Security Journal (2023) 36:712–731
https://doi.org/10.1057/s41284-022-00360-3

ORIGINAL ARTICLE

Detection of price manipulation fraud through rational choice theory: evidence for the retail industry in Taiwan

Chin Kuo[1] · **Seng-Su Tsang**[2]

Accepted: 25 October 2022 / Published online: 3 November 2022

Abstract

All enterprises, regardless of industry, are exposed to fraud risk. In the retail industry, the perpetrators of fraud manipulate the sales price of products instead of the quantity sold to avoid inventory discrepancies. Fraud examiners attempt to identify anomalous transactions using data analysis. This study analyzes the causes of fraudulent behavior, conceptualized based on the aspect of rational choice, and proposes an anomalous transaction detection model using variables identified as indicating fraud. A two-phase experiment analyzing a real-world data set from a retailer in Taiwan was designed to evaluate the performance of fraud variables. The findings demonstrate that these variables could slightly improve the results of the analysis and demonstrate that machine learning is applicable to fraud detection. In addition, this study contributes to rational choice theory to validate the applicability in fraud detection. Fraud examiners in the retail industry could reduce fraud losses by adopting the proposed approach, implemented using Weka software, to identify anomalies.

Keywords Occupational fraud · Cash register manipulation fraud · Rational choice theory · Fraud detection · Retail industry

Introduction

Fraud can be defined as intentional deception or misrepresentation for the benefit of the perpetrator (Rawte and Anuradha 2015). Generally, fraud can include any use of deception to profit. Occupational fraud, where an employee defrauds an employer, has caused considerable losses for governments and multinational companies. Kashyap

✉ Chin Kuo
D10816005@mail.ntust.edu.tw

[1] Graduate Institute of Management, National Taiwan University of Science and Technology, No. 43 Keelung Road Sec. 4, Taipei, Taiwan

[2] Department of Business Administration, National Taiwan University of Science and Technology, No. 43 Keelung Road Sec. 4, Taipei, Taiwan

(2019) identified three principal vulnerabilities to fraud in the retail industry: Personal data collection, cyber threats, and insider risks. Insider risk involves employees committing fraud by misusing internal company processes. Cash register manipulation, also known as point of sale (POS) fraud, is a common fraud scheme in the retail industry in which cashiers take advantage of their position to profit by altering the price of the merchandise. POS fraud is an example of occupational fraud. The typical schemes include perpetrators substituting low-price tags for high-price tags, altering the sales price, or entering the wrong quantity of merchandise for the spread, which had been sentenced in Taiwan (Judicial Yuan 2019, 2020, 2022).

In recent years, due to the increasingly complex systems used in the typical workplace and to frequently changing operational activities, perpetrators of fraud have improved their abilities to evade existing fraud detection technologies. Facing a massive volume of data, fraud examiners must apply their expertise and practical experience to prevent the occurrence and reduce the scope of incidents of fraud. Approximately 10% of the respondents to a survey by Deloitte (2018a) said that their company had more than four incidents of fraud every year, demonstrating the presence of fraud incidents in various enterprises in Taiwan. As many as 31% of the respondents were unsure as to whether incidents of fraud occurred, indicating that fraud may be much more common than is believed. The Association of Certified Fraud Examiners (ACFE), in their fraud and abuse survey, reported that total losses due to fraud reached over US $3.6 billion worldwide in 2021 with a plurality from asset misappropriation schemes (ACFE 2022). National Retail Federation (2021) indicated that experts adopted response measures to reduce stock shrink and prevent inventory loss due to a growing threat of organized retail crime events. However, limited research has investigated the potential correlation between price manipulation fraud and transaction behavior in the retail industry.

This study identified the variables, including holiday promotion, wholesale bulk products, multi-sale-item transactions, and continuing transactions, which were evaluated to determine the potential correlations with the price manipulation fraud based on rational choice theory. The Bayesian network, Logistic Regression, and Random Forest methods were applied to examine the detection model. A real-life data set was obtained from a retailer in Taiwan, and sensitive data were removed. The remaining parts of this article proceed as follows. First, we review the literature on fraudulent behavior in the retail industry. Subsequently, we review some literature on fraud detection and rational choice theory. The fraud-related variables proposed were based on rational choice theory. Empirical validation of the model was conducted using a real-life data set from a Taiwanese retailer. Finally, this study's findings identified points of vulnerability to fraud—specifically, holiday promotions, continuing transactions, multi-sale-item transactions, and wholesale bulk products—thus significantly improving the detection model. The managerial implications are presented at the end of this study.

Related research

Price manipulation fraud

According to ACFE (2019), occupational fraud can be understood as employees engaging in fraudulent activity such as asset misappropriation and unauthorized computer system alterations. The problem of occupational fraud is not restricted to any specific industry; all organizations are at risk of experiencing fraud. Holtfreter (2005) investigated the behavioral characteristics of perpetrators of occupational fraud in three categories: asset misappropriation, corruption, and fraudulent statements. Their results demonstrated that the type of fraud a person engages in depends on the person's characteristics. The frequency of asset misappropriation is highest in small organizations.

When shopping, a customer goes to a retail store first. The customer picks up the merchandise and waits in a queue to pay. The cashier then scans a barcode on the merchandise or enters the dollar amount into the cash register. Customers then pay in cash or by credit. However, the cashier has some opportunities to alter the price of merchandise by leveraging VIP discounts or clearance sales to profit by price manipulation illicitly. For example, a cashier buys expensive merchandise and checks out using a discount, or one of the cashier's friends asks him or her to lower the price in exchange for favors. Such a fraud scheme is different from that of keeping money after altering the cash register records (Dopson and Hayes 2015), and it leaves a voluminous audit trail on which data analytics can be applied to identify abnormal transactions (Gee 2014).

Studies investigating transaction fraud detection have focused on specific payment modes, such as credit cards and prepaid cards (Jurgovsky et al. 2018; Robinson and Aria 2018). Fraudulent activity through such modes usually occurs when the cardholders' information is stolen and then used to make unauthorized transactions. Therefore, before constructing a fraud detection model, Correa Bahnsen et al. (2016) considered typical customer consumption behavior to classify the data set accurately. According to the popularity of online shopping with consumers, transaction fraud in e-commerce has also been studied. Nevertheless, physical retail constitutes the largest market share in the consumer market. Retail sales in physical stores made up approximately 86% of global retail sales in 2019 (O'Connell 2020). Sales from brick and mortar stores in Taiwan made up over 90% of all sales in the 2019 global retail market (Ministry of Economic Affairs, R.O.C [MOEA] 2020a). Due to the varied operation methods of modern businesses, fraud has become less organized (Levi 2008). Because firms sustain substantial financial loss and reputational damage due to fraud, managers should understand how to investigate and detect fraud.

Causes of fraud

The fraud triangle theory is the well-renowned framework for fraud behavior analysis. Cressey (1973) first identified the reasons that trusted individuals become

trust violators: a non-shareable problem, an opportunity for trust violation, and a rationalization of the situation. Romney et al. (1980) conducted an empirical investigation to demonstrate that individual characteristics could be a significant factor in white-collar fraud. The International Auditing and Assurance Standards Board (2013) in ISA 240 identified three elements of fraud: pressures, perceived opportunity, and rationalization. However, these fraud-related theories are usually adapted to explain the psychological factors contributing to fraudulent behavior.

However, in the aspect of neoclassical criminology, researchers recognized the importance of criminals' characteristics and how they make a choice for crime. Therefore, the rational choice theory is considered one of the essential concepts to illustrate neoclassical thought (Schmalleger 2021). The concept of rational choice can be traced back to Clarke (1983) to describe an individual committing crimes relevant to the situation of crime and the considerations of the individual at the time. The author attempted to interpret the influence of perpetrators' decision-making on the displacement of crime. In other words, perpetrators would change their target to another time or place and change their fraud schemes after careful considerations. Therefore, Cornish and Clarke (1987) proposed the rational choice theory that consists of three components: opportunities, costs, and benefits, which were identified as the choice-structuring properties. During the involvement of criminals and the occurrence of criminal events, the time, area, target, and method were the key to influencing the decision (Clarke and Cornish 1985).

This theory was first introduced to examine the crime behavior were significantly varied due to the considerations among different crimes during the decision processes (Cornish and Clarke 1987, 1989). Rational choice theory was developed from classical criminology, assuming the crimes resulted from the individual's free will (Schmalleger 1999). Thus, perpetrators were rational to consider the possible consequences before taking action that maximizes benefits and minimizes costs (Akers 1990). Clarke and Harris (1992) adopted the choice-structuring properties from rational choice perspectives to identify why perpetrators made a choice and how the situation influenced the choice. Felson and Clarke (1998) proposed the new opportunity theory to explain the occurrence of crime from the three theories of crime opportunity: routine activity approach, crime pattern theory, and rational choice perspective. These three theories implied the opportunities facilitating crimes due to the change of society, local area, and perpetrators' consideration. Otu and Okon (2019) indicated that opportunity was a critical element among different fraud-related theories, such as rational choice theory and fraud triangle theory. Ultimately, the perspectives discussed above presume that the perpetrators' decisions were conscious (Wilcox 2015).

According to the fraudulent decision-making process involved the perpetrators' preparation and potential steps, Chan and Gibbs (2019) identified that the psychological and emotional factors influenced the decision of white-collar offenders. According to Junger et al. (2020), three specific types of fraudulent activities: C-level fraud, fraudulent contract, and fictitious invoices, were organized to interpret the occurrence of fraud. The authors adopted routine activity theory and rational choice theory to clarify that perpetrators used to consider the business size and seasonality before the fraudulent transaction schemes. Under the rational choice

theory framework, Ding and Zhai (2021) verified that pickpockets prefer to commit theft in the poor air quality, during rush hour and weaken police enforcement. In other words, people tend to take the bus due to the air pollution that the bus would be crowded. Therefore, pickpockets take the advantage to commit theft through the rational decision-making process.

In conclusion, the rational choice theory could be adopted to realize how perpetrators commit fraud under the favorite or unfavorite conditions; the possible consequences and risks would also be considered. Although most research investigated and planned the prevention methods based on rational choice theory (Meyer 2012; Piza et al. 2017). This study developed a fraud transaction detection model derived from the area, target, and modus operandi selection based on the rational choice theory proposed by Clarke and Cornish (1985).

Fraud indicators

Researchers extract features from each transaction to accurately describe an original dataset. Zheng et al. (2018) preprocessed and extracted features from the raw data based on historical consumer transaction records, and they classified the information in terms of five variables: transaction time, transaction location, category of goods, purchase amount, and shipping address. Carcillo et al. (2021) conducted fraud detection by considering the total amount of money spent by the consumer and the number of transactions made within 24 h. Van Vlasselaer et al. (2015) also postulated that fraudulent transactions share characteristics, such as the frequency or amount of consumption. Hence, an anomaly detection system was proposed using the recency, frequency, and monetary (RFM) value of transactions along with social network-related variables to examine e-commerce credit card transactions. Zhang et al. (2021) generated RFM variables based on behavioral analysis to identify fraudulent credit card transactions by using deep learning techniques. As indicated in the preceding discussion, the time, object, and frequency of transactions are key features for detecting fraudulent transactions.

Supervised machine learning

Along with technology and computing power development, analysis by machine learning has become popular. As a result, researchers attempted to investigate fraud behavior through related technologies, including various algorithms such as random forest, decision tree, Logistic Regression, and Naive Bayes (Mehbodniya et al. 2021). Compared with the classical statistics, machine learning results were evaluated by the past outcome in similar scenarios to recognize the possible pattern (Bzdok et al. 2018). In addition, supervised machine learning algorithms that a labeled dataset is adopted as a basis to find out the possible results have been identified as effective methods in prediction. Anomaly detection is a common research field that reveals a good prediction result by adopting supervised learning algorithms, such as cyberattacks on intrusion detection systems (D'hooge et al. 2020) and fake review identification on e-commerce platforms (Elmogy et al. 2021).

A Bayesian network is a causal graph model that analyzes the independent relationship between nodes. It represents the conditional probability distribution between each independent variable to process the inference of the dataset (Singh and Valtorta 1995). Masmoudi et al. (2019) introduced a new method based on Bayesian network to examine and build a credit risk assessment model. From this study, the outcome presented an excellent accuracy and significant classification result. Logistic Regression is considered a popular classification method to recognize the yes or no outcomes by the relationship between dependent variables and independent variables (Hosmer Jr et al. 2013). Guedes et al. (2022) considered the victimization, fears, and perception of victimization risk as the dependent variables to investigate online identity theft victimization through Logistic Regression. The dichotomy outcome easily identified what situation would be more victimized. Finally, Random Forest is an ensemble of decision trees classifier that processes the overfitting problem effectively. In other words, the algorithm results were derived from each decision tree's characteristics relationship (Breiman 2001). In order to improve the dataset's classification performance, Mqadi et al. (2021) adopted Random Forest algorithm to handle the data imbalance problem that reveals a better performance than the decision tree. Hence, this study adopts three supervised learning algorithms, namely Bayesian network, Logistic Regression, and Random Forest, to process the anomalous transaction detection model.

Our proposed approach comprises developing a fraud detection model based on the rational choice theory that identifies anomalous price markdown transactions. Figure 1 displays the fraud detection process: first, a transaction is an input into the database; then, the transaction is automatically analyzed by a fraud detection module; finally, transactions with an anomalous price markdown are manually reviewed by fraud examiners.

Materials and methods

Data description

The data set analyzed in this study covered physical transactions from a retail store in Taiwan that sells food, appliances, hardware, and groceries. The data covers the period from September 2019 to July 2020. Our analysis processed information on price markdowns with labeled fraudulent transactions. Each transaction data point included information on customer details and transaction characteristics, including the time, object, and amount of the transaction. The transactions were examined and then labeled as fraudulent or genuine by fraud examiners. The POS system recorded the transaction information after payment was made. We compared and removed paired data due to a number of recorded transactions that were later reversed. The combined unaltered data set had 212,792 transactions with 13 features each and with one labeled field.

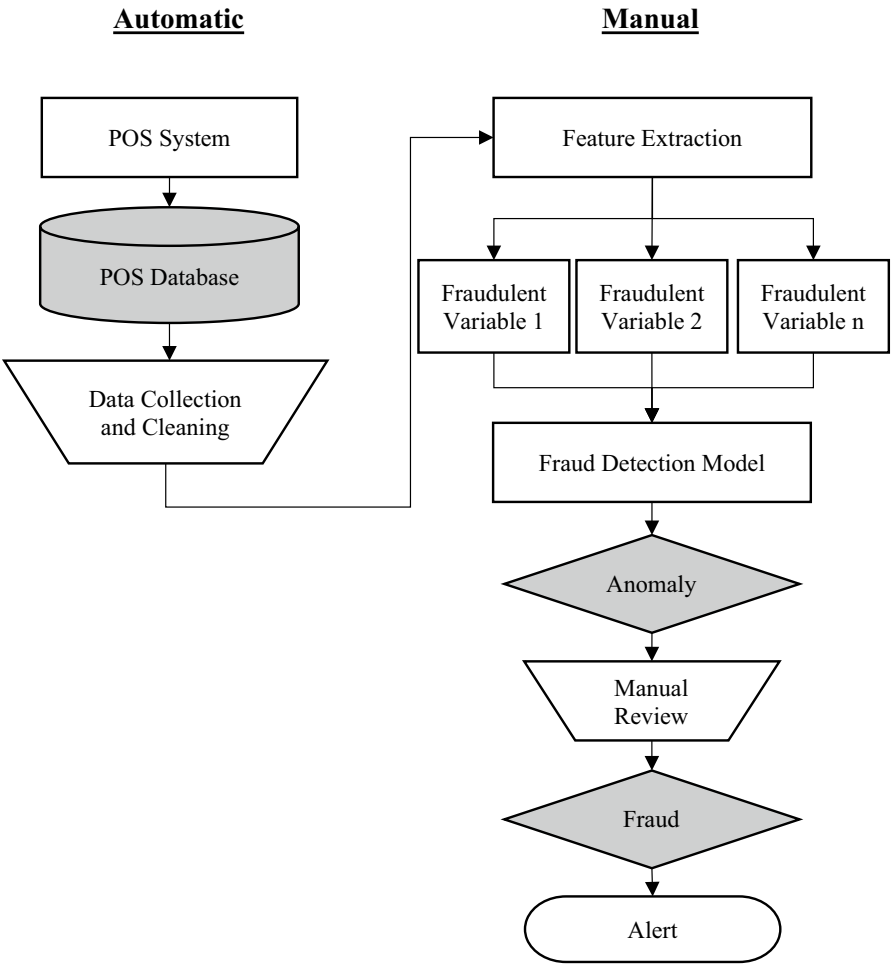

Fig. 1 Anomalous transaction detection process

Data scraping and aggregation

In practice, the perpetrators could be cashiers or consumers who were innocent. In addition, the transaction amount and transaction time record in the POS system were explicit that the user, transaction amount, and transaction time were mainly adopted in fraud detection (Singh and Best 2019; Zhang et al. 2018). However, we attempted to introduce the fraudulent variables to reflect the perpetrators' transaction behavior better. We adopted the rational choice theory framework to describe fraudulent behavior. We identified the variables due to the perpetrators deciding when to commit fraud (holiday promotions), against what to commit fraud (wholesale bulk products), and how to commit fraud (multi sale-item transactions and continuing transactions) that derived from the area, target, and modus operandi selection. We

hypothesized that the analysis would be accurate when these key fraud-related variables were used. The four fraud-related variables are described as follows.

Holiday promotions

Retail consumption increases during holiday seasons, such as Thanksgiving, Christmas, and Chinese New Year. Retailers often launch creative promotions to boost sales during this period (Oh and Kwon 2009). According to Tsoumakas (2019), both the weather and the occurrence of holidays are predictive of retail sales. According to the data from 2015 to 2020, the National Retail Federation (NRF) forecasted that holiday sales would increase even during the COVID-19 pandemic (NRF 2020). The consulting firm McKinsey reported that consumers prefer spending more on Black Friday or Amazon Prime Day (Charm et al. 2020). However, due to massive transaction volumes, anomalous transactions during these periods are less likely to be detected. According to Levy et al. (2010), price fluctuations are more common during the holiday season than at other periods. The analysis of the data set used in that study revealed that anomalous transactions are more frequent at specific periods. Therefore, we incorporated the transaction date into the holiday promotion variable to identify whether a transaction occurred on holiday.

Multi-sale-item transactions

Businesses analyze what customers purchase and when they purchase it to predict their behavior better. Guidotti et al. (2015) predicted customer behavior using the basket and spatiotemporal data from over 50,000 customer transactions. Consumers were found to be price sensitive. Chen and Li (2020) indicated that the intention of consumers to participate in a transaction is influenced by price-based promotions. Promotion can be seen as a factor increasing consumers' involvement and increasing perceived value (Hsia et al. 2020). Promotions were verified to stimulate consumption (Tang and Hu 2019). Accordingly, a person committing fraud by taking advantage of sale prices would likely buy multiple products simultaneously to take greater advantage of the promotion. Therefore, we combined invoice numbers and product names to identify transactions where multiple on-sale products were purchased, and these pieces of information were incorporated into a variable termed multi-sale-item transaction.

Continuing transactions

Individuals tend to spend time paying attention to an object they are interested in; people prefer to continue to consume when they see a lower price. However, the price of transactions is a key indicator in auditing. Gee (2014) indicated that individuals split transactions to avoid reaching the per-transaction limit that would attract the attention of a fraud examiner who used to check the continuing transactions through the invoice numbers. This scheme is known as order splitting (Stamler et al. 2014). Order splitting is a substantial problem for businesses confronting occupational fraud (Office of Inspector General [OIG] 2018). Consecutive invoice numbers

indicate that fraud perpetrators split orders over a short period. In this study, this behavior was defined to indicate a continuing transaction and, more generally, fraud.

Wholesale bulk products

As e-commerce has become more popular, more people have begun businesses based on retail arbitrage, in which they profit by purchasing products from a retailer at a low price and reselling those products at a higher price online. Mercer (2016) was the first mover to investigate retail arbitrage on Amazon. Such arbitrage primarily involved merchandise sold at a discount (e.g., as part of a clearance sale). According to the first sale doctrine, retail arbitrage is a legal business model (Tseng 2018). However, fraud can occur when individuals work in collusion with cashiers or insiders to reduce item prices. Merchandise with high market demand, such as groceries or electronics, is usually selected for fraudulent arbitrage. (Palmer and Richardson 2009). Individuals can devise methods to profit from specific products. In addition, these products are often sold in bulk. Therefore, this study postulated that sales of wholesale bulk products are more likely to be fraudulent, and this feature was included as a variable for transactions in the labeled data.

Methods

In fraud detection, researchers attempted to develop ML models to identify anomalous transactions and conduct further investigations immediately. Manual analysis and intervention thus expected reduction. In recent years, ML, such as the classification of algorithms, has been widely applied across several fields to enhance data analytic abilities. In this study, supervised learning algorithms were adopted, including Bayesian network, Logistic Regression, and Random Forest. Supervised learning algorithms were popular ML algorithms that relied on past instances with labels added manually. In other words, each instance with a corresponding label in the dataset indicates the transactions were genuine or fraudulent in this study. In the classification of algorithms, the labeled examples were used to train the machine learning model to identify the class of each instance. Therefore, researchers could conduct the prediction through the machine learning model according to the training results (Zhou 2018).

In addition, this real-life dataset in this study with a high imbalance label distribution could lead to the failure of classification accuracy (minority:majority = 1:29). Therefore, we applied synthetic minority oversampling techniques (SMOTE) to balance this dataset which was proposed by Chawla et al. (2002). SMOTE works by utilizing the nearest neighbors of the data to create synthetic examples. It generates the virtual training data by introducing linear interpolation for the minority class. The dataset was reconstructed after the oversampling process. SMOTE has been confirmed as an effective filtering method for handling imbalanced datasets (Blagus and Lusa 2013; Fernández et al. 2018).

Experiments and result analysis

Experimental setting

The goal of this study was to construct an anomaly transaction detection model on the basis of rational choice theory by using ML analysis. Three popular ML algorithms were used to measure the accuracy of the detection model, specifically the Bayesian network, Logistic Regression, and Random Forest methods. These are simple but efficient algorithms (Hooda et al. 2018, 2020; Lavanya et al. 2021; Lucas et al. 2020). We propose an approach with a two-phase design to investigate the significance of the identified fraud-related variables.

We used the Waikato Environment for Knowledge Analysis (Weka) 3.8.1 to implement the ML algorithms in this study. Weka is an efficient tool that includes many widely used ML algorithms and data mining methods. Weka is open-source software that offers a user-friendly interface, lowering the entry barrier for researchers. Each classifier used had specific parameters to improve the analysis results (Nafie Ali and Mohamed Hamed 2018; Soni et al. 2019). In a similar study investigating anomaly detection in a large data set, Cui and He (2016) used Weka to measure the accuracy and efficiency of cyberattack detection. Rajesh and Karthikeyan (2017) compared the results of different classification algorithms in Weka to discover patterns to predict the weather. Likewise, the identification of anomalous transactions was investigated using the data mining technologies included in Weka to discover useful variables (Dutta et al. 2017; Yee et al. 2018).

In this data set, we obtain 212,792 transactions in total with a high data set imbalance problem. 483 transactions were labeled as fraudulent in this data set, meaning that the fraud ratio of 0.23% should be getting attention. Therefore, we employed SMOTE to generate artificial data to handle class imbalance. Researchers have attempted to use numerous features to construct a perfect detection model. ML performance is influenced by the presence of unnecessary features or an excessive number of fields in the data set (Hajek and Henriques 2017). Firstly, we retained three native attributes from the transaction records: cash register number, sub-minor category, and quantity. Secondly, we convert four continuous variables into discrete values to illustrate the meaningful presentation: day of the week, business hour, cashier register location, and range of difference in price. Finally, based on rational choice theory, we extracted the key features of the transaction date to identify the holiday promotions attribute, calculated the number of transaction invoices to recognize the continuing and multi-sale-item transactions, and determined wholesale bulk products from the transaction items. Table 1 presents the essential attributes of fraudulent transactions.

This study conducts a two-phase experiment to evaluate the performance of the proposed fraud variables. The first phase determines the original performance with the native attributes collected from the transaction records, whereas the second confirms the increased measurement result by adopting the four fraud-related variables. Both phases use the same algorithms, including Bayesian network, logistic regression, and random forest. In addition, Pearson's correlation

Table 1 Essential attributes of transaction data

Data source	Attribute name	Description
Native	Cash register number	Cash register's serial number
Native	Sub-minor category	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	Cashier register location	Cashier register location
Discretization	Range of Difference in Price	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-Sale-item Transactions	The transaction only includes price-reduced products
Extraction	Wholesale Bulk Products	The products are bought in bulk
Label	Class	Anomalous or normal

coefficient was adopted to validate the strength of the relationship between each fraudulent variable.

Analysis and results

The research data set was provided by a retailer collecting markdown transactions from a POS system. The confidential information from the real-world data set was masked for this experiment. The ML process included two phases to measure the accuracy of the anomaly transaction detection model we constructed. We applied the three native features, four discrete features, and the anomaly label in phase 1. In phase 2, the four critical fraud features were added to improve accuracy: holiday promotion, wholesale bulk products, multi-sale-item transactions, and continuing transactions. The accuracy of different algorithms, specifically the Bayesian network, Logistic Regression, and Random Forest, was investigated. In order to retrieve more accurate insights, the accuracy, precision, recall, *F* measure, and Area Under the Curve (AUC) were used to compare the efficiency of the algorithms. In machine learning, accuracy indicates the ratio of observation of correct predictions to the total observations. The precision indicates the ratio of observation of correct positive predictions to the total positive observations. The recall represents the ratio of observation of correct positive predictions to the total positive observations in the actual class. The weighted average of the precision and the recall were conveyed as *F* measure. AUC represents the area under the ROC curve, which stands for a classification model's performance at all classification thresholds. In general, the higher values convey better performance (Witten et al. 2017). Tables 2 and 3 present the efficiency analysis results of all algorithms in each phase. According to the analysis results in Table 2, all algorithms had accuracy scores greater than 0.8. The Bayesian network had relatively high scores, with an accuracy of 0.863, precision of 0.863,

Table 2 Phase 1 algorithm evaluation performance

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

Table 3 Phase 2 algorithm evaluation performance

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

Table 4 Ranking of attributes

Ranking	Attribute name	*r*
1	Range of Difference in Price	0.52
2	Continuing Transactions	0.21
3	Sub-minor category	0.19
4	Holiday Promotions	0.13
5	Wholesale Bulk Transactions	0.07
6	Day of the Week	0.06
7	Cash register number	0.06
8	Multi-Sale-item Transactions	0.05
9	Cashier register location	0.04
10	Business Hour	0.03
11	Quantity	0.02

recall of 0.863, *F* measure of 0.863, and AUC of 0.991. The evaluation result in phase 1 represents a good performance.

As shown in Table 3, the efficiencies increased slightly when the four fraud-related variables were introduced. The accuracy, precision, recall, *F* measure, and AUC values slightly increased for each algorithm. We can therefore conclude that the identified fraud-related variables improve performance.

In order to measure the strength of the relationship between each variable, a further experiment was conducted to establish the significant coefficient through Pearson's correlation coefficient. This technique examines the strength of the relationship between variables that the coefficient values (r) range from -1 to $+1$, where $+1$ indicates the strongest association and -1 indicates the weakest association (Schober et al. 2018). Table 4 shows the result of the correlation analysis for the real-life dataset. The ranking of four added fraudulent variables were

taken into account: continuing transactions, holiday promotions, wholesale bulk transactions, and multi-sale-item transactions. For the price manipulation fraud, it can be seen that the range of difference in price, continuing transactions, sub-minor category, holiday promotions, and wholesale bulk transactions lie in the top ranking and can be considered as stronger predictors for this detection model. However, we tuned the parameters to unselect the attributes ranking below the average, and the values of accuracy, precision, recall, *F* measure, and AUC were decreased. In other words, these attributes have shown good efficacy in this study.

Discussion and conclusions

This study formulated an anomalous transaction detection model based on rational choice theory. We extracted the key fraud-related variables from the data set to improve the accuracy of the analysis. These variables were holiday promotions, continuing transactions, multi-sale-item transactions, and wholesale bulk transactions, and they were identified on the basis of the relatively limited literature on rational choice theory. An analysis was conducted on a data set of 212,792 retail transactions from Taiwan to identify anomalous transactions. Our results demonstrate that the extracted fraud-related variables slightly improved fraud detection. In addition, ML technologies can improve the efficiency of fraud examination by identifying potentially fraudulent transactions from large data sets.

Theoretical implications

Most studies investigating fraudulent behavior have been conducted on the basis of the fraud triangle theory (Huang et al. 2017; Schuchter and Levi 2016). However, fraudulent behavior has become more complex. Researchers must analyze fraud by using various fraud-related variables. The fraud triangle theory is not the only option (Lokanan 2015). We extracted fraud-related variables on the basis of rational choice theory, which is rarely used for deriving fraud variables. We notice that the four fraud-related variables could reflect perpetrators' interests, attention, and cognition through the rational choice theory. Rational choice theory was useful for constructing a behavioral model for analysis.

The performances of the newly identified fraud-related variables were demonstrated through the two-phase design of the study. The discrete values could be used to identify anomalous transactions in the first phase efficiently. In addition, the significant association of the attributes was validated through correlation analysis that the new fraud-related variables have a stronger association in the experiment. Researchers must infer the relevant features of the original data set and its corresponding information for more precise predictions. Most studies have investigated how variables can be adopted in ML but not why a particular variable ought to be adopted. The reasoning behind our specific choices of fraud-related variables is detailed in this study.

In conclusion, the fraud-detection model performed well when using the 13 features; the presence of more features increases computational cost with no improvement in accuracy. Domingos (2012) and Feng et al. (2018) have reported similar findings.

Managerial implications

In practice, detecting occupational fraud within a business is difficult due to internal collusion. Employees could circumvent internal controls and turn a blind eye to misconduct (PwC 2018). Collusion fraud occurs not only in brick-and-mortar shopping environments but also in e-commerce. Luo and Wan (2019) noticed that buyers and sellers colluded when giving reputation scores after each transaction on online shopping platforms. Fraud examiners and auditors preferred not to apply statistical sampling methods to avoid unstable audit quality. An overall analysis has also become a trend (Manita et al. 2020). Hence, enterprises are eager to identify anomalous transactions by using data analytics. Businesses can choose analytic data methods with Business Intelligence as an early warning system against fraud (Chang et al. 2015; Dilla and Raschke 2015).

In fraud detection, researchers proposed many novel detection models with various attributes to identify abnormal behavior. We identify the key features and introduce the fraud-related variables based on rational choice theory. The perpetrators' behavior was reflected by the four fraud-related variables: holiday promotions, multi sale-item transactions, continuing transactions, wholesale bulk transactions. In general, people used to increase consumption spending by credit card with the cashback or promotion provided by the card-issuing bank during holiday promotions. Therefore, anomalies would be carried out under the guise of regular transactions. In addition, individuals would attempt to keep buying goods with a wrong price or low price tag to conduct retail arbitrage. The individuals get greedy while the higher the goods turnover rate. Hence, the four fraud-related variables should be taken into consideration while developing a fraud detection model for price manipulation in retail.

Although physical transactions have conventionally been the focus of anti-fraud research, e-commerce is a rapidly growing economic sector, particularly during the COVID-19 pandemic. Since Shopee, the leading e-commerce platform in East and Southeast Asia, used a government subsidy program to expand their online shopping business to Taiwan in 2016, individuals have begun to buy and sell online due to its convenience. According to MOEA (2020b), e-commerce sales in 2017 increased by 8.2% on a yearly basis to USD $5.6 billion in Taiwan. In 2020, e-commerce sales reached a record high of USD $8 billion in Taiwan. Perpetrators of fraudulent retail arbitrage began to buy and sell products from famous brands to attract consumers (SIGNIFYD 2018). Fraud analysis has become more complex with the increasing variety of motivations behind the fraud. Fraud examiners and auditors can consider public data, such as weather data, social network opinions, and business competitors, to improve fraud analysis (Chen et al. 2019).

Because ML researchers prefer to use powerful ML tools such as Python, R, and TensorFlow, the barriers to conducting data analysis have been high. However, according to Deloitte (2018b) and PwC (2019), fraud examiners and auditors are eager to improve their data analytics capabilities. User-friendly software applications are believed to increase the use of analytics for auditing (Li et al. 2018). In addition, enterprises are expected to introduce advanced data analytics and big data analytics to improve their competence in the working environment (Krieger et al. 2021). In this study, we used ML algorithms in Weka software to analyze a real-world data set. Three popular algorithms were adopted, and all performed well in terms of their accuracy, precision, and recall values. In addition, the performance of Weka in analyzing credit card transactions and insurance claims has been evaluated (Itri et al. 2019; Kho and Vea 2017). Using Weka or similar software, fraud examiners can conduct data analytics without needing sophisticated coding skills. In the future, the digitalization of data will continue to affect fraud examiners and auditors. We believe that fraud detection can be more made efficient through the integration of these digital technologies.

Limitations and future research directions

Difficulties in data collection limited the empirical analysis of this anomalous transaction detection model. In this study, the data volume was relatively low because the data set was collected over less than one year. Future studies could collect more data over a longer period. Second, the data analysis was conducted after transactions were completed. Most enterprises have upgraded their information systems to perform real-time analysis and instantly identify anomalous transactions. Therefore, improving this study's anomalous transaction detection model to synchronize with real-time systems could be an avenue for future research.

References

ACFE. 2019. *2020 fraud examiners manual*. International. Austin: Association of Certified Fraud Examiners Inc.

ACFE. 2022. *2022 Report to the Nations*. I. Association of Certified Fraud Examiners. https://legacy.acfe.com/report-to-the-nations/2022/.

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.

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

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

Bzdok, D., N. Altman, and M. Krzywinski. 2018. Statistics versus machine learning. *Nature Methods* 15 (4): 233–234. https://doi.org/10.1038/nmeth.4642.

Carcillo, F., Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi. 2021. Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences* 557: 317–331. https://doi.org/10.1016/j.ins.2019.05.042.

Chan, F., and C. Gibbs. 2019. Integrated theories of white-collar and corporate crime. In *The handbook of white-collar crime*, 191–207. Hoboken: Wiley.

Chang, B., C. Kuo, C.-H. Wu, and G.-H. Tzeng. 2015. Using fuzzy analytic network process to assess the risks in enterprise resource planning system implementation. *Applied Soft Computing* 28: 196–207.
Charm, T., J. Perrey, F. Poh, and B. Ruwadi. 2020. *2020 Holiday season: Navigating shopper behaviors in the pandemic*. Atlanta: McKinsey & Company.
Chawla, N.V., K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. 2002. SMOTE: Synthetic minority oversampling technique. *Journal of Artificial Intelligence Research* 16: 321–357. https://doi.org/10.1613/jair.953.
Chen, C., and X. Li. 2020. The effect of online shopping festival promotion strategies on consumer participation intention. *Industrial Management & Data Systems* 120: 2375.
Chen, Y.-J., W.-C. Liou, Y.-M. Chen, and J.-H. Wu. 2019. Fraud detection for financial statements of business groups. *International Journal of Accounting Information Systems* 32: 1–23.
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., and D.B. Cornish. 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., and P.M. Harris. 1992. A rational choice perspective on the targets of automobile theft. *Criminal Behaviour and Mental Health* 2 (1): 25–42. https://doi.org/10.1002/cbm.1992.2.1.25.
Cornish, D.B., and R.V. Clarke. 1987. Understanding crime displacement: An application of rational choice theory. *Criminology* 25 (4): 933–948. https://doi.org/10.1111/j.1745-9125.1987.tb00826.x.
Cornish, D.B., and R.V. Clarke. 1989. Crime specialisation, crime displacement and rational choice theory. In *Criminal behavior and the justice system: Psychological perspectives*, ed. H. Wegener, F. Lösel, and J. Haisch, 103–117. Berlin, Heidelberg: Springer.
Correa Bahnsen, A., D. Aouada, A. Stojanovic, and B. Ottersten. 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*. Charleston: Patterson Smith.
Cui, B., and S. He. 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).
D'hooge, L., T. Wauters, B. Volckaert, and F. De Turck. 2020. Inter-dataset generalization strength of supervised machine learning methods for intrusion detection. *Journal of Information Security and Applications* 54: 102564. https://doi.org/10.1016/j.jisa.2020.102564.
Deloitte. 2018a. *2018a Taiwan Corporate Fraud Risk Management Survey and Future Outlook*. D. A. P. S. Limited. https://www2.deloitte.com/tw/tc/pages/risk/articles/2018a-fraud-report-press.html.
Deloitte. 2018b. *The innovation imperative: Forging Internal Audit's path to greater impact and influence* (Deloitte's 2018b Global Chief Audit Executive research survey Issue. D. T. T. Limited. https://www2.deloitte.com/global/en/pages/risk/articles/global-chief-audit-executive-survey-report.html.
Dilla, W.N., and R.L. Raschke. 2015. 2015/03/01/). 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/10.1016/j.accinf.2015.01.001.
Ding, N., and Y. Zhai. 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., and D.K. Hayes. 2015. *Food and beverage cost control*. Hoboken: Wiley.
Dutta, I., S. Dutta, and B. Raahemi. 2017. Detecting financial restatements using data mining techniques. *Expert Systems with Applications* 90: 374–393.
Elmogy, A.M., U. Tariq, M. Ammar, and A. Ibrahim. 2021. Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2021.0120169.
Federation, N.R. 2021. *National Retail Security Survey 2021*. N. R. Federation. https://nrf.com/research/national-retail-security-survey-2021.
Felson, M., and R.V. Clarke. 1998. Opportunity makes the thief. *Police Research* 98: 10.
Feng, Y., H. Akiyama, L. Lu, and K. Sakurai. 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., S. Garcia, F. Herrera, and N.V. Chawla. 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.

Gee, S. 2014. *Fraud and fraud detection: A data analytics approach*. Hoboken: Wiley.

Guedes, I., M. Martins, and C.S. Cardoso. 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., M. Coscia, D. Pedreschi, and D. Pennacchioli. 2015. Behavioral entropy and profitability in retail. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA).

Hajek, P., and R. Henriques. 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.

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., S. Bawa, and P.S. Rana. 2018. Fraudulent firm classification: A case study of an external audit. *Applied Artificial Intelligence* 32 (1): 48–64.

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

Hosmer, D.W., Jr., S. Lemeshow, and R.X. Sturdivant. 2013. *Applied logistic regression*, vol. 398. Hoboken: Wiley.

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

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

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

Itri, B., Y. Mohamed, Q. Mohammed, and B. Omar. 2019. *Performance comparative study of machine learning algorithms for automobile insurance fraud detection* . 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS).

Judicial Yuan, J.Y. 2019. *Taipei District Court 2019 No.159 Criminal Sentence* Law and Regulations Retrieving System. https://law.judicial.gov.tw/FJUD/data.aspx?ty=JD&id=TPDM%2c108%2c%e8%81%b2%e5%88%a4%2c159%2c20191230%2c1&ot=in.

Judicial Yuan, J.Y. 2020. *Taiwan Shilin District Court 2019 No. 1093 Criminal Sentence* Law and Regulations Retrieving System. https://law.judicial.gov.tw/FJUD/data.aspx?ty=JD&id=SLDM%2c108%2c%e5%af%a9%e7%b0%a1%2c1093%2c20200316%2c1&ot=in.

Judicial Yuan, J.Y. 2022. *Taiwan Taoyuan District Court 2022 No.560 Criminal Sentence* Law and Regulations Retrieving System. https://law.judicial.gov.tw/FJUD/data.aspx?ty=JD&id=TYDM%2c111%2c%e5%a3%a2%e7%b0%a1%2c560%2c20220429%2c1&ot=in.

Junger, M., V. Wang, and M. Schlömer. 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., M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen. 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.

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.

Kho, J.R.D., and L.A. Vea. 2017. *Credit card fraud detection based on transaction behavior* . In: TENCON 2017–2017 IEEE Region 10 Conference.

Krieger, F., P. Drews, and P. Velte. 2021. Explaining the (non-) adoption of advanced data analytics in auditing: A process theory. *International Journal of Accounting Information Systems* 41: 100511. https://doi.org/10.1016/j.accinf.2021.100511.

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

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

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

Li, H., J. Dai, T. Gershberg, and M.A. Vasarhelyi. 2018. Understanding usage and value of audit analytics for internal auditors: An organizational approach. *International Journal of Accounting Information Systems* 28: 59–76.

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

Lucas, Y., P.-E. Portier, L. Laporte, L. He-Guelton, O. Caelen, M. Granitzer, and S. Calabretto. 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems* 102: 393–402. https://doi.org/10.1016/j.future.2019.08.029.

Luo, S., and S. Wan. 2019. Leveraging product characteristics for online collusive detection in big data transactions. *IEEE Access* 7: 40154–40164.

Manita, R., N. Elommal, P. Baudier, and L. Hikkerova. 2020. The digital transformation of external audit and its impact on corporate governance. *Technological Forecasting Social Change* 150: 119751.

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

Mehbodniya, A., I. Alam, S. Pande, R. Neware, K.P. Rane, M. Shabaz, and M.V. Madhavan. 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.

Mercer, G. 2016. 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.

MOEA. 2020a. *Sales of Wholesale, Retail and Food Services in December 2019*. https://www.moea.gov.tw/Mns/dos/bulletin/Bulletin.aspx?kind=8&html=1&menu_id=6727&bull_id=6744.

MOEA. 2020b. *Wholesale, Retail and Catering Industry Statistics Survey*. https://dmz26.moea.gov.tw/GMWeb/investigate/InvestigateEA.aspx.

Mqadi, N.M., N. Naicker, and T. Adeliyi. 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.

Nafie Ali, F.M., and A.A. Mohamed Hamed. 2018. Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents. *Journal of Information Telecommunication* 2 (3): 231–245.

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

O'Connell, L. 2020. *In-store and e-commerce retail sales worldwide from 2019 to 2023*. https://www.statista.com/statistics/1095969/retail-sales-by-channel-worldwide/.

Oh, H., and K.N. Kwon. 2009. An exploratory study of sales promotions for multichannel holiday shopping. *International Journal of Retail Distribution Management* 37: 867.

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., and O.N. Okon. 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., and C. Richardson. 2009. *Organized retail crime: Assessing the risk and developing effective strategies*. An ASIS Foundation Research Council CRISP Report, Issue. I. ASIS Foundation.

Piza, E.L., J.M. Caplan, and L.W. Kennedy. 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.

PwC. 2018. *2018 Taiwan Economic Crime and Fraud Survey*. PwC. https://www.pwc.tw/zh/publications/topic-report/2018-taiwan-economic-crime-and-fraud-survey.html.

PwC. 2019. *Elevating internal audit's role: The digitally fit function* (2019 State of the Internal Audit Profession Study, Issue. https://www.pwc.com/us/en/services/consulting/risk-regulatory/library/internal-audit-transformation-study.html.

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

Rawte, V., and G. Anuradha. 2015. Fraud detection in health insurance using data mining techniques. In 2015 International Conference on Communication, Information & Computing Technology (ICCICT).

Robinson, W.N., and A. Aria. 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., W.S. Albrecht, and D.J. Cherrington. 1980. Auditors and the detection of fraud. *Journal of Accountancy* 149 (5): 63–69.

Schmalleger, F. 1999. *Criminology today: An integrative introduction*, 2nd ed. Upper Saddle River: Prentice Hall.

Schmalleger, F. 2021. *Criminology today: An integrative introduction*, 10th ed. London: Pearson.

Schober, P., C. Boer, and L.A. Schwarte. 2018. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia* 126 (5): 1763.

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

SIGNIFYD. 2018. *Ecommerce Fraud Index*. https://www.signifyd.com/ecommerce-fraud-index-2018/.

Singh, K., and P. Best. 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., and M. Valtorta. 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/10.1016/0888-613X(94)00016-V.

Soni, R., B. Kumar, and S. Chand. 2019. Optimal feature and classifier selection for text region classification in natural scene images using Weka tool. *Multimedia Tools Applications* 78 (22): 31757–31791.

Stamler, R.T., H.J. Marschdorf, and M. Possamai. 2014. *Fraud prevention and detection: Warning signs and the red flag system*. Boca Raton: CRC Press.

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

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.

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

Wilcox, P. 2015. Routine activities, criminal opportunities, crime and crime prevention. In *International encyclopedia of the social & behavioral sciences*, 2nd ed., ed. J.D. Wright, 772–779. Amsterdam: Elsevier.

Witten, I.H., E. Frank, M.A. Hall, and C. Pal. 2017. *Data mining: Practical machine learning tools and techniques*. Burlington: Morgan Kaufmann Publishers.

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

Zhang, X., Y. Han, W. Xu, and Q. Wang. 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., X. Zhou, X. Zhang, L. Wang, and P. Wang. 2018. A model based on convolutional neural network for online transaction fraud detection. *Security Communication Networks*. https://doi.org/10.1155/2018/5680264.

Zheng, L., G. Liu, C. Yan, and C. Jiang. 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.