國立臺灣科技大學

管理研究所博士班

博士學位論文

學號：D10816005

商業交易之舞弊偵測

On the Fraud Detection in Business Transactions

研 究 生：郭　勁

指導教授：曾盛恕　博士

中 華 民 國 一 百 一 十 二 年 六 月

D10816005

博士學位論文指導教授推薦書

Doctoral Dissertation Recommendation Form

系所： Department/Graduate Institute	管理研究所 Graduate Institute of Management
姓名： Name	郭勁 CHIN KUO
論文題目： (Dissertation Title)	商業交易之舞弊偵測 On the Fraud Detection In Business Transactions

係由本人指導撰述，同意提付審查。

This is to certify that the dissertation submitted by the student named above, has been written under my supervision. I hereby approve this dissertation to be applied for examination.

指導教授簽章：
Advisor's Signature

共同指導教授簽章（如有）：
Co-advisor's Signature (if any)

日期：
Date(yyyy/mm/dd)

2023/3/23

D10816005

博士學位考試委員審定書

Qualification Form by Doctoral Degree Examination Committee

系所：
Department/Graduate Institute　管理研究所 Graduate Institute of Management

姓名：
Name　郭勁 CHIN KUO

論文題目：
(Dissertation Title)　商業交易之舞弊偵測 On the Fraud Detection In Business Transactions

經本委員會審定通過，特此證明。

This is to certify that the dissertation submitted by the student named above, is qualified and approved by the Examination Committee.

學位考試委員會

Degree Examination Committee

委員簽章：
Member's Signatures

陳家祥　陳武忠

[illegible]　陳宗文

呂志豪　王惠芝

召集人簽章：
Committee Chair's Signature　陳家祥

指導教授簽章：
Advisor's Signature　陳武忠

共同指導教授簽章（如有）：
Co-advisor's Signature (if any)

系所（學程）主任（所長）簽章：
Department/Study Program/Graduate Institute Chair's Signature　[illegible]

日期：
Date(yyyy/mm/dd)　2023/5/5

中 文 摘 要

舞弊是指透過欺騙他人來獲取金錢的犯罪行為，例如：加密貨幣交易所 FTX 的破產事件、德國支付業者 Wirecard 的詐欺事件、中國河南省村鎮銀行存款消失事件、台灣台新銀行理財專員不當挪用客戶資產等，都再再顯示全球舞弊及詐騙行為層出不窮。是故政府機關或私人企業均持續投注資源於舞弊偵防。因此，本論文的目標旨在建立異常交易與投資舞弊行為的檢測模型，並試圖從數據中找出適當的解決方案。其中，本論文共包含了兩部分的研究，第一部分是以犯罪學的理性選擇理論為基礎，從海量交易數據中識別出異常交易的舞弊因子；第二部分則是透過情緒分析的方式，從聊天訊息內容中偵測出詐騙者的舞弊行為。兩項研究最後都透過機器學習的方式來評估檢測模型顯著的效能，進而提供研究人員及舞弊偵防從業人員有例可援。

關鍵字：價格操縱舞弊、職務舞弊、理性選擇理論、社交工程、情緒分析、投資舞弊生命週期、舞弊偵測

ABSTRACT

Fraud refers to the crime of obtaining money by deceiving people, such as the bankruptcy of the FTX cryptocurrency exchange, the Wirecard fraud in Germany, the disappearance of deposits in rural banks in Henan Province, China, and the misappropriation of customer assets by a financial consultant at Taiwan's Taishin Bank that illustrate the persistent issue of fraud and scams are still prevalent worldwide. As a result, government agencies and enterprises have invested significant resources in fraud prevention and detection. Therefore, this dissertation aims to establish fraud detection models and identify appropriate solutions from the data. This dissertation consists of two parts of research. The first part of the study is grounded in the rational choice theory of criminology and identifies fraudulent variables in abnormal transactions from massive transaction data. The second part detects the fraudulent behavior of scammers through emotion analysis of chat room messages. Both parts of this study evaluate the effectiveness of the detection model through machine learning, providing researchers and anti-fraud practitioners with valuable references.

Keywords: Price manipulation fraud; Occupational fraud; Rational choice theory (RCT); Social Engineering; Emotion Analysis; Investment Fraud Life Cycle (IFLC); Fraud detection

Acknowledgment

I would like to express my heartfelt gratitude to my dissertation advisor, Professor Seng-Su Tsang, for his invaluable guidance, encouragement, and patience throughout this process. His insights, feedback, and expertise have been instrumental in my success, and I could not have accomplished this without his support. I also want to extend my deepest appreciation to my family for their unwavering love and support. Their belief in me has provided the strength and motivation to pursue my dreams, and I am grateful for everything they have done for me.

Once again, I would like to thank my advisor and family for being there for me every step of the way.

Contents

List of Figures

List of Tables

1. Introduction

Since the Association of Certified Fraud Examiners (ACFE) published the "Report to the Nations" report twice a year in 1996, the general public has comprehensively understood occupational fraud and abuse. In real-life scenarios, various types of fraud exist, including financial statement fraud, asset misappropriation, corruption, consumer fraud, and cyber fraud (ACFE, 2021). According to Merriam-Webster's dictionary, "fraud" is derived from the Latin term "fraus," meaning deceit or deception. The term first appeared in Middle English in the mid-14th century and was used to describe dishonest or deceptive behavior. In Chinese, "fraud" refers to "弊," meaning cheating, badness, or decay, as defined in the Kangxi Dictionary. In essence, "fraud" denotes human behavior that leads to detrimental outcomes. However, ACFE (2022) identified typical fraud schemes, including financial statement fraud, asset misappropriation, identity theft, investment fraud, and cyber fraud.

Over the past few years, the rapid development of information technology has posed significant challenges and opportunities to various industries. Enterprises must promptly respond to the constantly changing market conditions in the current business environment. Fraudulent practices have been identified as one of the critical issues that enterprises must address. However, enterprises continuously seek innovative approaches, such as indispensable technologies like data analytics and machine learning, to enhance their capabilities. As technology advances, perpetrators' skills also constantly improve. With the growing complexity of workplace systems and ever-changing operational activities, fraud perpetrators have become increasingly adept at evading existing fraud detection technologies. Faced with substantial data, fraud investigators must employ their specialized knowledge and hands-on experience to mitigate and curtail the repercussions of fraudulent occurrences. A survey conducted by Deloitte (2018) found that approximately 10% of respondents reported experiencing more than four instances of fraud per year within their companies, illustrating the prevalence of fraud across various Taiwanese enterprises. Furthermore, 31% of respondents were uncertain about the occurrence of fraudulent incidents, suggesting that the actual incidence of fraud might be more widespread than commonly assumed.

The global economic recession caused by the COVID-19 pandemic has increased individuals seeking alternative investment opportunities, making them vulnerable to scammers (Fraud.org, 2021). In addition, the widespread use of mobile devices has made

it easier for scammers to contact and victimize individuals. In 2020, the Federal Trade Commission (FTC) reported over 2.2 million scam cases, resulting in a loss of US $3.3 million (FTC, 2021). Davies (2021) defines fraud as a pepretrator manipulating victims into giving them money through fraudulent means. Scammers use various methods to manipulate consumers' emotions, leading them to feel stress and confusion, ultimately resulting in their victimization. The occupational fraud case involving a financial consultant at Taiwan's Taishin Bank in 2022 provides a comprehensive illustration of the scheme. According to the civil judgment of the Taiwan High Court[1], the perpetrators abused their position of trust and high-level authority to deceive six high-net-worth individuals over a decade. Specifically, they obtained the seals and signatures of the six customers through fraudulent means, claiming that they could manage their financial affairs. The perpetrators then withdrew the victims' deposits for personal use, amounting to approximately 300 million New Taiwan dollars.

This study proposed two experiments to construct fraud detection models. The first experiment begins with the rational choice theory (RCT) in criminology, exploring the fraudulent behavior of perpetrators manipulating retail prices in stores, namely the anomalous transaction detection model. In the retail industry, Kashyap (2019) identified three sources of fraud: gathering personal information, cybersecurity threats, and internal risks. Internal risk involves employees using internal organizational procedures to perpetrate fraudulent acts. Point of sale (POS) fraud, also known as cash register manipulation, is a widespread fraudulent scheme in the retail sector, where cashiers exploit their roles to adjust merchandise prices for their benefit. This type of fraud belongs to the category of occupational fraud. Common schemes include substituting price tags, modifying sale prices, or entering incorrect merchandise quantities for monetary advantage. Therefore, this study uses real-world sales data from a Taiwanese retailer. First, we extract four fraudulent parameters, namely, holiday promotion, bulk goods, multi-product transactions, and continuing transactions, based on the benefits, risks, and costs that perpetrators will consider before committing fraud according to RCT. Subsequently, the four variables are used to develop an anomalous transaction detection model, and unsupervised learning algorithms such as Bayesian network, Logistic Regression, and Random Forest algorithms are used to evaluate its effectiveness. The results also show that the accuracy of the detection model can be slightly improved by incorporating the four fraudulent parameters. In addition, previous research on fraudulent behavior has revolved chiefly around the fraud triangle theory. However, according to the variable selection method of this study, it is demonstrated that the RCT

[1] The civil judgment of the Taiwan High Court Case of 111 Jin-Su-Zi No. 11

is useful for constructing an analytical model of fraudulent behavior.

The second experiment of this study presents a two-stage research analysis that explores investment fraud behavior from the emotion analysis perspective to construct an investment fraud detection model. Investment fraud is one of the most common types of fraud that governments continually warn individuals to be cautious about. These frauds typically guarantee high returns with no financial risk, similar to Ponzi schemes, which are illegal business practices (Chiluwa, 2019). However, frauds are continually evolving, and the National Police Agency (2021) publicly announced that online shopping and investment fraud were the top fraud in Taiwan in 2020. According to Trend Micro (2021), investment frauds were the most severe in Taiwan in the first quarter of 2021, including scammers pretending to be celebrities and offering penny stocks. Victims often fall into the trap through social media platforms. One of Taiwan's most popular instant messaging (IM) apps launched an open chat function in 2020, allowing strangers to join chat rooms anonymously. However, it is not easy to trace the source of messages, making open chats a hotbed for scammers to evade detection.

Despite the increasing prevalence of investment fraud, previous research has predominantly focused on victims' characteristics, psychology, or emotions, rather than exploring the emotional responses of scammers during the commission of scams (Buchanan & Whitty, 2014; Mueller et al., 2020; Whitty, 2020b). Furthermore, limited research has been conducted on detecting such scams in Mandarin Chinese, particularly in text-based communication in chat rooms. Most existing studies have focused on identifying and classifying scam words, creating a gap in the literature.Therefore, in the first stage, a complete investment fraud life cycle is constructed through a literature review divided into five stages: Lurking, Alluring, Catching, Executing, and Vanishing. Then, by analyzing seven sets of real-world investment fraud chat room messages using emotion analysis, the emotional fluctuations of scammers are examined, and emotions are classified into seven categories: Anger, Disgust, Fear, Sadness, Surprise, Good, and Happiness. The results show that scammers mostly display a "good" emotion. To confirm whether these seven sets of investment fraud chat rooms have the same emotional fluctuations, MANOVA is used for analysis, and the results indicate that the emotional fluctuations of the seven chat rooms have a similar pattern. In the second stage, based on these seven emotional variables, an investment fraud detection model is constructed, and its effectiveness is measured using Support Vector Machine (SVM), Decision Tree, and Random Tree, with an accuracy rate of over 70%. Additionally, the research results confirm that emotion analysis can be used as one of the analytical methods for detecting scam behavior.

2. Related Research

2.1 Price Manipulation Fraud

ACFE (2019) characterizes occupational fraud as the engagement of employees in deceitful actions, including unauthorized modifications to computer systems and asset misappropriation. The problem of occupational fraud transcends industry boundaries, making all organizations vulnerable to such illicit activities. Holtfreter (2005) conducted research on the behavioral traits of individuals engaged in occupational fraud. The study focused on three categories of fraud, namely asset misappropriation, corruption, and fraudulent statements. The study's findings indicated that the type of fraud an individual commits depends on their specific characteristics. Small organizations experience the highest frequency of asset misappropriation incidents.

While shopping, consumers generally go to a retail store, pick out the desired products, and then stand in line to make the payment. The cashier either scans the barcode on the merchandise or enters the monetary value into the cash register, after which the customer completes the transaction using cash or credit. However, there are opportunities for cashiers to illicitly manipulate prices by taking advantage of VIP discounts or clearance sales for personal gain. For instance, a cashier might purchase expensive merchandise at a discounted rate or lower the price for a friend in return for favors. This type of fraud scheme differs from embezzling money by modifying cash register data (Dopson & Hayes, 2015) Moreover, it creates an extensive audit trail that can be analyzed using data analytics to identify anomalous transactions (Gee, 2014).

Research studies on transaction fraud detection have primarily concentrated on specific payment methods, such as credit cards and prepaid cards (Jurgovsky et al., 2018; Robinson & Aria, 2018). Fraud involving these payment modes usually transpires when the cardholders' details are compromised and subsequently employed for unauthorized transactions. Consequently, Correa Bahnsen et al. (2016) assessed common customer spending patterns to precisely categorize the dataset before constructing a fraud detection model. Due to the diverse operational approaches of contemporary businesses, fraud has become less structured (Levi, 2008). As companies face significant financial losses and reputational damage from fraud, it is crucial for managers to understand how to investigate and detect fraudulent activities.

2.2 Anatomy of Investment Frauds

As digital products have become more convenient, scammers have used Internet and telecommunications vulnerabilities to plan different schemes. According to Tsai et al. (2009), the history of investment fraud can be divided into three phases: traditional face-to-face, written document, and communication network. Cash flows have shifted from cash and checks to Internet or mobile transfers. In addition, during the global economic turmoil, investors who were easily deceived poured into financial markets, accelerating the occurrence of investment fraud (Reurink, 2018).

Most investment fraud techniques involve social engineering and phishing approaches. Hatfield (2018) defined social engineering as individuals influencing others through a knowledge advantage in information asymmetry. In other words, social engineering is a tool used to gain victims' information and expose them to vulnerabilities (Hadnagy, 2010). Perpetrators who use social engineering must imply authority to convince victims, often using psychological manipulation, such as threatening or excitement (Bullée et al., 2018; Li et al., 2020). However, social engineering also involves taking advantage of victims' negligence to defraud their confidential information, often using insider lingo or terminologies (Thornburgh, 2004). Therefore, the contents of fraudulent phone calls can be analyzed to identify patterns of keywords and conversations, allowing for the identification of fraudulent phone calls (Peng & Lin, 2018; Zhao et al., 2018).

In cognitive research, individuals' judgment biases include the illusion of control, redundant information, failure to consider and seek out possible disconfirming evidence, and overconfidence in decision-making (Hogarth & Makridakis, 1981). Furthermore, the attention and memory processes can shift under specific stressors, such as noise, shock, or fatigue (Mendl, 1999). In addition, time pressure is considered an essential factor in speeding up fraud, as individuals' judgments can be influenced by the pressure to make a quick decision (Jones et al., 2019). From the victims' perspective, Whitty (2020a) noted that most individuals who have been deceived have impulsive and neurotic characteristics. Moreover, the higher the investment risk tolerance, the higher the probability of being deceived. In other words, if scammers provide a high-yield investment program with a high risk, aggressive investors are more likely to become cash cows. Furthermore, victims' acceptance of investment information from free seminars, communities, or advertisements that usually imply fraudulent schemes is high (Deliema et al., 2020).

2.3 The Perspectives of Criminology

A widely recognized framework for examining fraudulent behavior was the fraud triangle theory. Cressey (1973) identified the factors that drive individuals who are trusted to become violators of that trust: the existence of a non-shareable problem, an opportunity to commit the violation, and the ability to rationalize their actions. Romney et al. (1980) carried out an empirical study to illustrate that individual traits can considerably impact white-collar fraud. In International Standard on Auditing (ISA) 240, the IAASB (2013) has recognized the fraud-related components, including pressures, perceived opportunity, and rationalizing committing fraud. Nonetheless, these fraud-related theories are generally employed to elucidate the psychological factors driving fraudulent actions.

Researchers have acknowledged the significance of criminals' characteristics and their decision-making processes concerning crime within neoclassical criminology. The RCT has become a cornerstone of neoclassical thinking as a consequence (Schmalleger, 2021). The notion of rational choice originated with Clarke (1983), who examined individuals committing crimes in relation to the circumstances of the crime and the individual's considerations at that moment. The author sought to understand the impact of perpetrators' decision-making on crime behavior. In essence, perpetrators alter their target to a different time or location and modify their fraudulent schemes after careful deliberation. Cornish and Clarke (1987) subsequently introduced the RCT, encompassing three elements: opportunities, costs, and benefits, referred to as choice-structuring properties. The primary factors influencing decisions during the involvement of criminals and the unfolding of criminal events include time, area, target, and method (Clarke & Cornish, 1985).

The RCT was initially developed to examine the significant variations in criminal behavior due to differing considerations among various crimes during the decision-making process (Cornish & Clarke, 1987, 1989). This theory has its roots in classical criminology, positing that criminal acts stem from an individual's exercise of free will (Schmalleger, 1999). Consequently, perpetrators are deemed rational in weighing the consequences before engaging in actions that optimize benefits while minimizing costs (Akers, 1990). Clarke and Harris (1992) employed choice-structuring properties from a rational choice perspective to discern the reasons behind perpetrators' decisions and the influence of situational factors on these choices. Felson and Clarke (1998b) introduced a novel opportunity theory to elucidate the occurrence of crime, drawing on three

theories of crime opportunity: routine activity approach, crime pattern theory, and rational choice perspective. These theories propose that opportunities facilitating criminal acts arise from societal changes, local contexts, and perpetrators' deliberations. Otu and Okon (2019) asserted that opportunity constitutes a crucial element across various fraud-related theories, including rational choice and fraud triangle theories. Ultimately, the abovementioned perspectives presume that perpetrators' decisions are made consciously (Wilcox, 2015).

In the context of the fraudulent decision-making process, which encompasses perpetrators' preparation and potential actions, Chan and Gibbs (2019) determined that psychological and emotional factors significantly influenced white-collar offenders' choices. Junger et al. (2020) categorized three specific types of fraudulent activities (C-level fraud, fraudulent contracts, and fictitious invoices) to interpret the occurrence of fraud. The authors employed routine activity theory and RCT to elucidate that perpetrator often took into account factors such as business size and seasonality prior to engaging in fraudulent transaction schemes. Operating within the framework of RCT, Ding and Zhai (2021) established that pickpockets tend to commit theft in poor air quality conditions, during rush hour, and in the presence of weak police enforcement. In other words, individuals are more likely to use public transportation due to air pollution, resulting in crowded buses. Consequently, pickpockets capitalize on this situation and engage in theft through a rational decision-making process.

In summary, RCT can be employed to comprehend how perpetrators engage in fraudulent activities under favorable or unfavorable conditions, considering potential consequences and risks. While the majority of research has focused on investigating and devising prevention strategies based on RCT, this study presents a novel approach by developing a fraud transaction detection model grounded in the concepts of area, target, and modus operandi selection, as proposed by Clarke and Cornish (1985) within the framework of RCT (Meyer, 2012; Piza et al., 2017).

2.4 Fraud Indicators

From a scholarly perspective, researchers derive features from each transaction to precisely characterize an original dataset. Zheng et al. (2018) conducted preprocessing and feature extraction from raw data, drawing upon historical consumer transaction records, and categorized the information using transaction time, location, and amount, category of goods, and shipping address. Carcillo et al. (2021) carried out fraud detection by taking into account the aggregate sum of money expended by the consumer and the

number of transactions executed within a 24-hour period. Van Vlasselaer et al. (2015) further hypothesized that fraudulent transactions exhibit common features, including the frequency or magnitude of consumption. Consequently, they proposed an anomaly detection system that utilized recency, frequency, and monetary (RFM) values of transactions, in conjunction with social network-associated variables, to scrutinize e-commerce credit card transactions. Zhang et al. (2021) employed behavioral analysis to generate RFM variables, which were then utilized to detect fraudulent credit card transactions through deep learning approaches. As highlighted in the previous discourse, the temporal, object-related, and frequency aspects of transactions serve as crucial features for identifying fraudulent activities.

In cases of price manipulation fraud, perpetrators may include unsuspecting cashiers or customers. Furthermore, the transaction amount and transaction time recorded in the POS system explicitly indicate that the primary factors considered in fraud detection are the user, transaction amount, and transaction time (Singh & Best, 2019; Z. Zhang et al., 2018). Nonetheless, our aim was to incorporate fraudulent parameters to better represent the perpetrators' transaction behavior. We employed the RCT framework to depict fraudulent behavior. We pinpointed variables based on perpetrators' decisions regarding when to commit fraud (e.g., during holiday promotions), against what to commit fraud (e.g., bulk goods), and how to commit fraud (e.g., multi sale-item transactions and ongoing transactions). These variables were derived from the area, target, and modus operandi selection. We hypothesized that incorporating these crucial fraudulent parameters would yield a more accurate analysis. The four fraudulent parameters are elaborated as follows.

2.4.1 Holiday promotions

Consumer spending typically surges during holiday seasons, including Valentine's Day, Thanksgiving Day, and Chinese New Year. As a result, retailers frequently implement innovative promotional strategies to stimulate sales throughout these periods (Oh & Kwon, 2009). Tsoumakas (2019) posits that weather conditions and the presence of holidays can predict retail sales trends. Based on data from 2015 to 2020, the National Retail Federation (NRF) projected that holiday sales would continue to rise, even amidst the COVID-19 pandemic (NRF, 2020). In addition, the renowned consulting firm McKinsey discovered that consumer expenditure tends to increase significantly during events such as Black Friday or Amazon Prime Day (Charm et al., 2020).

Nonetheless, owing to the considerable volume of transactions during these periods,

the detection of anomalous transactions becomes increasingly challenging. Levy et al. (2010) observed that price fluctuations are more prevalent during holiday seasons than at other times. The dataset analysis utilized in their study indicated that specific periods exhibit a higher frequency of anomalous transactions. Consequently, the transaction date was integrated into the holiday promotion variable to determine if a transaction transpired during a holiday.

2.4.2 Multi-product transactions

Enterprises examine customer purchasing patterns and timings to enhance their understanding and forecasting of consumer behavior. Guidotti et al. (2015) forecasted customer behavior by analyzing the basket and spatiotemporal data from a massive volume of customer transactions. In addition, consumers have exhibited price sensitivity. Chen and Li (2020) posited that consumers' intent to engage in a transaction is affected by price-based promotions. Promotional efforts can be viewed as factors that enhance consumer engagement and elevate the perception of value (Hsia et al., 2020). Promotions have been demonstrated to encourage consumption (Tang & Hu, 2019). Consequently, an individual engaging in fraud by exploiting sale prices would likely purchase multiple items concurrently to maximize the benefits of the promotion. As such, we combined invoice numbers and product names to identify transactions involving the purchase of several discounted products. This information was then integrated into a variable referred to as multi-product transactions.

2.4.3 Continuing transactions

Individuals typically devote attention to items that interest them and are inclined to continue purchasing when they encounter lower prices. Nevertheless, the transaction price serves as a critical indicator in auditing processes. Gee (2014) noted that individuals might divide transactions to circumvent per-transaction limits, which could draw the scrutiny of fraud examiners who typically inspect ongoing transactions through invoice numbers. This scheme is referred to as order splitting (Stamler et al., 2014). The Office of Inspector General (2018), one of the United Stated deparment, comfirmed that order splitting poses a significant challenge for businesses dealing with occupational fraud. Successive invoice numbers suggest that fraud perpetrators divide orders over a brief time span. In this research, such behavior was considered indicative of a continuing transaction and, more broadly, fraudulent activity.

2.4.4 Bulk goods

With the growing popularity of e-commerce, an increasing number of individuals have started businesses focused on retail arbitrage, wherein they generate profits by acquiring products from retailers at low prices and subsequently reselling them at higher prices online. Mercer (2016) pioneered the investigation of retail arbitrage on Amazon, which primarily involves merchandise sold at discounted prices, such as during clearance sales. As per the first sale doctrine, retail arbitrage is considered a legal business model (Tseng, 2018). Nevertheless, fraud can transpire when individuals collude with cashiers or insiders to lower the prices of items. Merchandise with high market demand, such as groceries or electronics, is typically chosen for fraudulent arbitrage (Palmer & Richardson, 2009). As a result, individuals may develop strategies to profit from particular products. Moreover, these products are frequently sold in bulk. Consequently, this study hypothesized that sales of bulk goods are more prone to fraud, and this characteristic was incorporated as a variable for transactions in the labeled dataset.

2.5 Emotion Analysis

Emotion refers to an individual's mental state that reflects their attitudes, experiences, and related behavior (Hudlicka, 2011; Poria et al., 2019). The basic emotions are commonly classified as happiness, anger, sadness, disgust, surprise, and fear in different literate cultures (Ekman & Friesen, 1971). Plutchik's wheel of emotions developed based on eight emotions: acceptance, surprise, fear, sadness, disgust, expectancy, anger, and joy, adding color to express different intensities and emotional similarities (Plutchik, 1958; Plutchik, 2001). Emotion analysis aims to determine which emotions are present in a text and to what degree they are expressed. This method has been studied in social psychology, including public opinion analysis, recommendation systems, and data prediction. For example, Park et al. (2020) developed an emotion-embedding model through emotional words to classify emotions from story texts, and Qamar et al. (2021) identified the relationship between individuals through a text conversation by adopting an emotion analysis. The changes in emotions and behaviors were, therefore, able to observe. Xu et al. (2020) applied the convolutional neural network (CNN) algorithm to classify the emotion of microblogs with positive and negative sentences. In other words, it is possible to use the emotion analysis method to evaluate the individual’s attitude and response, whether good or bad, joy or sadness. Emotion analysis is typically more complex than simple sentiment analysis, as it requires identifying and understanding the nuances of emotional language.

Emotions have always played a significant role in marketing, such as in consumer experience or tourism marketing (Huang & Rust, 2022; Tsang et al., 2022). However, scammers manipulate emotions to design investment fraud schemes. Scammers invest time in designing schemes that appear less overtly fraudulent and strategically target victims with lower susceptibility to fraud. These victims possess specific characteristics, such as believing they can gain financial gain, information, or experiences, and lack knowledge and concern for information security risks. This type of victim selection enables scammers to exploit vulnerabilities in the victims' cognitive and emotional processes to achieve their fraudulent objectives (Steinmetz, 2020). Compliance with scammers' demands is one of the major reasons that victims fall prey to scams, such as responding to investment invitations from unknown individuals. This compliance can be driven by a variety of factors, such as the victims' susceptibility to persuasion and manipulation, their need for financial gain or the promise of other rewards, or a lack of knowledge and understanding of the risks involved (Shadel & Pak, 2017).

Investment objects in investment fraud are mostly securities fraud, including real estate, stocks, futures, and cryptocurrencies. Scammers design these objects as high-yield investment programs with ambiguous transaction strategies and high returns on unknown websites or platforms (Beals et al., 2015). Information asymmetry causes the parties to the transaction to have deception and trust concerns (Mavlanova et al., 2016). The information on sellers and transaction platforms can alleviate these concerns. Similarly, investment issues occur on crowdfunding platforms, and basic information about the funding, user comments, and discussions can distinguish the true from the false (Siering et al., 2016). People's speaking patterns and behavior can be measured to detect deceptive intentions (Li & Santos Jr, 2020). Investigation of malicious content from email datasets is a task to identify possible scam threats in organizations (Janjua et al., 2021).

2.6 Investment Fraud Life Cycle

With the outbreak of the COVID-19 pandemic, investment fraud complaints have increased globally. Scammers have taken advantage of people's usage of instant messaging and escalated their scam techniques by designing new schemes that use market uncertainty and investors' fear of missing out on opportunities to earn quick money. Scammers also fear that their scheme might collapse when victims stop contributing funds to the pool. Therefore, they become more proactive in designing scams (Springer, 2020).

Typically, investment fraud involve three stages: bait, hook, and catch (Torres et al., 2020). Scammers first search for victims and provide small benefits to gain their trust. For example, scammers create bogus profiles with attractive photographs to contact victims. Likewise, romance scams are used to create committed relationships to request money from easily impulsive victims (Whitty, 2013, 2015, 2019b). Secondly, scammers make promises or convince victims of the feasibility of their plans. For instance, victims may receive steady rewards at the beginning of a blockchain Ponzi scheme that eventually collapses (Bartoletti et al., 2020). Finally, scammers reel in victims slowly once they have joined the plan and handed over their funds (ACCC, 2016).

Phishing is a common method scammer uses to conduct investment fraud (Krombholz et al., 2015). Phishing is a popular social engineering method in which attackers pretend to be reputable figures who work in famous or Fortune 500 companies to gain victims' trust. However, the attack aims to lure victims into clicking on a malicious link or website to collect confidential information (Chaudhry et al., 2016). Mohammad et al. (2015) have illustrated the phishing attack life cycle, including planning the phishing attack, collecting victims' confidential information, and conducting fraud. Chiew et al. (2018) have indicated that phishing methods start from three channels: the Internet, short messaging service, and voice to attack victims.

Using Ponzi schemes as an example, fraudsters often target low-risk, high-return investment opportunities to attract unsuspecting victims who are unfamiliar with investments. For instance, Bernard Madoff promised stable investment returns and guaranteed no losses to perpetrate his Ponzi scheme. However, this is just the beginning of the fraud. Subsequently, fraudsters use the capital obtained from recruiting new members to pay interest to old members or even introduce different new products to encourage members to invest more, creating a seemingly plausible profit model (Henriques, 2011). Therefore, the entire fraudulent scheme is similar to a large project, starting from the initial stage, followed by planning, execution, monitoring, and finally, the closing stage. It resembles a complete life cycle. Given these points, we have outlined an Investment Fraud Life Cycle (IFLC) that can be broken down into five phases: Lurking, Alluring, Catching, Executing, and Vanishing.

2.6.1 Lurking

The IFLC begins with the lurking phase, during which scammers gather information about potential victims. As online storage and transmission technologies

have advanced, the increased reliance on digital storage and transmission technology has increased data breaches. Consequently, data breaches proliferated in cases such as Microsoft, Uber, and Flagstar Bank in 2021. Leaked data can include personal information such as individual's names, addresses, birth dates, and social security numbers, which scammers collect from different sources, such as databases or cloud services, using system vulnerabilities or malware (Guri et al., 2019; Shrivastava & Kumar, 2021).

2.6.2 Alluring

The second phase is "alluring," in which the scammer attempts to provide short-term profits to gain the victim's trust. Scammers often use psychological manipulation, such as in the case of romance scams, to build trust and request money from vulnerable victims (Salahdine & Kaabouch, 2019). In Ponzi schemes, scammers often use the guise of investment to offer high returns to attract victims. Thus, in the early stages of the scheme, they use the funds from later investors to pay interest to earlier investors, thereby establishing a false sense of trust among the victims (Chiluwa, 2019). The victims of these schemes hope to obtain more money or a better retirement life, even if they are already relatively wealthy (Trahan et al., 2005).

Additionally, research has also shown that the likelihood of victims participating in investments increases when they receive promotional messages from sources such as television advertisements or conferences. It indicates that exposure to such information can influence individuals to make investment decisions (Choi et al., 2021). According to Williams et al. (2018), the population between 18 and 25 is more susceptible to social engineering attacks, while in Taiwan, individuals between the ages of 30 and 40 are most vulnerable to investment fraud, according to the Ministry of the Interior (2023).

2.6.3 Catching

The third phase is called "catching," which means developing a convincing story or investment opportunity with unrealistic returns, guarantees, or limited-time offers to entice the victim to join the investment group and make a deposit to become a member. For example, in the membership grads, scammers have a silver membership with 10 thousand US dollars in investment and a 20% earning yield compounded monthly. A gold membership could gain a 35% earning yield compounded monthly with 150 thousand US dollars in investment. Furthermore, the scammer introduces authorities to continue providing investment-related information. Therefore, it is possible to convince

victims who doubt the investment opportunity is real (Jones et al., 2021). As indicated by Wang et al. (2021), such unbelievable investment returns can lead victims to be less critical in evaluating the legitimacy of the scam. In addition, implementing a tiered membership allows the scammer to manage the distribution of funds.

2.6.4 Executing

The fourth phase is “executing,” that the scammer deceives the victim into investing through a software or platform who unfamiliar. In other words, the scammer attempts to guide the victim into a more complicated environment and increases time pressure or investment difficulty to require the victim to invest (Lacey et al., 2020). From a psychological perspective, the psychological state and cognition of the victim can lead to different responses to fraudulent behavior. Therefore, victims who lack knowledge or experience in network security, socializing, and online transactions are more vulnerable to being defrauded (Wen et al., 2022). According to Ullah et al. (2022), victims fall into investment traps primarily because of greed, where unrealistic investment returns overshadow the perceived risks, ultimately impacting their investment decisions.

Concerning the time pressure, the victim's transfer to an unfamiliar investment environment can place them in a situation where they lack control. Additionally, the immediacy of financial instrument buys and sales can result in time pressure, further exacerbating the victim's tendency to make more mistakes when rapid responses are required (Butavicius et al., 2022; Jones et al., 2019). However, because the victim initially chose to believe the fraudulent claims of the scammer and the investment masters continually promoted the potential benefits of joining the investment scheme, the likelihood of the scheme being successful significantly increased.

2.6.5 Vanishing

While the victim would like to withdraw their fund after a succession of investments, the scammers then close the interaction called “vanishing.” In line with the typical characteristics of schemes, the scammer will disappear without a trace after obtaining the desired money or goods (Langenderfer & Shimp, 2001; Lee, 2021). Identifying the lurking phase in an investment fraud chat room from the beginning is challenging. Instead, the scheme often begins in the alluring phase after establishing the chat room. Therefore, in this study, we use real-life datasets to identify the expressions of emotion in the text to reveal the scammers' intentions during the IFLC from the alluring phase to the executing phase.

2.7 Supervised Machine Learning Algorithms

With technological advancements and computing power, machine learning has become a popular tool for analyzing fraud behavior. Consequently, researchers have attempted to investigate fraudulent behavior using various algorithms such as random forest, decision tree, Random Trees, Support Vector Machines (SVM), Logistic Regression, and Naive Bayes (Mehbodniya et al., 2021; Trivedi et al., 2020). Supervised learning is a popular approach in machine learning that involves training algorithms on labeled data to make predictions or classifications on new, unseen data. In the context of fraud detection, this means using labeled examples of anomaly and non-anomaly data to train algorithms to recognize patterns and features associated with the fraud. The advantages of supervised learning include its ability to leverage prior knowledge to improve accuracy and its suitability for structured datasets. In addition, supervised machine learning algorithms offer advantages for detecting and preventing fraud that traditional statistical methods may not be able to match, particularly in cases where the data is complex or nonlinear. Table 2.1 presents the algorithms' characteristics we adopted in this study.

Table 2.1 Algorithms adopts for the detection models

Algorithms	Characteristics
Bayesian network	Bayesian networks represent causal graphical models that evaluate independent relationships among nodes, using conditional probability distributions between independent variables to make inferences from datasets (Singh & Valtorta, 1995).
Logistic Regression	Logistic Regression is a prevalent technique for binary classification that models the relationship between the dependent and independent variables (Hosmer Jr et al., 2013).
Random Forest	Random Forest is a collection of decision tree classifiers that effectively tackle overfitting. The algorithm's results stem from the relationships among the features of each decision tree (Breiman, 2001).
SVM	SVM is a supervised learning algorithm suitable for classification and regression tasks. It identifies the ideal hyperplane that separates data points into distinct classes (Birzhandi et al., 2019).
Decision Tree	Decision Trees are algorithms that divide data into subsets based on chosen features, creating a tree-like structure consisting of decision rules (Askari & Hussain, 2020).
Random Tree	Random Trees, advanced versions of Decision Trees, employ ensemble learning methods to boost accuracy and minimize overfitting. In investment fraud detection, Random Trees demand fewer data preprocessing and feature selection and excel in handling large datasets and nonlinear challenges (Alghamdi & Alharby, 2019).

In summary, the algorithms mentioned above are promising tools for detecting fraud. The choice of algorithm depends on the specific problem, dataset size, and data properties. Nevertheless, the algorithms have proven effective in detecting fraud and scam behavior (Guedes et al., 2022; Li et al., 2021; Masmoudi et al., 2019; Mqadi et al., 2021). Hence, in the first experiment, we employed the Bayesian network, Logistic Regression, and Random Forest algorithms to determine their effectiveness in the anomalous transaction detecting model. Subsequently, in the second experiment, we utilized the SVM, Decision Trees, and Random Trees algorithms to evaluate the performance of the investment fraud detection model.

3. Materials and Methods

3.1 Data Description and Collection in Anomalous Transaction Detection Model

In this study, a physical transaction dataset from a retail store in Taiwan was analyzed, encompassing sales of food, appliances, hardware, and groceries during the period from September 2019 to July 2020. The dataset contained information on price markdowns and was labeled with fraudulent transactions. Each transaction data point contained details on customer information and transaction characteristics, such as the time, object, and amount of the transaction. Fraud examiners then labeled each transaction as genuine or fraudulent. Notably, the point-of-sale (POS) system recorded the transaction information after payment, and paired data was removed due to transaction reversals. The resulting unaltered dataset contained 212,792 transactions with 13 features and one labeled field.

In the context of fraud detection, researchers have increasingly turned to machine learning models as a means of identifying anomalous transactions and reducing the need for manual analysis and intervention. Supervised learning algorithms, which rely on labeled instances from past data to train models to classify instances as genuine or fraudulent, have become particularly popular in this field. This study employed three common supervised learning algorithms: Bayesian network, Logistic Regression, and Random Forest. By using labeled examples to train the machine learning model, these algorithms were able to predict the class of each instance in the dataset. As a result, researchers could leverage the predictive power of machine learning to enhance their fraud detection capabilities. (Zhou, 2018).

The highly imbalanced label distribution in the real-life dataset used in this study, with a minority to majority ratio of 1:29, could lead to the failure of classification accuracy. To address this issue, synthetic minority oversampling techniques (SMOTE), proposed by Chawla et al. (2002) , were applied to balance the dataset. SMOTE creates synthetic examples by utilizing the nearest neighbors of the data and introducing linear interpolation for the minority class, thus generating virtual training data. The dataset was reconstructed after the oversampling process. Previous studies have confirmed the effectiveness of SMOTE as a filtering method for handling imbalanced datasets (Blagus & Lusa, 2013; Fernández et al., 2018). Figure 3.1 displays the proposed methodology.

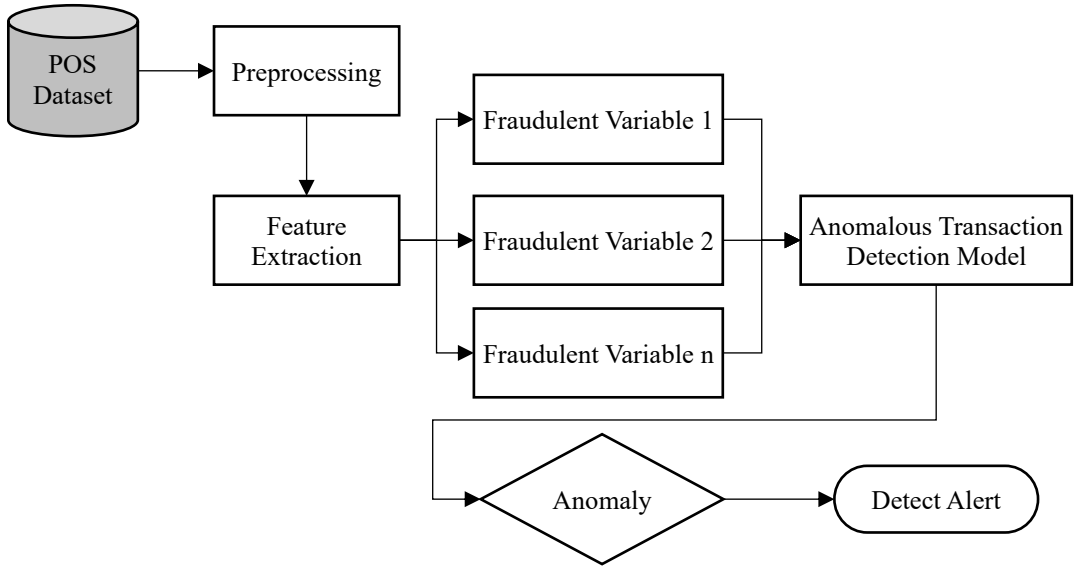


Figure 3.1 Anomalous transaction detection process

3.2 Data Description and Collection in Investment Fraud Detection Model

The real-life dataset we adopted was text message data from seven investment fraud chat rooms on the IM. Such chat rooms were created and closed within a short period of time. However, marketers preferred introducing multiple accounts that were planted in marketing campaigns. In other words, there would be a primary account used to be the Main Poster, and the others were Echoes. The Echoes pretended they were interested in the messages from the main poster and then showed a good interaction. We notice that there are similar roles in the investment fraud chat room. In addition, businesses create those accounts through automated software or manual operation, making it easy to send messages to different chat rooms simultaneously. We can see many accounts with the same names sending the same messages in different chat rooms simultaneously while we collect the text message data. This can be used as evidence of manipulation to identify mass text messaging by automated software and investigate sales talk associated with investment fraud that lure victims. Therefore, those conversations were confirmed as the scam label in the dataset.

According to this real-life dataset with 29,438 messages, we noticed that the scammers in the chat room adopted the following typical schemes (ACFE, 2019). Initially, scammers disguised as securities salespersons with beauty photos called "Fronter" would invite victims into the investment chat rooms. Then, they provided blue-

chip stocks, which guaranteed the stock price would rise in a few days. Subsequently, they began to launch different investment programs with high and steady profits, which were developed by wall street traders who ever created remarkable performance. They were also known as “Closers” to convince victims to join their investment fraud. Once individuals decide to invest, the “Verifier” would illustrate the program and teach how the program works, including opening accounts for foreign exchange and international stock and remitting security deposits to the brokers. However, these brokers were illegal, so the victims could not withdraw money as possible. Following the IFLC we proposed, fronter, closer, and verifier appeared sequentially in the alluring, catching, and executing phases. Therefore, we divided the messages into three parts to match with alluring, catching, and executing phases in the IFLC.

In sentiment analysis, researchers have adopted a rule-based approach based on a set of pre-defined rules and lexicons, such as a list of positive and negative words, to identify the sentiment or emotion expressed in a piece of text. Firstly, in order to analyze the emotion from Mandarin Chinese text, the emotion vocabulary word database was necessary. However, a handful of the emotion vocabulary word database could be applied in these studies, such as the National Taiwan University Sentiment Dictionary, the Dalian University of Technology (DUT) Sentiment Dictionary, and the HowNet Dictionary. These dictionaries present different combinations of sentimental strength and polarity to illustrate the emotional features of the text.

The DUT sentiment dictionary generally shows better performance mainly because it includes many more sentiment key terms that can be used in classification methods (Liu & Chen, 2015; Wei et al., 2022; S. Zhang et al., 2018). The dictionary contains a total of 27,466 words, including terms, types of emotional vocabulary, sentiment classification, and intensity. In the DUT sentiment dictionary, we can see the seven types of emotional dimensions, including happiness, surprise, fear, anger, sadness, and disgust, which humans construct. In addition, these dimensions can be classified into positive, negative, or neutral labels (Xu et al., 2008). Therefore, we adopted the DUT sentiment dictionary as the source of emotion analysis for Mandarin Chinese text to better classify each emotion in the text in the first stage. During data preprocessing, we removed a large number of emoji and meaningless modal particles from the messages. Secondly, we used data visualization to present the fluctuation of emotion to identify patterns and relationships in the data that are not immediately apparent (Karmy & Maldonado, 2019).

With the rapid advances in analytical techniques, researchers have applied machine learning algorithms to perform data analysis. One common approach is supervised learning, which utilizes labeled data to conduct classification problems. Supervised learning has been recognized as an effective tool for detecting fraud or scams (Kuo & Tsang, 2022). Therefore, in the second stage of this study, we applied supervised learning algorithms, including SVM, Decision Tree, and Random Tree, to detect the investment fraud detection model proposed in this study.

To summarize, this study employed a two-stage analysis. In the first stage, we utilized emotion analysis to identify emotional fluctuations in scam messages and employed MANOVA to detect patterns in these fluctuations across different chat groups. In the second stage, we constructed an investment fraud detection model based on emotion analysis results and evaluated its effectiveness using supervised learning algorithms. Figure 3.2 displays the proposed methodology.

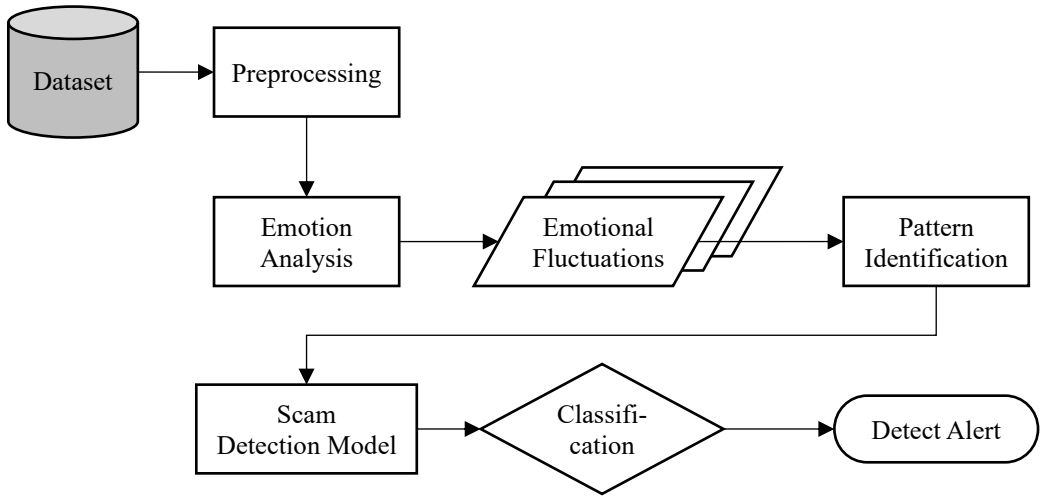


Figure 3.2 Investment fraud detection process.

3.3 Methods

In this study, we adopted Python and the Waikato Environment for Knowledge Analysis (WEKA) 3.8.1 to execute emotion analysis and machine learning algorithms. Python has been considered an efficient tool that provides multiple prevalent machine learning algorithms and data mining methods. Furthermore, according to the characteristics of open-source software, Python offers an extensible interface and simple scripting language that lowers researchers' entry barriers. Therefore, researchers apply

the algorithms in Python to conduct the linking, classification, matching, indexing, and visualization in emotion analysis (Yan, 2022). Likewise, anomaly behavior detection, such as credit card fraud and hacker attacks, was investigated through machine learning techniques to discover valuable behavior patterns through Python (Arya & Sastry G, 2020; Fang et al., 2021). Therefore, we adopt Python to conduct the emotion analysis of Mandarin Chinese text based on the DUT sentimental dictionary in this study.

WEKA, an open-source software, is a comprehensive tool containing various popular machine learning algorithms and data mining techniques. It provides a user-friendly interface that facilitates researchers' access to its functionalities. I n a related study on detecting anomalies in voluminous datasets, Cui and He (2016) employed WEKA software to evaluate the effectiveness and precision of detecting cyberattacks. Additionally, Rajesh and Karthikeyan (2017) explored the recognition of meteorological patterns using various classification algorithms in WEKA. Similarly, Yee et al. (2018) explored the use of data mining technologies within WEKA to detect anomalous transactions and identify valuable variables.

4. Experiments and Result Analysis

4.1 Anomalous Transaction Detection Model

4.1.1 Experimental Setting

The objective of the present study was to develop a machine learning-based model for anomalous transaction detection grounded on RCT. To this end, three well-known machine learning algorithms, namely Bayesian network, Logistic Regression, and Random Forest, were employed to evaluate the accuracy of the proposed detection model. These algorithms are straightforward yet effective in practice (Hooda et al., 2018, 2020; Lavanya et al., 2021; Lucas et al., 2020). We propose a two-phase design approach to examine the significance of the identified fraudulent parameters.

In this dataset, a high data set imbalance problem was observed, where the number of fraudulent transactions was only 0.23% of the total 212,792 transactions. Hence, to handle this class imbalance, SMOTE was employed to generate artificial data. To construct an efficient detection model, researchers have attempted to use numerous features. However, the presence of unnecessary features or an excessive number of fields in the data set can negatively impact the performance of machine learning algorithms. Thus, in this study, a total of seven features were selected, including three native attributes from transaction records, four continuous variables converted into discrete values, and four key features extracted based on RCT. Table 4.1 presents the essential attributes of fraudulent transactions (Hajek & Henriques, 2017). In this study, we employed a feature selection strategy that involves three main steps. Firstly, three native attributes including cash register number, subcategory, and quantity were retained from the transaction records. Secondly, four continuous variables, including day of the week, business hour, checkout location, and price variance range were converted into discrete values to provide a meaningful presentation. Lastly, based on RCT, the authors extracted key features of the transaction date, such as the holiday promotions attribute, the number of transaction invoices to recognize continuing and multi-product transactions, and identified bulk goods from the transaction items. Table 4.1 presents a summary of the essential attributes used to identify anomalous transactions.

Table 4.1 Essential features of transaction data.

Data Source	Attribute Name	Description
Native	Cash register number	Cash register's serial number.
Native	Subcategory	Lowest item category level.
Native	Quantity	Number of sales.
Discretization	Day of the Week	Day of the Week.
Discretization	Business Hour	Approximate time of transaction, such as morning, afternoon, or near closing time.
Discretization	Checkout location	Checkout location.
Discretization	Price Variance Range	Difference between the original price and sale price.
Extraction	Holiday Promotions	Price reduction during a holiday period.
Extraction	Continuing Transactions	The transaction numbers were sequential.
Extraction	Multi-product transactions	The transaction only includes price-reduced products.
Extraction	Bulk goods	The products are bought in bulk.
Label	Class	Anomalous or normal.

This study employs a two-phase experiment to assess the efficacy of the proposed fraud variables. The first phase involves measuring the original performance using the native attributes obtained from the transaction records. The second phase evaluates the enhanced performance obtained by incorporating the four fraudulent parameters. Both phases employ the same algorithms, namely, Bayesian network, logistic regression, and random forest. Additionally, Pearson's correlation coefficient is utilized to ascertain the degree of association between each fraudulent parameter.

4.1.2 Analysis and Results

A real-world data set was obtained from a retailer's POS system, which included markdown transactions. The confidential information was masked to ensure data privacy. The machine learning process consisted of two phases to evaluate the accuracy of the constructed anomalous transaction detection model. In phase 1, the model was trained using the three native features, four discrete features, and the anomaly label. In the second phase, we included the four critical fraud features (holiday promotion, bulk goods, multi-product transactions, and continuing transactions) to improve the accuracy of the anomalous transaction detection model.

The performance of three different algorithms, namely the Bayesian network, Logistic Regression, and Random Forest, was evaluated in terms of accuracy, precision, recall, F measure, and Area Under the Curve (AUC). Accuracy is the ratio of the number of correct predictions to the total number of predictions. Precision is the ratio of the number of true positive predictions to the total number of positive predictions. Recall is the ratio of the number of true positive predictions to the total number of positive observations in the actual class. F measure is the weighted average of precision and recall. AUC measures the performance of a classification model at all classification thresholds by calculating the area under the ROC curve. Higher AUC values generally indicate better performance (Witten et al., 2017).

Tables 4.2 and 4.3 exhibits the efficiency analysis outcomes of all the algorithms utilized in each phase. Based on the analysis results demonstrated in Table 4.2, all algorithms exhibited accuracy scores greater than 0.8. Notably, the Bayesian network algorithm produced relatively high scores with accuracy, precision, recall, F-measure of 0.863, and an AUC of 0.991. Consequently, the evaluation outcome from phase 1 portrays a good performance.

Table 4.2 Algorithm assessment results in Phase 1.

Classifier	Accuracy	Precision	Recall	*F* measure	AUC
Bayesian network	0.863	0.863	0.863	0.863	0.991
Logistic Regression	0.858	0.859	0.858	0.858	0.898
Random Forest	0.804	0.837	0.804	0.799	0.939

Based on the results presented in Table 4.3, it can be observed that the efficiencies increased slightly after the introduction of the four fraudulent parameters. The accuracy, precision, recall, F measure, and AUC values slightly increased for each algorithm, indicating that the inclusion of the identified fraudulent parameters improved the performance of the detection model.

Table 4.3 Algorithm assessment results in Phase 2.

Classifier	Accuracy	Precision	Recall	*F* measure	AUC
Bayesian network	0.881	0.882	0.881	0.881	0.926
Logistic Regression	0.887	0.887	0.887	0.887	0.907
Random Forest	0.812	0.852	0.812	0.806	0.960

In order to assess the strength of the relationship between each variable, a correlation analysis was conducted using Pearson's correlation coefficient. This technique measures the strength of the association between variables, with coefficient values (r) ranging from -1 to +1, where +1 indicates the strongest association and -1 indicates the weakest association (Schober et al., 2018). Table 4.4 presents the results of the correlation analysis for the real-life dataset. The added fraudulent variables, namely continuing transactions, holiday promotions, bulk goods, and multi-product transactions, were ranked based on their correlation coefficients. For price manipulation fraud, it can be observed that the price variance range, continuing transactions, subcategory, holiday promotions, and bulk goods have higher rankings and can be considered as stronger predictors for the detection model. However, the parameters were fine-tuned to exclude the attributes ranking below the average, and the accuracy, precision, recall, F measure, and AUC values decreased. In other words, these attributes have demonstrated good efficacy in this study.

Table 4.4 Ranking of attributes

Ranking	Attribute Name	r
1	Price Variance Range	0.52
2	Continuing Transactions	0.21
3	Subcategory	0.19
4	Holiday Promotions	0.13
5	Bulk goods	0.07
6	Day of the Week	0.06
7	Cash register number	0.06
8	Multi-product transactions	0.05
9	Checkout location	0.04
10	Business Hour	0.03
11	Quantity	0.02

4.2 Investment Fraud Detection Model

4.2.1 Experimental Setting

This study aims to develop an investment fraud detection model based on emotion analysis. In addition, we conducted a two-stage analysis to investigate the scam's intention and examine the performance of the detection model. In the first stage, we used Python 3.8.8 to measure the scam intention from seven open chat rooms to verify our

proposed alluring, catching, and executing phase. Then, we adopted WEKA 3.8.1 to execute the machine learning algorithms in the second stage. Three commonly used machine learning algorithms, including Support Vector Machine (SVM), Decision Tree, and Random Tree methods, were employed to assess the accuracy of the detection model.

In order to verify the fluctuation pattern of each investment fraud, MANOVA (Multivariate Analysis of Variance) was then applied in the second step to determine if there were significant differences in the emotion values among seven open chat rooms. MANOVA, an extension of ANOVA, is a statistical technique that is used to test whether two or more independent variables are equal across groups (Huang, 2019). Therefore, MANOVA is useful for analyzing the multiple independent variables, including anger, disgust, fear, sadness, surprise, good, and happiness, in this study.

The dataset had 29,438 messages in total and 15,503 messages belonging to scammers' accounts. Table 4.5 presents the essential features of investment fraud chat room messages. The number of messages in each phase is presented in Table 4.6. The analysis of the message count in three phases shows that the alluring phase has the highest number of interactive messages, followed by the catching phase, and the executing phase has the least number of interactive messages.

Table 4.5 Essential features of messages

Attribute Number	Attribute Name	Description
1	Date	Date of conversation.
2	Time	Time of conversation.
3	ID	User account name.
4	Document	Message content.
5	Phase	Alluring, Catching and Executing.
6	Class	Scam or non-scam.

Table 4.6 Numbers of scam messages

Phase	Group	Scam	Non-scam
Alluring	A	1,736	1,451
	B	419	404
	C	878	710
	D	884	1,018
	E	1,124	1,269
	F	149	131
	G	937	1,822

Phase	Group	Scam	Non-scam
Catching	A	2,311	1,545
	B	548	215
	C	340	332
	D	1,329	251
	E	635	520
	F	1,810	2,117
	G	15	569
Executing	A	1,060	625
	B	61	35
	C	169	166
	D	47	4
	E	489	210
	F	490	193
	G	72	348

4.2.2 Emotion Analysis and Results

The research dataset used in this study was collected from seven investment fraud chat rooms on the IM app, and the analysis was conducted in two stages. In the first stage, emotion analysis was performed on the text messages with scam labels to identify the sentiment in each message, including anger, disgust, fear, sadness, surprise, good, and happiness in each phase. The MANOVA technique was then applied to investigate if there were similar fluctuation patterns in each phase among the seven groups, supporting the development of the IFLC proposed in this study. In the second stage, machine learning algorithms were utilized to evaluate the effectiveness of the scam detection model.

The emotion analysis performed in this study allows for the identification of the emotional vocabulary present within each message, including categories such as anger, disgust, fear, and others. This information can be used to quantify the frequency of each emotional vocabulary within each message. Table 4.7 displays the number of occurrences of each emotional vocabulary across each group of the three phases of the study. According to the results shown in Table 4.8, the messages from scammers show a large amount of “good” sentiment in each phase. Such type of messages approximately accounts for about 60% of the total messages in each phase. The alluring phase contains information about current market conditions and high-quality investment opportunities.

In the catching phase, many messages congratulated members on their profits. In the final executing phase, there are still messages congratulating everyone on their steady profits and even encouraging people who have not joined yet to join quickly.

If we only consider scammers' messages, there are still some differences among different phases and groups, which are presented in Table 4.8. For example, in the alluring phase, the highest frequency words of emotion among all groups are "good," but the second and third highest appear in two situations. The first situation is that groups A, C, E, F, and G all have "disgust" and "happiness," and the second situation is that groups B and D have "disgust" and "fear" of emotions. Upon closer examination, it becomes clear that these two groups' "fear" emotions are primarily related to the impact of the Covid-19 pandemic on investment. Specifically, the fear investors or the market felt in response to the pandemic. In the catching phase, the order of frequency of emotional words also appeared in two categories. The first category is groups A, C, and G, with "good," "disgust," and "happiness," and the second category is groups B, D, E, and F, with "good," "disgust," and "fear." Again, upon closer examination, the "fear" emotions are similar to the first phase, mainly describing negative investment information. However, in the final executing phase, the frequency of emotional words among different groups is not quite the same. Groups A, E, and F have the order of "good," "disgust," and "happiness," group B has "good," "happiness," and "fear." Group C has "good," "happiness," and "disgust," group D has "good," "disgust," and "sadness," and group G has "good" and "fear."

Overall, in the alluring and catching phases, the most prevalent emotion among the seven groups was "good," accounting for approximately 60% of all instances. "Disgust" and "happiness" were the subsequent most prevalent emotions, accounting for approximately 20% and 10%, respectively. However, while "good" remained the most prevalent emotion across all groups in the executing phase, there were notable variations in the second and third most prevalent emotions among the groups.

Table 4.7 Frequency of emotional vocabulary

Phase	Group	Class	Anger	Disgust	Fear	Sadness	Surprise	Good	Happiness
Alluring	A	Scam	-	371	144	54	7	1,142	253
		Non- scam	-	111	38	23	4	281	96
	B	Scam	-	56	50	6	-	229	31
		Non- scam	-	19	8	4	-	59	13
	C	Scam	-	100	18	24	1	416	33
		Non- scam	-	49	16	11	2	87	14
	D	Scam	-	160	93	25	1	518	91
		Non- scam	-	3	2	2	-	38	10
	E	Scam	-	154	24	47	3	388	70
		Non- scam	-	54	23	17	4	203	44
	F	Scam	-	10	8	-	1	88	9
		Non- scam	-	2	1	-	-	10	4
	G	Scam	-	131	49	40	1	800	92
		Non- scam	-	146	68	43	6	418	50
Catching	A	Scam	-	284	95	60	13	1,155	271
		Non- scam	-	74	34	11	5	231	94
	B	Scam	-	125	71	28	-	350	53
		Non- scam	-	2	4	-	1	33	8
	C	Scam	-	41	8	9	-	173	16
		Non- scam	-	19	9	3	-	34	6

Phase	Group	Class	Anger	Disgust	Fear	Sadness	Surprise	Good	Happiness
	D	Scam	-	256	128	46	21	721	108
		Non- scam	-	5	2	2	-	9	1
	E	Scam	-	80	40	30	-	387	38
		Non- scam	-	31	9	13	-	105	20
	F	Scam	-	330	177	59	11	1,068	165
		Non- scam	-	138	50	26	13	301	91
	G	Scam	-	6	3	2	-	19	4
		Non- scam	-	46	14	3	3	63	31
Executing	A	Scam	-	131	57	27	5	624	112
		Non- scam	-	73	34	18	4	250	63
	B	Scam	-	4	6	-	-	43	7
		Non- scam	-	1	1	-	-	6	-
	C	Scam	-	17	3	2	-	97	18
		Non- scam	-	16	3	5	-	26	2
	D	Scam	-	6	1	6	-	26	2
		Non- scam	-	-	-	-	-	-	-
	E	Scam	-	98	26	24	1	262	34
		Non- scam	-	8	7	-	-	49	11
	F	Scam	-	110	28	15	4	275	45
		Non- scam	-	12	3	1	1	38	17
	G	Scam	-	-	1	-	-	7	-
		Non- scam	-	21	12	3	1	36	16

Table 4.8 The percentage of emotional word frequency used by scammer.

Phase	Group	Anger	Disgust	Fear	Sadness	Surprise	Good	Happiness
Alluring	A	0%	19%	7%	3%	0%	58%	13%
	B	0%	15%	13%	2%	0%	62%	8%
	C	0%	17%	3%	4%	0%	70%	6%
	D	0%	18%	10%	3%	0%	58%	10%
	E	0%	22%	3%	7%	0%	57%	10%
	F	0%	9%	7%	0%	1%	76%	8%
	G	0%	12%	4%	4%	0%	72%	8%
Catching	A	0%	15%	5%	3%	1%	62%	14%
	B	0%	20%	11%	4%	0%	56%	8%
	C	0%	17%	3%	4%	0%	70%	6%
	D	0%	20%	10%	4%	2%	56%	8%
	E	0%	14%	7%	5%	0%	67%	7%
	F	0%	18%	10%	3%	1%	59%	9%
	G	0%	18%	9%	6%	0%	56%	12%
Executing	A	0%	14%	6%	3%	1%	65%	12%
	B	0%	7%	10%	0%	0%	72%	12%
	C	0%	12%	2%	1%	0%	71%	13%
	D	0%	15%	2%	15%	0%	63%	5%
	E	0%	22%	6%	5%	0%	59%	8%
	F	0%	23%	6%	3%	1%	58%	9%
	G	0%	0%	13%	0%	0%	88%	0%

As noted previously, our preprocessing of messages revealed the presence of Echoes in Group A. We segmented the sources of scam messages from Group A into Primary Posters and Echoes, presented in Table 4.9. The results indicate that Echoes accounted for 50% of messages in each phase. Table 4.10 shows the frequency of emotional vocabulary used by Echoes, demonstrating their effective portrayal of their role. During the alluring phase, Echoes raised questions about joining the membership, inquired about the Primary Poster (expressing Disgust or Fear), or shared positive stock market information (expressing Good). In the catching phase, Echoes shared messages about stable profits after joining, explained the benefits of joining (expressing Good and Happiness), and even questioned and reinforced the Primary Poster's messages (expressing Disgust and Good). In the final executing phase, Echoes continued to express gratitude to the Primary Poster for providing investment opportunities (expressing Good).

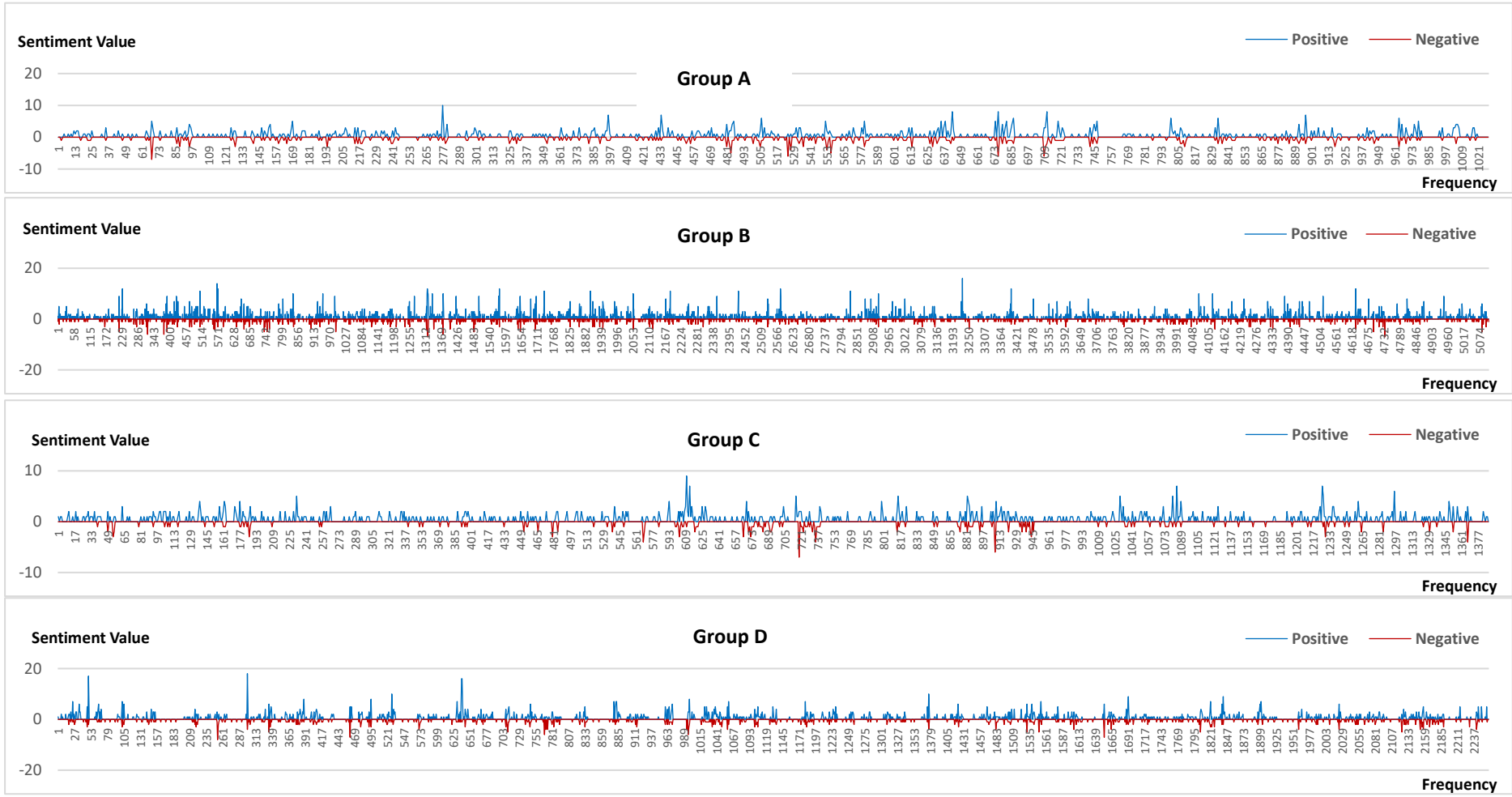
Table 4.9 Numbers of scam messages: Primary Poster and Echoes

Phase	Scam		Non-scam
	Primary Poster	Echoes	
Alluring	43	1,693	1,451
Catching	369	1,942	1,545
Executing	169	891	625

Table 4.10 Frequency of emotional vocabulary: Primary Poster and Echoes

Phase	Account	Anger	Disgust	Fear	Sadness	Surprise	Good	Happiness
Alluring	Primary Poster	0	2	2	2	0	42	111
	Echoes	0	369	142	52	7	1,100	24
Catching	Primary Poster	0	78	13	15	0	442	69
	Echoes	0	206	82	45	13	713	202
Executing	Primary Poster	0	48	21	9	0	333	43
	Echoes	0	83	36	18	5	291	69

This study further summarizes the frequency and polarity of scammers' positive and negative emotional vocabulary. Positive and negative emotion fold lines were plotted, and the numbers of positive and negative emotional vocabulary generated by different groups at various phases were collected, as shown in table 4.11 and figure 4.1. The study results indicate that the absolute value of positive emotions is significantly higher than that of negative emotions, and the number of positive emotional vocabulary is also higher than that of negative emotional vocabulary.


Sentiment Value
Positive
Negative
Group A
Frequency
Sentiment Value
Group B
Positive
Negative
Frequency
Sentiment Value
Group C
Positive
Negative
Frequency
Sentiment Value
Group D
Positive
Negative
Frequency

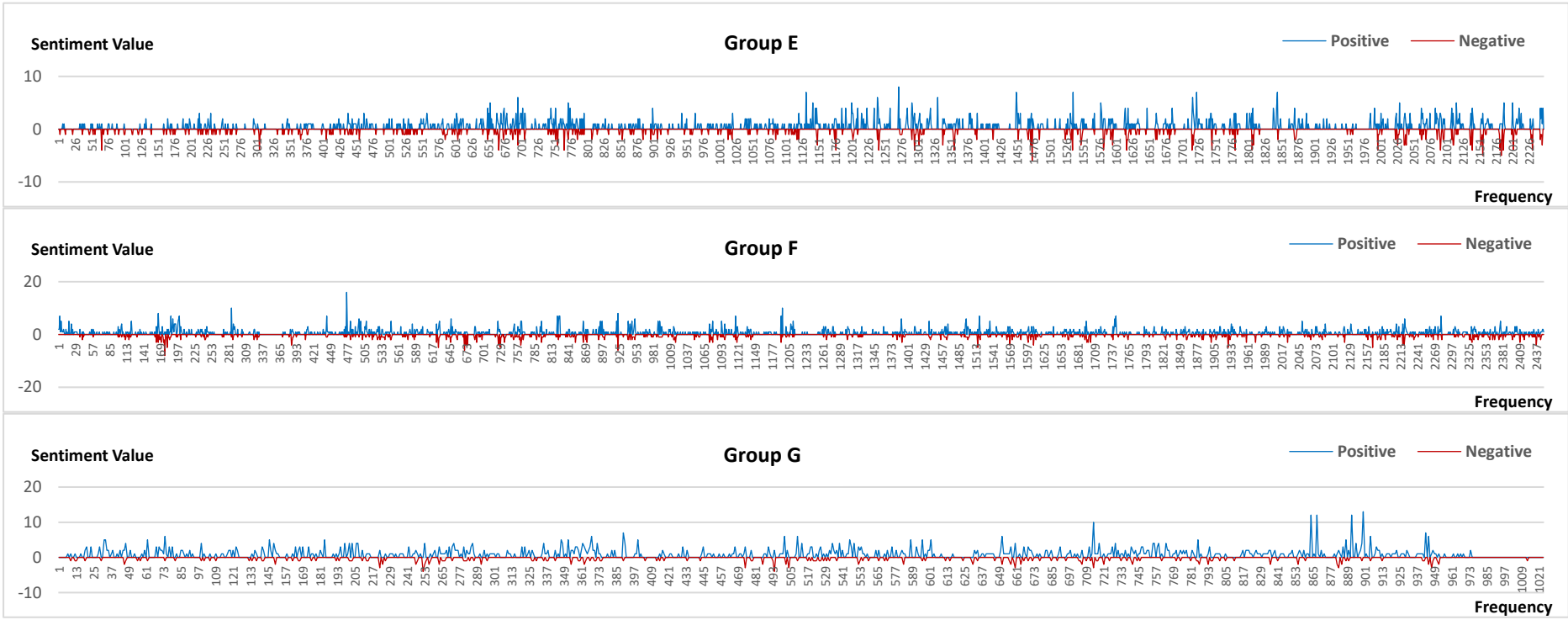


Figure 4.1 Time distribution of sentiment polarities

Table 4.11 Frequency of scammers' emotional vocabulary

Phase	Group	Positive	Negative
Alluring	A	1386	565
	B	254	112
	C	444	142
	D	598	277
	E	458	225
	F	95	18
	G	891	218
Catching	A	1434	436
	B	395	224
	C	189	58
	D	836	429
	E	421	150
	F	1227	562
	G	23	11
Executing	A	736	214
	B	50	10
	C	115	22
	D	28	13
	E	295	147
	F	315	151
	G	7	1

To test the hypothesis that emotional fluctuation patterns were similar across the seven groups in each phase of the IFLC framework, a MANOVA was conducted. The results, including Wilks' lambda, Pillai's trace, Hotelling-Lawley trace, and Roy's greatest root, are presented in Table 4.12. The p-values for the four test statistics were 0.44, 0.39, 0.48, and 0.28, respectively, all of which were greater than 0.05, indicating no significant difference in emotional fluctuation patterns among the seven groups. This finding provides empirical evidence supporting our proposed IFLC framework and characterizes the emotional fluctuations of scammers during the investment fraud process. In addition, these results have implications for developing scam detection models that can use emotional analysis to identify and prevent investment fraud.

Table 4.12 Multivariate linear models

		Value	Num DF	Den DF	F-Value	p-value
Phases	Wilks' lambda	0.46	12.00	26.00	1.04	0.44
	Pillai's trace	0.65	12.00	28.00	1.11	0.39
	Hotelling-Lawley trace	0.97	12.00	17.38	1.01	0.48
	Roy's greatest root	0.60	6.00	14.00	1.41	0.28

4.2.3 Supervised machine learning analysis and results

In the second stage of this study, SVM, Decision Tree, and Random Tree models were utilized to detect investment fraud messages. Four native features, a scam label, and seven emotion features were employed for the analysis. Several performance metrics were utilized to ensure accurate analysis, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). Accuracy represents the ratio of correct predictions to total observations, while precision indicates the ratio of correct positive predictions to the total number of positive observations. Recall represents the ratio of correct positive predictions to the actual number of positive observations, and the F-measure is the weighted average of precision and recall. Additionally, the AUC measures the performance of the classification model across various thresholds. Higher AUC values suggest superior performance (Witten et al., 2017).

Table 4.13 presents the detailed efficiency analysis results of the three algorithms. All utilized algorithms achieved an accuracy rate exceeding 0.7. In addition, the Decision Tree model achieved relatively high scores with an accuracy score of 0.865, precision of 0.893, recall of 0.865, F-measure of 0.862, and AUC of 1.000, indicating exceptional performance. The results of this study support the efficacy of the investment fraud detection model, suggesting that it could be a valuable tool for identifying fraudulent schemes in the financial domain.

Table 4.13 Algorithm evaluation results

Classifier	Accuracy	Precision	Recall	*F* measure	AUC
SVM	0.729	0.736	0.730	0.730	0.733
Decision Tree	0.865	0.893	0.865	0.862	1.000
Random Tree	0.715	0.715	0.715	0.714	0.817

5. Discussion and Conclusions

This study proposes two experiments to construct the anomalous transaction and investment fraud detection models based on RCT and emotion analysis. The first experiment involves extracting critical fraudulent parameters from the retail transaction dataset to improve the accuracy of the analysis, including holiday promotions, continuing transactions, multi-product transactions, and bulk goods. These variables were identified based on the limited literature on RCT. The analysis of 212,792 retail transactions from Taiwan shows that incorporating these fraudulent parameters enhances fraud detection. In the second experiment, we identified the emotional vocabulary from the dataset that revealed the emotional fluctuation among the three phases: alluring, catching, and executing. These phases were constructed based on fraud-related literature. The emotion analysis was conducted on the dataset of 29,438 messages from seven scam chat rooms on the IM app. Second, we adopted MANOVA to illustrate the pattern within the emotional fluctuation among the seven groups in each phase. Our results show a similar fluctuation pattern in the alluring, catching, and executing phases. In the last, we adopted supervised learning algorithms in both experiments to measure the performance of the detection models.

5.1 Theoretical Implications

Based on the theoretical foundation of criminology, there are various causes for the occurrence of crime and fraud. For example, (Cohen & Felson, 1979) routine activity theory illustrated that the opportunity for crime increases when three elements coincide. These three elements are individuals with the ability and motivation to commit the crime, suitable targets, and a lack of an appropriate monitoring environment. Crime pattern theory suggests that the spatial and temporal aspects of crime events are related to the crime's geography and activity frequency (Brantingham & Brantingham, 1984). The RCT holds that criminal behavior is the result of careful consideration and that different execution strategies are adopted in different contexts because criminals consider the benefits, risks, and costs of the crime (Cornish & Clarke, 1987). From these three theories, the similarity is that opportunity can promote the occurrence of crime. Felson and Clarke (1998a)integrated the three theories into the crime opportunity theory. They first used the routine activity theory to explore the changes in the opportunity for crime, then used the crime pattern theory to examine the spatial pattern of crime or differences that affect crime, and finally, used RCT to analyze how criminal actions result from rational choices. In

summary, crime is purposeful, as criminals observe their surroundings to identify areas with weak or vulnerable internal controls before committing a crime.

The fraud triangle theory has been the basis of many studies exploring fraudulent behavior (Huang et al., 2017; Schuchter & Levi, 2016). With the increasing complexity of fraudulent behavior, the analysis of various fraudulent parameters beyond the conventional fraud triangle theory has become necessary (Lokanan, 2015). In the initial experiment, we utilized RCT to identify fraudulent parameters, which are not commonly employed in this field of study. Our results showed that the four variables extracted from this theory were effective in representing the interests, focus, and cognition of fraud perpetrators. Thus, this approach produced a valuable behavioral model for analysis. The two-stage design employed in this study confirmed the efficacy of these newly identified fraudulent parameters. In the initial phase, the discrete values effectively identified anomalous transactions, and the correlation analysis validated the significant association of the attributes, indicating a stronger connection for the newly identified fraudulent parameters. Researchers need to deduce the original dataset's relevant parameters and corresponding information to obtain more precise predictions. Although many studies have explored the inclusion of variables in machine learning, they rarely explain the reasoning behind the selection of specific variables. Our research provides a detailed explanation of our selection of fraudulent parameters. In conclusion, the anomalous transactions detection model performed well using only 13 features, and the addition of more features would only increase computational cost without enhancing accuracy, a finding supported by Domingos (2012) and Feng et al. (2018).

In the second experiment, we noticed that many studies investigating fraudulent messages had utilized supervised machine learning algorithms to identify and classify such messages (Hanus et al., 2022; Shalke & Achary, 2022). However, the task of identifying fraudulent messages has become increasingly complex. Therefore, researchers have attempted to analyze scams through various aspects, such as how scammers obtain victims' information or exploit data (Esparza, 2019). In this study, we developed an IFLC that includes the lurking, alluring, catching, executing, and vanishing phases to illustrate various scam behaviors. Additionally, the IFLC was determined by analyzing seven chat room messages through emotion analysis. Similar to Garzia et al. (2022), we found that the frequency of emotional vocabulary could reflect scammers' intentions and behavior throughout the scam process. RCT in criminology posits that fraud or scams, and other criminal behaviors, are actions carried out after the criminal has thoroughly considered the right time, place, target, and criminal methods (Clarke & Cornish, 1985; Cornish & Clarke, 1987). Researchers have subsequently confirmed this

perspective (Ding & Zhai, 2021; Kuo & Tsang, 2022).

The emotion analysis results of this study found that approximately 60% of emotional words frequency in each phase belong to "good" emotions, 20% are "disgust," and the rest are distributed among "fear," "sadness," and "surprise." Based on the dataset used in this study, the IFLC begins in the lurking phase, where the scammer first collects the IM account of the victims and then invites them to the investment scheme chat room, providing short-term profitable stock information in the alluring phase. The messages include “recommending stock code 2476, which can be purchased at 56 NT dollars, or taking a stop-loss order, which can be sold at 59 NT dollars for a total profit of 4%. As Fischer et al. (2013) pointed out, trust can increase the willingness of the victim to participate in the scam. Next, in the catching phase, the scammer provides disproportionate high-return investment opportunities for specific members. The scammer even finds people who claim to be investment masters to join the chat room and provide investment courses on the online platform, analyzing the current stock market trend through the identity of the authority to promote. The main reason is that people are more likely to trust authority figures (Cialdini & James, 2009). In the online course platform's discussion board, exaggerated compliments to investment masters constantly appear, creating an atmosphere for more profitable investment following the famous master. Then influence other people in the community in this manner (Lacey et al., 2020). In the executing phase, the scammer continues to release limited membership quotas of the investment group or the deadline for membership up until today, making the victim eager to make a decision (Whitty, 2017). For individuals who have joined the member group, the scammer provides unfamiliar operating tools for investment or changes the investment target from stocks to foreign exchange, unconsciously exposing the victim to high-risk investment environments (Whitty, 2019a). When the victims consecutively fall into investment traps, the IFLC enters the vanishing phase, and the chat room is quickly closed. Overall, such message content reflects that most emotion vocabularies belong to "good." The MANOVA results verify that there are fluctuation patterns among the alluring, catching, and executing phases according to the messages we collect.

We constructed an investment fraud detection model using seven types of sentiment dimensions obtained from emotion analysis as analysis variables. Results from SVM, Decision Tree, and Random Tree showed that all three algorithms performed well in detecting investment fraud behavior, with Decision Tree achieving the highest accuracy. This suggests that our proposed detection model is effective in identifying scam behavior.

5.2 Managerial Implications

In recent years, the number of business fraud cases in Taiwan has been on the rise. However, as anti-fraud strategies become more widely known, the public's awareness and sensitivity towards this issue have grown. Moreover, advancements in new technologies such as big data, machine learning, and artificial intelligence have led to significant breakthroughs in the anti-fraud industry. Enterprises have a growing interest in detecting anomalous transactions using data analytics. By adopting Business Intelligence-based analytic methods, businesses can establish early warning systems against fraudulent activities (Chang et al., 2015; Dilla & Raschke, 2015). Numerous innovative detection models featuring various attributes have been proposed by researchers to identify abnormal behavior in fraud detection.

In the first experiment, we utilized RCT to identify essential attributes and incorporate fraudulent parameters into our analysis. Our goal was to represent the behavior of perpetrators through four fraudulent parameters: holiday promotions, multi-product transactions, continuing transactions, and bulk goods. During holiday promotions, individuals tend to increase their consumption spending with credit cards in order to take advantage of cashback or promotional offers provided by card-issuing banks. As a result, anomalous transactions may be disguised as regular transactions. In addition, fraud perpetrators may attempt to repeatedly purchase goods with incorrect or low price tags to engage in retail arbitrage. As the turnover rate of goods increases, greed may intensify. Therefore, when constructing an anomalous transactions detection model for price manipulation in retail, it is important to consider the four identified fraudulent parameters based on RCT.

In the second experiment, we identify the scammers' emotional fluctuation patterns based on emotion analysis, which is a valuable technique for uncovering individuals' underlying emotions, sentiments, and feelings as expressed through text (Peng et al., 2022). The scammers' intention was reflected in the three emotions vocabularies: good, disgust, and happiness. In general, scammers utilize various rhetorical strategies to lure victims, with a significant proportion promoting the advantages of investment opportunities. This study also identified the presence of Echoes, which play a multifaceted role throughout the scam scheme, including posing questions and providing the answer. We found that Echoes sometimes question and challenge Main Poster's messages and then uses another account to provide supportive answers to the Main

Poster's message. Thus, when developing an investment fraud detection model, the application of emotion analysis can demonstrate utility as a tool. In other words, emotion analysis helps the anti-scam industry achieve its goal of providing an effective method to researchers. Furthermore, machine learning can be implemented by using various software, such as Python and R. In this study, we used Python and WEKA as our analysis tools which were widely used in the field of research. The current popular low-code development method is to use a small amount of code to develop so that researchers can quickly conduct data analysis, thus improving the efficiency of scam detection. Ultimately, the IFLC illustrates the sequential phases of a scam scheme that is designed to defraud individuals. It can be used to understand the modus operandi of scammers and the methods they use to deceive individuals. Therefore, according to Fortinet (2022), The methods employed in cyber attacks are diverse and ever-changing. Thus, enterprises should adopt a multifaceted approach to addressing potential cyber security threats. One such approach is to understand the life cycle of cyber attacks. Similarly, when it comes to investment fraud, individuals need to grasp the motivations and tactics of scammers and the life cycle of an investment fraud to comprehend the modus operandi of these scammers.

Despite the long history of machine learning development, the advancement of hardware and software technologies such as computer chips and the internet has facilitated rapid progress in the overall development and application of artificial intelligence. In 2023, the emergence of killer applications such as ChatGPT, Midjourney, and Coplit will provide opportunities for people without programming experience to interact with AI. In addition, data-oriented issues have become increasingly important in order to improve decision-making accuracy and make more objective and comprehensive conclusions. The Anti-Fraud Technology Benchmarking Report issued by ACFE and SAS (2022) provides information on the use of data analytics for fraud prevention. The report shows that almost all surveyed enterprises agree that implementing anti-fraud analysis benefits the ability to audit transaction data and improves the real-time detection of anomalies. More than half of the enterprises indicate that data analytics is mainly used for anomaly detection and red flag monitoring. Additionally, due to the COVID-19 pandemic, enterprises have become more reliant on information technology, and therefore, 43% of enterprises have accelerated their use of data analytics as a fraud prevention tool. Like the WEKA software used in this study, which provides data preprocessing tools, classification tools, clustering tools, and various types of algorithms, they lower the threshold for researchers to enter the field of data analysis, allowing them to conduct data analysis without complex programming skills. It is believed that the efficiency of fraud detection can be improved due to technological advancement's progress.

5.3 Limitations and Future Research Directions

Challenges in data collection constrained the empirical analysis of both fraud detection models. Due to data collection spanning less than one year, the limited data volume impacted the performance of the first anomalous transaction detection model. Future research could benefit from gathering more data over an extended period. Moreover, the post-transaction data analysis contrasts with the real-time analysis capabilities of modern enterprise information systems. Consequently, future research should explore synchronizing these detection models with real-time systems for improved performance.

Furthermore, difficulties in collecting chat room messages hindered the development of the investment fraud detection model. Scammers' aim to expedite their plans often results in varying durations of the executing phase within different chat rooms, leading to fewer messages during this phase. Therefore, a larger dictionary of chat room messages would likely improve the model's effectiveness. Additionally, not all scams adhere to the IFLC proposed in this study, with many being more complex and involving extra phases. Therefore, future research should consider collecting a more diverse data set to refine the IFLC. Lastly, integrating real-time detection capabilities into scam detection models has recently gained traction. As such, future research could focus on enhancing the emotion analysis methodology used in this study to enable real-time detection through time series analysis.

Reference

ACCC. (2016). *The little black book of scams* https://www.accc.gov.au/publications/the-little-black-book-of-scams

ACFE. (2019). *2020 Fraud Examiners Manual* (International ed.). Association of Certified Fraud Examiners, Inc.

ACFE. (2021). *2022 Fraud Examiners Manual*. Association of Certified Fraud Examiners, Inc.

ACFE. (2022). *2023 Fraud Examiners Manual*. Association of Certified Fraud Examiners, Inc. https://fraudexaminersmanual.com/

ACFE, & SAS. (2022). *Anti-Fraud Technology Benchmarking Report*. SAS. https://www.sas.com/en/whitepapers/acfe-anti-fraud-technology-110652.html

Akers, R. L. (1990). Rational choice, deterrence, and social learning theory in criminology: The path not taken. *The Journal of Criminal Law and Criminology, 81*, 653.

Alghamdi, B., & Alharby, F. (2019). An intelligent model for online recruitment fraud detection. *Journal of Information Security, 10*(3), 155-176.

Arya, M., & Sastry G, H. (2020). DEAL – 'Deep Ensemble ALgorithm' Framework for Credit Card Fraud Detection in Real-Time Data Stream with Google TensorFlow. *Smart Science, 8*(2), 71-83. https://doi.org/10.1080/23080477.2020.1783491

Askari, S. M. S., & Hussain, M. A. (2020). IFDTC4.5: Intuitionistic fuzzy logic based decision tree for E-transactional fraud detection. *Journal of Information Security and Applications, 52*, 102469. https://doi.org/https://doi.org/10.1016/j.jisa.2020.102469

Bartoletti, M., Carta, S., Cimoli, T., & Saia, R. (2020). Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. *Future Generation Computer Systems, 102*, 259-277. https://doi.org/10.1016/j.future.2019.08.014

Beals, M., DeLiema, M., & Deevy, M. (2015). *Framework for a taxonomy of fraud*. S. C. o. Longevity. https://longevity.stanford.edu/framework-for-a-taxonomy-of-fraud/

Birzhandi, P., Kim, K. T., Lee, B., & Youn, H. Y. (2019). Reduction of Training Data Using Parallel Hyperplane for Support Vector Machine. *Applied Artificial Intelligence, 33*(6), 497-516. https://doi.org/10.1080/08839514.2019.1583449

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics, 14*(1), 106. https://doi.org/10.1186/1471-2105-14-106

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan New

York.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Buchanan, T., & Whitty, M. T. (2014). The online dating romance scam: causes and consequences of victimhood. *Psychology, Crime & Law, 20*(3), 261-283. https://doi.org/10.1080/1068316X.2013.772180

Bullée, J. W. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. (2018). On the anatomy of social engineering attacks—A literature-based dissection of successful attacks. *Journal of investigative psychology and offender profiling, 15*(1), 20-45. https://doi.org/10.1002/jip.1482

Butavicius, M., Taib, R., & Han, S. J. (2022). Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails. *Computers & Security, 123*, 102937. https://doi.org/https://doi.org/10.1016/j.cose.2022.102937

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. https://doi.org/https://doi.org/10.1016/j.ins.2019.05.042

Chan, F., & Gibbs, C. (2019). Integrated Theories of White-Collar and Corporate Crime. In *The Handbook of White-Collar Crime* (pp. 191-207). https://doi.org/https://doi.org/10.1002/9781118775004.ch13

Chang, B., Kuo, C., Wu, C.-H., & Tzeng, G.-H. (2015). Using Fuzzy Analytic Network Process to assess the risks in enterprise resource planning system implementation. *Applied Soft Computing, 28*, 196-207. https://doi.org/10.1016/j.asoc.2014.11.025

Charm, T., Perrey, J., Poh, F., & Ruwadi, B. (2020). *2020 Holiday Season: Navigating shopper behaviors in the pandemic*. P. b. Mckinsey.

Chaudhry, J. A., Chaudhry, S. A., & Rittenhouse, R. G. (2016). Phishing attacks and defenses. *International Journal of Security Its Applications, 10*(1), 247-256. https://doi.org/10.14257/ijsia.2016.10.1.23

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16*, 321-357. https://doi.org/10.1613/jair.953

Chen, C., & Li, X. (2020). The effect of online shopping festival promotion strategies on consumer participation intention. *Industrial Management & Data Systems*.

Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications, 106*, 1-20. https://doi.org/https://doi.org/10.1016/j.eswa.2018.03.050

Chiluwa, I. M. (2019). “Truth,” Lies, and Deception in Ponzi and Pyramid Schemes. In *Handbook of Research on Deception, Fake News, and Misinformation Online* (pp. 439-458). IGI Global.

Choi, J., Kruis, N. E., & Choo, K.-S. (2021). Explaining Fear of Identity Theft Victimization Using a Routine Activity Approach. *Journal of Contemporary Criminal Justice, 37*(3), 406-426. https://doi.org/10.1177/10439862211001627

Cialdini, R. B., & James, L. (2009). *Influence: Science and practice* (Vol. 4). Pearson education Boston.

[Record #682 is using a reference type undefined in this output style.]

Clarke, R. V. (1983). Situational Crime Prevention: Its Theoretical Basis and Practical Scope. *Crime and Justice, 4*, 225-256. https://doi.org/10.1086/449090

Clarke, R. V., & Cornish, D. B. (1985). Modeling Offenders' Decisions: A Framework for Research and Policy. *Crime and Justice, 6*, 147-185. https://doi.org/10.1086/449106

Clarke, R. V., & Harris, P. M. (1992). A rational choice perspective on the targets of automobile theft. *Criminal Behaviour and Mental Health, 2*(1), 25-42. https://doi.org/https://doi.org/10.1002/cbm.1992.2.1.25

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American sociological review, 44*(4), 588-608. https://doi.org/10.2307/2094589

Cornish, D. B., & Clarke, R. V. (1987). Understanding crime displacement: an application of rational choice theory. *Criminology, 25*(4), 933-948. https://doi.org/https://doi.org/10.1111/j.1745-9125.1987.tb00826.x

Cornish, D. B., & Clarke, R. V. (1989). Crime Specialisation, Crime Displacement and Rational Choice Theory. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal Behavior and the Justice System: Psychological Perspectives* (pp. 103-117). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-86017-1_7

Correa Bahnsen, A., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications, 51*, 134-142. https://doi.org/10.1016/j.eswa.2015.12.030

Cressey, D. R. (1973). *Other People's Money: A Study in the Social Psychology of Embezzlement.* Patterson Smith.

Cui, B., & He, S. (2016). Anomaly detection model based on hadoop platform and weka interface. 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS),

Deliema, M., Shadel, D., & Pak, K. (2020). Profiling victims of investment fraud: Mindsets and risky behaviors. *Journal of Consumer Research, 46*(5), 904-914. https://doi.org/10.1093/jcr/ucz020

Deloitte. (2018). *2018 Taiwan Corporate Fraud Risk Management Survey and Future Outlook*. D. A. P. S. Limited. https://www2.deloitte.com/tw/tc/pages/risk/articles/2018-fraud-report-press.html

Dilla, W. N., & Raschke, R. L. (2015). Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems, 16*, 1-22. https://doi.org/https://doi.org/10.1016/j.accinf.2015.01.001

Ding, N., & Zhai, Y. (2021). Crime prevention of bus pickpocketing in Beijing, China: does air quality affect crime? *Security Journal, 34*(2), 262-277. https://doi.org/10.1057/s41284-019-00226-1

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87.

Dopson, L. R., & Hayes, D. K. (2015). *Food and beverage cost control*. John Wiley & Sons.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), 124. https://doi.org/https://doi.org/10.1037/h0030377

Esparza, J. M. (2019). Understanding the credential theft lifecycle. *Computer Fraud & Security, 2019*(2), 6-9. https://doi.org/https://doi.org/10.1016/S1361-3723(19)30018-1

Fang, Y., Xie, M., & Huang, C. (2021). PBDT: Python Backdoor Detection Model Based on Combined Features. *Security and Communication Networks, 2021*, 9923234. https://doi.org/10.1155/2021/9923234

Felson, M., & Clarke, R. V. (1998a). Opportunity makes the thief. *Police research series, paper, 98*(1-36), 10.

Felson, M., & Clarke, R. V. (1998b). Opportunity Makes the Thief. *Police Research, 98*.

Feng, Y., Akiyama, H., Lu, L., & Sakurai, K. (2018). *Feature selection for machine learning-based early detection of distributed cyber attacks* 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech),

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research, 61*, 863-905. https://doi.org/10.1613/jair.1.11192

Fischer, P., Lea, S. E. G., & Evans, K. M. (2013). Why do individuals respond to fraudulent scam communications and lose money? The psychological

determinants of scam compliance. *Journal of Applied Social Psychology, 43*, 2060-2072. https://doi.org/10.1111/jasp.12158

Fortinet. (2022). *Cyber Threat Predictions for 2023*. I. Fortinet. https://www.fortinet.com/content/dam/maindam/PUBLIC/02_MARKETING/02_Collateral/WhitePaper/WP-threat-prediction-2023.pdf

Fraud. In. *Dictionary by Merriam-Webster*. Retrieved March 1, 2023, from https://www.merriam-webster.com/dictionary/fraud

Garzia, F., Borghini, F., Makshanova, E., Lombardi, M., & Ramalingam, S. (2022). Emotional analysis of safeness and risk perception of cybersecurity attacks during the COVID-19 pandemic. 2022 IEEE International Carnahan Conference on Security Technology (ICCST),

Gee, S. (2014). *Fraud and fraud detection: a data analytics approach*. John Wiley & Sons.

Guedes, I., Martins, M., & Cardoso, C. S. (2022). Exploring the determinants of victimization and fear of online identity theft: an empirical study. *Security Journal*. https://doi.org/10.1057/s41284-022-00350-5

Guidotti, R., Coscia, M., Pedreschi, D., & Pennacchioli, D. (2015). Behavioral entropy and profitability in retail. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA),

Guri, M., Puzis, R., Choo, K.-K. R., Rubinshtein, S., Kedma, G., & Elovici, Y. (2019). Using malware for the greater good: Mitigating data leakage. *Journal of Network and Computer Applications, 145*, 102405. https://doi.org/https://doi.org/10.1016/j.jnca.2019.07.006

Hadnagy, C. (2010). *Social engineering: The art of human hacking*. John Wiley & Sons.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods. *Knowledge-Based Systems, 128*, 139-152.

Hanus, B., Wu, Y. A., & Parrish, J. (2022). Phish Me, Phish Me Not. *Journal of Computer Information Systems, 62*(3), 516-526. https://doi.org/10.1080/08874417.2020.1858730

Hatfield, J. M. (2018). Social engineering in cybersecurity: The evolution of a concept. *Computers & Security, 73*, 102-113.

Henriques, D. B. (2011). *The wizard of lies: Bernie Madoff and the death of trust*. Macmillan.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management science, 27*(2), 115-138. https://doi.org/10.1287/mnsc.27.2.115

Holtfreter, K. (2005). Is occupational fraud "typical" white-collar crime? A comparison of individual and organizational characteristics. *Journal of Criminal Justice,*

33(4), 353-365. https://doi.org/10.1016/j.jcrimjus.2005.04.005

Hooda, N., Bawa, S., & Rana, P. S. (2018). Fraudulent firm classification: a case study of an external audit. *Applied Artificial Intelligence, 32*(1), 48-64.

Hooda, N., Bawa, S., & Rana, P. S. (2020). Optimizing fraudulent firm prediction using ensemble machine learning: a case study of an external audit. *Applied Artificial Intelligence, 34*(1), 20-30.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hsia, T.-L., Wu, J.-H., Xu, X., Li, Q., Peng, L., & Robinson, S. (2020). Omnichannel retailing: The role of situational involvement in facilitating consumer experiences. *Information & Management, 57*(8), 103390.

Huang, F. L. (2019). MANOVA: A Procedure Whose Time Has Passed? *Gifted Child Quarterly, 64*(1), 56-60. https://doi.org/10.1177/0016986219887200

Huang, M.-H., & Rust, R. T. (2022). A Framework for Collaborative Artificial Intelligence in Marketing. *Journal of Retailing, 98*(2), 209-223. https://doi.org/https://doi.org/10.1016/j.jretai.2021.03.001

Huang, S. Y., Lin, C.-C., Chiu, A.-A., & Yen, D. C. (2017). Fraud detection using fraud triangle risk factors. *Information Systems Frontiers, 19*(6), 1343-1356.

Hudlicka, E. (2011). Guidelines for Designing Computational Models of Emotions. *International Journal of Synthetic Emotions, 2*(1), 26-79. https://doi.org/10.4018/jse.2011010103

IAASB. (2013). International Standard on Auditing 240: The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements.

Janjua, F., Masood, A., Abbas, H., Rashid, I., & Khan, M. M. Z. M. (2021). Textual analysis of traitor-based dataset through semi supervised machine learning. *Future Generation Computer Systems, 125*, 652-660. https://doi.org/https://doi.org/10.1016/j.future.2021.06.036

Jones, H. S., Towse, J. N., Race, N., & Harrison, T. (2019). Email fraud: The search for psychological predictors of susceptibility. *PLOS ONE, 14*(1), e0209684. https://doi.org/10.1371/journal.pone.0209684

Jones, K. S., Armstrong, M. E., Tornblad, M. K., & Siami Namin, A. (2021). How social engineers use persuasion principles during vishing attacks. *Information & Computer Security, 29*(2), 314-331. https://doi.org/10.1108/ICS-07-2020-0113

Junger, M., Wang, V., & Schlömer, M. (2020). Fraud against businesses both online and offline: crime scripts, business characteristics, efforts, and benefits. *Crime Science, 9*(1), 13. https://doi.org/10.1186/s40163-020-00119-4

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection.

Expert Systems with Applications, 100, 234-245. https://doi.org/10.1016/j.eswa.2018.01.037

Karmy, J. P., & Maldonado, S. (2019). Hierarchical time series forecasting via Support Vector Regression in the European Travel Retail Industry. *Expert Systems with Applications, 137*, 59-73. https://doi.org/https://doi.org/10.1016/j.eswa.2019.06.060

Kashyap, A. (2019). *How digital transformation increases consumer and retail fraud risks*. https://www.ey.com/en_ro/how-digital-transformation-increases-consumer-and-retail-fraud-risks

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security applications, 22*, 113-122. https://doi.org/10.1016/j.jisa.2014.09.005

Kuo, C., & Tsang, S.-S. (2022). Detection of price manipulation fraud through rational choice theory: evidence for the retail industry in Taiwan. *Security Journal*. https://doi.org/10.1057/s41284-022-00360-3

Lacey, D., Goode, S., Pawada, J., & Gibson, D. (2020). The application of scam compliance models to investment fraud offending. *Journal of Criminological Research, Policy and Practice, 6*(1), 65-81. https://doi.org/10.1108/JCRPP-12-2019-0073

Langenderfer, J., & Shimp, T. A. (2001). Consumer vulnerability to scams, swindles, and fraud: A new theory of visceral influences on persuasion [https://doi.org/10.1002/mar.1029]. *Psychology & Marketing, 18*(7), 763-783. https://doi.org/https://doi.org/10.1002/mar.1029

Lavanya, P., Kouser, K., & Suresha, M. (2021). Effective feature representation using symbolic approach for classification and clustering of big data. *Expert Systems with Applications, 173*, 114658.

Lee, C. S. (2021). How Online Fraud Victims are Targeted in China: A Crime Script Analysis of Baidu Tieba C2C Fraud. *Crime & Delinquency, 68*(13-14), 2529-2553. https://doi.org/10.1177/00111287211029862

Levi, M. (2008). Organized fraud and organizing frauds: Unpacking research on networks and organization. *Criminology & Criminal Justice, 8*(4), 389-419.

Levy, D., Chen, H., Müller, G., Dutta, S., & Bergen, M. (2010). Holiday price rigidity and cost of price adjustment. *Economica, 77*(305), 172-198.

Li, C., Ding, N., Zhai, Y., & Dong, H. (2021). Comparative study on credit card fraud detection based on different support vector machines. *Intelligent Data Analysis, 25*, 105-119. https://doi.org/10.3233/IDA-195011

Li, D., & Santos Jr, E. (2020). Discriminating deception from truth and misinformation: an intent-level approach. *Journal of Experimental & Theoretical Artificial*

Intelligence, 32(3), 373-407. https://doi.org/10.1080/0952813X.2019.1652354

Li, T., Wang, X., & Ni, Y. (2020). Aligning social concerns with information system security: A fundamental ontology for social engineering. *Information Systems*, 101699.

Liu, S. M., & Chen, J.-H. (2015). A multi-label classification based approach for sentiment classification. *Expert Systems with Applications, 42*(3), 1083-1093. https://doi.org/https://doi.org/10.1016/j.eswa.2014.08.036

Lokanan, M. E. (2015). Challenges to the fraud triangle: Questions on its usefulness. *Accounting Forum, 39*(3), 201-224. https://doi.org/https://doi.org/10.1016/j.accfor.2015.05.002

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/https://doi.org/10.1016/j.future.2019.08.029

Masmoudi, K., Abid, L., & Masmoudi, A. (2019). Credit risk modeling using Bayesian network with a latent variable. *Expert Systems with Applications, 127*, 157-166. https://doi.org/https://doi.org/10.1016/j.eswa.2019.03.014

Mavlanova, T., Benbunan-Fich, R., & Lang, G. (2016). The role of external and internal signals in E-commerce. *Decision Support Systems, 87*, 59-68.

Mehbodniya, A., Alam, I., Pande, S., Neware, R., Rane, K. P., Shabaz, M., & Madhavan, M. V. (2021). Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques. *Security and Communication Networks, 2021*, 9293877. https://doi.org/10.1155/2021/9293877

Mendl, M. (1999). Performing under pressure: stress and cognitive function. *Applied Animal Behaviour Science, 65*(3), 221-244. https://doi.org/https://doi.org/10.1016/S0168-1591(99)00088-X

Mercer, G. (2016, 01/14). 4 Ways to Be a Successful Amazon Seller. https://www.linkedin.com/pulse/4-ways-successful-amazon-seller-greg-mercer/

Meyer, S. (2012). Reducing harm from explosive attacks against railways. *Security Journal, 25*(4), 309-325. https://doi.org/10.1057/sj.2011.23

Micro, T. (2021, 2021/04/01). *Top 3 Hot Sacms in 2021 investegated by Trend Micro, National Police Agency, and Consumers' Foundation* https://www.trendmicro.com/zh_tw/about/newsroom/press-releases/2021/2021-04-01.html

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2015). Tutorial and critical analysis of phishing websites methods. *Computer Science Review, 17*, 1-24. https://doi.org/https://doi.org/10.1016/j.cosrev.2015.04.001

MOI. (2023). *2022 Fraud techniques in Taiwan.* https://www.moi.gov.tw/News_Content.aspx?n=4&s=275532

Mqadi, N. M., Naicker, N., & Adeliyi, T. (2021). Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss. *Mathematical Problems in Engineering, 2021*, 7194728. https://doi.org/10.1155/2021/7194728

Mueller, E. A., Wood, S. A., Hanoch, Y., Huang, Y., & Reed, C. L. (2020). Older and wiser: age differences in susceptibility to investment fraud: the protective role of emotional intelligence. *Journal of Elder Abuse & Neglect, 32*(2), 152-172. https://doi.org/10.1080/08946566.2020.1736704

NPA. (2021). *Ministry of the Interior release the top scams in 2020* https://www.moi.gov.tw/News_Content.aspx?n=4&s=212607

NRF. (2020). *NRF expects holiday sales will grow between 3.6 and 5.2 percent*. National Retail Fedration. Retrieved 01/12 from https://nrf.com/media-center/press-releases/nrf-expects-holiday-sales-will-grow-between-36-and-52-percent

Oh, H., & Kwon, K. N. (2009). An exploratory study of sales promotions for multichannel holiday shopping. *International Journal of Retail Distribution Management*.

OIG. (2018). *Alleged Split Purchases at the VA St. Louis Health Care System*. O. o. I. G. Department of Veterans Affairs. https://www.va.gov/oig/pubs/VAOIG-16-02863-199.pdf

Otu, S. E., & Okon, O. N. (2019). Participation in Fraud/Cheat in the Buying and Selling of Meats Without Legal Metrology: A Theoretical and Empirical Investigations. *Deviant Behavior, 40*(2), 205-224. https://doi.org/10.1080/01639625.2017.1420458

Palmer, W. E., & Richardson, C. (2009). *Organized Retail Crime: Assessing the Risk and Developing Effective Strategies* (An ASIS Foundation Research Council CRISP Report, Issue. I. ASIS Foundation.

Park, S., Bae, B., & Cheong, Y. (2020, 19-22 Feb. 2020). Emotion Recognition from Text Stories Using an Emotion Embedding Model. 2020 IEEE International Conference on Big Data and Smart Computing (BigComp),

Peng, L., & Lin, R. (2018, 2-7 July 2018). Fraud Phone Calls Analysis Based on Label Propagation Community Detection Algorithm. 2018 IEEE World Congress on Services (SERVICES),

Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., Jia, W., & Yu, S. (2022). A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks, 8*(5), 745-762. https://doi.org/https://doi.org/10.1016/j.dcan.2021.10.003

Piza, E. L., Caplan, J. M., & Kennedy, L. W. (2017). CCTV as a tool for early police intervention: Preliminary lessons from nine case studies. *Security Journal, 30*(1), 247-265. https://doi.org/10.1057/sj.2014.17

Plutchik, R. (1958). Outlines of a new theory of emotion. *Trans N Y Acad Sci, 20*(5), 394-403. https://doi.org/10.1111/j.2164-0947.1958.tb00600.x

Plutchik, R. (2001). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist, 89*(4), 344-350. http://www.jstor.org/stable/27857503

Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. *IEEE Access, 7*, 100943-100953. https://doi.org/10.1109/ACCESS.2019.2929050

Qamar, S., Mujtaba, H., Majeed, H., & Beg, M. O. (2021). Relationship Identification Between Conversational Agents Using Emotion Analysis. *Cognitive Computation, 13*(3), 673-687. https://doi.org/10.1007/s12559-020-09806-5

Rajesh, P., & Karthikeyan, M. (2017). A comparative study of data mining algorithms for decision tree approaches using weka tool. *Advances in Natural Applied Sciences, 11*(9), 230-243.

Reurink, A. (2018). Financial fraud: a literature review. *Journal of Economic Surveys, 32*(5), 1292-1325. https://doi.org/10.1111/joes.12294

Robinson, W. N., & Aria, A. (2018). Sequential fraud detection for prepaid cards using hidden Markov model divergence. *Expert Systems with Applications, 91*, 235-251. https://doi.org/10.1016/j.eswa.2017.08.043

Romney, M. B., Albrecht, W. S., & Cherrington, D. J. (1980). Auditors and the detection of fraud. *Journal of Accountancy, 149*(5), 63-69.

Salahdine, F., & Kaabouch, N. (2019). Social Engineering Attacks: A Survey. *Future Internet, 11*(4).

Schmalleger, F. (1999). *Criminology today: an integrative introduction*. Prentice Hall Second edition. Upper Saddle River, NJ.

Schmalleger, F. (2021). *Criminology Today: An Integrative Introduction* (10th ed.). Pearson.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia, 126*(5). https://journals.lww.com/anesthesia-analgesia/Fulltext/2018/05000/Correlation_Coefficients__Appropriate_Use_and.50.aspx

Schuchter, A., & Levi, M. (2016). The fraud triangle revisited. *Security Journal, 29*(2), 107-121.

Shadel, D., & Pak, K. (2017). *AARP Investment Fraud Vulnerability Study*. AARP. https://www.aarp.org/research/topics/economics/info-2017/investment-fraud-survey.html

Shalke, C. J., & Achary, R. (2022, 28-30 April 2022). Social Engineering Attack and Scam Detection using Advanced Natural Langugae Processing Algorithm. 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI),

Shrivastava, G., & Kumar, P. (2021). Android application behavioural analysis for data leakage [https://doi.org/10.1111/exsy.12468]. *Expert Systems, 38*(1), e12468. https://doi.org/https://doi.org/10.1111/exsy.12468

Siering, M., Koch, J.-A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems, 33*(2), 421-455. https://doi.org/10.1080/0952813X.2019.1652354

Singh, K., & Best, P. (2019). Anti-money laundering: using data visualization to identify suspicious activity. *International Journal of Accounting Information Systems, 34*, 100418. https://doi.org/10.1016/j.accinf.2019.06.001

Singh, M., & Valtorta, M. (1995). Construction of Bayesian network structures from data: A brief survey and an efficient algorithm. *International Journal of Approximate Reasoning, 12*(2), 111-131. https://doi.org/https://doi.org/10.1016/0888-613X(94)00016-V

Springer, M. (2020). *The Politics of Ponzi Schemes: History, Theory and Policy*. Routledge.

Stamler, R. T., Marschdorf, H. J., & Possamai, M. (2014). *Fraud prevention and detection: warning signs and the red flag system*. CRC Press.

Steinmetz, K. F. (2020). The Identification of a Model Victim for Social Engineering: A Qualitative Analysis. *Victims & Offenders*, 1-25.

Tang, T., & Hu, P. (2019). Quantitative standard of promotion strategy and analysis on the influence of consumer purchase behavior. *Cluster Computing, 22*(2), 4949-4955.

Thornburgh, T. (2004). *Social engineering: the "Dark Art"* Proceedings of the 1st annual conference on Information security curriculum development, Kennesaw, Georgia. https://doi.org/10.1145/1059524.1059554

Torres, C. F., Baden, M., & State, R. (2020). Towards Usable Protection Against Honeypots. 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC),

Trahan, A., Marquart, J. W., & Mullings, J. (2005). Fraud and the American Dream: Toward an Understanding of Fraud Victimization. *Deviant Behavior, 26*(6), 601-620. https://doi.org/10.1080/01639620500218294

Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology, 29*(5), 3414-3424.

Tsai, T.-M., Chou, W.-Y., & Chen, Y.-S. (2009). *A Study on the Characteristics of Fraud Victims*. R. O. C. T. Ministry of the Interior. https://www.grb.gov.tw/search/planDetail?id=1895111&docId=313819

Tsang, S.-S., Kuo, C., Hu, T.-K., & Wang, W.-C. (2022). Exploring impacts of AR on group package tours: Destination image, perceived certainty, and experiential value. *Journal of Vacation Marketing*, 13567667221078244. https://doi.org/10.1177/13567667221078244

Tseng, S. C. (2018). An analysis of first sale rule in the Trademarks Act. *Taiwan Bar Journal, 12*, 24-44.

Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review, 52*(1), 441-447.

Ullah, I., Ahmad, W., & Ali, A. (2022). Determinants of investment decision in a Ponzi scheme: Investors' perspective on the Modaraba scam. *Journal of Financial Crime, 29*(4), 1172-1190. https://doi.org/10.1108/JFC-02-2020-0027

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems, 75*, 38-48.

Wang, L., Cheng, H., Zheng, Z., Yang, A., & Zhu, X. (2021). Ponzi scheme detection via oversampling-based Long Short-Term Memory for smart contracts. *Knowledge-Based Systems, 228*, 107312. https://doi.org/https://doi.org/10.1016/j.knosys.2021.107312

Wei, Z., Liu, W., Zhu, G., Zhang, S., & Hsieh, M.-Y. (2022). Sentiment classification of Chinese Weibo based on extended sentiment dictionary and organisational structure of comments. *Connection Science, 34*(1), 409-428. https://doi.org/10.1080/09540091.2021.2006146

Wen, X., Xu, L., Wang, J., Gao, Y., Shi, J., Zhao, K., Tao, F., & Qian, X. (2022). Mental States: A Key Point in Scam Compliance and Warning Compliance in Real Life. *International Journal of Environmental Research and Public Health, 19*(14).

Whitty, M. T. (2013). The Scammers Persuasive Techniques Model: Development of a Stage Model to Explain the Online Dating Romance Scam. *The British Journal of Criminology, 53*(4), 665-684. https://doi.org/10.1093/bjc/azt009

Whitty, M. T. (2015). Anatomy of the online dating romance scam. *Security Journal, 28*(4), 443-455. https://doi.org/10.1057/sj.2012.57

Whitty, M. T. (2017). Do You Love Me? Psychological Characteristics of Romance

Scam Victims. *Cyberpsychology, Behavior, and Social Networking, 21*(2), 105-109. https://doi.org/10.1089/cyber.2016.0729

Whitty, M. T. (2019a). Predicting susceptibility to cyber-fraud victimhood. *Journal of Financial Crime, 26*(1), 277-292. https://doi.org/10.1108/JFC-10-2017-0095

Whitty, M. T. (2019b). Who can spot an online romance scam? *Journal of Financial Crime, 26*(2), 623-633. https://doi.org/10.1108/JFC-06-2018-0053

Whitty, M. T. (2020a). Is there a scam for everyone? Psychologically profiling cyberscam victims. *European Journal on Criminal Policy, 26*(3), 399-409. https://doi.org/10.1007/s10610-020-09458-z

Whitty, M. T. (2020b). Is There a Scam for Everyone? Psychologically Profiling Cyberscam Victims. *European Journal on Criminal Policy and Research, 26*(3), 399-409. https://doi.org/10.1007/s10610-020-09458-z

Wilcox, P. (2015). Routine Activities, Criminal Opportunities, Crime and Crime Prevention. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 772-779). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.45080-4

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies, 120*, 1-13. https://doi.org/https://doi.org/10.1016/j.ijhcs.2018.06.004

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers. https://doi.org/10.1016/C2015-0-02071-8

Xu, D., Tian, Z., Lai, R., Kong, X., Tan, Z., & Shi, W. (2020). Deep learning based emotion analysis of microblog texts. *Information Fusion, 64*, 1-11. https://doi.org/https://doi.org/10.1016/j.inffus.2020.06.002

Xu, L., Lin, H., Pan, Y., Ren, H., & Chen, J. (2008). Constructing the Affective Lexicon Ontology. *Journal of the China Society for Scientific and Technical Information*(2), 180-185.

Yan, Q. (2022). Real-Time Analysis of Youth Emotion Based on Python Language and Smart Sensor Network. *Mobile Information Systems, 2022*, 8635787. https://doi.org/10.1155/2022/8635787

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 Computer Engineering, 10*(1-4), 23-27.

Zhang, S., Wei, Z., Wang, Y., & Liao, T. (2018). Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Future Generation Computer Systems, 81*, 395-403. https://doi.org/https://doi.org/10.1016/j.future.2017.09.048

Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences, 557*, 302-316. https://doi.org/10.1016/j.ins.2019.05.023

Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. *Security Communication Networks, 2018*. https://doi.org/10.1155/2018/5680264

Zhao, Q., Chen, K., Li, T., Yang, Y., & Wang, X. (2018). Detecting telecommunication fraud by understanding the contents of a call. *Cybersecurity, 1*(1), 8. https://doi.org/10.1186/s42400-018-0008-5

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.

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National science review, 5*(1), 44-53. https://doi.org/10.1093/nsr/nwx106