EXPLORING THE IMPACT OF ARTIFICIAL INTELLIGENCE OF FINANCIAL TECHNOLOGY : A USED-CASE OF CREDIT CARD FRAUD DETECTION

GAN JIA SHENG

BACHELOR OF INTERNATIONAL BUSINESS (HONOURS)

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF ACCOUNTANCY AND MANAGEMENT
DEPARTMENT OF INTERNATIONAL BUSINESS

MAY 2024

EXPLORING THE IMPACT OF ARTIFICIAL INTELLIGENCE OF FINANCIAL TECHNOLOGY : A USED-CASE OF CREDIT CARD FRAUD DETECTION

BY

GAN JIA SHENG

A final year project submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF INTERNATIONAL BUSINESS (HONS)

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF ACCOUNTANCY AND MANAGEMENT
DEPARTMENT OF INTERNATIONAL BUSINESS

MAY 2024

DECLARATION

I hereby declare that:

(1) This undergraduate FYP is the end result of my own work and that due acknowledgement has been given in the references to ALL sources of information be they printed, electronic, or personal.
(2) No portion of this FYP has been submitted in support of any application for any other degree or qualification of this or any other university, or other institutes of learning.
(3) Sole contribution has been made by me in completing the FYP.
(4) The word count of this research report is ____9641_________________.

Name of student:	Student ID:	Signature:
Gan Jia Sheng	2102078	

Date: 1 May 2024

ACKNOWLEGEMENT

I would like to to sincerely thank everyone who helped navigate this study project through to its conclusion.

First and foremost, I want to sincerely thank Dr. Seah Choon Sen, my research supervisor, for his insightful counsel, unwavering encouragement, and motivational remarks. Their suggestions and counsel have been very helpful in shaping my study plan and direction.

I would like to sincerely thank my second examiner, Dr. Farah Waheeda binti Jalaludin, for her help when my research is not up to par. Without their willingness to freely share their thoughts and experiences, this research project would not have been achieved.

Nevertheless, I also want to express my gratitude to Universiti Tunku Abdul Rahman (UTAR), for giving me access to the tools and resources I needed to complete this research. I would want to thank my friends for their assistance, since they have given me insightful suggestions and constructive criticism during this investigation.

Finally, we would like to thank our families for their steadfast understanding and support during this effort. I appreciate all of your thoughtful contributions.

DEDICATION

This Final Year Project is dedicated to my respected supervisor, Dr. Seah Choon Sen, and my respected second examiner, Dr. Farah Waheeda binti Jalaludin, who provided me with invaluable advice, direction, understanding, and motivation during my Final Year Project Journey. I would like to thank all of my loved ones—friends, instructors, seniors, for their invaluable help with my final year project. It would be difficult for me to complete this final year project without all of your encouragement, supports, and assistance.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

PREFACE

For each student to graduate with a Bachelor of International Business (Honours), they must complete the Universiti Tunku Abdul Rahman (UTAR) Final Year Project, "UKMZ2016 Research Project." Understanding and being well-aware of global issues is a major source of competitive advantage for a Bachelor of Science student focusing in International Business. However, there are many facets to the complex and wide-ranging topic of international business. Among the many worldwide issues, credit card fraud is a modern problem that has been recognized as a significant trend that has seriously altered the structure of the economy. Every industry has intensified their efforts to counteract fraudulent activities. Nonetheless, among businesses that have recently begun to invest in machine learning, fraud is still a minor concern. Consequently, this motivates the author to investigate the variables influencing credit card fraud.

ABSTRACT

The detection of credit card fraud remains a critical challenge in the digital age, prompting extensive research into effective methodologies and techniques. This study contributes to the field by employing logistic regression and analyzing a dataset comprising 1,754,155 transactions from Axis Bank in India. Through Pearson and Spearman correlations, it identifies Transaction Amount as a significant predictor of fraud, underscoring its pivotal role in fraud detection. Furthermore, the study explores the implications of threshold setting in machine learning models for fraud detection, emphasizing the balance between false positives and false negatives. It also highlights the importance of diverse datasets and the adoption of multiple analysis methods to enhance the accuracy and reliability of fraud detection systems. The findings provide valuable insights for regulators, financial institutions, and researchers, aiding in the development of evidence-based policies and the refinement of fraud detection models to combat evolving fraud threats effectively.

Chapter 1 Research Overview

1.0 Introduction

This chapter primarily explains the study's background, problem statement, research questions, aims, scope, and significance. It also discusses the study's limitations and potential future directions.

1.1 Research Background

Individuals are enthusiastically pursuing a lifestyle marked by increased convenience as the globe continues to progress. This desire has resulted in the establishment of Financial Technology (Fintech). According to (Fama, 1980), a bank acts as an intermediary, but the Internet has changed the way financial service providers carry out their responsibilities. This revolution is fundamentally redefining the character of banking itself, influencing both the type and mode of delivery of banking services. (Brei et al., 2020) discovered that many banks are putting more emphasis on fee-generating services to ensure their long-term profitability. Moreover, the impact of the COVID-19 epidemic and the growing demand for digital innovations drove significant growth in the fintech sector between 2020 and 2021. The growth of Malaysia's fintech sector can be attributed to several factors, including unwavering government support, a large market characterized by a population inclined toward digital solutions, a regulatory framework conducive to growth, and a heightened appetite for financial services. As a result of the convergence of these factors, the sector has experienced rapid and widespread expansion. The Fintech News Malaysia Fintech Report 2021 highlights Malaysia's fintech landscape with over 250 enterprises. These entities are leveraging mobile banking and digital transformation trends to align with consumers' lifestyles. The trend is expected to persist, with related sectors like retail and e-commerce leveraging advancements in fintech (Noah, 2022).

Businesses and governmental institutions in the modern landscape are confronted with an increasing incidence of fraudulent endeavours. As a result, it is critical to build automated systems for the goal of detecting fraud. The inherent difficulties involved with human analysts seeking to spot fraudulent trends within transaction databases make these automated solutions indispensables. These datasets are usually distinguished by their massive sample count, sophisticated multidimensional structures, and constant influx of online updates. In the context of credit card transactions, 'fraud' refers to the unlawful and inappropriate use of an account by someone other than the rightful account holder. To avoid and reduce such misuse, it is critical to implement necessary preventive measures. Furthermore, investigating the methods of these fraudulent operations provides significant information that can be used to reduce their recurrence and strengthen safeguards against any re-occurrences (Vishal Kumar, 2023). A credit card is a card given to a customer that allows them to buy products and services up to a certain credit limit or acquire cash advances. It provides the temporary benefit of deferred repayment till the following billing cycle. Credit card fraud, on the other hand, poses a substantial vulnerability because it permits huge quantities to be withdrawn secretly without the card owner's knowledge. Fraudsters alter their behaviours to make fraudulent transactions appear legitimate, complicating fraud detection. The lack of inherent risk and the ability of fraudsters to exploit these transactions complicate the process even further (Dornadula & Geetha, 2019).

In the late 1970s and early 1980s, the scope of artificial intelligence research shifted away from algorithmic techniques, targeting on logical, knowledge-based approaches.

This period saw the abandonment of neural network research, creating a schism between artificial intelligence and machine learning, where the latter had previously served as a training ground for AI (Artificial Intelligent) systems. Subsequently, the machine learning industry underwent a significant overhaul, reorganizing itself into a separate field with a renewed focus on solving practical problems and providing services. This transformation marked a departure from AI-centric approaches to an emphasis on methods rooted in probability theory and statistics (Foote, 2023). Machine learning is an aspect of artificial intelligence characterized by its ability to autonomously learn from experience and construct models without explicit user programming (Clarke et al., 2009). Operating independently, this method can pick out and recognize patterns and make decisions based on pre-existing data, all without direct user guidance. As an integral component of artificial intelligence, machine learning is fundamentally a learning system that not only adapts to new inputs but also autonomously generates automatic actions to arrive at informed decisions. By learning from data together with automating decision-making processes, machine learning significantly add a helping hand towards human roles in decision-making, offering a valuable tool in scenarios where the system can independently analyse information and arrive at appropriate conclusions (Kane, 2017). Machine learning finds diverse applications across computing, enabling the design and implementation of high-performance algorithms in various domains. Examples of its application range from email spam filtering and fraud detection on social networks to online stock trading, face and shape detection, medical diagnosis, traffic prediction, character recognition, and product recommendations, among others. Additionally, machine learning plays a pivotal role in credit card fraud detection. These examples underscore the versatility of machine learning in enhancing computational tasks across a wide spectrum of fields (Alzubi et al., 2018).

1.2 Problem Statement

Amid the ongoing COVID-19 pandemic, heightened concerns about health and safety have driven a significant shift towards cashless payments. Recognizing the potential transmission of viruses and bacteria through physical currency, individuals are increasingly opting for digital payment methods. Visa, a global leader in digital payments, recently revealed findings indicating a substantial change in consumer behaviour in Malaysia. According to the Visa Consumer Payment Attitudes 2021 study, most Malaysian consumers (55%) can now sustainably live for over a week without using cash, marking a 13% increase compared to the previous year. Notably, the pandemic has influenced long-term preferences, with 28% of respondents expressing an inclination towards digital payments even post-pandemic. Furthermore, the study suggests that Malaysian consumers envision a potential shift to a cashless society by 2025 (The Star, 2022).

A credit card typically refers to a thin rectangular piece of plastic or metal, which is issued by banks or financial services companies (Bloomenthal, 2023). Major credit cards which including Visa, Mastercard, Discover, and American Express, are commonly issued by banks, credit unions, or other financial institutions. The vulnerability of credit card information is a notable concern due to its relatively fragile nature. With just the 16-digit card number on the front and the 3-digit CVC number on the back, transactions can be completed online at any time. This simplicity in transaction processing, however, poses a security risk, as hackers may exploit system vulnerabilities to carry out unauthorized payments without the need for Transaction Authentication Codes (TAC).

The impact of credit card fraud is multifaceted, imposing significant costs on banks and card issuers while concurrently jeopardizing their reputation (Soltani Halvaiee & Akbari, 2014). This threat arises when transactions on individuals' credit cards are initiated by unauthorized persons, posing substantial risks to businesses and organizations, especially in a competitive environment lacking preventive systems (Phua et al., 2010). The escalating number of credit card transactions further exacerbates the challenge, leading to a surge in fraudulent activities. The expense associated with analysing the legitimacy of transactions, determining whether they are authorized by the actual cardholder or not, adds an additional layer of complexity to the issue (Gadi et al., 2008).

The insecurity of credit cards has led to substantial industry losses in recent years, with banks and credit card companies bearing the brunt of fraudulent transactions. The escalating significance of credit card fraud detection is evident as these entities grapple with financial repercussions. In response, developers are increasingly turning to AI to combat fraud, employing algorithms to scrutinize datasets and enhance machine learning capabilities efficiently (TuxCare, 2023). The efficiency of AI in analysing large transaction volumes within seconds surpasses human capabilities, positioning it as a superior tool for fraud identification in the dynamic landscape of credit card security. This underscores the critical need for robust and efficient preventive measures to mitigate the detrimental effects of credit card fraud on financial institutions and their clients.

While similar studies have investigated credit card fraud detection, our aim is to assess whether the factors influencing credit card fraud have undergone any changes. According to Chaudhary et al. (2012), suspicious behavior in credit card transactions, such as sudden transactions for large amounts or high frequency of usage without the cardholder's knowledge, can be identified through break-point analysis. Other research, such as that by Panigrahi et al. (2009), Bolton and Hand (2001), Khan et al. (2014a), and Correia et al. (2015), has also highlighted the importance of analyzing transaction patterns, particularly sudden increases in transaction amounts, to classify transactions as fraudulent or suspicious. Our concern lies in understanding how the dynamics of digitalization, particularly in the post-COVID-19 era, may have impacted these factors and their relevance in detecting credit card fraud.

1.3 Research Question

The current study seeks to provide answers to the following research questions considering each of the previously given issue statements:

1. What are the factors affecting credit card fraud detection?
2. Which transactions are in fraud?

1.4 Research Objective

Considering the research questions, the goal of the study may be further explained as follows:

1. To investigate the factors affecting credit card fraud detection
2. To identify the numbers of transaction that are in fraud.

1.5 Scope of study

Given the use of secondary data, the study's scope is constrained by the selected dataset. Nevertheless, the breadth of transaction data sourced from various global banks and credit card companies ensures a comprehensive and diverse perspective within the study's limitations.

1.6 Research Significant

This particular study serves as a wise opportunity for the public sector, particularly the government, to invest in advanced credit card fraud detection processes. The implementation of these measures not only holds the significantly contribute to national stability but also plays a crucial role in enhancing economic stability, fortifying national security, and fostering an environment that attracts foreign investors. Recognizing its responsibility to safeguard citizens, the government is urged to adopt a distinct approach in response to the evolving landscape of AI technology and societal requirements (Pi, 2021). Part of this distinctive approach involves strengthening the national financial security system, allowing the government to redirect attention and resources towards critical areas such as Research and Development (R&D), thereby contributing further to the nation's progress. Amidst the dual challenge of shielding citizens from the detrimental impact of algorithms and responding to the demands of a rapidly evolving society, governments can leverage Machine Learning technologies to enhance efficiency and meet societal expectations (Kuziemski & Misuraca, 2020). The research outcomes have the potential to instil confidence in industries within the country, facilitating increased investment and expansion while mitigating apprehensions surrounding credit card fraud. As a ripple effect, this is anticipated to trigger economic growth, exerting a positive influence on the country's Gross Domestic Product (GDP). Moreover, private industry is encouraged to explore investments in AI and Machine Learning specifically tailored for fraud detection purposes. The transformative impact of AI on work culture and efficiency is substantial, unlocking vast potential. Businesses across diverse sectors can harness AI technology for financial and accounting tasks, ushering in a new era of enhanced productivity and operational efficiency (Donepudi, 2019).

Academic institutions stand poised to harness the insights derived from this study, enabling a deeper exploration of machine learning as well as Artificial Intelligence. The mastery of these processes extends far beyond the financial sector, promoting advancements in such areas like the prevention of money laundering and enriching the broader educational landscape. In today's educational landscape, college students are presented with an exceptional opportunity for interactive and personalized learning experiences. AI, leveraging insights from vast datasets, holds the potential to provide students with tailored educational endeavours. To maximize the benefits of these technologies, colleges, universities, and other educational institutions must embrace innovative approaches, paving the way for a competitive edge in the evolving educational landscape (Kuleto et al., 2021). This study delves into the environmental and ethical considerations associated with the implementation of advanced fraud detection processes. An essential aspect of this exploration involves pre-emptively addressing privacy issues when designing machine learning-based systems. The inherent unpredictability in data processing strategies, intrackability of public data downloads, and the diverse sources of data acquisition amplify the potential threat to privacy posed by machine learning technologies. However, it is noteworthy that the judicious and responsible application of ML technology can also play a pivotal role in safeguarding privacy (Sun et al., 2020). Emphasizing the importance of aligning implementation with ethical standards, this ensures sustainability and fosters trust in the technology, thereby striking a balance between innovation and privacy protection.

1.7 Summary

In this chapter, the background study is meticulously outlined, emphasizing the imperative need for credit card fraud detection in the contemporary era. The rationale behind the study is articulated, unveiling the pressing reasons for addressing this issue. The chapter meticulously identifies the research questions and objectives, offering a clear roadmap for the investigation. Significantly, the potential benefits that would be derived from this study across various sectors are underscored, establishing the research's broader relevance and impact.

Chapter 2 Literature Review

2.0 Introduction

In this chapter, the explanation of the variables towards credit card fraud will be covered. Prediction model is applied to investigate the variables that influencing the correlation with credit card fraud using machine learning.

2.1 Credit Card Fraud

Credit card fraud manifests when an unauthorized party acquires your credit card details to carry out purchases or misappropriate funds. This pervasive issue has afflicted individuals globally for an extended period, impacting not only the victims but also credit card companies and merchants (Barker et al., 2008). To gain a comprehensive understanding, this study will explain several prominent schemes: Point of Sale Fraud, Phishing and Vishing, Keystroke Logging and Application Fraud. Each of these methods will be explored in detail to shed light on their distinct mechanisms within the realm of credit card fraud.

2.1.1 Point-Of-Sale Fraud

In this form of fraud, inconspicuous skimming devices are affixed to standard Point-of-Sale (Pos) terminals, surreptitiously capturing card data during swipe transactions. Typically orchestrated by a merchant or store employee, these pilfered details may be shared with malicious actors. A comparable tactic involves attaching such devices to ATM card slots, enabling the cloning of card information, while an inconspicuously positioned camera captures the unsuspecting user's PIN. Although Point-of-Sale (POS) transactions, facilitated by credit cards or smartphones, offer heightened payment convenience, the integration of technologies like near-field communication (NFC) and radio-frequency identification (RFID) has augmented the vulnerability of POS payments to fraudulent activities (Limited, 2020).

2.1.2 Phishing and Vishing

Phishing constitutes an email attack wherein perpetrators, masquerading as a trustworthy entity, fraudulently seek sensitive user information through electronic communication. Meticulously designed emails are employed to target specific groups, enticing recipients to click on embedded links that, once clicked, deploy malicious code on the victim's computer (GeeksforGeeks, 2022). Vishing, on the other hand, is a cyber-attack leveraging voice communication to extract confidential data from a collective audience. In vishing attempts, attackers employ voice calls, posing as employees from ostensibly reputable organizations, to deceive targets into divulging sensitive information (GeeksforGeeks, 2022). Both methods involve the impersonation of official bank communications, creating a deceptive lure that prompts users to click on fraudulent links. Subsequently, victims are directed to seemingly authentic websites where, unwittingly, they input their card details. This information is then exploited by fraudsters for their illicit gains.

2.1.3 Keystroke logging

Keyloggers, also known as keystroke logging, are applications designed to record keystrokes on a device. While keyloggers can have legitimate uses, a significant number of applications are deployed for malicious purposes. In a keylogger attack, the software captures every keystroke made on the victim's device and transmits this sensitive information to the attacker. These tools have the capability to record every keystroke entered via a computer or mobile keyboard, enabling the tracking of credit card details, visited websites, and passwords (CrowdStrike, 2023). Unfortunately, contemporary hackers increasingly resort to keystroke logging using malicious software to illicitly obtain credit card information. This often initiates after a user clicks on a dubious link, unwittingly installing malware that facilitates the unauthorized capture of sensitive data on their system.

2.1.4 Application fraud

Application fraud occurs when a malicious actor utilizes a stolen or synthetic ID to apply for a loan or line of credit, having no intention of repaying the lender. The fraudster meticulously crafts a facade of authentic-looking credit and account activity to progressively secure more loans and elevated credit limits. Some perpetrators employ bust-out fraud, a comprehensive strategy involving multiple instances of application fraud. Over time, the fraudster systematically establishes numerous lines of credit, strategically maxing them out within a condensed timeframe before disappearing (Data Visor, 2023). This form of identity theft involves fraudulent actors assuming the identity of a genuine customer by using stolen or counterfeited documents to procure a credit card. Although this activity may be discerned through comprehensive background checks, if successful, it enables criminals to wield a valid credit card that leaves behind a false paper trail.

2.2 Underlying Theory

In this research, the adoption of Credit Risk Theory serves as a foundational framework, specifically addressing the intricacies tied to lending, credit transactions, and the financial exposure inherent in borrower relationships. As defined by Anderson and Salas, & Saurina (2002), credit risk encapsulates the potential that a borrower may default on any form of debt, neglecting obligatory payments. This risk primarily pertains to the lender, encompassing the jeopardy of losing principal and interest. Disruption in payment can manifest as complete or partial loss, with instances ranging from an insolvent bank unable to reimburse funds to a depositor.

The escalating instances of credit card fraud contribute to an elevated credit risk for banks, resulting in heightened interest rates for borrowed funds. This creates a potential vicious cycle, adversely affecting the overall market through the rise in interest rates. In response to this challenging scenario, the implementation of effective AI tools becomes essential for identifying and preventing credit card fraud. By curbing credit risk through AI-driven fraud detection, banks can not only save resources expended on fraud-related expenses but also allocate these funds to other sectors, fostering efficiency and financial stability.

2.3 Variables

2.3.1 Terminal ID

A terminal ID number, commonly referred to as "terminal identification number" or "TID," consists of an eight-digit character sequence crucial for financial institutions to monitor the terminal involved in processing a transaction. This identifier also facilitates merchants in swiftly locating transactions, particularly in scenarios involving refunds or disputes. Terminal ID numbers play a vital role in aiding banks and processors to identify your equipment within the network. They serve as distinctive markers for transaction processing hardware associated with your business (Decorte, 2022). Moreover, these terminal ID numbers contribute significantly to the differentiation between legitimate and fraudulent transactions. By cross-referencing Terminal ID Numbers across transactions, it becomes possible to identify inconsistencies or discrepancies that may signify fraudulent activities. This proactive approach to detection serves as a preventive measure against chargebacks and potential financial losses, enhancing overall transaction security (Chargeflow, 2023).

2.3.2 Transaction Amount

The Transaction Amount is defined as the monetary value associated with a submitted transaction. In the context of new contracts, it represents the total contract amount, while for contract amendments, it signifies the value of the change—whether an increase, decrease, or zero. This amount encompasses both the actual Balance to be transferred and any fees and taxes levied to complete the transaction. In practical terms, the Transaction Amount is the sum debited from the Balance during your Card usage. This includes not only the principal amount being transferred but also any associated fees and taxes required to finalize the transaction. It's essential to note that the term "Merchant" or "Retailer" refers to an authorized retail establishment capable of accepting the Card for transactions (Transaction Amount Definition: 211 Samples | Law Insider, n.d.).

2.3.3 Transaction Time (Second)

A transaction, defined as a business interaction involving buying or selling, is essentially the process of executing a particular task, like government business transactions. Time, measured globally, signifies the temporal dimension of when an event occurs. When we delve into seconds, they represent occurrences succeeding the first in a series (OxfordLearnersDictionaries, n.d.).

In the context of coding, the amalgamation of these terms reflects the duration in seconds taken by a transaction. Specifically, Transaction Per Second emerges as a pivotal metric, gauging the efficiency of a system or platform in executing transactions promptly. This metric holds significance across various domains, including technology and finance, impacting user experience and operational efficiency (Caringal, 2023).

2.3.4 Transaction Time (Day)

A day is characterized by a span of 24 hours, encompassing the duration from morning light to evening darkness (OxfordLearnersDictionaries, n.d.). In the coding context, the fusion of these terms signifies the measurement of time in days for transactional processes. This information is instrumental in revealing the frequency of transactions occurring within a single day.

This data becomes particularly valuable for back-end users as it offers insights into the occurrence of transactions over the course of a day. The ability to discern the number of transactions within this time frame serves as a crucial metric, allowing for the evaluation of any unusual or noteworthy patterns in transactional activity.

2.3.5 Customer ID

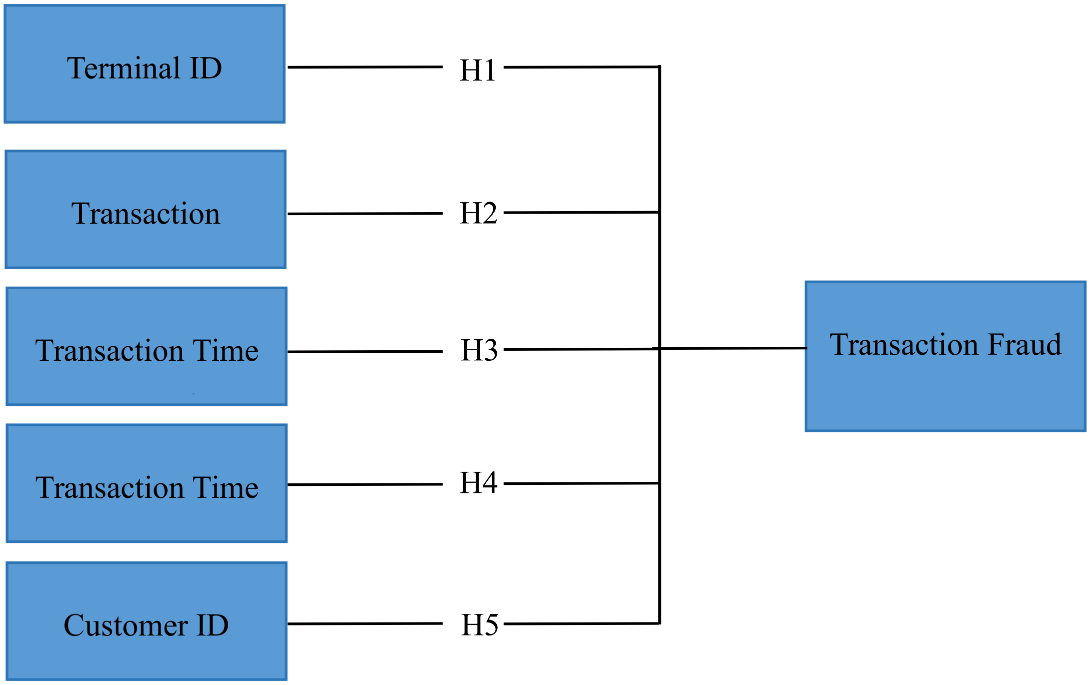
Customer ID, or Customer Identification, is a unique code assigned to individuals or entities by companies. It distinguishes one customer from another in a company's database or system, aiding in personalized services and efficient communication. Widely used across industries like banking and e-commerce, Customer IDs streamline customer management and support tracking of interactions. Generated during onboarding or provided by the company, it's crucial to keep Customer IDs confidential for authentication when accessing services or interacting with customer support (Customer ID Definition: 108 Samples | Law Insider, n.d.).

2.3.6 Transaction Fraud

Fraud which represents as the criminal act of deceiving someone with the intent to unlawfully acquire money or goods, involves intentionally deceptive actions for illegal gain or to deny rightful entitlement to a victim (OxfordLearnersDictionaries, n.d.). Various types of fraud encompass tax fraud, credit card fraud, wire fraud, securities fraud, and bankruptcy fraud (Chen, 2022). This deceptive conduct can be orchestrated by an individual, a group of individuals, or an entire business entity, spanning from large organizations to small groups or even individual actors (What Is Fraud? Definition of Fraud, Fraud Meaning - the Economic Times, n.d.).

2.4 Conceptual Framework

Figure 2.4: Conceptual Framework



Source : Developed from the Research

2.5 Hypothesis Development

2.5.1 Terminal ID

Terminal ID emerges as a crucial element in distinguishing between legitimate and fraudulent transactions. Through the comparison of Terminal ID Numbers across transactions, inconsistencies or discrepancies that may indicate potentially fraudulent activity can be readily identified.

In support of this, Ghosh Dastidar et al. (2020) emphasizes the significance of merchant category and terminal ID, along with time features, contributing significantly to the determination of transaction similarity. This observation suggests that focusing on these two key aspects holds substantial value in the realm of credit card fraud detection. Furthermore, insights from Lucas et al. (2020) underscore the importance of terminal ID by highlighting that a sequence of transactions occurring at a fixed terminal can unveil valuable patterns for fraud detection. This acknowledgment reinforces the pivotal role of terminal ID in uncovering meaningful patterns and enhancing the effectiveness of credit card fraud detection algorithms.

H_1: There is a relationship between Terminal ID and Transaction Fraud

2.5.2 Transaction Amount

The transaction amount stands out as a direct and significant indicator in the context of credit card fraud detection. Abdallah et al. (2016) highlight that the strategy employed in credit card fraud detection revolves around pattern recognition, particularly through the automatic analysis of user spending behaviours. User spending behaviour, encompassing factors such as transaction amount, time gap since the last purchase, day of the week, item category, and customer address, serves as a rich source of information for fraud detection algorithms.

Anomaly-based fraud detection, as indicated by Malekian and Hashemi (2013), is a prevalent approach in credit card fraud detection. This method involves creating a cardholder's profile by scrutinizing spending behaviours patterns. Any incoming transaction that deviates from the established cardholder profile is flagged as suspicious, forming a proactive and effective strategy in identifying potentially fraudulent activities.

H_2 : There is a relationship between Transaction Amount and Transaction Fraud

2.5.3 Transaction Time - Seconds

Transaction Per Second act as a critical factor in fraud detection, playing a pivotal role in evaluating the efficiency of systems and platforms. From a machine learning perspective, normal credit card transactions exhibit a characteristic range of transactions per time, providing a baseline for algorithms to identify legitimate patterns. A gradual increase in transaction per second can trigger suspicions, prompting fraud detection mechanisms to scrutinize anomalous behaviour. This heightened vigilance is essential for identifying activities that deviate from established norms, signifying potential threats. Importantly, the dynamic nature of fraudulent activity necessitates continuous monitoring of transaction per second, as fraudsters adapt their strategies over time.

H_3 : There is a relationship between Transaction Time (Second) and Transaction Fraud

2.5.4 Transaction Time - Days

Transaction Per Day emerges as a pivotal factor in the landscape of fraud detection, offering valuable insights for machine learning algorithms. By delving into the unique patterns of each credit card user, artificial intelligence accumulates extensive data on users' spending behaviours. This data includes an average of a user's weekly expenditures, the frequency of transactions, and the habitual times of credit card usage. This wealth of information establishes a baseline, enabling the AI to discern normal usage patterns for each individual. Consequently, when a user's transaction patterns deviate significantly from historical norms, it triggers an alert within the system. This alert mechanism is crucial for identifying potentially fraudulent activities that may manifest as irregularities in spending habits.

H_4 : There is a relationship between Transaction Time (Day) and Transaction Fraud

2.5.5 Customer ID

Verifying the identity of customers can significantly reduce credit card fraud risk. It ensures that the person making a purchase is the rightful owner of the credit card, a crucial step in protecting businesses from potential losses. Confirming a customer's identity helps validate the legitimacy of the credit card and ensures the transaction is not fraudulent. Methods for verification may include requesting a government-issued ID or validating the billing address linked to the credit card. Additionally, some e-commerce platforms offer built-in identity verification systems to streamline this process (Kostin, 2023).

H_5 : There is a relationship between Customer ID and Transaction Fraud

2.6 Summary

Chapter 2 introduces credit card fraud, encompassing Point-of-Sale Fraud, keystroke logging, and related methods. Both independent and dependent variables are being introduced in this chapter. The hypothesis development establishes their relationships, detailing how each variable influences credit card fraud detection. This chapter forms a robust theoretical framework, guiding the empirical investigation into the impact of artificial intelligence on credit card fraud detection in subsequent chapters.

Chapter 3 Methodology

3.0 Introduction

In this chapter, the methodology for this research is elucidated. The research utilizes secondary datasets, publicly shared on Kaggle. These datasets consist of transaction information crucial for credit card fraud detection. It involves employing statistical analyses such as Logistic Regression, Pearson, and Spearman. These analyses aim to uncover patterns and relationships within the dataset towards credit card fraud detection mechanisms.

3.1 Accessing Secondary Data

The methodology adopted for this study centres on the utilization of secondary data. Employing secondary analysis, the research leverages existing datasets to address inquiries distinct from the original research objectives (Tripathy, 2013). The dataset is obtained from Kaggle, a reputable data science platform, offering a rich source of information. The decision to use secondary data is strategic, streamlining access to extensive datasets that would otherwise be challenging or time-consuming to collect data independently. Data from the source are raw data which they required pre-processing in advance.

For fraud detection datasets, it's crucial to adhere to several mandatory requirements when selecting the appropriate data. Firstly, the total transaction count must surpass one million, ensuring an adequate volume for robust analysis. Additionally, the provided transaction period should extend over a minimum of three months, allowing for comprehensive trend analysis and pattern recognition. Moreover, the dataset should include essential variables such as transaction amount, transaction time, customer ID, and terminal ID, which are imperative for accurate fraud detection algorithms.

3.1.1 Source of the Secondary Data

The secondary data for this study is sourced from Kaggle, a prominent platform for data science competitions. Kaggle provides an arena where data scientists and machine learning professionals engage in competitions to develop the most effective models for specific problem-solving or data analysis challenges. Kaggle offers access to public datasets, machine learning notebooks, and tutorials, serving as a valuable resource for learning and honing skills in data science and machine learning (What is Kaggle and what is it used for? 2023). Established in 2010, Kaggle became a part of Google Cloud following its acquisition by Google in 2017. The data contributor for this study is Old Monk, a student at the University of Texas, Austin.

3.2 Sampling design

While sampling design typically involves considerations of the target population, sampling frame, and sampling technique, the use of secondary data in this study precludes the determination of these factors. Given the intricate nature of fraud detection, demanding an in-depth analysis of transaction patterns, a dataset covering a six-month period was selected from Kaggle. This dataset comprises an extensive volume of transaction data sourced from Axis Bank Limited, a prominent financial institution situated in India. Axis Bank stands as the third-largest private sector bank in India by assets and the fourth largest by market capitalization (Axis Bank Limited, n.d.). With its headquarters in Mumbai, Maharashtra, the bank offers financial services to a diverse range of entities, including large and mid-size companies, SMEs, and retail businesses.

The dataset comprises a total of 1,754,155 transactions, containing variables such as TRANSACTION_ID, CUSTOMER_ID, TERMINAL_ID, TX_AMOUNT, TX_TIME_SECONDS, TX_TIME_DAYS, and TX_FRAUD. Licensed under the Database Contents License (DbCL) v1.0, this dataset is openly accessible for public use, enabling individuals to utilize and redistribute it freely. Open data, characterized by its availability for unrestricted use and redistribution, is a valuable resource that encourages collaboration and innovation across various domains, promoting transparency and knowledge sharing.

3.3 Data Processing

This chapter delves into the intricate process of data processing and normalization undertaken by the researcher. From the original pool of eight variables, a meticulous selection of six crucial ones was made to unravel their profound impact on the dependent variable. This careful curation ensures a refined dataset, eliminating irrelevant elements and focusing on those pivotal to the research objectives. Through adept data transformation, irregularities were addressed, elevating the dataset's quality. Notably, the transition from pkl files to Excel, facilitated by Python, opens avenues for detailed analysis and exploration in the upcoming phases of the study.

3.4 Proposed Data Analysis Tool

3.4.1 Descriptive Analysis

Descriptive analytics involves examining current and past data to uncover trends and connections, often considered the foundational level of data analysis due to its focus on illustrating trends without delving into deeper insights. It serves as a valuable tool for conveying changes over time and serves as a starting point for more in-depth analyses that inform decision-making processes (What is descriptive analytics? 5 examples: HBS Online 2021). In product analysis, descriptive methods offer meticulous, accurate, and unbiased sensory information, utilizing humans as measuring instruments in controlled conditions to generate precise data (Kemp et al., 2018).

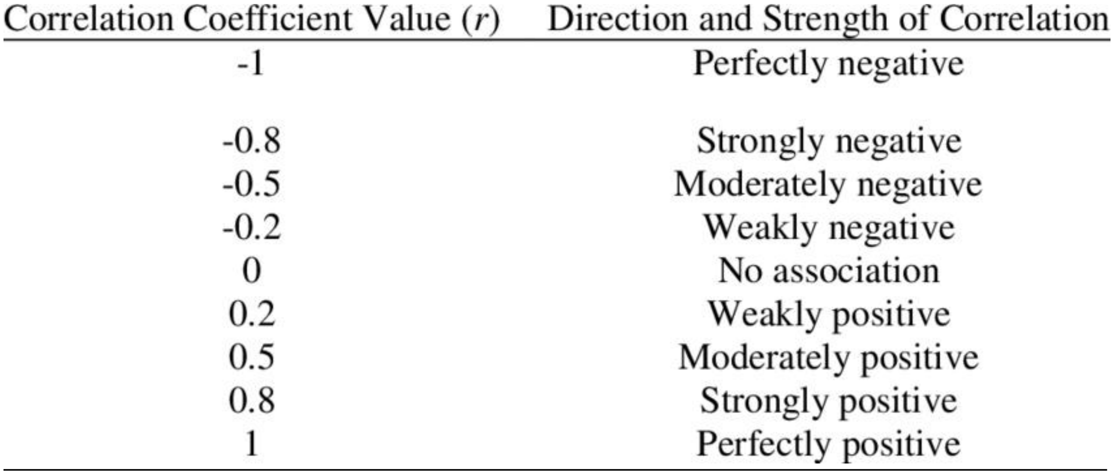
3.4.2 Inferential Analysis

In this study, Pearson Correlation, Spearman Correlation and Logistic regression will be used to examine the secondary data.

3.4.2.1 Pearson Correlation

The Pearson correlation coefficient stands as a statistical measure gauging the strength and direction of a linear connection between two random variables (Zhou et al., 2016). Acknowledged as the premier method for assessing the association between variables, it relies on covariance principles. It provides insights into the magnitude and direction of the correlation (Pearson's correlation coefficient 2021). A negative value indicates an inverse relationship, whereas a positive value signifies a direct relationship between the variables. This metric proves valuable in comprehending the nature and extent of relationships within datasets. The Pearson correlation is a metric quantifying the strength of the linear relationship between two variables. Ranging from -1 to 1, a value of -1 indicates a complete negative linear correlation, 0 signifies no correlation, and +1 represents a total positive correlation. This coefficient provides insights into the direction and intensity of the linear association between the variables (ScienceDirect.com, n.d.).

Figure 3.4.2.1: Rules of thumb about Pearson Correlation



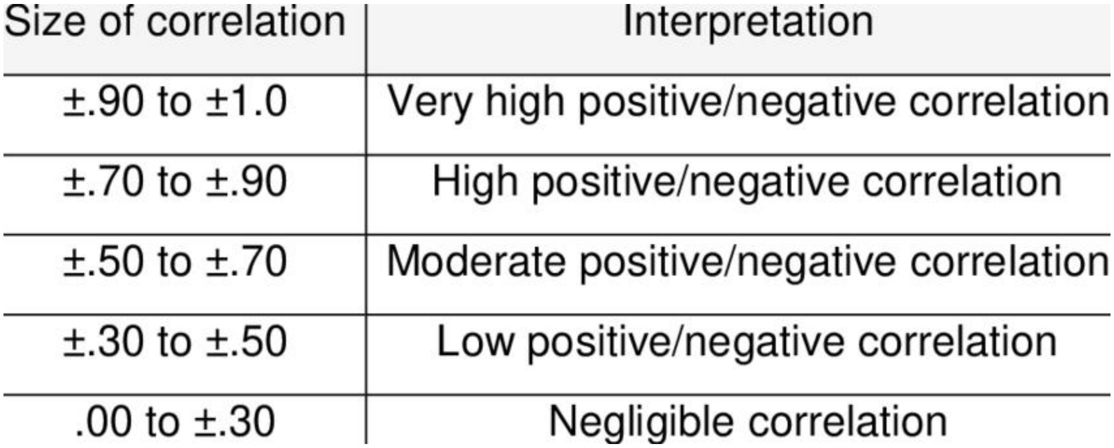
Correlation Coefficient Value (*r*)	Direction and Strength of Correlation
-1	Perfectly negative
-0.8	Strongly negative
-0.5	Moderately negative
-0.2	Weakly negative
0	No association
0.2	Weakly positive
0.5	Moderately positive
0.8	Strongly positive
1	Perfectly positive

Source: (Rule of Thumb for Correlation Coefficient, n.d.)

3.4.2.2 Spearman Correlation

Spearman’s correlation coefficient serves as a statistical measure evaluating the strength of a monotonic relationship between paired data. As a nonparametric rank statistic, it offers a distribution-free approach to quantify the association between two variables. This method proves advantageous when the dataset's distribution makes Pearson’s correlation coefficient inappropriate. Unlike Pearson’s coefficient, Spearman’s does not assume a linear relationship. Instead, it assesses the extent to which an arbitrary monotonic function can depict the association (Hauke & Kossowski, 2011). The closer "rs" is to ±1, the stronger the monotonic relationship, providing a reliable effect size for correlation strength.

Figure 3.4.2.2: Rules of thumb about Spearman Correlation



Size of correlation	Interpretation
±.90 to ±1.0	Very high positive/negative correlation
±.70 to ±.90	High positive/negative correlation
±.50 to ±.70	Moderate positive/negative correlation
±.30 to ±.50	Low positive/negative correlation
.00 to ±.30	Negligible correlation

Source: (Rule of Thumb for Interpreting Spearman’s Correlation Value [41], n.d.)

3.4.2.3 Logistic Regression

Logistic regression, a versatile modeling technique, is particularly well-suited for scenarios involving binary or dichotomous outcome variables, utilizing a mix of continuous and categorical independent variables. It also extends to the multinomial logistic regression when dealing with outcomes with multiple classes, referred to as the "multivariate case." This statistical method proves instrumental in addressing modeling and discrimination challenges within the marketing domain. Despite its efficacy in statistical analysis over the years, logistic regression has garnered relatively little attention in marketing literature compared to other regression applications. However, it stands out for its ability to generate more appropriate and accurate findings, emphasizing model fit and analysis correctness (Akinci et al., 2007).

Logistic regression facilitates the calculation of odds ratios, especially beneficial when dealing with more than one explanatory variable. This method closely resembles multiple linear regression, with the key distinction being the binomial nature of the response variable. By focusing on the odds ratio of the observed event of interest, logistic regression enables the assessment of each variable's impact. A notable advantage lies in its capacity to avoid confounding effects by concurrently analyzing the association of all variables. This makes logistic regression a valuable tool in navigating the complex landscape of marketing analytics (Sperandei, 2014).

Figure 3.4.2.3: Logistic Regression Overview

$$\frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$

Source: (Logistic Regression Explained – Learn by Marketing, n.d.)

3.5 Summary

This chapter outlines the methodology, emphasizing the use of secondary data. It covers the access and source of this data, employing descriptive and inferential analyses, including Pearson correlation, Spearman correlation, and Logistic regression. The study specifically focuses on six key variables, streamlining the research for a more precise investigation.

Chapter 4 Result and Findings

4.0 Introduction

The outcomes of the data analysis using the regression model that was mentioned in the previous chapter will be the main topic of this chapter. It involves carrying out statistical analyses like Pearson, Spearman, and logistic regression. The purpose of these analysis is to find patterns and connections in the dataset that relate to methods for detecting credit card fraud.

4.1 Data Screening

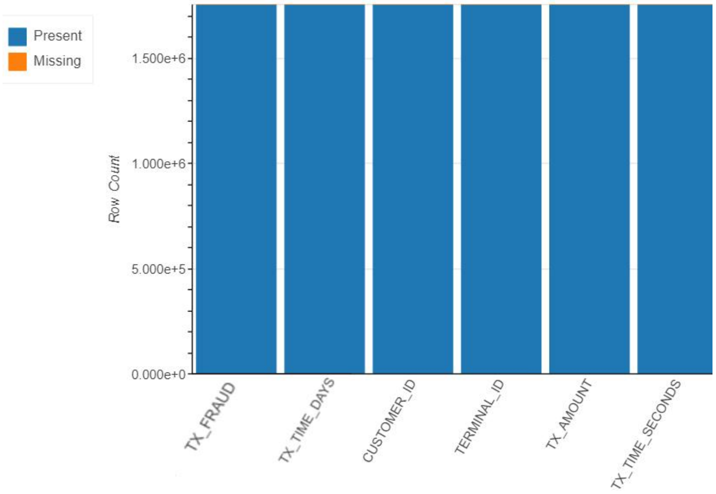
In this study, the author collected 183 days of daily transaction data from Axis Bank in India. All transactions gathered are included for analysis. However, out of the six variables initially considered, only integer variables will be discussed, resulting in a focus on five specific variables in the study.

4.1.1 Data Filtering

Filtering data involves selecting or excluding specific information from a dataset based on predefined criteria. This process is essential for isolating relevant data, eliminating extraneous information, and enhancing overall data quality. Implementing robust filtering strategies is paramount in research to ensure the extraction of accurate and meaningful insights aligned with the research objectives (Collaborators, 2023).

Figure 4.1.1 demonstrates that there are no missing values in the secondary data utilized for this study. All 1,754,155 transaction records have been fully utilized for the research. Furthermore, outliers with extreme values have been excluded from this study to prevent their significant influence on statistical analyses and to maintain the accuracy of hypothesis testing results. Additionally, removing outliers aids in mitigating bias, especially when they stem from measurement errors. This process reduces the potential for bias, resulting in a more accurate and representative analysis of the data.

Figure 4.1.1: Missing Value

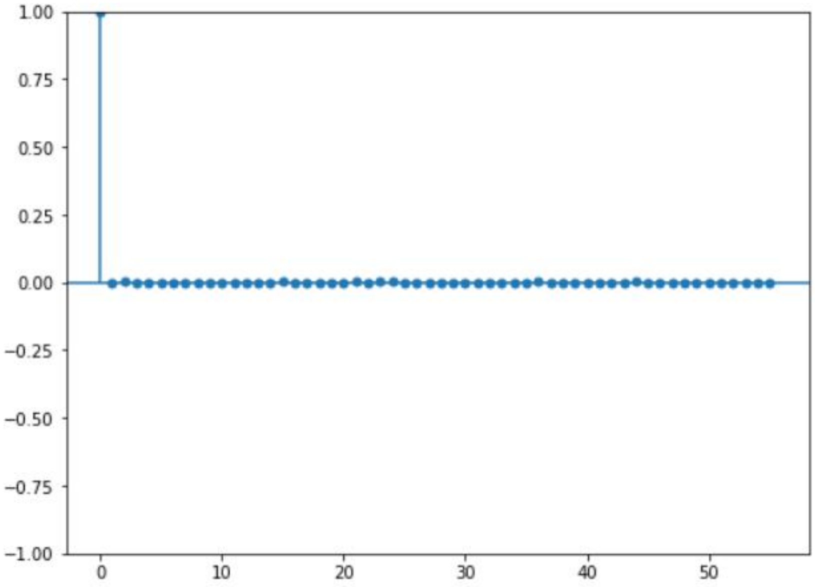


Source: Developed from this research

4.1.2 Autocorrelation

The correlation between a time series and its lag values is measured by autocorrelation. It can be determined whether there are meaningful correlations between the time series values at various time points by looking at the autocorrelation graphic. To identify any underlying patterns, trends, or seasonality in the data, this can be useful. To determine whether the data show serial correlation of any kind—which is crucial for many times series analysis and forecasting methods—autocorrelation charts are frequently utilized. It assists in ensuring that the facts satisfy specific presumptions or qualities required for precise modelling and analysis (Taylor, 2023).

Figure 4.1.2: Autocorrelation



Source: Developed from this research

4.2 Descriptive Analysis

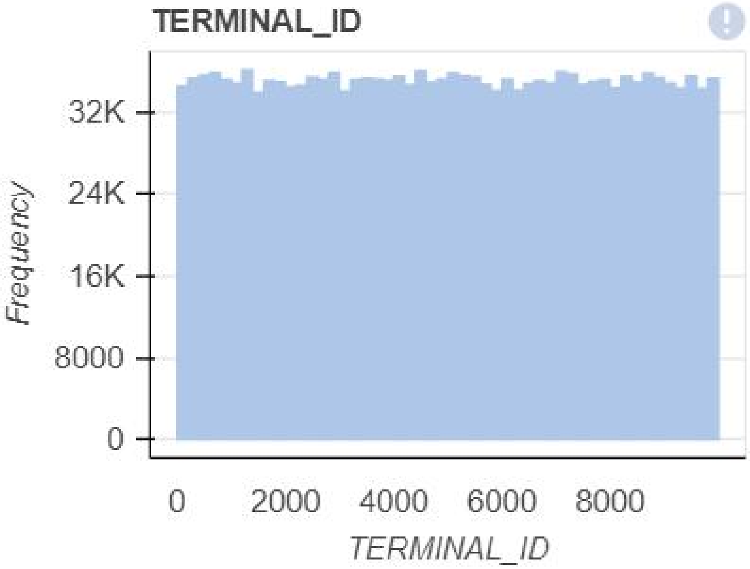
In this research, descriptive analysis will focus on examining five variables: Terminal ID, Transaction Amount, Transaction Time (Second), Transaction Time (Day), and Customer ID. This analysis will entail interpreting their Mean, Standard Deviation, Variance, Sum, and skewness. Descriptive analysis plays a pivotal role in statistical data analysis, as it offers insights into data distribution, aids in identifying errors and outliers, and facilitates the detection of patterns among variables. This phase of analysis is instrumental in preparing for future statistical investigations.

Table 4.2: Descriptive Analysis of the Variables

Variables / Descriptive Statistic	Terminal ID	Transaction Amount (TX Amount)	Transaction Time - Seconds (TX Time Second)	Transaction Time - Days (TX Time Day)	Customer ID
Mean	4996.7332	53.6323	7.9032×10^{06}	90.9726	2504.0114
Standard Deviation	2886.1005	42.3265	4.5652×10^{06}	52.8371	1445.9869
Variance	8.3296×10^{06}	1791.5319	2.0841×10^{13}	2791.7582	2.0909×10^{06}
Sum	8.765×10^{09}	9.4079×10^{07}	1.3863×10^{13}	1.5958×10^{08}	4.3924×10^{09}
Skewness	0.00038577	2.4439	0.00097223	0.00097849	0.001931
Kurtosis	-1.1993	35.2392	-1.2008	-1.2008	-1.2019
Coefficient of Variance	0.5776	0.7892	0.5776	0.5808	0.5775

Source: Developed from this research

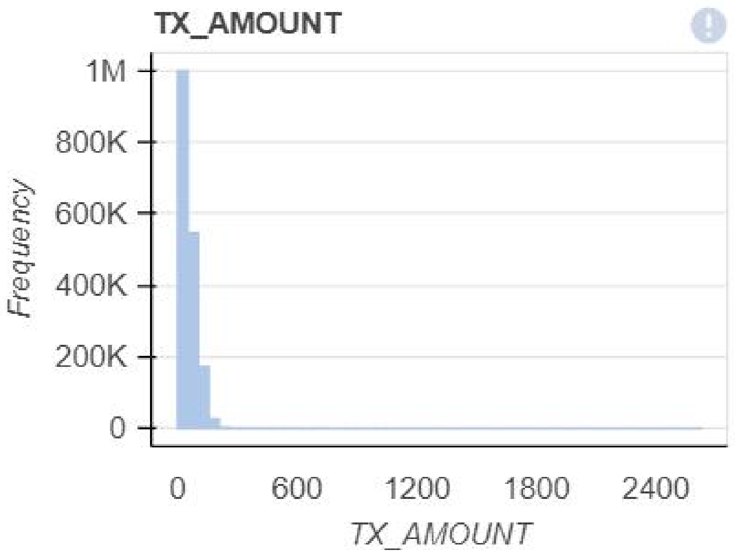
Figure 4.2: Terminal ID



Source: Developed from this research

The quantile statistic and descriptive statistic for the variable terminal ID are displayed in Table 4.2 and Figure 4.2.1. The sampling variability of the parameter, or the standard deviation, is 2886.1005, whereas the mean is 4996.7332. The variable is right skewed.

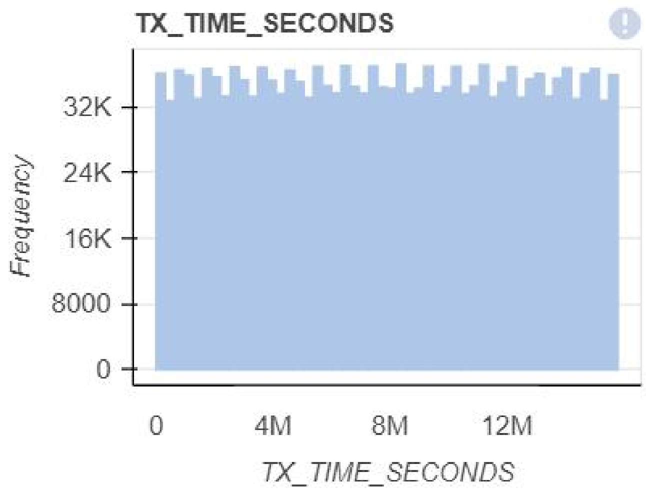
Figure 4.2: Transaction Amount



Source: Developed from this research

The quantile statistic and descriptive statistic for the variable transaction amount are displayed in Table 4.2 and Figure 4.2.2. The sampling variability of the parameter, or the standard deviation, is 42.3265, whereas the mean is 53.6323. The variable is right skewed.

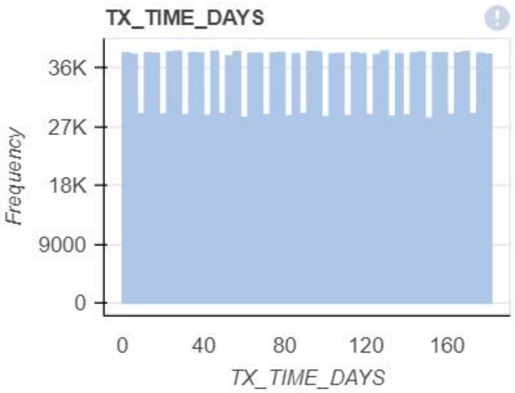
Figure 4.2: Transaction Time – Second



Source: Developed from this research

The quantile statistic and descriptive statistic for the variable transaction time - second are displayed in Table 4.2 and Figure 4.2.3. The sampling variability of the parameter, or the standard deviation, is 4.5652 × 10^06, whereas the mean is 7.9032 × 10^06. The variable is right skewed.

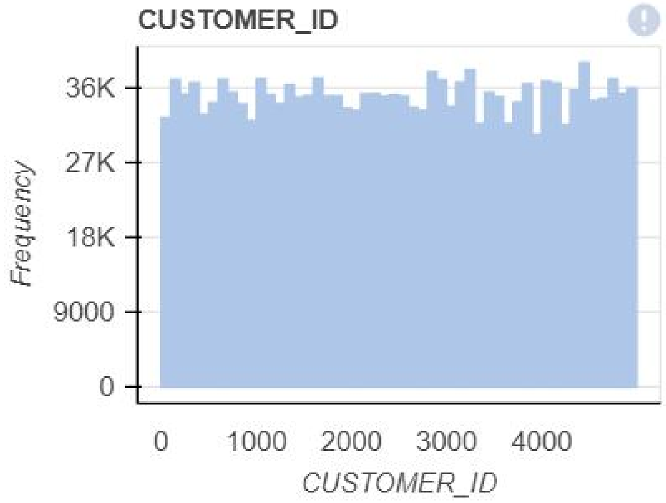
Figure 4.2: Transaction Time – Day



Source: Developed from this research

The quantile statistic and descriptive statistic for the variable transaction time - day are displayed in Table 4.2 and Figure 4.2.4. The sampling variability of the parameter, or the standard deviation, is 52.8371, whereas the mean is 90.9726. The variable is right skewed.

Figure 4.2: Customer ID

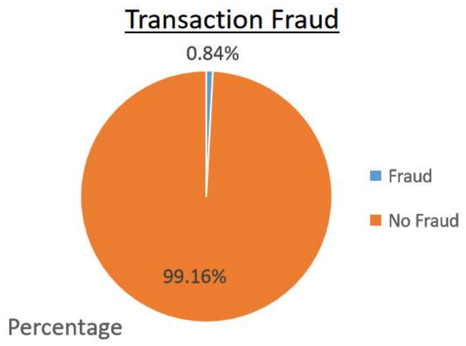


Source: Developed from this research

The quantile statistic and descriptive statistic for the variable Customer ID are displayed in Table 4.2 and Figure 4.2.5. The sampling variability of the parameter, or the standard deviation, is 1445.9869, whereas the mean is 2504.0114. The variable is right skewed.

4.2.1 Transaction Fraud

Figure 4.2.1: Pie Chart of Transaction Fraud



Source: Developed from this research

Figure 4.2.6 illustrates that out of the total transactions, 1,739,474 (99.16%) are not flagged as fraudulent, while 14,681 (0.84%) of the total transactions are flagged as fraudulent.

4.3 Inferential Analysis

4.3.1 Pearson Correlation

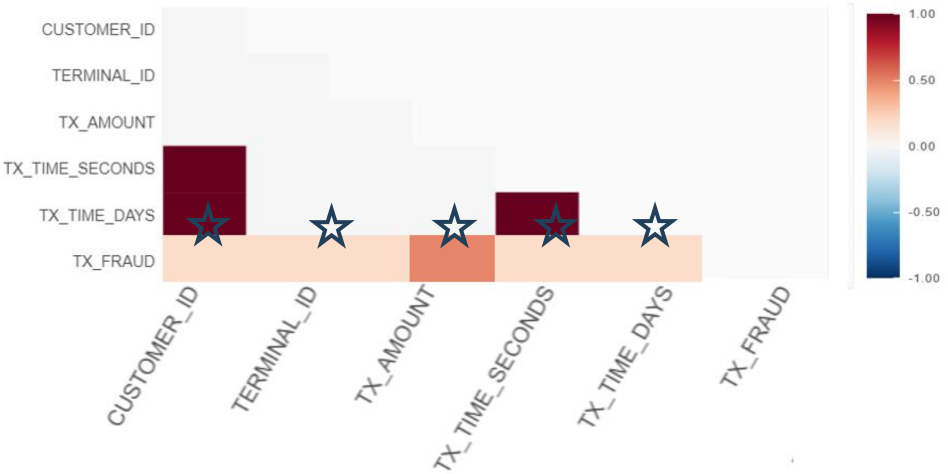
The Pearson correlation coefficient (r) measures the linear correlation between two variables, ranging from -1 to +1. A value of -1 indicates a perfect negative linear correlation, 0 represent no linear correlation, and 1 signifies a perfect positive linear correlation. Importantly, r remains unchanged under separate changes in the location and scale of the variables, meaning that variations in the angle to the x-axis do not affect its value. To compute r for two variables X and Y, one divides their covariance by the product of their standard deviations.

Table 4.3.1: Pearson Correlation

		Terminal ID	Transaction Amount (TX Amount)	Transaction Time - Seconds (TX Time Second)	Transaction Time - Days (TX Time Day)	Customer ID
Transaction Fraud	Pearson Correlation	0.19	0.5	0.13	0.15	0.18
	N	1754155	1754155	1754155	1754155	1754155

Source: Developed from this research

Figure 4.3.1: Pearson Correlation Heat Map



Source: Developed from this research

4.3.2 Spearman Correlation

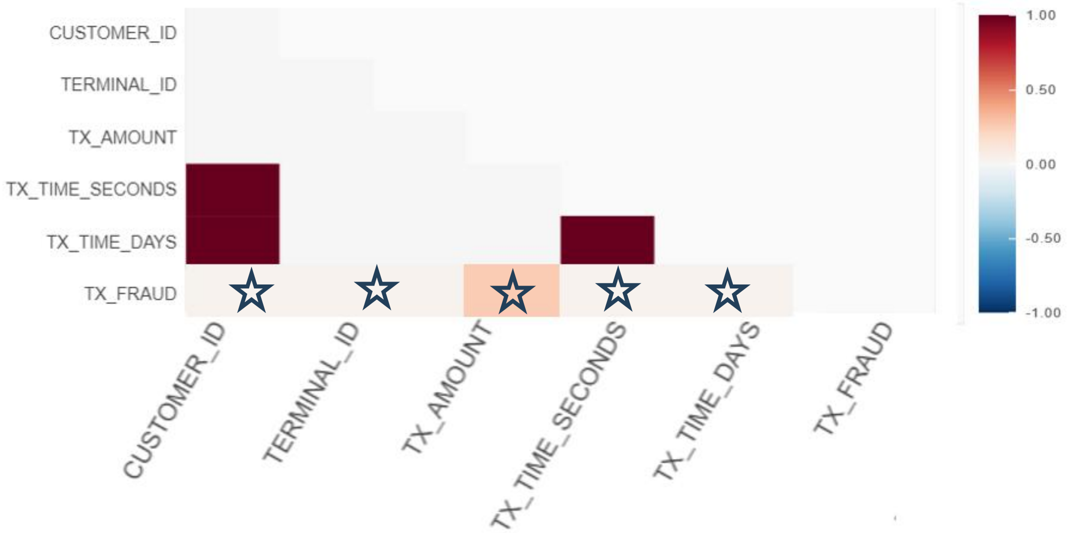
The Spearman's rank correlation coefficient (ρ) assesses the monotonic correlation between two variables, making it more adept at capturing nonlinear monotonic correlations compared to Pearson's r. Its range spans from -1 to +1, with -1 signifying a complete negative monotonic correlation, 0 indicating no monotonic correlation, and 1 denoting a full positive monotonic correlation. To compute ρ for two variables X and Y, the covariance of their rank variables is divided by the product of their standard deviations (Looney & Hagan, 2011).

Table 4.3.2: Spearman Correlation

		Terminal ID	Transaction Amount (TX Amount)	Transaction Time - Seconds (TX Time Second)	Transaction Time - Days (TX Time Day)	Customer ID
Transaction Fraud	Spearman Correlation	0.07	0.3	0.03	0.05	0.06
	N	1754155	1754155	1754155	1754155	1754155

Source: Developed from this research

Figure 4.3.2: Spearman Correlation Heat Map



Source: Developed from this research

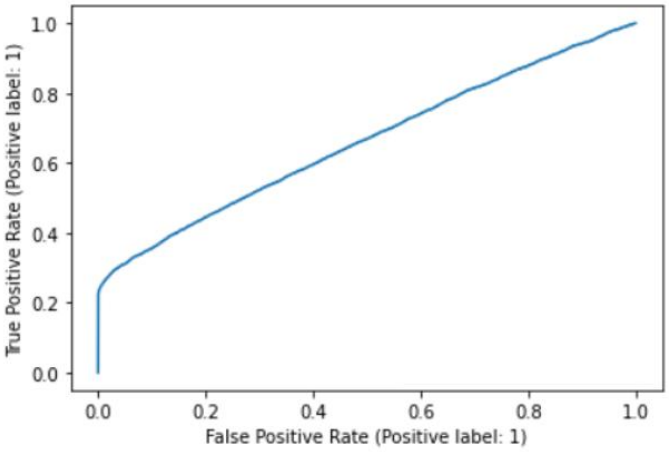
4.3.3 Logistic Regression

Logistic regression is a statistical technique used to predict the likelihood of a binary outcome based on one or more predictor variables. In hypothesis testing, logistic regression allows researchers to examine the relationship between these predictors and the binary outcome, helping to determine whether there is a statistically significant association. By formulating hypotheses about the impact of specific variables on the outcome, researchers can use logistic regression to assess these hypotheses and draw conclusions about the relationship between variables and the likelihood of the outcome occurring (Hoffman, 2019).

Logistic regression was originally developed to classify binary outcomes using multiple categorical or continuous independent variables. It predicts probabilities based on maximum-likelihood estimations, where each probability (denoted as 'py') falls between 0 and 1. A cutoff is then applied to these probabilities to classify samples into different groups, with the cutoff value depending on the specific objectives of the analysis (Rezaei & Jabbari, 2022).

For logistic regression analysis, all five variables—Terminal ID, Transaction Amount, Transaction Time (Second), Transaction Time (Day), and Customer ID—were selected based on their Pearson and Spearman correlation coefficients exceeding 0. Given the dataset's size and the relatively low incidence of fraudulent transactions, a train-test ratio of 60/40 will be employed to ensure an adequate representation of fraud cases in the test set. Logistic regression was chosen for classification due to its simplicity in parameterization and suitability for binary outcomes.

Figure 4.3.3: Logistic Regression Result



Source: Developed from this research

Table 4.3.3: Logistic Regression Result

Threshold	0.9582 (95.82%)
Test Set Transaction	701662
Number of Fraudulent Transaction	5870
Number of Fraudulent Transactions Flagged	5430
Number of Non-Fraudulent Transactions Flagged	26
Percent of Fraud Correctly Flagged	92.50%

Source: Developed from this research

The model's performance is notable, with a probability threshold of 95.82% correctly flagging approximately 92% of fraudulent transactions. While this outcome is satisfactory for the current analysis, it's essential to recognize that fraud detection is a multifaceted issue that may require the application of various supervised and unsupervised machine learning models for comprehensive resolution.

4.4 Hypothesis Testing

Hypothesis Testing	Pearson & Spearman Correlation	Conclusion
H_1: There is a relationship between Terminal ID and Transaction Fraud	> 0	Supported
H_2: There is a relationship between Transaction Amount and Transaction Fraud	> 0	Supported

H_3: There is a relationship between Transaction Time (Second) and Transaction Fraud	> 0	Supported
H_4: There is a relationship between Transaction Time (Day) and Transaction Fraud	> 0	Supported
H_5: There is a relationship between Customer ID and Transaction Fraud	> 0	Supported

H_1: There is a positive relationship between Terminal ID and Transaction Fraud

Supported H_1 as both Spearman and Pearson Correlation value is larger than 0. Therefore, H_1 is supported.

H_2: There is a positive relationship between Transaction Amount and Transaction Fraud

Supported H_2, as both Spearman and Pearson Correlation value is larger than 0. Therefore, H_2 is supported.

H_3: There is a positive relationship between Transaction Time (Second) and Transaction Fraud

Supported H_3, as both Spearman and Pearson Correlation value is larger than 0. Therefore, H_3 is supported.

H_4: There is a positive relationship between Transaction Time (Day) and Transaction Fraud

Supported H_4, as both Spearman and Pearson Correlation value is larger than 0. Therefore, H_4 is supported.

H_5: There is a positive relationship between Customer ID and Transaction Fraud

Supported H_5, as both Spearman and Pearson Correlation value is larger than 0. Therefore, H_5 is supported.

4.5 Summary

In chapter 4, the result and findings of the particular variables have been showcased by utilizing descriptive analysis. Inferential analysis helps to provide a better understanding of the relationship between each variable.

Chapter 5 Discussion and Conclusion

5.1 Discussion on 1st Research Objective – To investigate the factors affecting credit card fraud detection

Based on the findings from the previous chapter, all variables included in this study demonstrate statistical significance. However, the strength of their relationship with the dependent variable is determined by their correlation coefficients. Among the five variables examined, Transaction Amount (TX-Amount) emerges as the most influential predictor of Transaction Fraud. This finding underscores the importance of Transaction Amount in identifying fraudulent transactions within the dataset based on its robust correlation with the dependent variable.

Transaction Amount serves as a reliable indicator for detecting fraudulent transactions due to its correlation with spending patterns and the purchasing behavior of genuine credit card holders. Anomalies or inconsistencies in spending patterns can often reveal potential instances of credit card fraud, making Transaction Amount a crucial factor in fraud detection. While various factors contribute to identifying credit card fraud, Transaction Amount remains a significant component in the detection process.

The announcement by the Bank Negara Governor regarding the reduction of the threshold for banks' daily cash threshold report (CTR) from January 1, 2019, underscores the importance of stringent reporting measures to combat fraudulent activities. This initiative mandates banks to report any cash transaction exceeding RM25,000 in their CTR, while customers must provide additional information for transactions involving cash amounts exceeding this threshold, including the source of funds and purpose of the transaction (Bank Negara set new limit for transactions, 2019).

Heryadi and Warnars (2017) utilized various models to identify fraud, ultimately concluding that the CNN model exhibits high performance in recognizing fraudulent data. This success suggests that the transaction features proposed by the authors effectively capture most financial transaction patterns, particularly those characterized by short-term relationships. The prevalence of strong short-term financial transactions implies that typical fraudulent actors exploit stolen cards promptly, underscoring the importance of transaction amounts in fraud detection.

The remaining variables in this study serve as supportive factors in identifying fraudulent transactions, highlighting the necessity of a multifaceted approach to fraud detection. Sole reliance on transaction amount may not suffice in determining the validity of credit card transactions. Fang et al. (2019) introduces additional factors contributing to credit card fraud detection, including customer age, zip code, consumption category, and more. This underscores the complexity of fraud detection, indicating that a single factor or machine learning model may not be adequate. Increasing the utilization of machine learning models and incorporating diverse factors could ideally enhance the accuracy of fraud detection methodologies.

5.2 Discussion on 2nd Research Objective - To identify the numbers of transaction that are in fraud.

Table 5.2: Logistic Regression Result

Threshold	0.9582 (95.82%)
Test Set Transaction	701662
Number of Fraudulent Transaction	5870
Number of Fraudulent Transactions Flagged	5430
Number of Non-Fraudulent Transactions Flagged	26
Percent of Fraud Correctly Flagged	92.50%

Source: Developed from this research

In this study, 5430 fraudulent transactions out of 5870 were successfully flagged, resulting in a commendable accuracy rate of 92.50% using the logistic regression model. However, 26 non-fraudulent transactions were erroneously flagged. This outcome suggests potential for improvement by incorporating additional factors into the model and adjusting the threshold value, thereby enhancing the precision of fraud detection.

5.3 Implications

5.3.1 Regulator

This study serves as a valuable resource for regulators seeking to comprehend the vulnerabilities within the contemporary digital spending landscape. Credit card fraud not only jeopardizes the public economy but has also emerged as a pervasive global issue expected to persist in the future. Instances of credit card fraud can swiftly inflict financial losses on cardholders, potentially leading to unexpected liabilities arising from unauthorized transactions (Can Being the Victim of Credit Card Fraud Lead to Bankruptcy? 2023).

The evolution of digitalization, tracing back to the 1950s, has revolutionized nearly every aspect of modern life, including work, shopping, banking, and more (Press, 2015). Despite its widespread adoption, the efforts to combat credit card fraud have been notably insufficient. Traditionally, fraud detection methods relied heavily on manual techniques such as discovery sampling, as seen in the works of Tennyson and Forn (2002), which were complex and time-consuming. These methods span multiple disciplines, including economics, finance, law, and business practices. However, the invention of computerized and automated Fraud Detection Systems (FDS) aimed to enhance detection effectiveness. Nevertheless, early FDS implementations were limited by their reliance on predefined rules established by experts (Li et al., 2008). To overcome these limitations, there's a growing need for more sophisticated FDS incorporating diverse data mining methods for robust fraud detection (Abdallah et al., 2016).

This study offers valuable insights into credit card fraud, which can inform regulators, such as the Malaysian Government and Bank Negara, in setting stringent anti-fraud standards. These measures could include allowing transactions only from trusted sources and accounts, as well as imposing daily transaction limits to enhance credit card user protection. With five variables identified as significant contributors to credit card fraud in this study, regulators could propose comprehensive rules and regulations targeting these variables to effectively monitor and mitigate credit card fraud in the future. Based on the findings of this study, it appears that the current methods employed for credit card fraud detection can continue to be utilized, as the factors or parameters influencing fraud remain unchanged even after the COVID-19 period.

5.3.2 Academia

This research contributes to educational resources in the fields of computing and business within academia. The insights and data presented offer valuable learning materials for students, financial institutions, and researchers interested in fraud detection. By providing real-world examples and case studies, this research enhances understanding of statistical methods and their applications in financial security. As digital payment methods continue to mature, ensuring financial security becomes increasingly crucial in safeguarding the public and users.

Additionally, this research contributes to the formulation of government policies and regulatory standards. The insights into the significance of specific variables in fraud detection can inform policymakers and regulatory bodies, prompting the development of more robust regulations and standards within the financial industry. Academics across the nation can play a crucial role in shaping evidence-based policies aimed at preventing credit card fraud and safeguarding consumers. According to Patel (2023), the integration of automated compliance tools into workflows can monitor and ensure adherence to regulations, reducing compliance risks for institutions. Continuous education and training of staff enhance their ability to adapt to evolving regulatory environments, further strengthening compliance measures.

This research significantly contributes to the advancement of fraud detection techniques, particularly in exploring the effectiveness of logistic regression in detecting credit card fraud. The findings serve as valuable insights for academics in refining and developing fraud detection models, enabling the incorporation of new variables, and enhancing predictive accuracy. Additionally, the utilization of machine learning, especially with larger datasets, mitigates biased decisions and improves the accuracy of fraud detection models. With sufficient data, machine learning models can make highly accurate decisions when deployed, ensuring better performance and reliability in real-world applications. Conversely, limited data during model training may result in lower accuracy and potentially irrelevant outcomes during testing, highlighting the importance of comprehensive datasets for effective fraud detection ("How to Improve Accuracy in Machine Learning Models - Audio AI - Medium," 2023).

5.3.3 Financial Institutions

This research holds significant benefits for financial institutions, particularly in enhancing the detection of credit card fraud. As key players in the financial ecosystem, institutions like licensed banks in Malaysia are at the forefront of monitoring and identifying fraudulent transactions. The insights provided by this research offer critical factors that can aid financial institutions in recognizing and addressing instances of credit card fraud effectively. Given their direct involvement in managing credit card transactions, banks can leverage this research to enhance their fraud detection capabilities, thereby safeguarding both their own interests and the financial security of their customers.

The Bank Negara Malaysia article underscores the imperative for licensed banks to uphold rigorous security standards, especially concerning internet and mobile banking services. Mandates from BNM necessitate ongoing security assessments and the implementation of measures to fortify existing controls, ensuring robust protection against emerging threats while ensuring seamless service delivery to customers. Given the dynamic nature of fraud risk, characterized by evolving tactics employed by scammers to deceive individuals into disclosing banking details or installing malicious software, both financial institutions and regulatory bodies remain vigilant in combating new modus operandi (Financial fraud alerts - Bank Negara Malaysia 2022). According to the results obtained from this study, the existing practices employed in credit card fraud detection seem to remain effective even in the aftermath of the COVID-19 period. This suggests that the factors or parameters influencing fraud have not undergone significant changes despite the disruptions caused by the pandemic so that financial institution can still pay high attention towards the vital factor affecting credit card fraud.

5.4 Limitation and Recommendation

5.4.1 Accuracy of the Machine Learning

The classification threshold in machine learning sets a crucial boundary, acting as a cut-off point to determine the predicted class for each object. Especially in probabilistic machine learning models, where labels are not directly assigned, thresholds play a pivotal role. Instead of assigning labels outright, these models predict the probability of a specific class or outcome, typically outputting scores ranging from 0 to 1. By establishing a threshold, typically above which a data point is classified into a particular category, these probability scores are translated into actionable classification decisions.

In this specific scenario, both false positives and false negatives carry significant costs. False positives, mistakenly flagging legitimate transactions as fraudulent, can inconvenience customers and disrupt their transactions. Conversely, false negatives, failing to detect actual fraudulent transactions, can lead to financial losses and erode trust among customers. Therefore, future studies must aim to strike a balance between precision and recall. While maximizing the detection of fraudulent transactions is crucial, it's equally important to minimize the occurrence of false positives. This underscores the importance of carefully setting the threshold for identifying fraudulent transactions, considering the potential costs associated with both types of errors (How to Use Classification Threshold to Balance Precision and Recall, n.d.).

5.4.2 Sample Source

The source of the sample data significantly influences the reliability of fraud detection using machine learning. While this research incorporates a substantial dataset comprising 1,754,155 transactions, sourced exclusively from Axis Bank in India, it may encounter limitations in representing a broader spectrum of real-world scenarios. As Kezmann (2023) emphasizes, diversifying the data sources is crucial to ensure the dataset's representativeness of real-world challenges. This entails gathering data from various locations, demographics, and timeframes, thereby mitigating biases towards specific groups or domains. A diverse dataset facilitates more comprehensive learning for the model, reducing the risk of overfitting to data patterns.

In future studies, employing a more diverse dataset could help address the limitations encountered in the current study. Researchers stand to attain greater accuracy and representation, leading to improved understanding and predictive capabilities regarding credit card fraud in machine learning. Moreover, diverse data sources offer access to a variety of information, including user behaviour, spending patterns, and geographic data. This expanded dataset facilitates the identification of narrow patterns and correlations that may not be readily visible in more restricted datasets.

5.4.3 Analysis Method

Furthermore, the analysis methods used in this study may be considered inadequate. Specifically, Pearson correlation, Spearman correlation, and logistic regression were employed to investigate five hypotheses with a sample size of 1,754,155 observations. It is recommended to incorporate additional analysis methods to thoroughly examine the data. Employing multiple analysis techniques is advantageous as it fosters a more comprehensive understanding of the dataset and helps validate the results. Each analytical approach has its own strengths and limitations, and by using a variety of methodologies, researchers can strengthen their findings and enhance the reliability of their conclusions.

As per Davis et al. (2010), employing multiple methods in research provides a promising solution to address criticisms related to method bias in marketing research. Authors conducting studies using multiple methods should consider adopting a uniform approach in reporting their findings. This consistency can help mitigate any potential biases or limitations associated with relying solely on a single method. Ultimately, incorporating diverse analysis methodologies can enhance the robustness and credibility of study outcomes.

5.5 Summary

In the last chapter, we revise the research objectives that we stated in the first chapter. List of implications of this study are provided as well as the limitations and recommendations towards the future studies.

REFERENCES

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007

Akinci, S., Kaynak, E., Atilgan, E., & Aksoy, Ş. (2007). Where does the logistic regression analysis stand in marketing literature? *European Journal of Marketing*, *41*(5/6), 537–567. https://doi.org/10.1108/03090560710737598

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of Physics: Conference Series*, *1142*, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

Barker, K. J., D’Amato, J., & Sheridon, P. (2008). Credit Card Fraud: Awareness and Prevention. *Journal of Financial Crime*, *15*(4), 398–410. https://doi.org/10.1108/13590790810907236

Bloomenthal, A. (2023, May). *Credit card: What it is, how it works, and how to get one*. Investopedia. https://www.investopedia.com/terms/c/creditcard.asp

Bolton, R. J., & Hand, D. J. (2002). Unsupervised Profiling Methods for Fraud Detection. Department of Mathematics Imperial College London . http://procon.bg/article/unsupervised-profiling-methods-fraud-detection

Brei, M., Jacolin, L., & Noah, A. (2020). Credit risk and bank competition in Sub-Saharan africa. *Emerging Markets Review*, *44*, 100716. https://doi.org/10.1016/j.ememar.2020.100716

Can being the victim of credit card fraud lead to bankruptcy?. AprilRandleLaw. (2023, September 18). https://www.aprilrandlelaw.com/can-being-the-victim-of-credit-card-fraud-lead-to-bankruptcy

Cao, B., Mao, M., Viidu, S., & Yu, P. S. (2017). Collective Fraud Detection Capturing Inter-Transaction Dependency. *Knowledge Discovery and Data Mining*.

Caringal, C. (2023, October 9). *What is transaction per second (TPS)? impact and significance*. HeLa is a Modular. https://helalabs.com/blog/what-is-transaction-per-second-impact-and-significance/

Carminati, M., Caron, R., Maggi, F., Epifani, I., & Zanero, S. (2014). Banksealer: An online banking fraud analysis and decision support system. ICT Systems Security and Privacy Protection, 380–394. https://doi.org/10.1007/978-3-642-55415-5_32

Chargeflow. (2023, July 7). *Terminal ID number: Track transactions and prevent disputes*. RSS. https://www.chargeflow.io/blog/terminal-id-number#:~:text=Authentication%20and%20Verification%3A%20Terminal%20ID,comes%20from%20a%20genuine%20source.

Chaudhary, K., Yadav, J., & Mallick, B. (2012). A review of Fraud Detection Techniques: Credit Card. International Journal of Computer Applications. https://research.ijcaonline.org/volume45/number1/pxc3878991.pdf

Chen, J. (2022). *What is fraud? definition, types, and consequences*. Investopedia. https://www.investopedia.com/terms/f/fraud.asp

Clarke, B., Fokoue, E., & Zhang, H. H. (2009). Principles and theory for Data Mining and machine learning. *Springer Series in Statistics*. https://doi.org/10.1007/978-0-387-98135-2

Collaborators, Q. (2023, November 27). *Data filtering: What it is, uses, benefits and example*. https://www.questionpro.com/blog/data-filtering/#:~:text=Data%20filtering%20is%20a%20key,obtaining%20accurate%20and%20insightful%20metrics.

Correia, I., Fournier, F., & Skarbovsky, I. (2015). The uncertain case of credit card fraud detection. Proceedings of the 9th ACM International Conference on Distributed Event-Based Systems. https://doi.org/10.1145/2675743.2771877

CrowdStrike. (2023, June 8). *Keyloggers: How they work & how to detect them* . crowdstrike.com. https://www.crowdstrike.com/cybersecurity-101/attack-types/keylogger/

Customer ID definition: 108 samples. Law Insider. (n.d.). https://www.lawinsider.com/dictionary/customer-id

DataVisor. (2023, May 10). *Application fraud*. https://www.datavisor.com/wiki/application-fraud/#:~:text=Application%20fraud%20is%20where%20a,and%20higher%20lines%20of%20credit.

Davis, D. F., Golicic, S. L., & Boerstler, C. N. (2010). Benefits and challenges of conducting multiple methods research in marketing. *Journal of the Academy of Marketing Science*, *39*(3), 467–479. https://doi.org/10.1007/s11747-010-0204-7

Decorte, D. (2022, November 10). *How transaction IDS help merchants prevent fraud*. Chargebacks911. https://chargebacks911.com/transaction-id/

Donepudi, P. K. (2019). Automation and machine learning in transforming the financial industry. *Asian Business Review*, *9*(3), 129–138. https://doi.org/10.18034/abr.v9i3.494

Fang, Y., Zhang, Y., & Huang, C. (2019). Credit card fraud detection based on machine learning. *Computers, Materials &amp; Continua*, *61*(1), 185–195. https://doi.org/10.32604/cmc.2019.06144

Financial fraud alerts - bank Negara Malaysia. (2022, June). https://www.bnm.gov.my/financial-fraud-alerts

Foote, K. D. (2023, May 4). *A brief history of machine learning*. DATAVERSITY. https://www.dataversity.net/a-brief-history-of-machine-learning/

Gadi, M. F., Wang, X., & do Lago, A. P. (2008). Credit card fraud detection with artificial immune system. *Lecture Notes in Computer Science*, 119–131. https://doi.org/10.1007/978-3-540-85072-4_11

GeeksforGeeks. (2022, July 22). *Difference between phishing and Vishing*. https://www.geeksforgeeks.org/difference-between-phishing-and-vishing/

Ghosh Dastidar, K., Jurgovsky, J., Siblini, W., He-Guelton, L., & Granitzer, M. (2020). Nag: Neural feature aggregation framework for credit card fraud detection. *2020 IEEE International Conference on Data Mining (ICDM)*. https://doi.org/10.1109/icdm50108.2020.00018

Gupta, D. (2023, July 18). *“performance testing terminologies.”* Medium. https://medium.com/@guptadiksha88/performance-testing-terminologies-e0293c9aafbd

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *QUAGEO*, *30*(2), 87–93. https://doi.org/10.2478/v10117-011-0021-1

Heryadi, Y., & Warnars, H. L. (2017). Learning temporal representation of transaction amount for fraudulent transaction recognition using CNN, stacked LSTM, and CNN-LSTM. *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*. https://doi.org/10.1109/cyberneticscom.2017.8311689

Hoffman, J. I. E. (2019). Logistic regression. *Basic Biostatistics for Medical and Biomedical Practitioners*, 581–589. https://doi.org/10.1016/b978-0-12-817084-7.00033-4

How exponential smoothing forecast works. How Exponential Smoothing Forecast works-ArcGIS Pro | Documentation. (n.d.). https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/learnmoreexponentialsmoothingforecast.htm#:~:text=Exponential%20smoothing%20is%20one%20of,given%20number%20of%20time%20steps.

How to use classification threshold to balance precision and recall. Evidently AI - Open-Source ML Monitoring and Observability. (n.d.). https://www.evidentlyai.com/classification-metrics/classification-threshold#:~:text=Adjusting%20this%20threshold%20can%20affect,increases%20recall%20but%20decreases%20precision.

http://proceedings.mlr.press/v71/cao18a/cao18a.pdf
https://www.oxfordlearnersdictionaries.com/definition/english/second1_1?q=Second

Kane. (2017). *Hands-On Data Science and Python Machine Learning: Perform Data Mining and Machine Learning Efficiently Using Python and Spark, Packt Publishing, Birmingham.*

Kemp, S. E., Ng, M., Hollowood, T., & Hort, J. (2018). Introduction to descriptive analysis. *Descriptive Analysis in Sensory Evaluation*, 1–39. https://doi.org/10.1002/9781118991657.ch1

Kezmann, J. M. (2023, May 8). *The DOS and don'ts of dataset selection for machine learning you have to be aware of*. Medium. https://medium.com/mlearning-ai/the-dos-and-donts-of-dataset-selection-for-machine-learning-you-have-to-be-aware-of-8b14513d94a

Kostin, A. (2023, May 4). *Importance of customer identity verification in preventing credit card fraud and protecting your business*. LinkedIn. https://www.linkedin.com/pulse/importance-customer-identity-verification-preventing-credit-kostin#:~:text=Why%20is%20Customer%20Identity%20Verification,the%20credit%20card%20being%20used.

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in Higher Education Institutions. *Sustainability*, *13*(18), 10424. https://doi.org/10.3390/su131810424

Kuziemski, M., & Misuraca, G. (2020). AI governance in the Public Sector: Three tales from the frontiers of Automated Decision-making in Democratic settings. *Telecommunications Policy*, *44*(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976

Limited, I. (2020). *Why does POS fraud affect you now more that ever: Infosys BPM*. https://www.infosysbpm.com/blogs/fraud-retail/why-does-pos-fraud-affect-you-now-more-than-ever.html

Looney, S. W., & Hagan, J. L. (2011). Statistical methods for assessing biomarkers and analyzing biomarker data. *Essential Statistical Methods for Medical Statistics*, 27–65. https://doi.org/10.1016/b978-0-444-53737-9.50005-0

Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., & Calabretto, S. (2020). Towards Automated Feature Engineering for credit card fraud detection using multi-perspective hmms. *Future Generation Computer Systems*, *102*, 393–402. https://doi.org/10.1016/j.future.2019.08.029

Mann, D. (2023, October 20). *What is a transaction ID & how can they help merchants?*. PaymentCloud Blog. https://paymentcloudinc.com/blog/transaction-id/#:~:text=A%20transaction%20ID%20(T%2DID,to%20a%20cryptocurrency%20wallet%20transfer.

Medium. (2023, October 25). *How to improve accuracy in machine learning models*. Medium. https://medium.com/@brainbox.space/how-to-improve-accuracy-in-machine-learning-models-41f26b4a6b11

OxfordLearnersDictionaries. (n.d.).

Panigrahi, S., Kundu, A., Sural, S., & Majumdar, A. K. (2009). Credit Card Fraud Detection: A fusion approach using Dempster–Shafer Theory and Bayesian learning. Information Fusion, 10(4), 354–363. https://doi.org/10.1016/j.inffus.2008.04.001

Patel, K. (2023). Credit Card Analytics: A review of Fraud Detection and risk assessment techniques. *International Journal of Computer Trends and Technology*, *71*(10), 69–79. https://doi.org/10.14445/22312803/ijctt-v71i10p109

Pearson correlation. Pearson Correlation - an overview | ScienceDirect Topics. (n.d.). https://www.sciencedirect.com/topics/computer-science/pearson-correlation#:~:text=The%20Pearson%20correlation%20measures%20the,meaning%20a%20total%20positive%20correlation.

Pearson's correlation coefficient. Statistics Solutions. (2021, June 9). https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/

Personal banking: Internet banking: Corporate, Nri Banking Services Online - Axis Bank. Personal Banking | Internet Banking | Corporate, NRI Banking Services Online - Axis Bank. (n.d.). https://www.axisbank.com/

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*. https://doi.org/10.1109/icicta.2010.831

Pi, Y. (2021). Machine learning in governments: Benefits, challenges and future directions. *JeDEM - eJournal of eDemocracy and Open Government*, *13*(1), 203–219. https://doi.org/10.29379/jedem.v13i1.625

Rezaei, N., & Jabbari, P. (2022). Linear and logistic regressions in R. *Immunoinformatics of Cancers*, 87–125. https://doi.org/10.1016/b978-0-12-822400-7.00004-x

Rule of thumb for correlation coefficient - researchgate. (n.d.). https://www.researchgate.net/figure/Rule-of-Thumb-for-Correlation-Coefficient_tbl1_351236539

Soltani Halvaiee, N., & Akbari, M. K. (2014). A novel model for credit card fraud detection using artificial immune systems. *Applied Soft Computing*, *24*, 40–49. https://doi.org/10.1016/j.asoc.2014.06.042

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12–18. https://doi.org/10.11613/bm.2014.003

Sun, Y., Liu, J., Wang, J., Cao, Y., & Kato, N. (2020). When machine learning meets privacy in 6G: A survey. *IEEE Communications Surveys &amp; Tutorials*, *22*(4), 2694–2724. https://doi.org/10.1109/comst.2020.3011561

Taylor, S. (2023, November 21). *Autocorrelation*. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/data-science/autocorrelation/#:~:text=Autocorrelation%20refers%20to%20the%20degree,it%20in%20a%20time%20series.

The Star Online. (2019, January). *Bank Negara set new limit for transactions*. SIDREC. https://www.sidrec.com.my/announcements/bank-negara-set-new-limit-for-transactions/

The Star. (2022, February 16). *Pandemic has accelerated switch to cashless payment, study finds*. https://www.thestar.com.my/metro/metro-news/2022/02/16/pandemic-has-accelerated-switch-to-cashless-payment-study-finds

Transaction amount definition: 211 samples. Law Insider. (n.d.). https://www.lawinsider.com/dictionary/transaction-amount?cursor=Cl8SWWoVc35sYXdpbnNpZGVyY29udHJhY3RzcjsLEhpEZWZpbml0aW9uU25pcHBldEdyb3VwX3Y0MSIbdHJhbnNhY3Rpb24tYW1vdW50IzAwMDAwMDBhDKIBAmVuGAAgAA%3D%3D

Tripathy, J. P. (2013). *Secondary Data Analysis: Ethical Issues and challenges*. Iranian journal of public health. https://pubmed.ncbi.nlm.nih.gov/26060652

TuxCare. (2023, November 6). *What role does artificial intelligence have in fraud detection?* https://tuxcare.com/blog/what-role-does-artificial-intelligence-have-in-fraud-detection/#:~:text=By%20using%20advanced%20algorithms%2C%20AI,tackling%20fraud%2C%20improving%20detection%20accuracy.

Wang, S. (2010). A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*. https://doi.org/10.1109/icicta.2010.831

What is descriptive analytics? 5 examples: HBS Online. Business Insights Blog. (2021, November 9). https://online.hbs.edu/blog/post/descriptive-analytics

What is Fraud? Definition of Fraud, Fraud Meaning - The Economic Times. The Economic Times. (n.d.). https://economictimes.indiatimes.com/definition/fraud

What is Kaggle and what is it used for?. Coursera. (2023). https://www.coursera.org/articles/kaggle

Zhou, H., Deng, Z., Xia, Y., & Fu, M. (2016). A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing*, *216*, 208–215. https://doi.org/10.1016/j.neucom.2016.07.036