputs of all decision trees in the forest through voting process or averaging process depending on the type of task. Figure 3.6 demonstrate a simple overview of the ensemble trees utilized in RF algorithm. RF is known for its robustness and capacity to handle high-dimensional data. Further, it can estimate the importance of each feature which can be used to select most important features to the model.

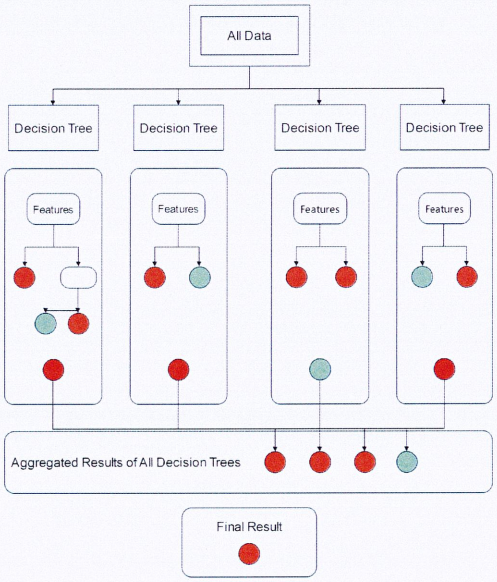


Figure 3.6: Random Forest Prediction Overview.

3.8.2 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boost also known as XGboost is an extended implementation of gradient boosting algorithm that is widely used in both the industry and academia due to its efficiency, scalability, and performance. XGBoost is a flexible algorithm that can be used for a wide range of supervised learning applications including regression, classification, and ranking. It includes with additional features designed to improve the performance such as regularization, pruning, and parallel processing which helps in overfitting. In addition, XGBoost automatically handle missing values in the dataset (P. Gupta et al., 2023).

XGBoost uses an ensemble of decision trees to make predictions. Sequential trees, also known as gradient boosted trees which are a key component of XGBoost. In XGBoost, each tree is trained sequentially instead of training each tree independently, with each subsequent tree attempting to correct the errors of the previous tree. The output of each tree is added to the predictions of the previous trees, creating a strong ensemble model. The sequential nature of the trees also allows XGBoost to handle non-linear relationships between features and the target variable (Ashfaq et al., 2022).

3.8.3 Support Vector Machine

Support Vector Machine (SVM) is a ML algorithm utilized for classification and regression analysis that determines the optimal hyperplane to divide the data into distinct classes. (Saheed, Hambali, Arowolo, & Olasupo, 2020). SVM is a powerful algorithm that can handle both linearly separable and nonlinearly separable data by using kernel functions that map the data into a higher-dimensional space to simplify the process of classification. SVM has some advantages over other classification algorithms, such as its ability to handle high-dimensional data, its ability to handle small datasets, and

its robustness to overfitting (S. Kumar et al., 2022). However, SVM also has some limitations, such as its sensitivity to the choice of kernel function and hyperparameters, its high computational complexity, and its difficulty in handling noisy and overlapping data. The most popular kernel functions include linear, polynomial, Radial Basis Function (RBF), and sigmoid kernels (Almaiah et al., 2022).

- Linear kernel: This is the simplest kernel function and is used for linearly separable data. It is defined as the dot product of the input features.
- Polynomial kernel: This kernel function is used for data that is not linearly separable. It maps the data into a higher-dimensional space using a polynomial function. The degree of the polynomial can be adjusted using a hyperparameter.
- Radial Basis Function (RBF) kernel: This is a popular kernel function that maps the data into an infinite-dimensional space using a Gaussian function. It is commonly used for nonlinearly separable data and has two hyperparameters: gamma and C.
- Sigmoid kernel: This kernel function is used for data that is not linearly separable and maps the data into a higher-dimensional space using a sigmoid function. It has two hyperparameters: gamma and C.

3.9 Hyper-Parameter Tuning

Hyper-Parameter tuning refers to the process of selecting the best set of hyper-parameters for a machine learning algorithm to improve its performance. Hyper-Parameter tuning is a crucial step in the machine learning pipeline to optimize the model's performance. There are various techniques

to find the best hyper-parameters including grid search, random search, and Bayesian optimization. In this research, we implement the use of different techniques depending on the ML algorithm. To find the optimal hyper-parameters, the hyper-parameter tuning is combined with cross validation using Stratified K-Fold to ensure identify optimal hyper-parameters that work well on multiple validation sets (Tayebi & El Kafhali, 2021).

3.9.1 Random Forest Hyper-Parameters Selection

In RF model, HalvingRandomSearchCV technique is used. HalvingRandomSearchCV is a hyper-parameter optimization technique that combines elements of both random search and iterative halving. It is a variant of randomized search, which is a type of hyper-parameter optimization algorithm that randomly samples hyper-parameters from a given distribution and evaluates their performance on a validation set. However, HalvingRandomSearchCV goes one step further by iteratively halving the data and the hyper-parameter search space to focus on the most promising candidates, making it more computationally efficient than traditional random search (A. A. Ali, Khedr, El-Bannany, & Kanakkayil, 2023).

3.9.2 XGBoost Hyper-Parameters Selection

In XGBoost model, Bayesian optimization technique is used through the BayesSearchCV within the scikit-learn library. BayesSearchCV is a hyper-parameter optimization algorithm that search for the optimal set of hyperparameters. It is a more efficient alternative to grid search and random search, as it uses a probabilistic model to guide the search towards promising regions of the hyperparameter space (Huang, 2020).

3.9.3 Support Vector Machine Hyper-Parameters Selection

In SVM model, to find the optimal kerner and the C value, GridSearchCV is used. GridSearchCV is a technique for hyperparameter tuning in machine learning. It is a simple but exhaustive approach to search over a predefined set of hyperparameters and find the best combination that maximizes the performance of a machine learning model (Chugh & Malik, 2022).

Chapter 4

Findings and Results

4.1 Introduction

In this chapter, the results of our study on CCF detection is presented. The performance of each model on the test dataset is demonstrated using classification report, confusion matrix, the ROC curve plot and the precision-recall curve plot. Moreover, the strength and limitation of each approach is discussed. Finally, insights into the factors that contribute to fraudulent activities in electronic funds transfer transactions are highlighted.

4.2 Random Forest Results

This section showcases the results of the experiments that utilized the RF algorithm. These include evaluations on both the original and oversampled datasets, along with the outcomes of cross-validation.

4.2.1 Original Dataset

Initially, the RF model performance is evaluated using the original dataset without any resampling techniques. Figure 4.1 illustrates the four evaluation metrics chosen to validate the performance of the model. From the results we can see that the model performed perfectly in classifying the legitimate transactions, with only 89 instances reported as false positives out of over 500,000. In term of fraudulent classification, the model achieved an acceptable precision score of 94.76%, but the model had deficiencies in its recall result and achieved a PR AUC score of 88.1 as shown in Table 4.1. Moreover, the model attained an ROC AUC score of 98%.

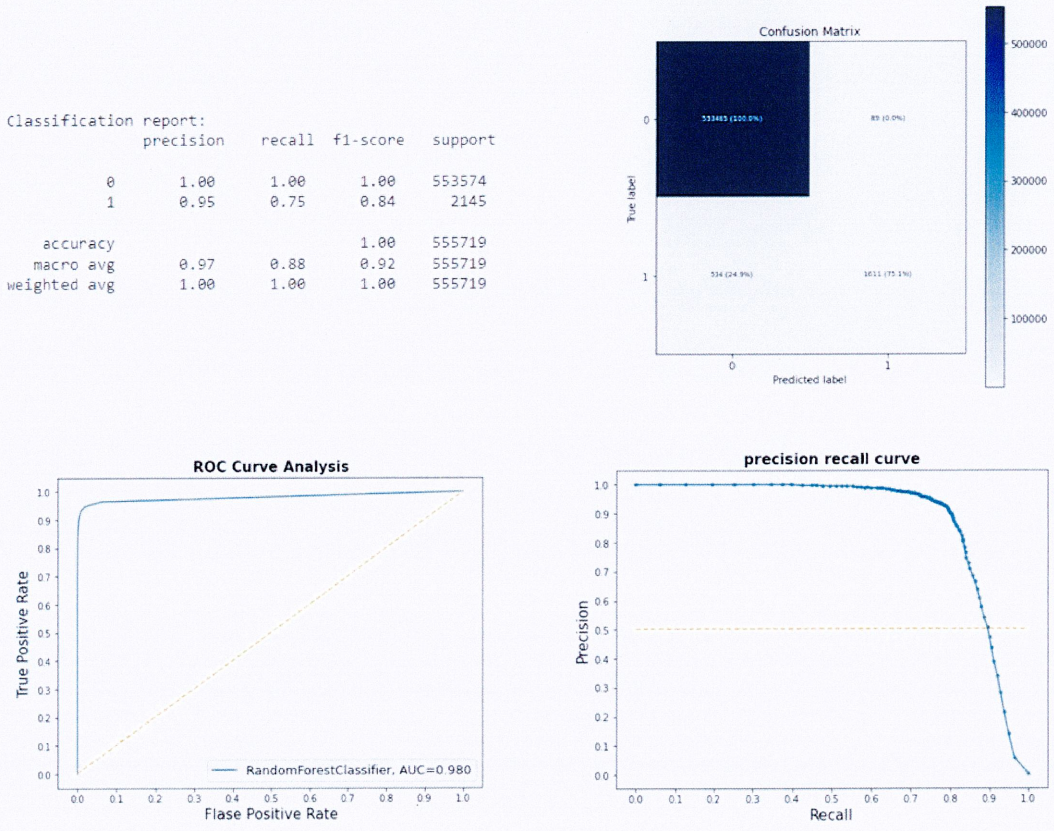


Figure 4.1: Random Forest Original Dataset Results.

Table 4.1: Random Forest Original Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
Random Forest	94.76	75.10	83.79	98	88.1	267.86

4.2.2 Balanced Dataset using SMOTE

Since the dataset is imbalanced and the legitimate instances make the majority proportion of the dataset, it is unreliable to validate the performance of the ML model since it may be biased toward the majority class and perform irrespectively to the minority class which was observed in the previous results. As a result, the model may learn to predict the majority class more accurately and neglect the minority class, leading to a high false negative rate or false positive rate for the minority class.

Thus, in this attempt, the test portion of the dataset has been balanced using SMOTE technique to create synthetic samples of the fraudulent instances. In addition, the training set has been resampled using SMOTE. Figure 4.2 present the results of the model using the balanced dataset. From the results we can see the model managed to achieve a promising and improved ROC AUC and PR AUC scores of 98.8 and 99.1 respectfully as shown in Table 4.2. Furthermore, the recall score increased to 82% in predicting fraudulent class. Nevertheless, the rise in the model's recall score came with a cost as the precision score for the legitimate class decreased to 85%. Furthermore, the model failed to detect a considerable number of fraudulent transactions, amounting to 33,707 instances.

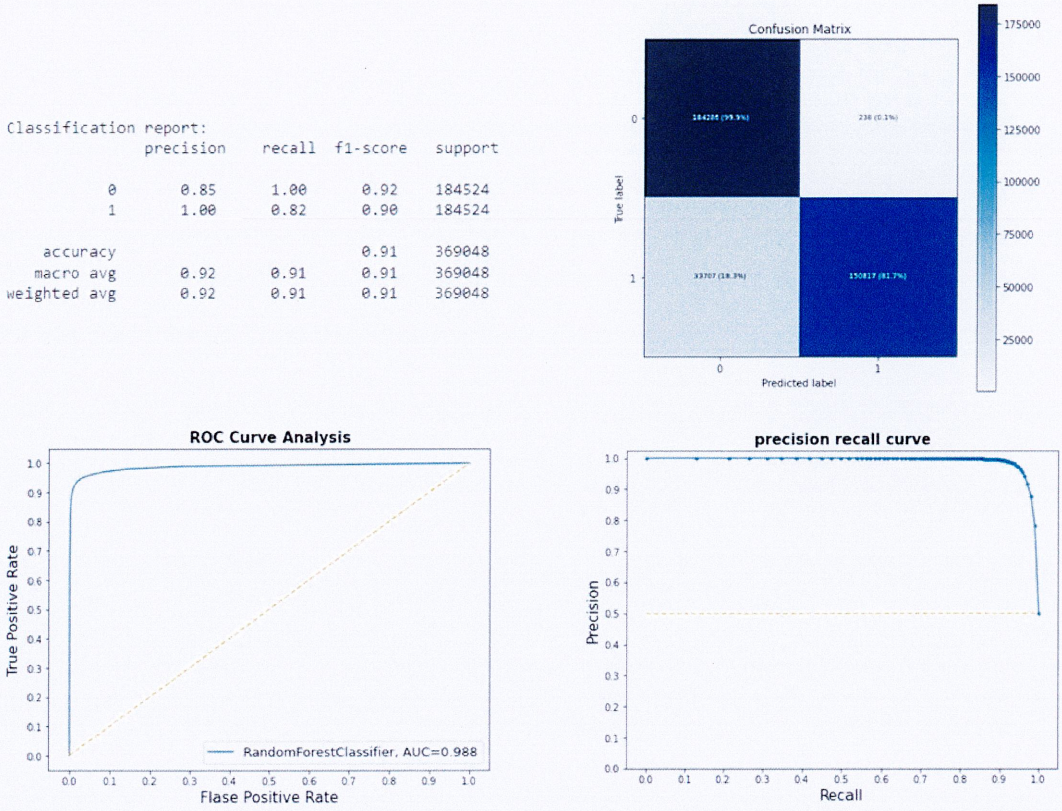


Figure 4.2: Random Forest Resampled Dataset Results.

Table 4.2: Random Forest Resampled Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
Random Forest	99.84	81.73	89.88	98.8	99.1	280.37

4.2.3 Cross Validation

To evaluate the model's performance comprehensively, cross validation is performed to ensure the model is not over-fitting from the given training and testing proportion. The training and testing datasets were concatenated

into single dataset to perform the cross validation. Stratified K-fold with 5 splits is used to ensure that each fold of the cross-validation process has a representative sample of the fraudulent transactions data, and that each fold has a similar proportion of the target class labels as the whole dataset. Table 4.3 presents the result of the cross validation of the RF model.

Table 4.3: RF Model Performance

Index	Fit Time	Score Time	Accuracy	Precision	Recall	F1-Score
0	290.46	4.25	0.998	0.962	0.782	0.863
1	292.12	4.27	0.998	0.974	0.765	0.857
2	293.58	4.11	0.998	0.972	0.779	0.865
3	290.86	4.53	0.998	0.969	0.765	0.855
4	291.78	4.20	0.998	0.970	0.788	0.869
Mean	291.14	4.29	0.998	0.969	0.776	0.862

4.3 Extreme Gradient Boosting Results

The following section demonstrates the results of several experiments utilizing the XGBoost algorithm. These experiments involved evaluating both the original and oversampled datasets, as well as performing cross-validation.

4.3.1 Original Dataset

The evaluation of XGBoost on the original dataset has revealed enhancements in several evaluation metrics. Figure 4.3 demonstrates that the classification report for the fraudulent class exhibited a higher precision and recall, specifically 96% and 82%, respectively. This led to an improvement in the PR AUC, resulting in a score of 94.1 as shown in Table 4.4. Additionally, the model attained an ROC AUC score of 99.7. Despite obtaining

better overall results, the model's recall outcome remains unsatisfactory and the rate of false negative is still high.

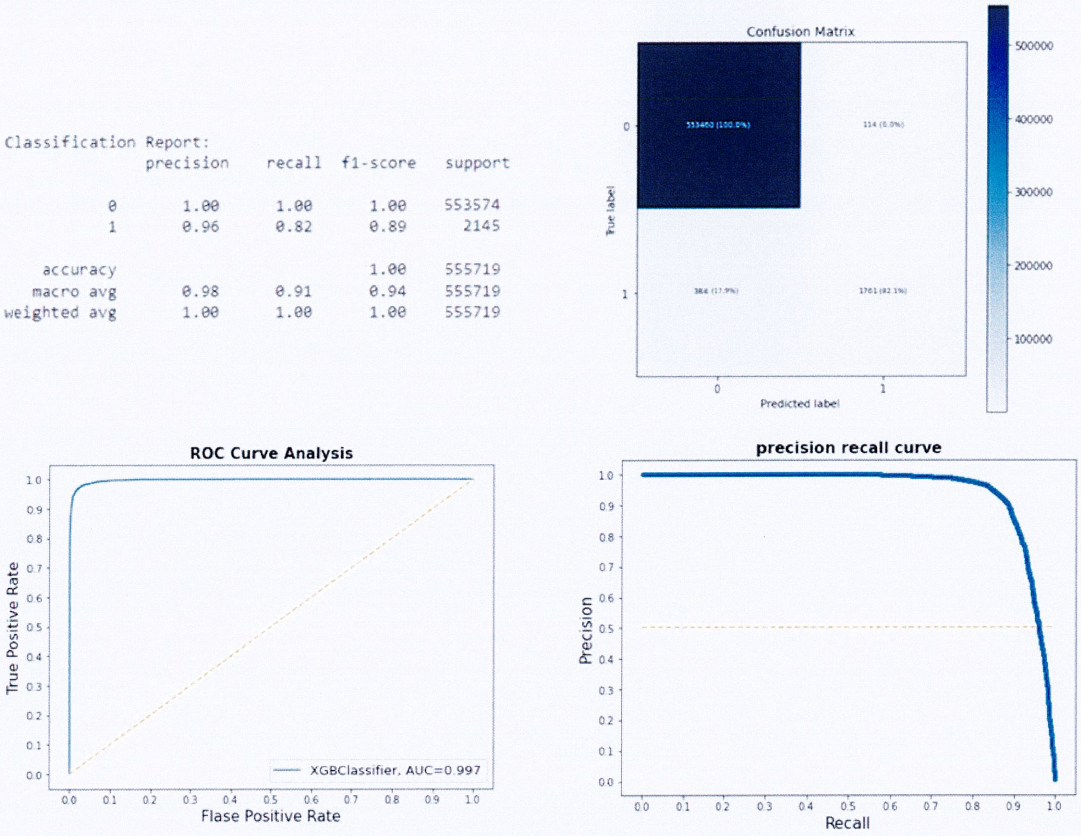


Figure 4.3: XGBoost Original Dataset Results.

Table 4.4: XGBoost Original Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
XGBoost	95.92	82.14	89.50	99.7	94.1	277.76

4.3.2 Balanced Dataset using SMOTE

To further validate the model result, the model was evaluated using a balanced dataset. To balance the test dataset, the dataset was split into third to

avoid heavy processing. From the results demonstrated in Figure 4.4 and Table 4.5, we can see tremendous improvements across all evaluation metrics. The model attained a score exceeding 99% for recall, and F1-score. In addition, the model identified only 195 instances as false positives which indicates the model has sustained high precision score while improving the recall score. Regarding the ROC AUC, the model reached a flawless score of 1, while the PR AUC score escalated to 0.994.

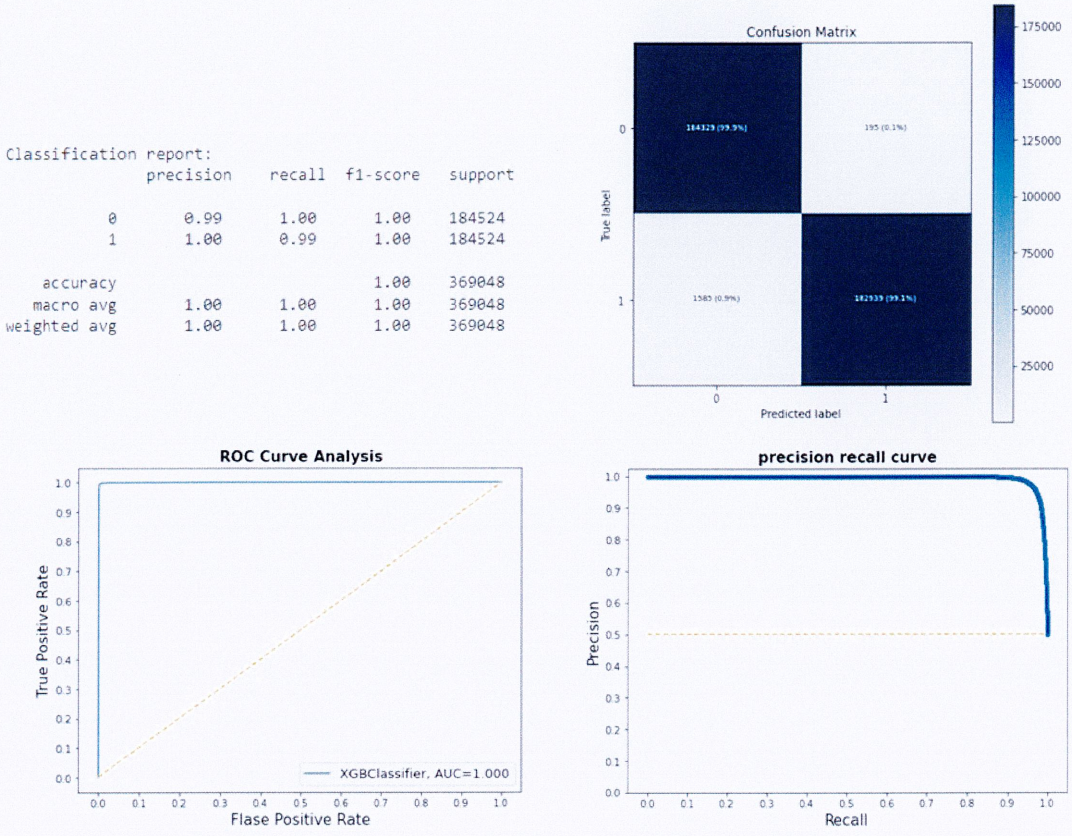


Figure 4.4: XGBoost Resampled Dataset Results.

4.3.3 Cross Validation

The results of the cross-validation of the XGB model are presented in Table 4.6. The results demonstrate consistent performance of the XGBoost model with a mean accuracy of 99.9%. The mean precision, recall and F1-score

Table 4.5: XGBoost Resampled Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
XGBoost	99.89	99.14	99.51	100	99.4	299.03

of the model is 95.3%, 82.2%, and 88.4%, respectively. The model also exhibits notable efficiency in terms of fitting and scoring time.

Table 4.6: XGBoost Model Performance

Index	**Fit Time**	**Score Time**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
0	5.56	0.63	0.999	0.955	0.822	0.884
1	1.69	0.62	0.999	0.961	0.814	0.881
2	1.77	0.62	0.999	0.953	0.819	0.881
3	1.77	0.65	0.999	0.946	0.829	0.884
4	2.62	1.05	0.999	0.955	0.829	0.888
Mean	2.68	0.71	0.999	0.953	0.822	0.884

4.4 Support Vector Machine Results

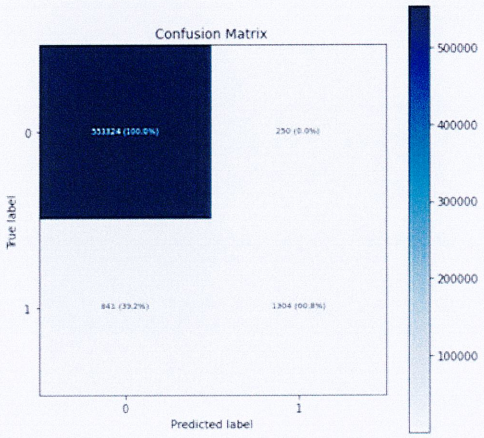
In this section, the results of different experiments utilizing the SVM algorithm are presented. To determine the optimal kernel and hyperparameters for the model, GridSearchCV was performed on the used dataset with 5-fold cross validation. Based on the result of the GridSearchCV, the RBF kernel is selected for this experiment. Similarly, the experiment incorporated evaluating both the original and oversampled datasets. Additionally, cross-validation was performed to estimate the average performance of the model.

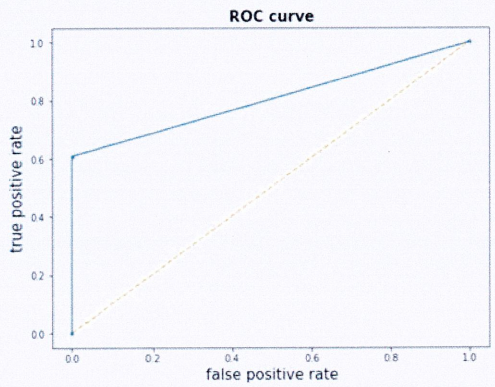
4.4.1 Original Dataset

The results in Figure 4.5 and Table 4.7 show the performance of SVM algorithm that was examined on the original dataset. Although the model demonstrated competent performance in categorizing regular transactions, its performance in detecting fraudulent transactions revealed its inadequacy for fraud detection as shown from the classification report as well as the number of false negative reported. The performance results also indicate that the SVM algorithm is not proficient in dealing with imbalanced dataset. The ROC AUC and PR AUC scores are 0.804 and 0.724 respectively which clearly shows the performance deficiency of the model. Thus, the model is further assessed using a balanced dataset in the next experiment.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	553574
1	0.84	0.61	0.71	2145
accuracy			1.00	555719
macro avg	0.92	0.80	0.85	555719
weighted avg	1.00	1.00	1.00	555719





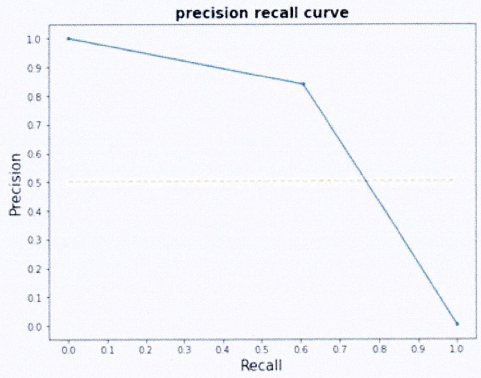


Figure 4.5: SVM Original Dataset Results.

Table 4.7: SVM Original Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
SVM - RBF Kernel	83.91	60.79	70.50	80.4	72.4	225.10

4.4.2 Balanced Dataset using SMOTE

Figure 4.6 and Table 4.8 presents the scores of the SVM model using the resampled datasets which shows better results in term of precision and recall reported. Furthermore, The PR AUC score has significantly increased to achieve a score of 0.908. However, the improvements in fraudulent classification comes along with a sharp decrease in the precision score for the genuine transactions. The ROC AUC score didn't attain any noticeable improvement and reported an area of 0.816.

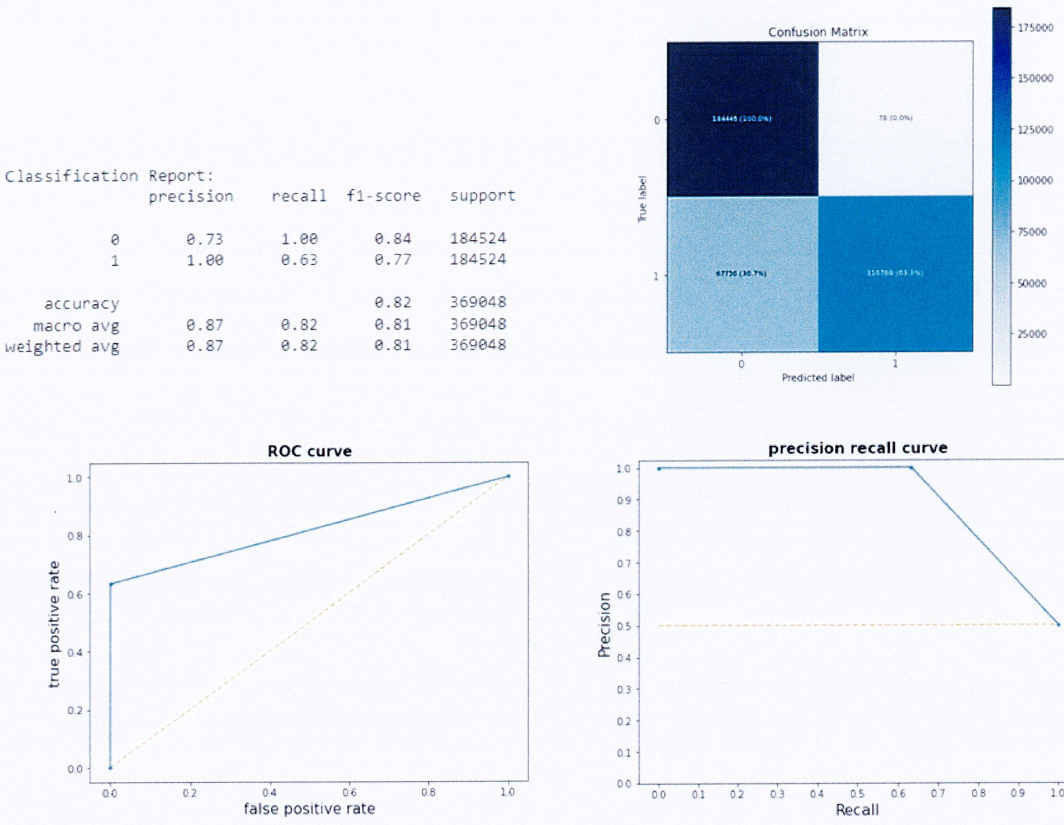


Figure 4.6: SVM Resampled Dataset Results.

Table 4.8: SVM Resampled Dataset Results

Model	Precision	recall	F1-score	ROC AUC	PR AUC	Sort weight
SVM - RBF Kernel	99.93	63.28	77.49	81.6	90.8	244.81

4.4.3 Cross Validation

The results in Table 4.9 present the SVM model cross validation performance. The metrics include the time taken to fit the model, the time taken

to score the model, the accuracy, precision, recall, and F1-score. The model achieved consistent results among all the folds with mean accuracy of 99.7%. The mean precision and recall of the model are 0.908 and 0.620, respectively. While the mean F1-score, which is a harmonic mean of precision and recall, is 0.737.

Table 4.9: SVM Model Performance

Index	Fit Time	Score Time	Accuracy	Precision	Recall	F1-Score
0	101.48	1.49	0.997	0.920	0.606	0.731
1	99.83	1.05	0.997	0.902	0.618	0.733
2	102.36	1.04	0.997	0.893	0.631	0.739
3	102.26	1.49	0.997	0.912	0.625	0.742
4	102.71	1.12	0.997	0.917	0.623	0.742
Mean	101.94	1.23	0.997	0.908	0.620	0.737

4.5 Comparative Analysis

Based on the evaluation results of our model, the ensemble tree algorithms demonstrated superior performance compared to the SVM algorithm. Table 4.10 presents the performance of each model's Accuracy (ACC), True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), and False Negative Rate (FNR). The results show that all models have very low FPR (0.2%) and high TNR (99.8%), indicating that they are effective at correctly identifying non-fraudulent transactions. However, the XGBoost model on the balanced dataset has the highest TPR (99.1%) and the lowest FNR (0.9%), outperforming the other models. On the original dataset, RF and XGBoost have high specificity (99.9%) and negligible false alarm rate (Fall-out) (0.1%) which makes them suitable for certain contexts where false positives are particularly costly. But have considerably high FNR com-

pared to their balanced dataset counterparts. SVM on the balanced dataset has the lowest accuracy (81.61%) and the highest FNR (36.8%). Overall, the results suggest that balancing the dataset improves the performance of the models.

Table 4.10: All Experiment Results

Model	Resampling Technique	ACC	TPR	FPR	TNR	FNR
RF	No Resampling	99.88%	0.751	0.001	0.999	0.249
RF	SMOTE	90.80%	0.817	0.002	0.998	0.183
XGB	No Resampling	99.88%	0.821	0.001	0.999	0.179
XGB	SMOTE	99.51%	0.991	0.001	0.999	0.009
SVM	No Resampling	99.80%	0.607	0.001	0.999	0.393
SVM	SMOTE	81.61%	0.632	0.001	0.999	0.368

Figures 4.7 and 4.8 illustrates the performance results in term of accuracy, precision, recall and F1-score on both the original dataset and resampled dataset. Similarily, XGBoost outperforms RF and SVM in terms of precision, recall, and f1-score, indicating that XGBoost has better classification performance overall.

On the contrary, even though balancing the dataset through resampling slightly reduces the accuracy of the model, it leads to better outcomes in terms of precision, recall, and f1-score.

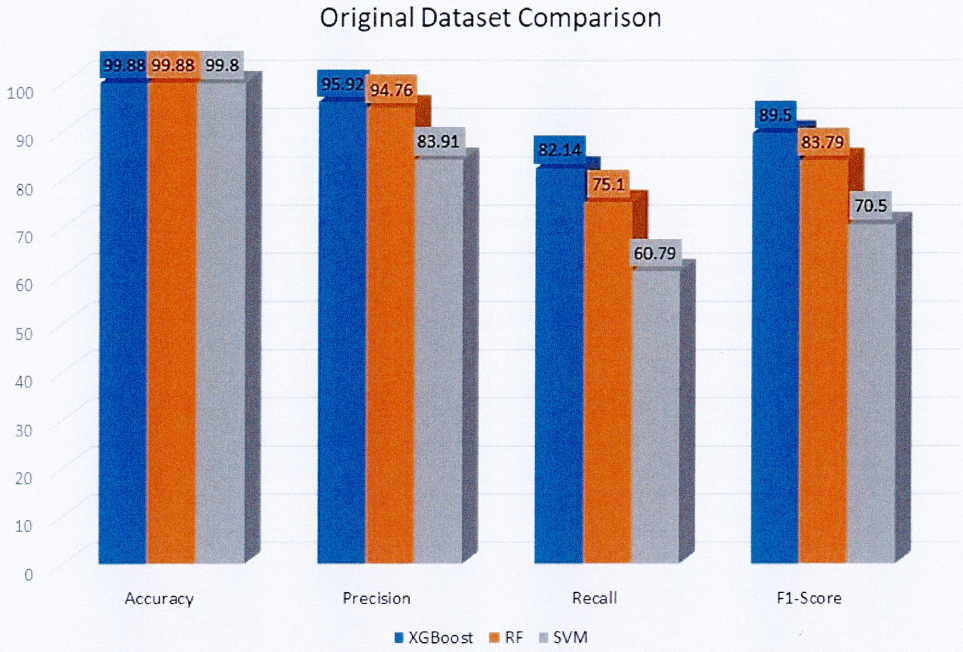


Figure 4.7: Original Dataset Results Comparison.

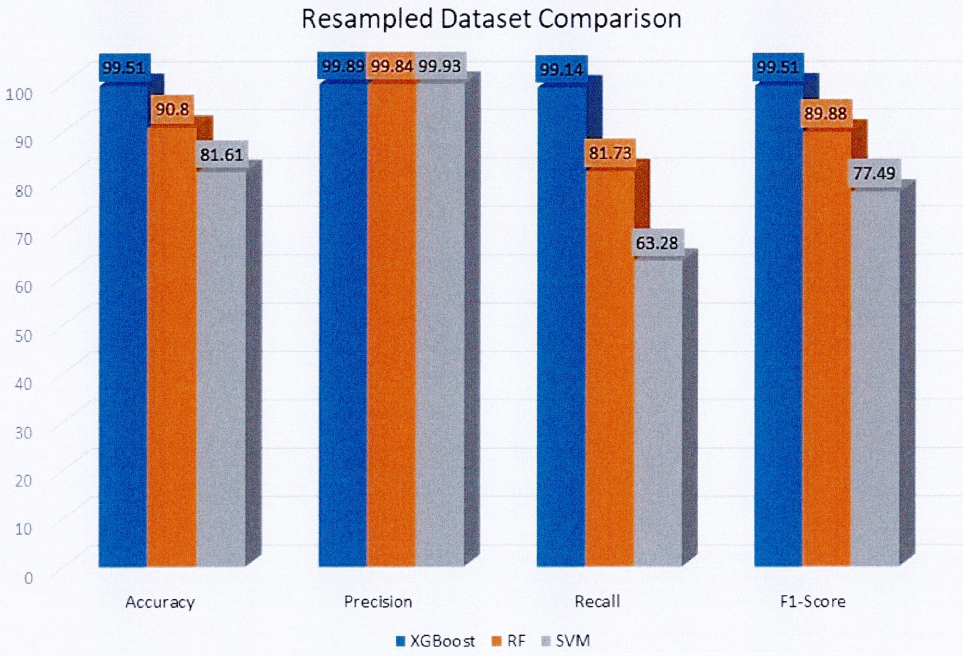


Figure 4.8: Resampled Dataset Results Comparison.

Finally, it is important to compare the models in terms of the ROC AUC

and PR AUC scores to provide a more comprehensive evaluation of the performance of a classification model beyond simple accuracy. ROC curves plot the TPR against the FPR at different classification thresholds. ROC curves are typically used when the class distribution is balanced and where both the false positives and false negatives costs are approximately the same.

On the other hand, PR curves, on the other hand, plot the precision (positive predictive value) against the recall (sensitivity) at different classification thresholds and it is typically used when the class distribution is highly imbalanced and when the costs associated with false positives and false negatives are not equal. Since the comparison includes both unbalanced and balanced dataset, we have used both ROC curves and PR curves to present a complete picture of the model performance. Figure 4.9 demonstrates the ROC Curves for all the models and Figure 4.10 shows the PR Curve results for all the experiments.

XGBoost again outperforms RF and SVM on both datasets, and achieved a perfect ROC AUC score of 1 and a PR AUC score of 0.994 with the balanced dataset. Random Forest has a good ROC AUC score of 0.988 and PR AUC score of 0.991, while SVM has the lowest ROC AUC score of 0.816 and PR AUC score of 0.908.

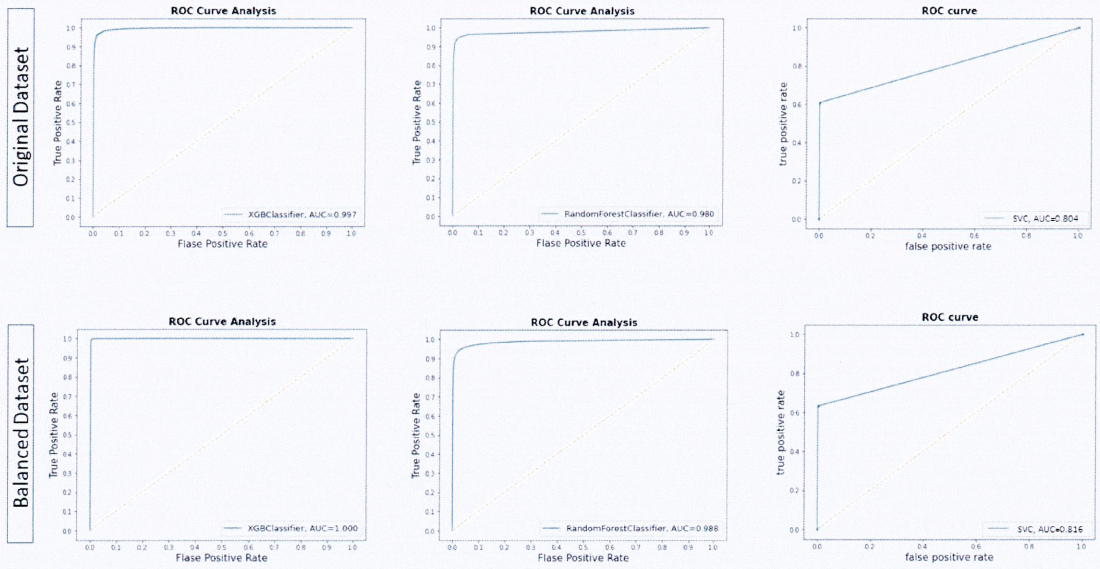


Figure 4.9: ROC AUC Results Comparison.

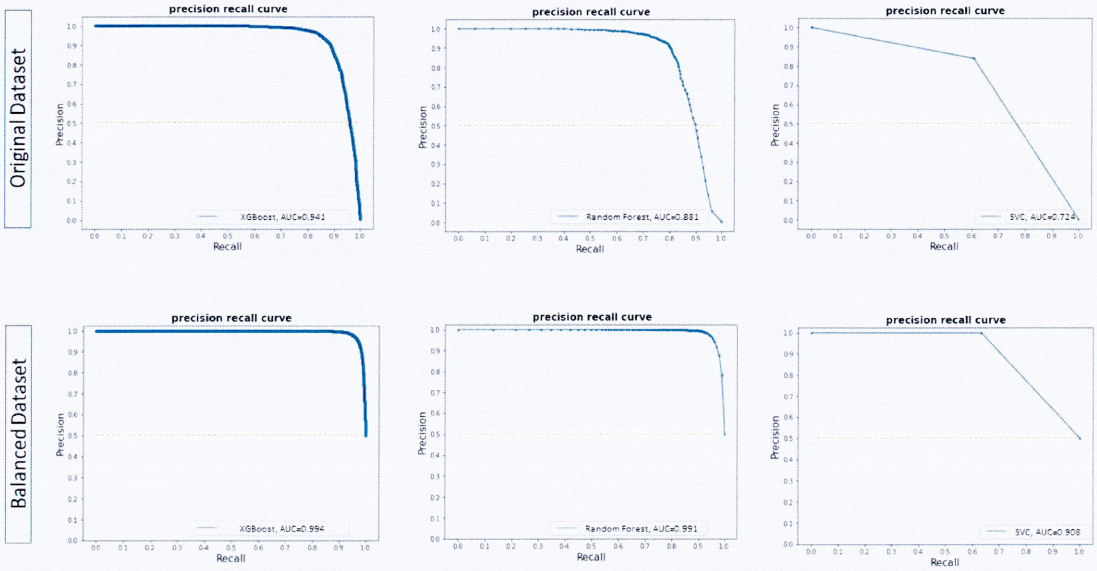


Figure 4.10: PR AUC Results Comparison.

Overall, the XGBoost model is the most effective at detecting CCF on the balanced dataset, with the highest accuracy, precision, recall, f1-score, TPR, ROC AUC, and PR AUC scores among the three models. The RF model is also relatively effective, but its recall is lower than XGBoost. The

SVM model is the least effective among the three, with the lowest accuracy, recall, f1-score, TPR, ROC AUC, PR AUC.

4.5.1 Time Efficiency

One of the most disregarded criteria in evaluating ML models for fraud detection is the time efficiency. Table 4.11 presents the training and scoring time consumed for 10 thousand rows for each of the three models. The results is collected on a google colab environment running Intel Xeon CPU @ 2.00GHz, 12.7 GB RAM, and Tesla T4 16 GB RAM GPU.

Table 4.11: All Experiment Results

Model	**Fitting Time (10000 row)**	**Scoring Time (10000 row)**
RF	1964.85 ms	115.89 ms
XGB	180.64 ms	19.19 ms
SVM	688.01 ms	33.23 ms

The results show that XGBoost has achieved the best time which makes it the most effective and efficient model of all three. Furthermore, the low required scoring time XGBoost makes it suitable for real-time implementation and allow for quick analysis of incoming transactions as it is received.

4.5.2 Comparison with Previous Works

Based on the evaluation metrics and experimental results, it can be confidently stated that the best proposed model is the XGBoost model along with SMOTE oversampling. To put this into perspective, Table 4.12 provides a comparison between the proposed model and previous works that utilized

similar approaches, giving us a better understanding of how the proposed model performs in relation to existing models.

Table 4.12: Comparison with State of the Art

Reference	Dataset	Model	ROC AUC
(Niu et al., 2019)	European Cardholder dataset	RF	0.988
(Meng et al., 2020)	European Cardholder dataset	XGBoost+SMOTE	0.985
(Taha & Malebary, 2020)	European Cardholder dataset - UCSD-FICO	OLightGBM	0.928
(Alarfaj et al., 2022)	European Cardholder dataset	CNN	0.980
(Hajek et al., 2022)	PaySim dataset	XGBoost+RUS	0.996
(Alfaiz & Fati, 2022)	European Cardholder dataset	Catboost+AIIKNN	0.979
(Ileberi et al., 2022)	European Cardholder dataset	RF+GA	0.950
(Hsin et al., 2022)	National Police Agency transactions data	XGBoost+WGAN	0.969
(Esenogho et al., 2022)	European Cardholder dataset	LTSM+SMOTE-ENN	0.990
This Study	Simulated dataset	XGBoost+SMOTE	**1.00**

The proposed model is outperforming other proposed models in term of ROC AUC. Furthermore, the model not only achieved high detection accuracy, but it also demonstrated great robustness and reliability in term of TPR and FPR. The algorithm and the methods employed in the model was able to effectively capture the underlying patterns and dependencies in the dataset, leading to better generalization and more accurate predictions. Overall, the proposed model showed a significant improvement over existing models and

maintained a low FPR while increasing the recall score, making it a promising solution for the fraud detection problem.

Chapter 5

Conclusion, Recommendations and Future Work

This chapter summarizes the main findings of the study, the contributions to the literature, and the implications for practice. It also provides recommendations for improving fraud data mining and fraud detection as well as future work summary.

5.1 Conclusion

The world sudden and tremendous digitization has simulated exponential growth and rapid acceleration toward digital payments and online transactions with growth forecasts continue to rise globally. The successful evolution of digital payment however is impeded by ensuring an adequate system in place to combat fraud. In this work, we proposed a real-time fraud detection model for electronic funds transfer transactions. The model utilized the use of the most widely used classification supervised learning algorithms

namely Random Forest, Extreme Gradient Boosting, and Support Vector Machine. Each model is then validated using the original dataset, the resampled dataset, and cross-validation. The evaluation of the model has taken a holistic approach. Thus, the assessment incorporated the use of several metrics including the confusion matrix, ROC AUC, PR AUC, precision, recall and F1-score. A comparative analysis is conducted to select the best model which embraced different metrics such as specificity, FPR, FNR, and time efficiency. The proposed XGBoost model with SMOTE oversampling technique outperformed all the other models and achieved the highest precision (99.89%), Recall (99.14%), ROC AUC (1.00), and PR AUC (0.994). In addition, XGBoost's highly time-efficient characteristics allow for the use of data oversampling techniques to improve the model's fraud detection rate.

5.2 Recommendations and Future Work

The unprecedented level of sophistication and complexity exhibited in the fraudulent activities is raising significant concerns. The latter indeed entails for collaboration from the initiating and receiving ends of transactions to bring the fraud level under control. The projected potential advances in the financial industry such as decentralization and open banking are potential vectors for related risks and criminals to support their activities more freely. In contrast, mining the data from these advances can bring insights into consumer behavior which can enable more opportunities to leverage Artificial Intelligence (AI) and enhance fraud detection and management.

As observed in this study, advanced technologies such as ML and data analytics can empower financial institutions and payment processors to analyze large volumes of transaction data in real-time and identify patterns of behavior that are indicative of fraud, enabling quick detection and prevention of fraudulent activities. However, the development of ML models

for fraud detection is obstructed by several challenges. These challenges includes the lack of diversity and limited scope of the available datasets. Furthermore, the constraints of data privacy regulations tend against the development in public research.

Conversely, the technology that is currently accessible has some challenges for implementing fraud detection model as well such as the lack of domain specific metrics, and potential biases due to the distribution of the dataset. Although conventional metrics can offer certain observations regarding the efficacy of a fraud detection model, metrics that are specific to the domain are crucial for appraising the model's efficiency in identifying other aspects such as number of accounts detected and amount of money saved. Therefore, the implementation of ML models require a good understanding of the various payment transactions, fraud types, feature engineering, model selection, model evaluation, and real-time implementation considerations. Moreover, real-world data can change over time and may experience new fraud patterns or type, and the ML model need to be updated or retrained to adapt to these changes and allow the model to detect certain shifts in fraud patterns.

In the future, this work may be extended to employ a real-world fraud transactions dataset. Furthermore, other algorithms may be employed or combined to address the drawbacks of single technique and capitalize the performance. In addition, future work may incorporate exploring methods to preserve privacy of transactions information such as Blockchain technology and decentralization.

References

Addison Howard, I. C. i. J. L. L. M. P. H. A., Bernadette Bouchon-Meunier. (2019). *Ieee-cis fraud detection.* Kaggle. Retrieved from `https://kaggle.com/competitions/ieee-fraud-detection`

Adepoju, O., Wosowei, J., Jaiman, H., et al. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. In *2019 global conference for advancement in technology (gcat)* (pp. 1–6).

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, *10*, 39700–39715.

Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *The Knowledge Engineering Review*, *34*, e1.

Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, *11*(4), 662.

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*, 100402.

Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A powerful predicting model for financial statement fraud based on optimized xgboost ensemble learning technique. *Applied Sciences*, *13*(4),

2272.

Ali, M. A., Azad, M. A., Centeno, M. P., Hao, F., & van Moorsel, A. (2019). Consumer-facing technology fraud: Economics, attack methods and potential solutions. *Future Generation Computer Systems*, *100*, 408–427.

Almaiah, M. A., Almomani, O., Alsaaidah, A., Al-Otaibi, S., Bani-Hani, N., Hwaitat, A. K. A., ... Aldhyani, T. H. (2022). Performance investigation of principal component analysis for intrusion detection system using different support vector machine kernels. *Electronics*, *11*(21), 3571.

Al-Shabi, M. (2019). Credit card fraud detection using autoencoder model in unbalanced datasets. *Journal of Advances in Mathematics and Computer Science*, *33*(5), 1–16.

Altman, E. (2021). Synthesizing credit card transactions. In *Proceedings of the second acm international conference on ai in finance* (pp. 1–9).

Anirudh, G., & Talukdar, U. (2022). An experimental analysis for credit card fraud detection with imbalanced and machine learning techniques. In *Edge analytics: Select proceedings of 26th international conference—adcom 2020* (pp. 539–550).

Arfeen, A. A., & Khan, B. M. A. (2022). Empirical analysis of machine learning algorithms on detection of fraudulent electronic fund transfer transactions. *IETE Journal of Research*, 1–13.

Arora, B. (2022). A review of credit card fraud detection techniques. *Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1*, 485–496.

Ashfaq, T., Khalid, R., Yahaya, A. S., Aslam, S., Azar, A. T., Alsafari, S., & Hameed, I. A. (2022). A machine learning and blockchain based efficient fraud detection mechanism. *Sensors*, *22*(19), 7162.

Banerjee, R., Bourla, G., Chen, S., Kashyap, M., & Purohit, S. (2018).

Comparative analysis of machine learning algorithms through credit card fraud detection. In *2018 ieee mit undergraduate research technology conference (urtc)* (pp. 1–4).

Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, *557*, 317–331.

CBB. (2020, Feb). *Pos, efts and digital wallet statistics for 2019.* Retrieved from `https://www.cbb.gov.bh/media-center/pos-efts-and-digital-wallet-statistics-for-2019/`

Chaudhary, K., Yadav, J., & Mallick, B. (2012). A review of fraud detection techniques: Credit card. *International Journal of Computer Applications*, *45*(1), 39–44.

Chugh, B., & Malik, N. (2022). Machine learning classifiers for detecting credit card fraudulent transactions. In *Information and communication technology for competitive strategies (ictcs 2021) ict: Applications and social interfaces* (pp. 223–231). Springer.

Cynthia, P., & Thomas George, S. (2021). An outlier detection approach on credit card fraud detection using machine learning: a comparative analysis on supervised and unsupervised learning. In *Intelligence in big data technologies—beyond the hype: Proceedings of icbdcc 2019* (pp. 125–135).

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems*, *29*(8), 3784–3797.

Dang, T. K., Tran, T. C., Tuan, L. M., & Tiep, M. V. (2021). Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems. *Applied Sciences*, *11*(21),

10004.

Dastidar, K. G., Granitzer, M., Siblini, W., et al. (2022). The importance of future information in credit card fraud detection. In *International conference on artificial intelligence and statistics* (pp. 10067–10077).

de Sá, A. G., Pereira, A. C., & Pappa, G. L. (2018). A customized classification algorithm for credit card fraud detection. *Engineering Applications of Artificial Intelligence*, *72*, 21–29.

Dhankhad, S., Mohammed, E., & Far, B. (2018). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In *2018 ieee international conference on information reuse and integration (iri)* (pp. 122–125).

Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia computer science*, *165*, 631–641.

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, *10*, 16400–16407.

Feng, H., Dan, T., Wang, W., Gui, R., Liu, J., & Li, Y. (2022). A combination of resampling method and machine learning for text classification on imbalanced data. In *Artificial intelligence and mobile services–aims 2021: 10th international conference, held as part of the services conference federation, scf 2021, virtual event, december 10–14, 2021, proceedings* (pp. 3–17).

GROUP, M. L. (2020). *Credit card fraud detection.* Retrieved from `https://www.kaggle.com/datasets/mlg-lb/creditcardfraud`

Grover, P., Li, Z., Liu, J., Zablocki, J., Zhou, H., Xu, J., & Cheng, A. (2022). Fdb: Fraud dataset benchmark. *arXiv preprint arXiv:2208.14417*.

Gupta, K., Singh, K., Singh, G. V., Hassan, M., Sharma, U., et al. (2022).

Machine learning based credit card fraud detection-a review. In *2022 international conference on applied artificial intelligence and computing (icaaic)* (pp. 362–368).

Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2023). Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. *Procedia Computer Science*, *218*, 2575–2584.

Hajek, P., Abedin, M. Z., & Sivarajah, U. (2022). Fraud detection in mobile payment systems using an xgboost-based framework. *Information Systems Frontiers*, 1–19.

Harris, B. (2022). *Sparkov data generation.* Retrieved from `https://github.com/namebrandon/Sparkov`$_{Data}$$_{Generation}$

Hasan, I., & Rizvi, S. (2022). Ai-driven fraud detection and mitigation in e-commerce transactions. In *Proceedings of data analytics and management: Icdam 2021, volume 1* (pp. 403–414).

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud:: A review of anomaly detection techniques and recent advances.

Hsin, Y.-Y., Dai, T.-S., Ti, Y.-W., & Huang, M.-C. (2021). Interpretable electronic transfer fraud detection with expert feature constructions. In *Cikm workshops.*

Hsin, Y.-Y., Dai, T.-S., Ti, Y.-W., Huang, M.-C., Chiang, T.-H., & Liu, L.-C. (2022). Feature engineering and resampling strategies for fund transfer fraud with limited transaction data and a time-inhomogeneous modi operandi. *IEEE Access*, *10*, 86101–86116.

Huang, K. (2020). An optimized lightgbm model for fraud detection. In *Journal of physics: Conference series* (Vol. 1651, p. 012111).

Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using smote and adaboost. *IEEE Access*, *9*, 165286–165294.

Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the ga algorithm for feature selection. *Journal of Big Data*, *9*(1), 1–17.

Itoo, F., & Singh, S. (2021). Comparison and analysis of logistic regression, naïve bayes and knn machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, *13*, 1503–1511.

Jain, Y., Tiwari, N., Dubey, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. *Int J Recent Technol Eng*, *7*(5S2), 402–407.

Jovanovic, D., Antonijevic, M., Stankovic, M., Zivkovic, M., Tanaskovic, M., & Bacanin, N. (2022). Tuning machine learning models using a group search firefly algorithm for credit card fraud detection. *Mathematics*, *10*(13), 2272.

Kaur, G., Habibi Lashkari, Z., Habibi Lashkari, A., Kaur, G., Habibi Lashkari, Z., & Habibi Lashkari, A. (2021). Cybersecurity threats in fintech. *Understanding Cybersecurity Management in FinTech: Challenges, Strategies, and Trends*, 65–87.

Khatri, S., Arora, A., & Agrawal, A. P. (2020). Supervised machine learning algorithms for credit card fraud detection: a comparison. In *2020 10th international conference on cloud computing, data science & engineering (confluence)* (pp. 680–683).

Khosla, C., & Saini, B. S. (2020). Enhancing performance of deep learning models with different data augmentation techniques: A survey. In *2020 international conference on intelligent engineering and management (iciem)* (pp. 79–85).

Kumar, M. S., Soundarya, V., Kavitha, S., Keerthika, E., & Aswini, E. (2019). Credit card fraud detection using random forest algorithm. In *2019 3rd international conference on computing and communica-*

tions technologies (iccct) (pp. 149–153).

Kumar, S., Gunjan, V. K., Ansari, M. D., & Pathak, R. (2022). Credit card fraud detection using support vector machine. In *Proceedings of the 2nd international conference on recent trends in machine learning, iot, smart cities and applications: Icmisc 2021* (pp. 27–37).

Lopez-Rojas, E., Elmir, A., & Axelsson, S. (2016). Paysim: A financial mobile money simulator for fraud detection. In *28th european modeling and simulation symposium, emss, larnaca* (pp. 249–255).

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020a). Data reduction for big data. In *Big data preprocessing: Enabling smart data* (pp. 81–99). Springer International Publishing.

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020b). Dimensionality reduction for big data. In *Big data preprocessing: Enabling smart data* (pp. 53–79). Springer International Publishing.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data preprocessing and data augmentation techniques. *Global Transitions Proceedings*.

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, *7*, 93010–93022.

Maniraj, S., Saini, A., Ahmed, S., & Sarkar, S. (2019). Credit card fraud detection using machine learning and data science. *International Journal of Engineering Research*, *8*(9), 110–115.

Mansour, A. Z., Ahmi, A., Popoola, O. M. J., & Znaimat, A. (2022). Discovering the global landscape of fraud detection studies: a bibliometric review. *Journal of Financial Crime*, *29*(2), 701–720.

Mansourifar, H., & Shi, W. (2020). Deep synthetic minority over-sampling

technique. *arXiv preprint arXiv:2003.09788.*

Megdad, M. M., Abu-Naser, S. S., & Abu-Nasser, B. S. (2022). Fraudulent financial transactions detection using machine learning.

Mekterović, I., Karan, M., Pintar, D., & Brkić, L. (2021). Credit card fraud detection in card-not-present transactions: Where to invest? *Applied Sciences*, *11*(15), 6766.

Meng, C., Zhou, L., & Liu, B. (2020). A case study in credit fraud detection with smote and xgboost. In *Journal of physics: Conference series* (Vol. 1601, p. 052016).

Mittal, S., & Tyagi, S. (2019). Performance evaluation of machine learning algorithms for credit card fraud detection. In *2019 9th international conference on cloud computing, data science & engineering (confluence)* (pp. 320–324).

Mittal, S., & Tyagi, S. (2020). Computational techniques for real-time credit card fraud detection. *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, 653–681.

Mrozek, P., Panneerselvam, J., & Bagdasar, O. (2020). Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets. In *2020 ieee/acm 13th international conference on utility and cloud computing (ucc)* (pp. 426–433).

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. *Ieee Access*, *9*, 78658–78700.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), 559–569.

Niu, X., Wang, L., & Yang, X. (2019). A comparison study of credit card fraud detection: Supervised versus unsupervised. *arXiv preprint*

arXiv:1904.10604.

Padurariu, C., & Breaban, M. E. (2019). Dealing with data imbalance in text classification. *Procedia Computer Science, 159*, 736–745.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119.*

Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems, 133*, 113303.

Priya, G. J., & Saradha, S. (2023). Global fraud prevention leveraging artificial and machine learning technologies. In *Aip conference proceedings* (Vol. 2523, p. 020174).

Raghavan, P., & El Gayar, N. (2019). Fraud detection using machine learning and deep learning. In *2019 international conference on computational intelligence and knowledge economy (iccike)* (pp. 334–339).

Rai, A. K., & Dwivedi, R. K. (2020). Fraud detection in credit card data using unsupervised machine learning based scheme. In *2020 international conference on electronics and sustainable communication systems (icesc)* (pp. 421–426).

Report, N. (2021, Dec). *Card fraud worldwide - nilson report.* Retrieved from `https://nilsonreport.com/upload/NilsonReportIssue1209`

Reurink, A. (2018). Financial fraud: a literature review. *Journal of Economic Surveys, 32*(5), 1292–1325.

Rodrigues, V. F., Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., da Costa, C. A., Barbosa, J. L. V., ... Arcot, T. (2022). Fraud detection and prevention in e-commerce: A systematic literature review. *Electronic Commerce Research and Applications*, 101207.

Roy, P., Rao, P., Gajre, J., Katake, K., Jagtap, A., & Gajmal, Y. (2021).

Comprehensive analysis for fraud detection of credit card through machine learning. In *2021 international conference on emerging smart computing and informatics (esci)* (pp. 765–769).

Sadgali, I., Sael, N., & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia computer science*, *148*, 45–54.

Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020). Application of ga feature selection on naive bayes, random forest and svm for credit card fraud detection. In *2020 international conference on decision aid sciences and application (dasa)* (pp. 1091–1097).

Şahin, Y. G., & Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines.

Salekshahrezaee, Z., Leevy, J. L., & Khoshgoftaar, T. M. (2023). The effect of feature extraction and data sampling on credit card fraud detection. *Journal of Big Data*, *10*(1), 1–17.

Sanad, Z., & Al-Sartawi, A. (2021). Financial statements fraud and data mining: a review. *Artificial Intelligence Systems and the Internet of Things in the Digital Era: Proceedings of EAMMIS 2021*, 407–414.

Shenoy, K. (2020). *Credit card transactions fraud detection dataset.* Retrieved from `https://www.kaggle.com/datasets/kartik/fraud-detection`

Shirgave, S., Awati, C., More, R., & Patil, S. (2019). A review on credit card fraud detection using machine learning. *International Journal of Scientific & technology research*, *8*(10), 1217–1220.

Singh, A., & Jain, A. (2020). A novel framework for credit card fraud prevention and detection (ccfpd) based on three layer verification strategy. In *Proceedings of icetit 2019: Emerging trends in information technology* (pp. 935–948).

Society, I. C. I. (2019). *Ieee-cis fraud detection.* Retrieved from https://www.kaggle.com/competitions/ieee-fraud-detection

Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. (2008). Credit card fraud detection using hidden markov model. *IEEE Transactions on dependable and secure computing*, *5*(1), 37–48.

Strelcenia, E., & Prakoonwit, S. (2023). Improving classification performance in credit card fraud detection by using new data augmentation. *AI*, *4*(1), 172–198.

Taha, A. A., & Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, *8*, 25579–25587.

Tamtama, A. S., & Arifudin, R. (2022). Increasing accuracy of the random forest algorithm using pca and resampling techniques with data augmentation for fraud detection of credit card transaction. *Journal of Advances in Information Systems and Technology*, *4*(1), 60–76.

Tasca, P., Aste, T., Pelizzon, L., & Perony, N. (2016). Introduction. In *Banking beyond banks and money: A guide to banking services in the twenty-first century* (pp. 1–4). Springer.

Tayebi, M., & El Kafhali, S. (2021). Hyperparameter optimization using genetic algorithms to detect frauds transactions. In *Proceedings of the international conference on artificial intelligence and computer vision (aicv2021)* (pp. 288–297).

Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019). Real-time credit card fraud detection using machine learning. In *2019 9th international conference on cloud computing, data science & engineering (confluence)* (pp. 488–493).

Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., & Singh, A. K. (2021). Credit card fraud detection using machine learning: a study. *arXiv preprint*

arXiv:2108.10005.

Tran, T. C., & Dang, T. K. (2021). Machine learning for prediction of imbalanced data: Credit fraud detection. In *2021 15th international conference on ubiquitous information management and communication (imcom)* (pp. 1–7).

Udeze, C. L., Eteng, I. E., & Ibor, A. E. (2022). Application of machine learning and resampling techniques to credit card fraud detection. *Journal of the Nigerian Society of Physical Sciences*, 769–769.

Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019). Credit card fraud detection-machine learning methods. In *2019 18th international symposium infoteh-jahorina (infoteh)* (pp. 1–5).

Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 23–27.

Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

Zhang, C., Soda, P., Bi, J., Fan, G., Almpanidis, G., Garcia, S., & Ding, W. (2023). An empirical study on the joint impact of feature selection and data resampling on imbalance classification. *Applied Intelligence*, *53*(5), 5449–5461.

Zheng, L., Liu, G., Yan, C., & Jiang, C. (2018). Transaction fraud detection based on total order relation and behavior diversity. *IEEE Transactions on Computational Social Systems*, *5*(3), 796–806.

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., & Li, J. (2021). Intelligent financial fraud detection practices in post-pandemic era. *The Innovation*, *2*(4), 100176.

Appendix A

Interview Questions

Q1: What are the challenges of getting financial transactions datasets?

In the modern age, its important to not disclose any records that have personally identifying information of any individual or customers. There are regulations now which are different in different countries and in different part of the world. They are different in Europe and North America and in Asia pacific. These rules strictly oblige that any information passed in a dataset must not reveal the identity of an individual. There are other rules as well for transferring data between other companies and that would be kind of a big concern for creating a synthetic dataset and make it publicly available for other people's research.

Creating a synthetic dataset in a way that preserves the properties of the categorical data is a useful way and getting the individual distributions correctly. For instance, an aspect of the problem that need to be considered for generating a synthetic dataset is having a data that would represent a real-world scenario and reflect a realistic information.

Q2: What are the algorithms used in practical applications for fraud detection?

There's a large amount of good work going on from specialized people in algorithms where some are expert in one particular algorithm that they

use. In practice, its best to choose selections of their methods rather than developing algorithms from scratch unless there is a very specific idea or something that require to be tested in that part of the problem.

Internally, the method that we have been using most recently are called random forest method and its core is a decision tree. The reason behind that is that you can put multiple decision trees and put them together in a statistically rigorous and sensible way to give much better results than a single decision tree that would be used. Random forest is one of the good, solid, and stable methods being used frequently. There are some other more advanced methods such as support vector machine methods and XGBoost where we have done some research using them in the company internally and we got some decent result with them. However, we haven't used those for a model that we actually deployed.

Q3: What are the features or fields of a financial transaction message is being focused on for fraud detection systems?

The base fields that we would start with are basically the transaction amount, the transaction date and time, the merchant code (Merchant code has gone out in almost every model I have seen). Then, the POS entry mode and the condition code are very important now because of two reasons. Number 1, with a chip card coming around with better technologies, there is definitely some channels that are easier to do fraud than others and the POS entry mode code tells you that. It also tells you whether the card present versus card not present which is important now because it tells you what is going on and the type of fraud (instore or online) where in online fraud there is no need to have the physical card and it requires only the card number.

Additionally, there is usually some sort of information about the terminal. Either what its location or the terminal postal code. When building features and taking what type of feature to put in the dataset, that type of in-

formation is often used more in what we call a relative sense than we would in an absolute sense. We don't necessary want to know exactly what the terminal code was. What we want to know is that does the person who's using this code, has made transaction on that code previously by the card holder.A very simple example, in an ATM machine, a feature that would investigate the amount and the terminal code and check if the same amount has been used at the same ATM machine in the past. That greatly reduces the risk of fraud.

A lot of basic features are not necessarily required for an effective predictive model. The basic features do not tend to look for fraud patterns knowing that fraud is behavioural kind of activities. Therefore, having features that are looking for potential fraud patterns to eliminate or assign a high fraud score to is something necessary for an effective fraud detection system.

Q4: What are the practical limitations for employing machine learning techniques for fraud detection?

The practical limitations lay with two aspects which are the database storage limitation and the time consumed for the computations. Financial systems must function very fast, and we are talking about milliseconds to decide there. Its not always good to use as much history as a data scientist would like to have. A data scientist would like to be able to look at 2 years if transaction history and look at all kind of ways to reduce that data as well as perform intense calculations to summarize if the person is engaged in normal type of transaction and performing the same type of pattern or not.

Unfortunately, the architects and the performance consultants tend to limit us to something that is closer to about 30 days of history which has been the standard for us in the company. Where that is a little bit short because ideally you would like to be able to look at cycles that repeat over a few months to identify when does somebody pays their bills? What bills

they have? How many times a month they transact? What is the average spend over the course of a couple of pay checks cycles?

We have done experimentation in the past (A long time ago), and we didn't see huge differences between making a model for 60 days instead of 30 days. But again, this might have been because of the methods used back then where the feature library at the time wasn't optimized to take advantage of the longer time period.

While from data science perspective, the size of the dataset doesn't add a lot of complexity to the model because all they need to do is change the length of the time windows. But it adds a whole bunch of complexity for the database administrator and the architects to provide storage and computation resources to come up with something that is capable of handling the computational processes needed with the speed required.

Q5: There several type of financial institutions such as banks, acquires, merchants etc. Each provide different services (e.g. cash withdrawal, goods and services, cashback, refunds) an these variant type of transactions have different characteristics and different transaction cycle. How these points are being considered when building a fraud detection system?

When you are doing card fraud or merchant fraud, you need to always focus on the authorization piece with the logic being that the goal is to stop the fraud transaction when its happening because when financial institutions perform cost benefit analysis on how to stop crimes. Unless the amount of fraud is huge, its usually not worth to do the full investigation to find out who did it and prosecuted through the legal system. Financial institutions aim to stop fraud while its happening and stop transaction from ever occurring. So that the merchant or the bank doesn't lose the money in the first place.

On the other hand, when doing money laundering investigations, that is a different problem and you are trying to come up with list of things to turn

performing exactly the same as you want it to be but the differences in the result might illusion a sort of problem.

The second practical problem in this business is the parameter to which the statistics are calculated (The number of accounts versus the number of transactions). For example, when having two fraud customers in your database where 1 perform 99 fraud transactions of 1$ each and the other did only 1 fraud transaction of 1000$. A system which detects the transactions from the first fraudster would result in a model with accuracy of 99% which looks effective. However, if you look at the amount of transaction that have been missed, the system misses most of the money. Therefore, for choosing evaluation metrics, a metric that consider the money that is being lost from missed transactions is necessary. This is a little bit of a subtle problem there because of the complexity to count the number of transactions and the amount of the transactions.

In addition, for the customer with the 99 fraud transactions, if the system detects the 1st transaction, then the others one would not occur. Where if the system detects the last transaction, it would still be a valid fraud detection, but it did not detect nearly as much activity from the same account. Therefore, a system can't be assessed just in term of the number of transactions that have been detected. The efficiency and the performance of the model need to have a factor that accounts for the transactions which might have occurred based on the detected transactions.

Receiver operator characteristics is of the most technical evaluation metric that is being used for evaluating the performance of machine learning model for fraud detection systems. As it answers the question of what the model is doing independent of the data characteristics which may ultimately change. In the end, the objective is about saving money, therefore the selected metric needs to be the best thing that describe the amount of money been saved by the system.

ملخص الدراسة

تطورت طرق الدفع الإلكتروني بشكل كبير وأصبحت السائدة في النظام المالي المعاصر، مما توفر من مجموعة متنوعة من القنوات السلسة للمعاملات الإلكترونية. لقد شهد عدد المعاملات الإلكترونية زيادة كبيرة خلال العقد الأخير، حيث يتم معالجة ملايين المعاملات يومياً. ومع ذلك، فإن الزيادة الكبيرة في المعاملات الرقمية ترافقت بعدد كبير من الأنشطة الاحتيالية الرهيبة التي تؤدي إلى خسائر مالية كبيرة للمؤسسات المالية وعملائها. وبالإضافة إلى ذلك، فإن التقدم التكنولوجي السريع يلعب دورًا حاسمًا في زيادة عدد عمليات الاحتيال ويتوقع أن تزداد هذه المخاطر بشكل ملحوظ. وبالتالي، أصبح الوقاية والكشف عن الاحتيال ضرورة لضمان موثوقية ومصداقية المؤسسات المالية.

لقد كانت تقنيات التنقيب عن البيانات والتعلم الآلي هي الاستراتيجيات الرائدة التي اعتمدها الباحثون لتطوير طرق للكشف عن الاحتيالات وتقليل المخاطر. في هذا السياق، واجه تصميم الخوارزميات العملية الى العديد من التحديات بما في ذلك عدم وجود تفسيرات لتعريف الاحتيال، وعدم توفر بيانات الحوالات الاحتيالية بسبب السرية، والخصائص المعقدة والصعبة التي تمييز بها عمليات الاحتيال. في هذه الأطروحة، نقترح نموذج تعلم آلي للكشف عن عمليات الاحتيال في تحويلات الأموال الإلكترونية.

لتقييم فعالية النموذج، تم إجراء تجارب على مجموعة بيانات محاكاة ضخمة تمتد على مدى عامين. والتي شملت تقنية إعادة عينة الأقلية الاصطناعية SMOTE ، بالإضافة الى التحقق المتقاطع باستخدام Stratified K-fold . شمل التقييم استخدام مقاييس متعددة لتقييم النموذج بشكل شامل، بما في ذلك مصفوفة الالتباس، ومنحنى دقة التنبؤ، ومنحنى معدل الخطأ، والدقة والاستدعاء والنتيجة الكاملة وكفاءة الوقت. تشير نتائج الأداء إلى أن استخدام تقنية SMOTE يحسن من معدل اكتشاف النموذج، وأثبت النموذج المقترح XGBoost أفضل كفاءة وفعالية، حيث حقق علامة كاملة لمنحنى دقة التنبؤ وأعلى علامة لمنحنى معدل الخطأ.

جـامعة البحـرين
كلية تقنية المعلومات



إستخدام التعلم الآلي نحو الكشف عن الاحتيال في التحويل الإلكتروني للأموال

أطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير
في علوم الأمن الإلكتروني

إعـــداد

حسين إسماعيل
٢٠٢١٠٠٠٤٨

إشـــراف
د. وائل محمد المدني
أستاذ مشارك
جامعة البحرين

مملكـة البحـرين
مايو ٢٠٢٣م