2022

A Prototype for detecting procurement fraud using data mining techniques: case of banking industry in Kenya.

Muriithi, Francis Wachira
School of Computing and Engineering Sciences
Strathmore University

A Prototype for Detecting Procurement Fraud Using Data Mining Techniques: Case of Banking Industry in Kenya

Francis W. Muriithi

Master of Science in Information Technology

2022

A Prototype for Detecting Procurement Fraud Using Data Mining Techniques: Case of Banking Industry in Kenya

Francis W. Muriithi

008096

Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University

School of Computing and Engineering Sciences

Strathmore University

Nairobi, Kenya

October 2022

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

MURIITHI, FRANCIS WACHIRA

Sign: _____ **Date:** 16 June 2022

Approval

The thesis of Muriithi, Francis Wachira was reviewed and approved for examination by the following:

PROF. ISMAIL ATEYA
School of Computing & Engineering Sciences,
Strathmore University

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University

Dr. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Fraud is a million-dollar business, and it is increasing every year. The numbers are shocking, all the more because over one third of all frauds are detected by 'chance' means. Given that the procurement process is part of the expenditure cycle that culminates with the payment of cash, it is rife with potential for exposing an organization to fraud and embezzlement. Today, whistle blowing, is the most common fraud detection method. However, this method does not proactively search for misconduct. As a result, a fraud detected through this means tends to be caught too late and after the organization has already lost millions of dollars. In this study, we propose a data driven fraud detection prototype to reduce the duration and cost of procurement fraud in Kenya's banking industry. To achieve this, electronic data from the HR and ERP systems was analysed by the prototype using data mining techniques to identify potential fraud misconduct. The data mining techniques applied included rule-based, fuzzy string-matching, and z-score outlier analytics to crossmatch the data against procurement fraud red flag indicators. Thereafter, the prototype generated potential frauds notifications to the organization's audit, risk, or forensic department for further investigation. The outcome of the investigation done by the audit team was also captured by the prototype to increase the accuracy of fraud detection and reduce future false positive alerts.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

ACFE	Association of Certified Fraud Examiners
AI	Artificial Intelligence
CRB	Credit Reference Bureau
DDL	Data Definition Language
EACC	Ethics and Anti-Corruption Commission
ERP	Enterprise Resource Planning
ES	Expert System
NACOSTI	National Commission for Science, Technology & Innovation
OAUG	Oracle Applications Users Group
OFAC	The Office of Foreign Assets Control
PCAOB	Public Company Accounting Oversight Board
PTP	Purchase to Pay
RAD	Rapid Application Development
REST	Representational State Transfer
SDN	Specially Designated Nationals
SOD	Segregation of Duties
SQL	Structured Query Language

Definitions of Terms

Backward Chaining: A strategy for searching the rule base in an expert system that acts like a problem solver by beginning with a hypothesis and seeking out more information until the hypothesis is either proved or disproved (Laudon & Laudon, 2014).

Data Mining: Analysis of large pools of data to find patterns and rules that can be used to guide decision making and predict future behaviour (Laudon & Laudon, 2014).

Enterprise Resource Planning (ERP): A process that integrates the information processing of all routine activities inside an organization (e.g., ordering, billing, production scheduling, budgeting, and staffing) and among business partners (Turban & Aronson, 2001).

Expert System: A computer system that applies reasoning methodologies to knowledge in a specific domain to render advice or recommendations, much like a human expert (Turban & Aronson, 2001).

Forward Chaining: A strategy for searching the rule base in an expert system that begins with the information entered by the user and searches the rule base to arrive at a conclusion (Laudon & Laudon, 2014).

Fraud: Fraud involves all deceptive ways in which one individual obtains an advantage over another by false representations. Fraud always involves confidence and trickery. Fraud is different than robbery where force is used (Albrecht, Albrecht, & Albrecht, Fraud Examination, 2006).

Fraud Detection: Fraud detection involves activities to determine whether or not it is likely that fraud is occurring. Fraud detection allows companies to identify suspicions or predications of fraud (Albrecht S. W., Albrecht, Albrecht, & Zimbelman, 2012).

Fraud Red Flag: A red flag is a set of circumstances that are unusual in nature or vary from the normal activity. It is a signal that something is out of the ordinary and may need to be investigated further (DiNapoli, 2008).

Homogeneous Sampling: A purposive sampling method which focuses on selecting cases from particular subgroup in which all the members are similar (Saunders, Lewis, & Thornhill, 2012).

Inference Engine: The strategy used to search through the rule base in an expert system; can be forward or backward chaining (Laudon & Laudon, 2014). It is the part of an expert system that actually performs the reasoning function (Turban & Aronson, 2001).

Inference Tree: A schematic view of the inference process showing the order in which rules are being tested (Turban & Aronson, 2001).

Knowledge base: A collection of facts, rules, and procedures organized into schemas. The assembly of all the information and knowledge about a specific field of interest (Turban & Aronson, 2001).

Machine Learning: Machine learning is the study of how computer programs can improve their performance without explicit programming (Laudon & Laudon, 2014).

Outliers: Data objects which are grossly different from or inconsistent with the remaining set of data (Han & Kamber, 2001).

Outlier mining: Given a set of n data points or objects, and *k*, the expected number of outliers, find the top *k* objects that are considerably dissimilar, exception, or inconsistent with respect to the remaining data (Han & Kamber, 2001).

Purposive Sampling: Non-probability sampling procedure in which the judgement of the researcher is used to select the cases that make up the sample. This can be done on the basis of extreme cases, heterogeneity (maximum variation), homogeneity (maximum similarity), critical cases, theoretical cases, or typical cases (Saunders, Lewis, & Thornhill, 2012).

Procurement: Procurement also referred as Purchase to Pay (PTP) entails purchasing the goods and services you need, when you need them, and paying for them in a timely manner. It involves purchasing inventory and non-inventory items such as fixed assets and current assets (Hurt, 2010).

Procurement Fraud: Procurement fraud or contract fraud is a deliberate deception intended to influence any stage of the procurement lifecycle in order to make a financial gain or cause a loss. It can be perpetrated by contractors or sub-contractors external to the organization, as well as staff within the organization (National Fraud Authority, 2011). It comes in two main varieties: (1) fraud perpetrated by vendors acting alone and (2) fraud perpetrated through collusion between buyers and vendors. The fraud usually results in either an overcharge for purchased goods, the shipment of inferior goods, or the non-shipment of goods even though payment was made (Albrecht, Albrecht, & Albrecht, Fraud Examination, 2006).

Prototyping: The process of building an experimental system quickly and inexpensively for demonstration and evaluation so that users can better determine information requirements (Laudon & Laudon, 2014).

Rapid Application Development (RAD): Process for developing systems in a very short time period by using prototyping, fourth-generation tools, and close teamwork among users and systems specialists (Laudon & Laudon, 2014).

Rapid Prototype: In expert systems development, an initial version of an expert system (usually one with 25 to 200 rules) that is quickly developed to test the effectiveness of the proposed knowledge representation and inference mechanisms in solving a particular problem (Turban & Aronson, 2001).

Rule: A formal way of specifying a recommendation, directive, or strategy, expressed as an IF premise and a THEN conclusion and possibly an ELSE conclusion (Turban & Aronson, 2001).

Rule Based Analytics: Rules-based analytics involve a series of business rules that use conditional statements to address logical questions (Davenport & Harris, 2007).

Time-series analysis: A technique that analyses historical data over several time periods and then makes a forecast (Turban & Aronson, 2001).

Acknowledgments

There is no doubt that this research thesis in its present form would not have been possible without the gracious gift of time given to me by several people and institutions to whom I am externally grateful. I would therefore wish to express my sincere gratitude and give special thanks to the following persons and institutions.

Prof. Ismail Ateya, my supervisor, who listened intently to my original proposal and took it forward with vision and commitment. My sincere appreciation for his thoughtful guidance on the thesis's structure and content, as well as his constant enthusiasm, prompt feedback, unwavering support, and honest critique throughout the period of this research. He truly played a critical role in advancing my ideas much further than they might otherwise have gone.

The thesis coordinator, Dr. Vincent Omwenga, who provided direction, vital feedback, and invaluable insights throughout the semester. The insights he shared during the classroom sessions played a pivotal role in moving this research thesis forward and in the right direction.

The Strathmore University Institutional Ethics Review Committee (SU-IERC) and the National Commission for Science, Technology & Innovation (NACOSTI), for granting the Institutional Ethical Approval and Research License respectively that permitted me to undertake this research.

I would also like to extend special thanks to all the interviewees who agreed to diligently participate in this research amid their busy work schedules and willingly shared their insights and also tested the prototype. Their efforts helped to make this thesis a reality and I cannot thank you enough for your expertise and insights.

Dedication

This research thesis is dedicated to my loving parents Mr. Christopher Muriithi and Mrs. Eunice Muriithi for their constant encouragement to complete this research thesis. They have been a constant source of love and support throughout my life.

I also dedicate this research thesis to my remarkable and loving wife Wangari and our three beautiful daughters Natalia, Ruby, and Gianna for their unconditional support and bearing with me during the demanding times when I had to be absent in order to complete this research thesis. So, thank you Wachira Girls—the research thesis is finally done.

From the bottom of my heart, I thank you and God Bless you all.

Chapter 1: Introduction

1.1. Background of Study

Procurement is an area that has long been recognized as being particularly vulnerable to fraud and corruption. This is true in the private sector, where the typical problem is kickbacks to the contract officer. It is also true in public sector procurement where the fraud can be initiated by either the public contracting official or the private supplier. The World Bank has estimated that roughly $1.5 trillion in public contract awards are influenced by corruption (Passas, 2007).

Several studies and surveys conducted have concluded that cases of procurement fraud are on the rise and are increasingly becoming a threat to economic development. According to the PwC Global Economic Crime study, procurement fraud is the second most frequently reported form of economic crime behind asset misappropriation (Green, 2014). The Oracle Applications Users Group (OAUG) ranked procurement, as the business process most vulnerable to fraud, waste, and errors (OAUG, 2011). The ACFE found that 22.3% of reported occupational fraud cases were due to employees causing their employers to issue payments by submitting invoices for fictitious goods or services, inflated invoices, or invoices for personal purchases (e.g., employee creates a shell company and bills employer for services not actually rendered) with a median loss of $100,000 (ACFE, 2020). Albrecht et al. (2012) further states that procurement fraud will either be committed by vendors, employees, or a collusion between the vendor and company employees. The fraud will be perpetrated using various schemes such as, cost mischarging, multiple claims, inflated claims, phantom vendors, purchases for personal use, bid manipulation, bid splitting, unjustified sole source awards, and change order abuse. The 2014 Kenya, Global Economic Crime Survey conducted by PwC revealed that 31% of organizations reported having experienced procurement fraud which was lower than the African average of 43% (PwC, 2014).

The three main trends driving procurement fraud were an increase in public tender processes, companies altering their global supply chains, and a rise in outsourcing (Green, 2014). However, few organizations have comprehensive or robust procedures in place to track, monitor and report fraud. Therefore, frauds go on unnoticed for months and years (Oracle Applications Users Group (OAUG), 2011). A tip by whistle blowers is the most common fraud detection method. However,

a tip is not an active fraud detection method that proactively and deliberately searches for misconduct. It is often a passive detection method. Frauds that are detected through active methods tend to be caught sooner and cause smaller losses than frauds that are detected passively. Thus, organizations might be able to reduce the duration and cost of fraud by implementing controls or processes that will increase the likelihood of active detection, such as active management review, attentive account reconciliation, and surveillance or monitoring techniques (ACFE, 2020).

Therefore, there is need to investigate fraud in Kenya's banking industry procurement processes and develop a data driven fraud detection prototype that will rely on data mining techniques. The prototype will help organizations reduce the duration and cost of procurement fraud.

1.2. Problem Statement

It takes organizations an average of 372 days to detect fraud, at which point 89% of all proceeds are unrecoverable (KPMG, 2013). In 2015, the Ethics Anti-Corruption Commission received and analysed a total of 5,660 corruption reports which represented an increase of 41% from the previous year. 50% of the 5,660 cases received involved middle level officers such as Inspectors and Procurement Officers (EACC, 2015). With procurement fraud gaining prominence in Kenya (PwC, 2014), 65% of victim organizations do not have effective anti-fraud controls such as proactive data monitoring and analysis to detect procurement fraud leading to unrecoverable financial losses (ACFE, 2020). Therefore, there is need for a scalable ICT solution that can help organizations to proactively detect procurement fraud thereby minimizing losses.

1.3. Research Objectives

i. To identify schemes and red flag indicators that exist in procurement fraud.
ii. To analyse the challenges in detecting procurement fraud.
iii. To review the architectures and methods used to detect fraud.
iv. To develop a prototype using data mining techniques to detect procurement fraud.
v. To test the prototype.

1.4. Research Questions

i. What schemes and red flag indicators exist when perpetrating procurement fraud?
ii. What are the challenges in detecting procurement frauds?
iii. How are the current architectures and methods used in detecting fraud?
iv. How can the prototype be developed?
v. How can the prototype be tested?

1.5. Justification

Without proactive fraud detection solutions, it will take organizations an average of 372 days to detect fraud, at which point 89% of all proceeds will be unrecoverable (KPMG, 2013). Given that the procurement process is part of the expenditure cycle that culminates with the payment of cash, it is rife with potential for exposing an organization to fraud and embezzlement (Gelinas & Sutton, 2002).

This research assisted organizations in Kenya's banking industry to proactively detect procurement fraud and flag anomalous behaviour thereby minimizing fraud related losses. It also aided audit, forensic, security and compliance professionals by continuously monitoring procurement transaction data and advising them of potential fraud cases to investigate.

1.6. Scope and Limitations

Occupational fraud schemes target a variety of functions and operations within a business or government entity. However, as part of the value delivery chain, every organization must purchase raw materials or inventory and therefore has a procurement process (whether manual or automated). Consequently, the research focused on creating a prototype for detecting procurement fraud in automated cases.

Given that the procurement process tends to be generic with few variations across organizations. The research also narrowed down to procurement fraud affecting banking institutions in Kenya. Due to limitations of time and other resources, the final product was a prototype covering only the basic functionalities.

Chapter 2: Literature Review

2.1. Introduction

Organizations are constantly modifying business processes to be more efficient and effective to gain market share, increase profitability, and streamline operations. All these changes can affect an organization's bottom line and expose it to external and internal business risks (PwC, 2012). Therefore, making the procurement process efficient, cost-effective, and secure remains a pivotal focus area for every organization (Cronie, 2008). However, given that the procurement process is part of the expenditure cycle that culminates with the payment of cash, it is rife with potential for exposing an organization to fraud and embezzlement (Gelinas & Sutton, 2002). Therefore, this process is prime target by fraudsters.

Unlike financial and accounting fraud, procurement fraud is slightly less likely to be driven by senior management; it is more likely to be found in organizations with weaker controls, culture and/or morale. Procurement fraud is found more often in companies that are more geographically remote; where divisions, operations and processes are not central to a company's main business; and in places where the "tone at the top" is not effectively communicated (Caulfield, 2010).

2.2. Empirical Literature

Fraud is global and rampant as suggested by the following recent numbers: A typical organization loses 5 percent of its revenues to fraud each year; the total cost of insurance fraud in the United States is estimated to be more than $40 billion per year; fraud is costing the United Kingdom £73 billion a year; and credit card companies lose approximately seven cents per every hundred dollars of transactions due to fraud (Baesens, Vlasselaer, & Verbeke, 2015).

Further, not all fraud can be prevented. Even in the most secure organizations, it is likely that some type of employee fraud will eventually occur. Consequently, quick detection of fraud is vital to protecting an organization from potential damage. The Association of Certified Fraud Examiners (2020) found that the median duration of a fraud—that is, the typical time between when a fraud begins and when it is detected—is 14 months. Additionally, the longer a fraud remains undetected, the greater the financial losses.

2.2.1 Procurement Related Fraud Cases in Kenya

Studies outlined by the Ethics Anti-Corruption Commission (2015) showed that the EACC received and analysed a total of 5,660 corruption reports which represented an increase of 41% from the previous year. The study reports that 50% of the 5,660 cases received involved middle level officers such as Inspectors and Procurement Officers (EACC, 2015). Table 2.1 highlights selected procurement related fraud cases under investigation by the EACC.

Table 2.1: Selected Procurement Fraud Cases Under Investigation (EACC, 2015)

No.	Nature of Inquiry/Allegation	Institution	Amount Involved (KES)
1	Allegations of irregularities in the tendering process of leasing motor vehicles for the National Police Service.	National Treasury	3,400,000,000.00
2	Allegation of irregular award of contract for supply and installation of multi-channel security system for the Administration Police	National Treasury	5,362,900,000.00
3	Allegation of irregular award of contract for modernization of Police equipment and accessories	National Treasury	4,240,000,000.00
4	Allegations of double payment of rent	Judiciary	44,000,000.00
5	Allegations of irregular procurement of training services	Cabinet Office	57,000,000.00
6	Allegations of corrupt conduct in respect of supply of drugs to Thika District Level Five Hospital	Kiambu County	100,000,000.00
7	Inquiry into allegations of irregular procurements of works	Machakos County Assembly	29,000,000.00
8	Inquiry into allegations of irregular acquisition of land and d building for a referral hospital	Trans Nzoia County Government	500,000,000.00
9	Allegations of double payments made for the renovation of the Deputy President's Karen Residence	Office of the Deputy President	154,000,000.00
10	Allegations of procurement irregularities and abuse of office.	Nairobi City	4,600,000,000.00

2.2.2 Fraud Schemes in the Banking Industry

Identifying the most common fraud schemes within industries can help organizations design controls to guard against their most significant threats. According to a global study on occupational fraud and abuse, ACFE (2020), the banking and financial services industry had the highest number of fraud schemes reported (i.e., 364 fraud cases) with corruption representing the highest risk (i.e., 40% of cases). The study also found that the median loss resulting in the fraud schemes affecting the banking industry was US$ 100,000.00. Table 2.2 shows the most common occupational fraud schemes in the various industries.

Table 2.2: Most common occupation fraud schemes per industry (ACFE, 2020)

No.	Industry	Cases	Billing	Check and payment tampering	Corruption	Expense reimbursement	Noncash	Register disbursements
1.	Banking and financial services	364	8%	9%	40%	8%	10%	2%
2.	Government and public administration	189	18%	4%	48%	17%	17%	0%
3.	Manufacturing	177	23%	8%	50%	20%	23%	2%
4.	Health care	145	33%	14%	40%	22%	24%	6%
5.	Energy	89	24%	6%	66%	11%	25%	1%
6.	Retail	89	22%	11%	37%	17%	20%	7%
7.	Insurance	82	24%	9%	43%	16%	9%	2%
8.	Education	82	30%	18%	30%	22%	17%	1%
9.	Construction	77	22%	17%	47%	9%	13%	4%
10.	Transportation and warehousing	64	13%	5%	52%	9%	23%	0%

2.2.3 Behavioural Red Flags Displayed by Perpetrators

Reviewing the results of the global study on occupational fraud and abuse, ACFE (2020) concludes that the typical occupational fraud scheme lasts 14 months before it is detected; during this time, the perpetrator will often display certain behavioural traits that tend to be associated with fraudulent conduct. Figure 2.1: Behavioural Red Flags of Fraud shows the relative frequency of 17 common behavioural red flags. Significantly, all of these red flags had been identified by someone in the respective victim organizations before the frauds were detected.

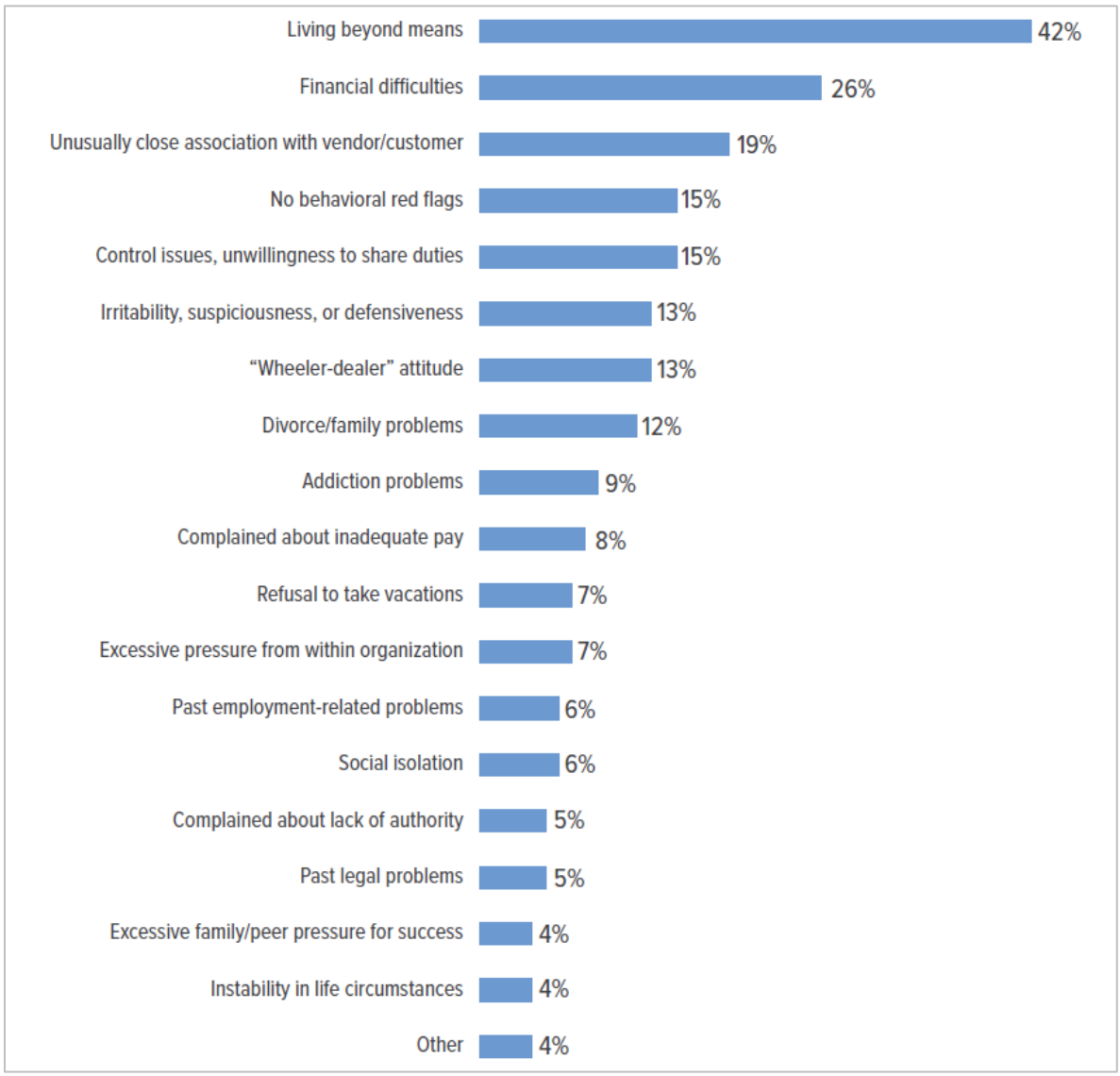


Figure 2.1: Behavioural Red Flags of Fraud (ACFE, 2020)

2.2.4 Approaches for Detecting Fraud

The following fraud detection approaches have been implemented by organizations to detect fraud:

i. **External Audit of Financial Statements:** An external audit of financial statements by independent audit firms is the most common implemented anti-fraud control. However, the ACFE (2020) found that such audits are not designed specifically to find fraud and were not effective.

ii. **Whistle-blowing System:** Albrecht et al. (2006) argue that for a whistle-blowing system to work effectively, it must have the following elements: Anonymity; Independence; Accessibility and Follow up. Studies by the ACFE (2020) show that tips being the most common method used to detect fraud, and by a wide margin, with 43% of cases being uncovered by tips. However, whistle-blower tips are a passive fraud detection method and schemes detected using this method tended to last longer and are associated with the highest median losses relative to all other detection methods.

iii. **Background Checking:** Purcell (2016) maintains that employee background checks provide the assurance that the employee has not lied on the job application, and that the employee does not have something in his past that could be used to blackmail or otherwise compromise him in his position. He further states that the checks should be carried out on new employees and also on existing employees when job responsibilities increase. However, recent studies by the ACFE (2020), found that most fraudsters had no prior criminal history before they commit their crimes due to organizations not reporting these perpetrators to law enforcement. Only 4% of the perpetrators had been previously convicted of a fraud-related offense which reduced the effectiveness of background checks.

iv. **Job Rotation / Mandatory Leave:** Purcell (2016) contends that an employee working alone or in concert with others to defraud the organization is more likely to be caught when a new person examines the system's work processes and notices irregularities. He further asserts that in many fraud schemes, the attacker must be present each day to carry out some action to commit the fraud or cover his tracks so he will not be caught. But while on vacation, it is more likely the illegal activity will be detected. He therefore proposes that for sensitive positions (such as system administrator), many organizations should schedule audits of the employees' system activities while they are on vacation.

Table 2.3: Summary of Fraud Detection Approaches

Approach	Merits	Demerits
External Audit of Financial Statements	Examines the company's financial statements and provides a written report that contains an opinion as to whether the financial statements are fairly stated and comply in all material respects with GAAP.	The typical external audit of the financial statements is not primarily designed to look for fraud. As a result, external audit is one of the least effective methods of detecting fraud cases.
Whistle-blowing System	Best suited to detect fraud at the conversion stage. Co-workers, friends, and managers have a reference point from which to see these changes in lifestyle e.g., they can recognize certain changes such as wearing designer clothes, taking expensive vacations, buying expensive jewellery, or buying a new home with stolen funds.	Many of the calls made through hotlines do not involve fraud at all. Some represent non-fraud issues such as employee work-related concerns; some are motivated by grudges, or a desire to do harm to an organization or individual. It is reactive and identifies the fraud after funds have been lost / stolen.
Background Checking	Background checks are especially important for employees in trusted positions (such as system administrators). Further, they can be carried out on new employees and on existing employees when job responsibilities increase	The search is extremely manual and may not uncover all the pertinent details. Further, Most fraudsters have no prior criminal history before they commit their crimes due to organizations not reporting these perpetrators to law enforcement.
Job Rotation / Mandatory Leave	Detects frauds where the attacker must be present each day to carry out some action to commit the fraud or cover his tracks so he will not be caught.	Job rotation is difficult in small organizations with limited staff. Over time, an employee gains knowledge of enough business processes to make it easier for the employee to commit fraud.

2.3. Theoretical Literature

2.3.1 Understanding the Procurement Process

Although the procurement process can have slight differences across organizations and industries, it has two distinct elements: **sourcing** and **purchasing** (KPMG, 2010).

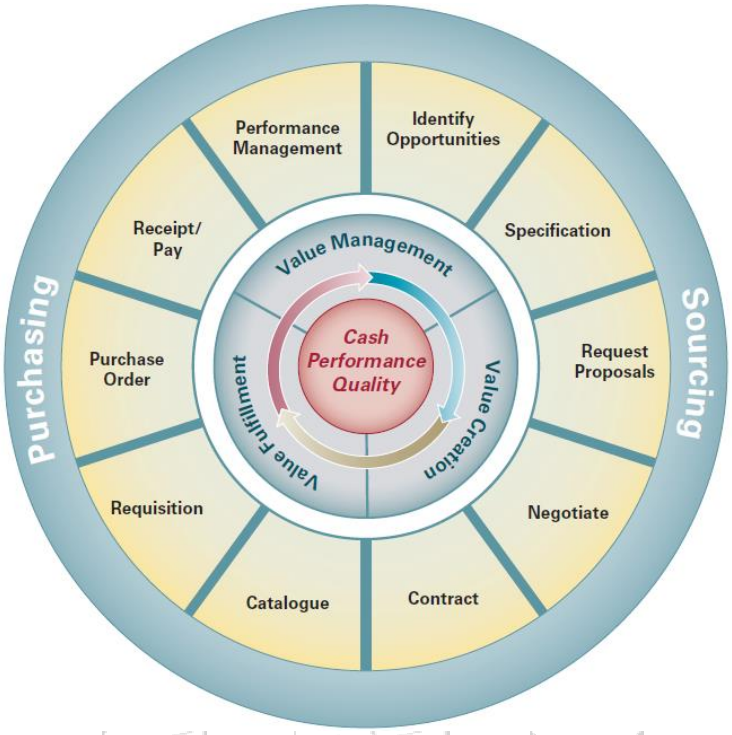


Figure 2.2: The procurement value cycle (KPMG, 2010)

Sourcing is where an organization identifies specific purchasing opportunities, embarks on a competitive bid process that culminates in negotiated contracts with approved suppliers, delivering lower costs (KPMG, 2010). The National Fraud Authority (2011) mentions that the sourcing element involves the core stages of pre-tendering i.e., defining the requirement, developing the specification, tendering, bidder selection and evaluation and awarding of a contract.

Purchasing is the process of sending an approved purchase request to a supplier from the approved suppliers list; receiving the goods/services; receiving the supplier invoice; and processing payment (Hollander, Denna, & Cherrington, 2000).

2.3.2 Procurement Fraud Schemes

Procurement fraud schemes include those committed by vendors alone, by procuring officials alone, and the most dangerous and damaging of all, those schemes in which both parties of the procurement process collude and conspire to defraud (Caulfield, 2010).

Albrecht et al. (2012) states that procurement fraud schemes **committed by vendors** include:

i. **Cost Mischarging Schemes:** Occurs when a vendor charges the procuring entity for costs that are not allowable, not reasonable, or cannot be allocated to the contract.
ii. **Multiple Claims:** Single incurred cost submitted on different contracts or multiple times to the same contract.
iii. **Inflated Claims:** Higher costs that actually incurred.
iv. **False Claims:** No services or goods were provided.
v. **Non-Conforming Goods or Services / Product Substitution:** Attempts by vendors to deliver goods or services to the procuring entity that do not conform to the contract.

Albrecht et al. (2012) states that procurement fraud schemes **committed by employees** include:

i. **Phantom Vendor:** An employee submits invoices from a non-existent vendor.
ii. **Purchases for Personal/Resale:** An employee purchases items that are intended for his/her own use, or for resale.

Albrecht et al. (2012) states that procurement fraud schemes **committed by vendor and company employees in collusion** include:

i. **Bid Manipulation:** Corrupt vendor persuades a purchasing company employee to ensure that one or more of the vendor's competitors cannot bid on the contract, thereby improving that vendor's chances of winning the contract.
ii. **Bid Splitting:** Employee splits large contracts into smaller contracts, thus avoiding the scrutiny required for large-dollar-value contracts. Once the contract is split, the employee can award some or all the component parts to a vendor with whom he is conspiring.
iii. **Unjustified Sole Source Awards:** Sole source contracting occurs when the goods or services are available only from a single source, when exigent circumstances will not permit delay resulting from a competitive bid, secrecy etc. Therefore, in cases of sole

source awards, the vendor charges a much higher price than the company could have obtained through bidding.

iv. **Change Order Abuse:** The corrupt vendor submits a low bid to ensure that they win the contract award, but after the procuring entity awards the contract, the corrupt vendor increases their price with subsequent change orders.

2.3.3 Procurement Fraud Red Flags

DiNapoli (2008) argues that fraud itself is rarely seen; only its symptoms, red flags, or indicators are observed. According to Albrecht et al. (2012), there are several red flags that should alert companies to potential procurement fraud. These red flags need to be identified to facilitate proactive data mining techniques to detect and prevent a fraud. Albrecht et al. (2012, pp. 137-153) separates the fraud red flags or symptoms into six groups: accounting anomalies; internal control weaknesses; analytic anomalies; extravagant lifestyles; unusual behaviour; and Tips and complaints.

i. **Accounting anomalies:** Result from unusual procedures in the accounting system. Examples of accounting anomalies that may indicate procurement fraud include: Missing or alterations on purchase documents, Excessive goods return, Common names or addresses of payees, Duplicate payments, Document sequences that do not make sense, Journal entries that do not balance and Journal entries made by unauthorized individuals.

ii. **Internal Control Weaknesses:** Includes cases when internal controls are absent or overridden, providing opportunity of fraud. Examples of internal control weaknesses that may indicate procurement fraud include: lack of segregation of duties; lack of independent checks; lack of proper authorization; and overriding of existing controls.

iii. **Analytical Anomalies:** Procedures or relationships that are unusual or too unrealistic to be believable. Examples of analytical anomalies that may indicate procurement fraud include: excess purchases; unexplained inventory shortages or adjustments; cash shortages or overages; excessive late charges; and significant changes in account balances or ratios.

iv. **Extravagant Lifestyle:** Fraud perpetrators often live beyond their means since their income does not support their lifestyle.

v. **Unusual Behaviour:** No particular behaviour signals fraud, but changes in behaviour are signals. Perpetrating a fraud will often be accompanied by stress brought on by fear and

guilt. This stress will often cause perpetrators to act in abnormal ways. People who are normally nice may become intimidating and belligerent.

vi. **Tips and Complaints**: Many frauds are detected when an employee, a friend, a manager, a customer, or other untrained person provides a tip or complaint that something is wrong. Employees other than auditors are usually in the best position to detect fraud.

2.3.4 Challenges in Detecting Fraud in the Procurement Process

i. **Manual Nature of the Process:** KPMG (2013) found that the sourcing process is manually conducted making it prone to abuse and difficult to detect cases of bribery/kickbacks and conflict of interest between an employee and the vendor company during vendor selection.

ii. **Collusion Fraud:** Collusion fraud is that it is more difficult to detect and leads to larger loss compared to fraud committed by an individual (KPMG, 2013). According to recent studies by the ACFE (2020), 51% of frauds were committed by two or more fraudsters working in collusion and losses tended to increase with multiple. The reason collusive frauds might be more costly is that multiple fraudsters working together might be able to undermine the systems of separated duties and independent verification that are at the heart of many anti-fraud controls (ACFE, 2020).

iii. **Employees Who Can Detect Fraud Lack Training:** Albrecht et al. (2012) state that there are three elements of fraud; the theft act, concealment act and conversion act and fraud can be detected in all three elements. They further argue that although co-workers and managers are in the best position to detect fraud, they are usually the least trained to recognize fraud. In addition, Caulfield (2010) also asserts that majority of co-workers do not have any training in identifying fraud indicators, along with what to do once they have found one.

iv. **Unrealistic Expectation on External Auditors to Detect Fraud:** Recent studies outlined by the ACFE (2020), suggest that the typical external audit of the financial statements is not primarily designed to look for fraud. As a result, it concludes that external audit is one of the least effective methods of detecting fraud cases. Further, Albrecht et al. (2012, pp. 152) also noted that external auditors are often criticized for not detecting more frauds. Yet, because of the nature of fraud, auditors are often in the worst position to detect fraud throughout the three elements of fraud i.e., the theft, concealment, and conversion acts. In

the theft act, auditors are rarely present when funds are stolen, or fraud is committed. Rather, they spend one or two weeks on periodic audits and thefts usually stop during the audit periods. Instead, co-workers, managers, and other employees are in the best position to detect fraud in the theft act stage. In concealment act, auditors do have a chance to detect fraud. If audit samples include altered documents, miscounts, or other concealment efforts, auditors may detect fraud. Similarly, they may see internal control weaknesses or analytical relationships that don't make sense. However, company accountants and even co-workers are probably in a better position to detect fraud in concealment. At the conversion act, the auditors are not in the best position to detect fraud e.g., auditors could never recognize certain changes such as wearing designer clothes, taking expensive vacations, buying expensive jewellery, or buying a new home with stolen funds. Auditors do not have a reference point from which to see these changes in lifestyle. Again, it is co-workers, friends, and managers who should detect fraud in conversion.

v. **Technological Advancements:** Caulfield (2010) promotes the idea that with today's technology, it is easy to fabricate legitimately looking source documentation (bidding documents, change orders, invoices, shipping documents) making procurement fraud detection difficult. This was further exacerbated by the fact that in today's business environment, we frequently use electronic authorization and electronic signatures which may be easily altered in the absence of strong internal controls.

vi. **Veil of Trust:** Caulfield (2010) states that a procurement fraudster is usually connected to the acquisition process in some way; therefore, the fraudster is currently, or has in the past, interacted at some level with the employees of the victimized organization. As a result, the degree of interaction between the fraudster and the employee reinforces this veil of trust, and therefore the fraudster is often considered to be a "trusted agent" by the same people whose trust has been violated.

vii. **Flexibility of Procurement Fraud:** Caulfield (2010) argues that each stage of the procurement process can be corrupted in one of a thousand different ways within the generally recognized fraud schemes and is only limited to the imagination of the fraudster, and the type of access the fraudster has to the procurement process.

2.3.5 Data Mining in Fraud Detection

It is stated that the fraud data analytics methodology is a circular process of analysing data and continually refining the search process as we learn more about the data and the existence of a fraud scenario in the core business system (Vona, 2017). Vona (2011) states that data mining will analyse transactional data and descriptive data that are consistent with a fraud scenario thereby locating and recognizing fraud. Once identified, the fraud scenario should be followed by building a fraud data profile. Thereafter, a data mining routine is constructed and performed on this data to uncover any red flags associated with the fraud scenario. However, data mining effectiveness is directly correlated to the integrity and availability of the data residing in the database. The following eight steps are needed for effective fraud data mining, (Vona, 2011, pp. 117-129):

i. **Understanding the "What," "Where," and "How Much" of Data:** Know "what" data fields exist in the database, then determine if a field is to be used, which may be accomplished by building a listing of the data required and working with the data administrator to identify the name of the data element in the table. know "how much" data is being worked with through counting, by data element, the number of records that are actually populated.

ii. **Mapping the Data Fields to the Fraud Scenario:** Data mapping is the process of connecting the data to the fraud scenario. The process starts with the fraud scenario, then each data element is linked to the fraud scenario.

iii. **Understanding the Integrity of the Data:** Identify the data inconsistencies, errors, blanks caused by data input or system changes, then develop strategies to fix, group, isolate, or eliminate them. If the problems with the data cannot be fixed, then a decision should be made on how to resolve them within the context of the search routine.

iv. **Applying Inclusion/Exclusion Theory:** The inclusion portion of the theory starts with a database of transactions where the data is categorized into like groups. The purpose of doing this categorization is twofold. For one, examining a smaller database is easier, and for another, an anomaly is easier to spot when all the transactions are in common. The grouping of data is dependent on the fraud scenario.

v. **Understanding False Positives:** A false positive transaction is one that meets the fraud data profile but is not in and of itself a fraudulent transaction. These false positives occur for a variety of reasons such as data integrity issues.

vi. **Understanding the "Norm" of the Data:** The word anomaly is defined as an extreme deviation from the norm, an outlier. Therefore, one should understand the norm before proceeding to the anomaly. A report providing statistical information by entity structure as to the entity creation date, number of transactions, aggregate dollar value of transactions, maximum dollar value of a transaction, minimum dollar value of a transaction, and average dollar amount is a useful one to have in understanding the norm from the anomaly.

vii. **Data Correlations:** The process of making a connection between the fraud data profile and an entity or an individual. Correlation of the data can be considered a four-tier process:

First-Tier Analysis: Analyse data from the master file fields for the following: missing data, duplicate data, matching records in the primary database to records in a secondary one, identifying a data filed that has been changed one or more times, and identifying nondescriptive data that are intended to obscure the identity and location of an entity.

Second-Tier Analysis: Analyse data within the transactional data that correlates to an entity structure. Items to be analysed include, searching for a specific data pattern, frequency occurrences of the event in correlation to the norm or the fraud scenario, logical order of transactions, logical range of data items, off-period transactions, circumvention by structuring a transaction to avoid a control level such as the perpetrator intentionally splitting a transaction into two or more, or insufficient characters in the field.

Third-Tier Analysis: Correlate the frequency of the pattern to the fraud perpetrator. The transactions should be correlated to the logical creator or the transactional anomaly.

Fourth-Tier Analysis: Correlate the dollar activity to the entity structure and the fraud perpetrator, with the absence of dollars indicating the exception is a data integrity issue.

viii. **Entity Structures and Search Routines:** Prior to implementing the search routines, a decision needs to be made as to whether all entities or just the active ones are to be examined. Active entities can be identified by a code or dollar activity. If just active entities are used, then a report should be generated to ensure all inactive entities have no financial activity or recent changes to key fields, such as address or bank account number.

2.3.6 Process of Developing a Fraud Profile

Deloitte (2021) argues that anyone in the organisation can present a potential fraud risk regardless of their position, age, gender, or length of service. Therefore, the fight against fraud requires us to understand fraud typologies and common profiles of a fraudster. McNeal (2014) asserts that individuals who perpetrate fraud look just like everyone else. They are our co-workers, our acquaintances and even our friends and family members. He further states that many appear outwardly honest and ethical, making it particularly difficult to suspect that they would breach their employers' trust. And yet, effectively detecting fraud requires us to do just that.

In order to identify something, you have to know what it looks like. Research indicates that most employee fraud is not sophisticated and, generally, not well concealed. All that is required, is to identify what it would look like, and it will be found. The process outlined does not provide an absolute catch-all solution to fraud detection but an incremental improvement in the odds against an organization becoming a victim of fraud (Padgett, 2015). As shown in Figure 2.3, there are three fundamental phases in the process of developing a fraud profile for your organization.

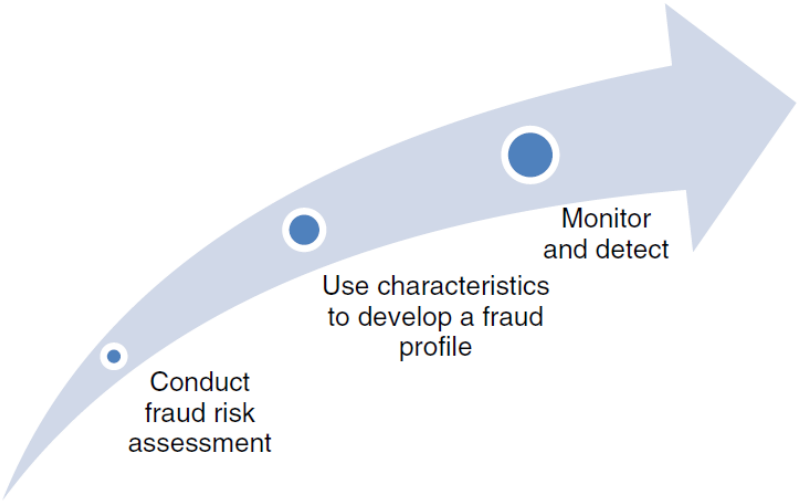


Figure 2.3: Process of Developing a fraud Profile (Padgett, 2015)

Padgett (2015, pp. 213-214) argues that when people commit fraud, there are normally identifiable generic behavioural patterns that are consistent with fraudulent activity such as employees working long hours, not taking vacations, being protective and secretive about their work, and being generally uncooperative. They are generic by default largely because the particular personalities

of individuals are not usually sufficiently well known for them to be more specific. These behavioural characteristics manifest themselves as patterns in data which may indicate a characteristic of fraudulent activity, particularly when those patterns are anomalies or do not follow a standard trend. These data patterns could be a key indicator of fraud and can be captured to form part of a fraud profile. The output of this process is a profiling schedule that identifies the potential perpetrator or class of perpetrators, the type of behavioural characteristics, and the controls or enhanced controls that should be in place to reduce fraud risk.

Studies outlined by the ACFE provides some insight into the people behind the crimes and comes up with statistical fraudster profile. However, McNeal (2014) argues that simply meeting the profile outlined by the ACFE does not mean that an individual is going to commit fraud, nor should someone who falls outside the profile be immune from suspicion. Rather, he suggests that trends in the characteristics of fraudsters should be examined to provide context and perspective to those charged with identifying high-risk areas and individuals. He further maintains that the statistical profile of a fraudster will help inform anti-fraud professionals regarding occupational fraud perpetrators as it can be a valuable training and awareness tool. He concludes that the more we understand and help educate others about the human face of fraud, the better prepared we all are to prevent and detect fraudulent behaviour.

Padgett (2015) states that forensic profiling is generally conducted using data mining and analysis technology as a means by which relevant patterns are examined and profiles generated from sometimes large quantities of data. The main techniques used for extracting, analysing, and reporting this data include (Padgett, 2015, p. 192):

i. Data mining to classify, cluster, and segment the data and automatically find associations and rules in the data that may signify interesting patterns (supervised learning).
ii. Systems to encode expertise for detecting fraud in the form of rules.
iii. Pattern recognition to detect approximate classes, clusters, or patterns of suspicious behaviour either automatically (unsupervised learning) or to match given inputs.
iv. Machine learning techniques to automatically identify characteristics of fraud.
v. Neural networks that can identify and learn suspicious patterns from samples and use this information later to detect them.

2.4. Models for Fraud Detection

2.4.1 Data-driven Fraud Detection Model

According to a study by the ACFE (2020), the presence of proactive data monitoring / data analysis as an anti-fraud control reduces the duration taken to detect fraud by over 30%. Despite this, only 31% of organizations in Sub-Saharan Africa had implemented proactive data monitoring/analysis as an anti-fraud control. Albrecht et al. (2012) states that data-driven fraud detection can be applied to detect existing frauds hidden within an organization's data, inform fraud risk assessments, and identify process and control weaknesses. Furthermore, Singh et al. (2019) found that data analytics is recommended by regulatory bodies as a rigorous auditing tool to detect and reduce fraud incidents, and the PCAOB has urged auditing companies to leverage data analytics in the audit process. Albrecht et al. (2012) states that with data-driven fraud detection, the fraud investigator no longer waits for a tip to be received; instead, he brainstorms on the fraud's red flags that might exist and looks for them. Figure 2.4 shows the proactive (data-driven) fraud detection method.

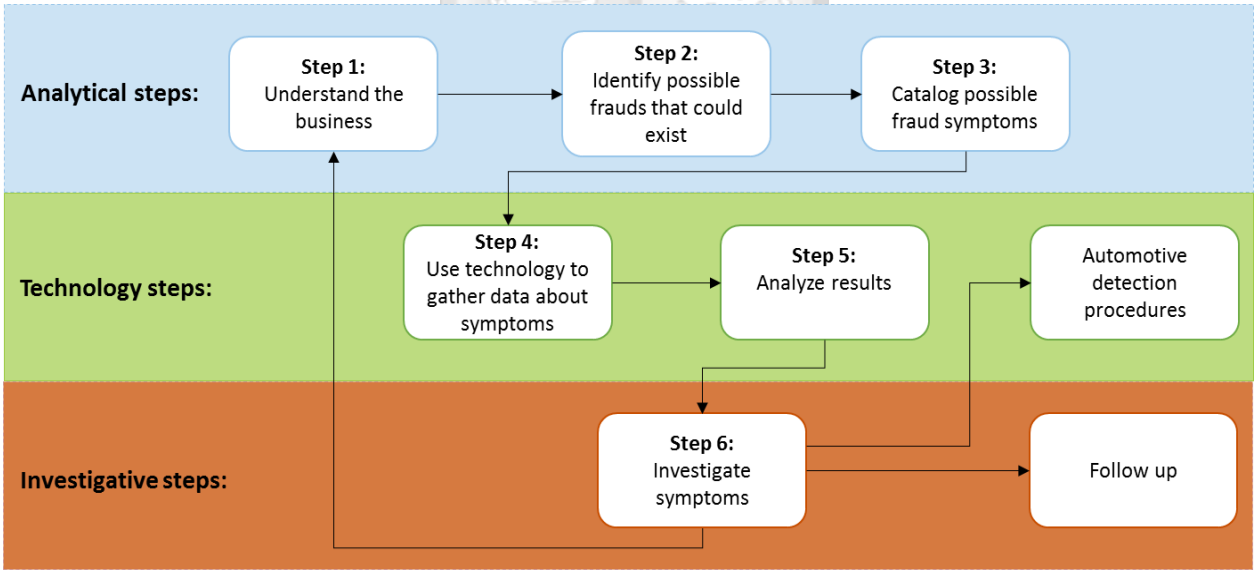


Figure 2.4: The Proactive Method of Fraud Detection (Albrecht S. W., Albrecht, Albrecht, & Zimbelman, 2012)

Rule-based modelling may be embedded in the data driven model to keep up with new fraud schemes when they are detected, as the inclusion of new rules in an existing model is comparably simple (Singh, Lai, Vejvar, & Cheng, 2019).

2.4.2 Classification Based Fraud Detection Model

Singh et al. (2019) argue that fraud schemes keep on changing and are dynamic as somebody seeking to perpetrate a fraud would find newer schemes. So, newer rules would have to be developed continuously to keep track of new fraud schemes. In such a scenario, they suggest that it is vital to have a classification-based approach which is able to identify flag-off anomalous transactions. Single classification methods (or classifiers) develop a (single) classification model using logistic regression, neural networks, and decision trees.

Guttag (2017) states that in supervised learning, we start with a set of feature vector/value pairs and the goal is to derive from these pairs a rule that predicts the value associated with a previously unseen feature vector. Classification models associate one of a finite number of labels with each feature vector. According to IBM Cloud Education (2020), they use an algorithm to assign test data into specific categories by recognizing specific entities within the dataset and attempts to draw some conclusions on how those entities should be labelled or defined.

Classification models are broadly used in practice for such tasks as detecting fraudulent use of credit cards (Guttag, 2017). Logistical regression is selected when the dependent variable is categorical, meaning they have binary outputs, such as "true" and "false" or "yes" and "no." (IBM Cloud Education, 2020). In fraud detection, the outcome of a logistic regression classifier would therefore be a binary dependent variable (i.e., is “fraud”’ or “no fraud”) that will be classified from the underlying data which is transactional accounting information (Singh, Lai, Vejvar, & Cheng, 2019).

Singh et al. (2019) state that in classification models, correctly predicted flags are called true positives, while correctly predicted non-flags are referred to as true negatives. There are two potential errors for logistic regressions, namely a flag is not flagged appropriately (false negative), and a non-flag is falsely flagged (false positive). Although accuracy is a reasonable way to evaluate a classifier when two classes are of roughly equal size, there are other statistics about classifiers shown in Equation 2.1, Equation 2.2, Equation 2.3 and Equation 2.4 that can be used to shed light when classes are imbalanced (Guttag, 2017):

i. **Sensitivity/Recall** is the true positive rate, i.e., the proportion of positives that are correctly identified as such.

$$sensitivity = \frac{true\ positive}{true\ positive\ + false\ negative}$$

Equation 2.1: Sensitivity calculation

ii. **Specificity/Precision** is the true negative rate, i.e., the proportion of negatives that are correctly identified as such.

$$specificity = \frac{true\ negative}{true\ negative\ + false\ negative}$$

Equation 2.2: Specificity calculation

iii. **Positive Predictive Value** is the probability that an example classified as positive is truly positive.

$$positive\ predictive\ value = \frac{true\ positive}{true\ positive\ + false\ positive}$$

Equation 2.3: Positive Predictive Value calculation

iv. **Negative Predictive Value** is the probability that an example classified as negative is truly negative.

$$negative\ predictive\ value = \frac{true\ negative}{true\ negative\ + false\ negative}$$

Equation 2.4: Negative Predictive Value Calculation

Both the sensitivity and specificity of a classification model are heavily influenced by the cut-off value (or classification threshold) of the dependent variable, that is, how decimals between 0 and 1 are rounded up or down in the classification. A high cut-off value generally lowers the number of flags, whereas a low cut-off value increases both the number of flags and false positives (Singh, Lai, Vejvar, & Cheng, 2019).

2.5. Data Mining Algorithms used to Detect Fraud

2.5.1 Rule-Based Analytics

Lemon (2017) argues that rules-based analytics are commonly seen as sets of procedures that a procurement specialist or contracting official follow, such as checking established databases for illegal or suspicious behaviour prior to awarding a contract. This allows a procurement specialist to raise red flags that require further investigation within the organization's acquisition program. However, problems may result from relying solely on standalone rules for an analytical program as it is not possible to create a rule for preventing all types of possible fraud. Some legitimate behaviours or transactions may raise suspicion and create a large backlog in working operations.

2.5.2 Fuzzy String Matching

Albrecht et al. (2012) state that fuzzy string matching of values is common technique in fraud investigation with the classic use being the matching of employee and vendor addresses, phone numbers or other personal information. This is because most employees who set up dummy companies use their home address as the company address. Although such cross-correlations might seem simple, they are complicated because personal data contains so many inconsistencies. Santhi (2016) states that the Levenshtein distance approach is used for fuzzy string matching where it calculates the number of edits i.e., insertion and deletion required to match the two strings. The distance of a match is then measured in terms of the number of basic operations required to convert the string into an exact match. The number of edits obtained from fuzzy string matching is converted to match percentage using Equation 2.5.

$$strDist = 1.0 - \frac{\left(LevenshteinDist(s1, s2)\right)}{\max\left(s1, s2\right)}$$

Equation 2.5: Levenshtein Distance Equation

Where, LevenshteinDist (s1, s2) = edit distance between s1, s2, s1 = length of string1, s2 = length of string2 (Santhi, Acharjya, & Ezhilarasan, 2016). Cervo and Allen (2011) suggest that when performing data matching, exact and fuzzy matching should be combined to achieve maximum accuracy, efficiency, and throughput.

2.5.3 Outlier Analysis

Albrecht et al. (2012) found that one of the most common analyses that fraud investigators perform is identification of outliers. By focusing on outliers, investigators can easily identify cases that do not match the norm. According to Baesens et al. (2015), outliers may indicate fraudulent activity, such as detecting unusual usage of credit cards. Albrecht et al. (2012) state that outliers may be identified using the statistical z-score calculation method that converts data to a standard scale and distribution, regardless of the amounts and variances in the data. The calculation for a z-score is shown in Equation 2.6.

$$Z - score = \frac{(Value - Mean)}{Standard\ Deviation}$$

Equation 2.6: Z-score Equation

The numerator of the calculation pulls the result to centre on 0 (rather than on the true average). The denominator standardizes the result to a standard deviation of 1. Statistical theory predicts that 68 percent of the data have scores between – 1 and 1, 95 percent will have scores between – 2 and 2, and 99.7 percent will have scores between – 3 and 3. In most outlier analysis, any case greater than 2 or 3 should be investigated (Albrecht S. W., Albrecht, Albrecht, & Zimbelman, 2012).

2.5.4 Predictive Analytics

Predictive analytics looks at past behaviours to predict what variables had the most prolific impact on fraud prevention. Focusing on existing variables in instances of fraud occurrence provides the ability to predict future threats and risks of fraud (Akerkar, 2013). Lemon (2017) found that predictive analytics looks at fraud schemes such as change order abuse, product substitutions, and defective pricing situations that have occurred with vendors in the past, and "predicts" the possibility of the events occurring in the future. For predictive analytics to be successful, fraudulent behaviours must already be known. This provides a limited ability in adapting to new fraud prevention schemes. This method is not the most efficient way to detect a trail of involved individuals within a fraud scheme and has not proven to be effective in detecting or preventing events that have not already occurred (Lemon, 2017).

2.6. Design and Architecture of Fraud Detection Solutions

Banks and other financial institutions such as card issuers and insurers have always used patterns presented in historical data to attempt to detect fraud. It is suggested that transactions, once found to be fraudulent, are logged, and characteristics are flagged as possible indicators that can be used to suggest future transactions may also be fraudulent (Marr & Ward, 2019).

Expert systems capture the knowledge of skilled employees in the form of a set of rules in a software system that can be used by others in the organization. They model human knowledge as a set of rules that collectively are called the knowledge base. The rules are interconnected; the number of outcomes is known in advance and is limited; there are multiple paths to the same outcome; and the system can consider multiple rules at a single time. The strategy used to search through the knowledge base is called the inference engine. The inference engine works by searching through the rules and "firing" those rules that are triggered by facts gathered and entered by the user (Laudon & Laudon, 2014). The key parts of an expert system is depicted in Figure 2.5.

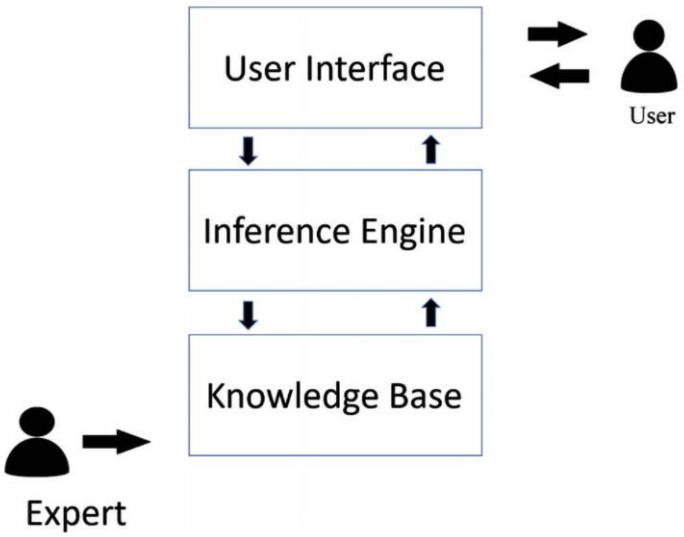


Figure 2.5: Key parts of an expert system (Taulli, 2019)

Machine learning is also adopted when you know enough about your data through supervised learning. It is suggested that in supervised learning you show the machine the connection between a known outcome and the variables that affect that outcome (Rose, 2020).

2.7. Conceptual Framework

From this literature, the researcher has conceptualized the use of data mining techniques (i.e., Rule-Based Analytics, Fuzzy String Matching, and Z-Score Outlier Analysis) to proactively detect procurement fraud. The conceptual framework is summarized in Figure 2.6.

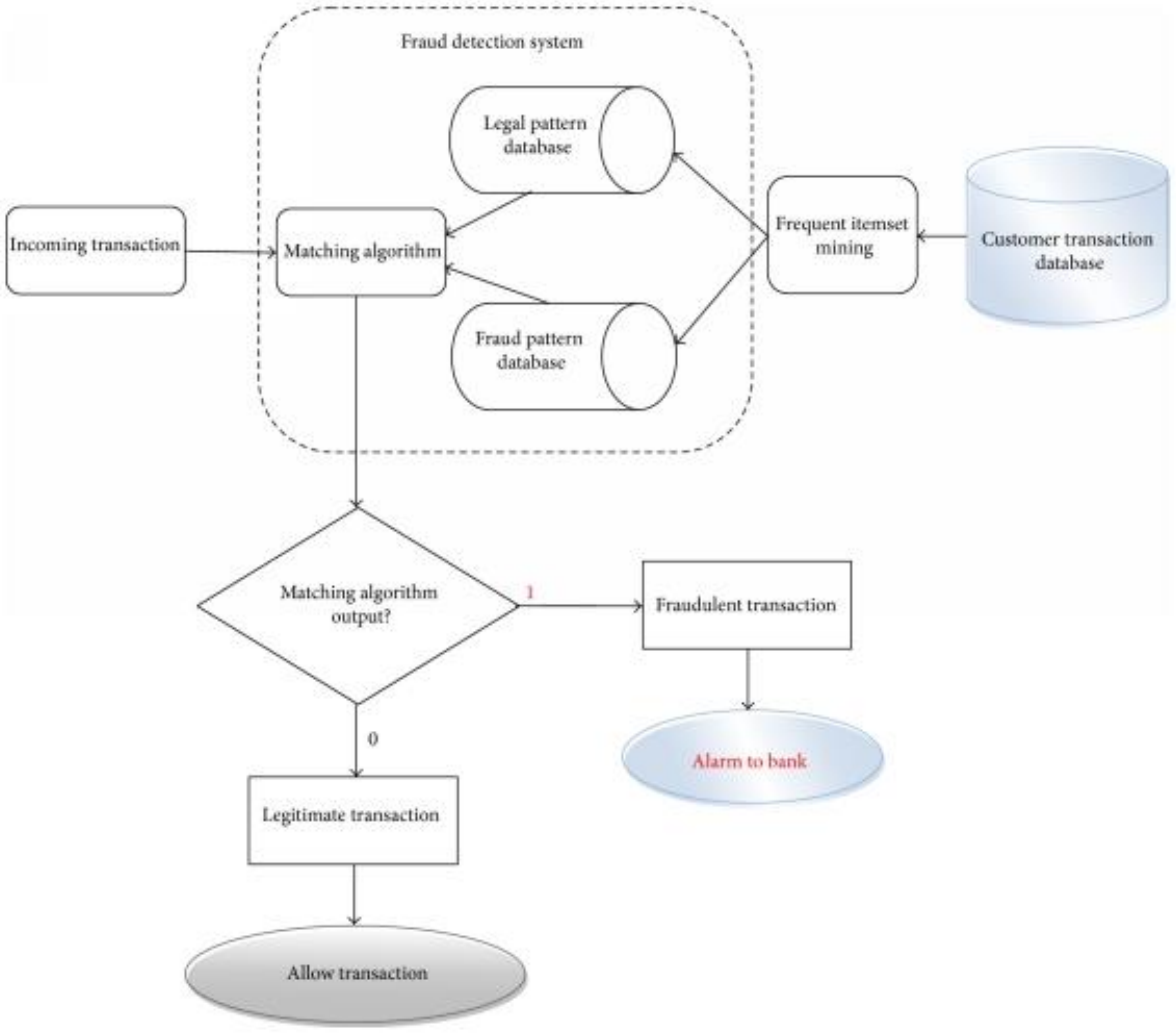


Figure 2.6: Conceptual Framework for the Proposed Prototype

Chapter 3: Research Design and Methodology

3.1. Introduction

This was an applied research whose aim was to improve the understanding of procurement fraud resulting in a potential solution that will enable proactive detection of the said fraud. The research techniques used to obtain data for this study included in-depth interviews, structured online questionnaires, and review of secondary data. The primary data was collected from the 'Large Peer Group' / Tier I licensed commercial banks in Kenya. Once the data was collected, it was analysed using quantitative (statistical) analysis techniques which helped the researcher make decisions on the development of the prototype. The development of the prototype followed the expert system development life cycle methodology and utilized the structured system design techniques.

3.2. Research Design

The research relied on the **case study research strategy** to answer the research questions identified in this research. The case study strategy enabled the researcher to gain a rich understanding of the procurement fraud domain and processes being enacted to proactively detect it. The case study strategy has considerable ability to generate answers to the 'why?', 'what?' and 'how?' research questions and it may use quantitative or qualitative methods or use a mix of these methods to collect and analyse data. These may include, interviews, observation, documentary analysis and questionnaires (Saunders, Lewis, & Thornhill, 2012). However, for this research, the researcher used quantitative research methods to collect, analyse data and developed a conceptual framework.

3.3. System Development Methodology

Prototyping is the generally accepted approach to developing expert systems (Turban & Aronson, 2001). The prototype was developed using the Rapid Application Development (RAD) methodology. This methodology is appropriate due to its iterative nature and the limitation of time and other resources for the researcher. This approach provides a reduction in development time, ability to garner constant customer feedback and allows progress to be accurately and quickly measured. This inherently reduced the overall risk as it broke the prototype into smaller

manageable subtasks (Stiner, 2016). The development of the prototype followed the expert system development life cycle shown in Figure 3.1.

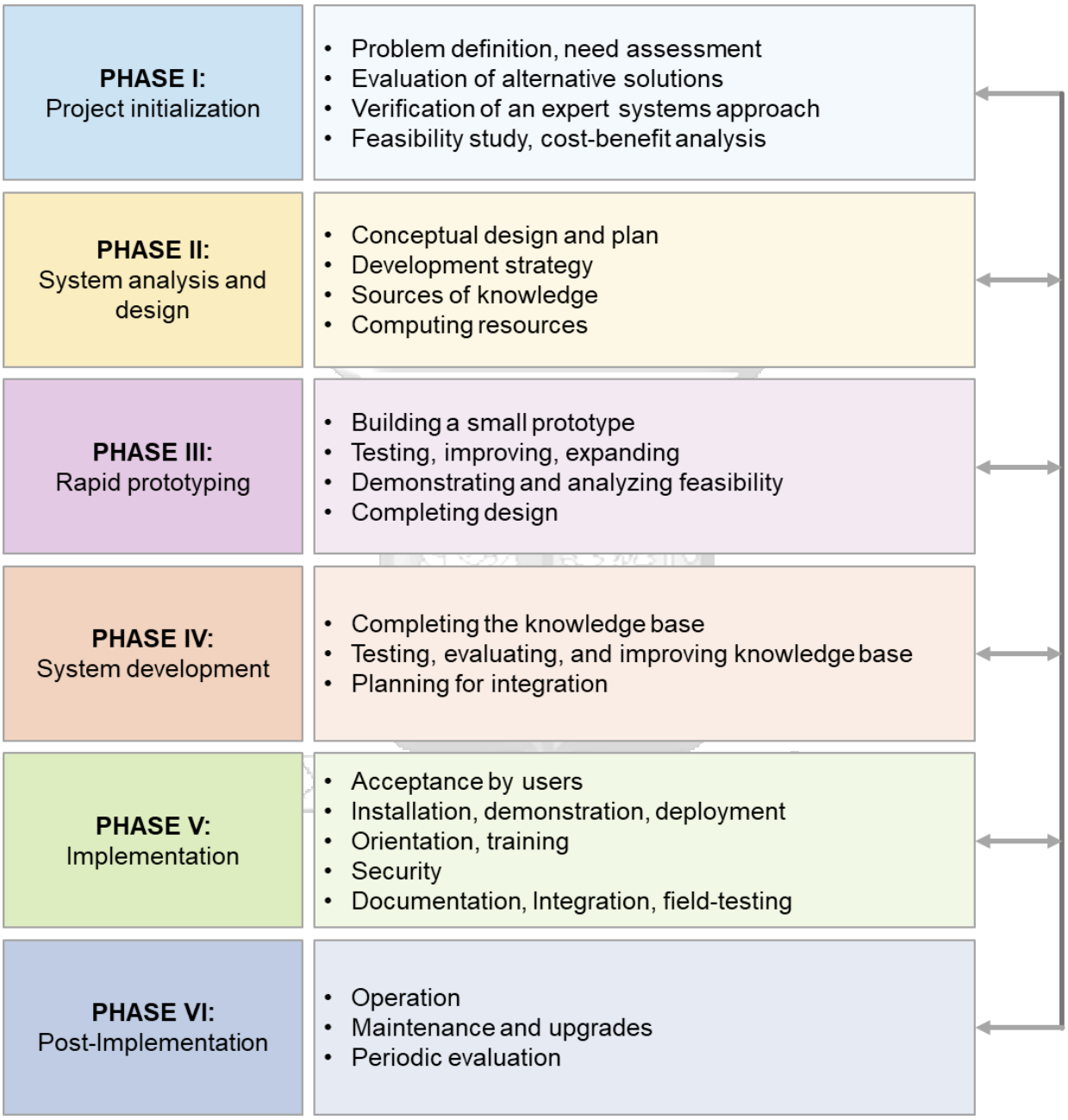


Figure 3.1: Schematic View of the Expert System Development Life Cycle (Turban & Aronson, 2001)

3.4. System Development Tools and Technologies

The prototype was developed using the following tools and technologies:

i. **SQLite database**, a C-language library that implements a small, fast, self-contained, high-reliability, full-featured, SQL database engine. The SQLite file format is stable, cross-platform, and backwards compatible. SQLite database files are commonly used as containers to transfer rich content between systems and as a long-term archival format for data (The SQLite Consortium, 2021).

ii. **Django**, a Python-based web framework, that follows the model–template–views architectural pattern. The model layer provides an abstraction layer (the "models") for structuring and manipulating the data of your web application. The template layer provides a designer-friendly syntax for rendering the information to be presented to the user. The view layer encapsulates the logic responsible for processing a user's request and for returning the response (Django Software Foundation, 2021).

iii. **Django REST framework**, a powerful and flexible toolkit for building Web APIs and rendering the database file in JSON so that the frontend can consume this data. It provides serialization and authentication policies (Encode OSS Ltd, 2021).

iv. **Pandas**, a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language (The Pandas Development Team, 2021).

v. **React JS**, a front-end JavaScript library for building interactive user interfaces based on UI components. It composes complex UIs from small and isolated pieces of coded called "components". React will efficiently update and render just the right components when your data changes (Meta Platforms, Inc., 2021).

vi. **Material UI**, a robust, customizable, and accessible library of foundational and advanced React-based components used to build faster, beautiful, and more accessible React applications (Material-UI SAS, 2021).

vii. **Visual Studio Code,** a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git (Microsoft Corp., 2021).

3.5. Population and Sampling

This research was based in Kenya's banking industry. As of 31 December 2020, Kenya had 42 licensed commercial banks regulated by the Central Bank of Kenya. 9 out of the 42 commercial banks were categorized under the 'Large Peer Group' or Tier I category which had a 93.05% market share in the number of deposit accounts and 96.78% market share in the number of loan accounts (Central Bank of Kenya, 2020). Therefore, the target population was composed of the 9 'Large Peer Group' / Tier I licensed commercial banks in Kenya.

Table 3.1: Banking Sector Market Share - December 2020 (Central Bank of Kenya, 2020)

#	Large Peer Group / Tier I – Commercial Banks	% Market share in number of loan accounts	% Market share in number of deposit accounts
1	KCB Bank Kenya Ltd	16.39	12.82
2	Equity Bank Kenya Ltd	6.09	15.36
3	NCBA Bank Kenya Plc	63.91	55.42
4	Co-operative Bank of Kenya Ltd	7.03	5.18
5	Absa Bank Kenya Plc	2.23	2.63
6	Standard Chartered Bank (K) Ltd	0.44	0.34
7	Diamond Trust Bank Kenya Limited	0.12	0.75
8	I & M Bank Limited	0.13	0.25
9	Stanbic Bank Kenya Ltd	0.45	0.30
I	**Large Peer-Group Sub-Total**	**96.78**	**93.05**
II	**Medium Peer-Group Sub-Total**	**2.31**	**5.57**
III	**Small Peer-Group Sub-Total**	**0.91**	**1.38**
IV	**Grand Total**	**100.00**	**100.00**

3.5.1 Sample Size

The researcher applied the Slovin's formula in Equation 3.1 to determine the desired sample size of the 'Large Peer Group' licensed commercial banks in Kenya that was selected for the research.

$$n = \frac{N}{1 + N(e)^2}$$

Equation 3.1: Slovin's Sample Size Equation (Anderson, Sweeney, Williams, Freeman, & Eddie, 2010)

Where:

n = Sample size

N = Total Population (9 'Large Peer Group' Banks)

e = Margin of error (10%)

Based on the Sample Size Equation 3.1, the number of Tier I licensed commercial banks in Kenya that was sampled for this research was eight (8).

$$\mathbf{8.25} = \frac{9}{1 + 9\ (10\%)^2}$$

Equation 3.2: Sample Size Calculation

3.5.2 Target Group

Due to the technical nature of the research, the research focused on participants working in the 'Large Peer Group' Banks across Kenya whose primary occupation / profession belonged to one of the following categories:

i. Fraud examiner/investigator working in the Bank's Forensic department.

ii. Internal auditor working in the Bank's Internal Audit department.

iii. Risk and controls professional working in the Bank's Risk department.

iv. External/independent auditor or consultant providing assurance services to the Bank.

According to Saunders et al. (2012), homogenous sampling focuses on one subgroup in which all the sample members are similar, such as a particular occupation or level in an organization's hierarchy. For this research, the participants were selected using the purposive non-probability sampling design, specifically the homogenous sampling technique since the research focused on one subgroup of participants (i.e., professionals working in Internal/External Audit, Risk or Forensic departments).

The researcher applied the minimum non-probability sample size limited guidance available for different types of study (Saunders, Lewis, & Thornhill, 2012) summarized in Table 3.2. For this research where the aim was to understand commonalities within a homogeneous group, **8 in-depth interviews** were conducted.

Table 3.2: Minimum non-probability sample size (Saunders, Lewis, & Thornhill, 2012)

Nature of Study	Minimum sample size
Semi structure/in-depth interviews	5 – 25
Ethnographic	35 – 36
Grounded theory	20 – 35
Considering a homogeneous population	4 – 12
Considering a heterogenous population	12 – 30

3.6. Data Collection and Requirements Gathering

During this research, the researcher aimed at collecting both primary and secondary data. This involved:

i. **Interviews**—this was used to gather user requirements and information from forensic, risk and audit professionals in the Banking industry. Due to the technical and sensitive nature of the topic, interviews offered a better approach to understanding the process of fraud detection and also provided an avenue for more explanations which helped the researcher to gather user requirements. The interview questions included open-ended questions, list questions, category questions and rating questions which used the **Likert-style rating**.

ii. **Acceptance testing questionnaires**—these were administered after developing the prototype to test and evaluate the readiness of the developed prototype. The questionnaire covered four main areas, namely: functionality, usability, performance, and data security.

iii. **Secondary data collection**— given the sensitive and confidential nature of fraud cases in organizations, it would not be appropriate to collect this data using primary data collection techniques. Therefore, secondary data was also relied upon to obtain and analyse data of prior fraud cases and gather data on the risk profiles of those who committed occupation fraud to better understand red flags of fraud in organizations. In addition, secondary data on a listing of firms and individuals that are ineligible to participate in procurement tenders was used to enhance the fraud rules.

3.7. Data and Requirements Analysis and Presentation

This is involved organizing the data collected and breaking it further into smaller parts which could be easily understood. The **Likert-style rating** was used to rate the responses to the interview questions (Saunders, Lewis, & Thornhill, 2012) thereby enabling quantitative data analysis to be done using Microsoft Excel. Data analysis results were presented using the following tools:

i. **Tables**—to summarize the significant variables.

ii. **Charts**—to provide a visual representation of the quantitative data and to facilitate comparison and correlations within the data.

iii. **Structured software development techniques**—to analyse user requirements and create system design models. These included contextual diagrams, dataflow diagrams (DFDs), Entity Relationship Diagrams (ERD), relational database schemas, and wireframes.

3.8. Research Validity and Reliability

To ensure the research validity, the researcher ensured that the data collected from the respondents was relevant and sensitive questions that would reveal confidential information in the Bank were not asked to the respondents. The researcher's academic supervisor validated the study to ensure that it matches the research objectives.

The researcher ensured the research reliability by following the thesis guideline provided by the university and international research standards. Test cases were also developed and used to test the

functionalities of the prototype. Due to time constraints, only integration and system testing was done.

3.9. Ethical Considerations

To uphold ethical standards, the researcher had to obtain consent from the selected participants before the survey. The data gathered was treated with a high degree of confidentiality and was solely used for the purpose of this research. All the interview schedules also had a disclaimer.

3.10. Dissemination and Proposed Utilization of the Research Results

The results of the research were disseminated with the participants through presentations to raise awareness and inform them on the findings of the research. The results from the research were also utilized to design, develop, and test the fraud detection prototype.

The prototype will be utilized to assist institutions in Kenya's banking industry to proactively detect procurement fraud and flag anomalous behaviour thereby minimizing fraud related losses. It will aid audit, forensic, security and compliance professionals to continuously monitor procurement transaction data and advise them of potential fraud cases to investigate.

3.11. Ethical Approval

The Institutional Ethical Approval was obtained from the Strathmore University Institutional Ethics Review Committee (SU-IERC), Application Reference Number: **SU-IERC1278/22** refer to Appendix B for the approval letter. In addition, a research license to conduct the research in Nairobi, Kenya was also obtained from the National Commission for Science, Technology & Innovation (NACOSTI), License No: **NACOSTI/P/22/16583** refer to Appendix C for the research license.

Chapter 4: System Analysis, Design and Architecture

4.1. Introduction

The main purpose of this study is to design and develop a prototype that will utilize data mining techniques to detect procurement fraud. Using the expert system development life cycle methodology and **Structured System Analysis and Design techniques**, a thorough requirements analysis and system design was performed. The analysis involved data collection, understanding, clarifying, and documenting the requirements of the system. Thereafter, the system process models, data models, user interface and system architecture were designed to enable the development and implementation of the system.

4.2. Requirements Gathering and Analysis

4.2.1 Red Flag Indicators of Procurement Fraud

To extensively gather user requirements, in-depth interviews were conducted to audit, risk and forensic professionals working in Banks in Kenya to understand how procurement fraud was being perpetrated using interview questions in Appendix D. Respondents provided information on the most common red flags indicators for procurement fraud based on their experience, knowledge, and exposure. For each red flag indicator, the respondents provided a score between 0.5 to 1.0 where indicators with a score of 0.5 represented a possible fraudulent scenario and a score greater than 0.8 represented a potential fraudulent scenario. This data helped to build the data mining fraud detection rulesets that were incorporated in the prototype. Table 4.1 summarizes the procurement fraud red flags provided by the respondents during the interviews.

Table 4.1: Red Flag Indicators of Procurement Fraud

No.	Category	Red Flag Indicator	Score
1.	**Analytical Anomalies**	Employee identifier document and number (e.g., ID Number) matches with the identifier numbers of vendor directors	1.0
		Employee mobile number matches the mobile number of vendors	1.0
		Employee residential address matches vendor's office address	0.5
		Employee postal address matches vendor's postal address	1.0
		Employee personal email matches the email address of vendors	1.0

No.	Category	Red Flag Indicator	Score
		Employee bank details matches vendor's bank details	1.0
		Employee name matches the names of vendor directors	0.8 – 1.0
		Employee next of kin details matches the names of vendor directors	1.0
		Employee next of kin mobile number matches the mobile number of vendor directors	1.0
		Employee residential address matches the residential address of vendor directors	0.5
		Employee postal address matches the postal address of vendor directors	1.0
		Employee personal email matches the email address of vendor directors	1.0
		Employee mobile number matches the mobile number of vendor directors	1.0
		Vendors without physical addresses	0.5
		Duplicate vendors and duplicate employees	0.5
		High volume of purchases from new vendors	0.5
		Vendor and Vendor Director details (i.e., name and address) matches the World Bank Listing of Ineligible Firms and Individuals	0.8 – 1.0
		Slight variation of vendor names	0.8 – 1.0
2.	**Accounting Anomalies**	Multiple duplicate payments of the same or similar amounts to a vendor	1.0
		Multiple payments for the same invoice number	1.0
		Lack of invoices to support payments	1.0
		Payments to vendors who aren't on an approved vendor list	0.5
		Purchase document details alterations	0.5
3.	**Internal Control Weaknesses**	Unauthorized changes to vendor name, address, and bank details	1.0
		Employee records created or modified after normal working hours or on Sundays	0.5
		Vendor records created or modified after normal working hours or on Sundays	0.5
		Procurement related transactions posted after working hours or on weekends	0.5
		Procurement transactions created and approved by the same user	1.0
		Deletion of Employee, vendor or PTP data in the system	1.0
		Procurement related transactions posted by former staff	1.0

4.2.2 World Bank Listing of Ineligible Firms and Individuals

Secondary data from the World Bank containing a listing of ineligible firms and individuals was utilized to cross check whether a vendor exists in the list as per the gathered user requirements. This list is maintained and published by the World Bank to ensure that misconduct perpetrated by vendors are subject to sanctions. These *"Sanctionable Practices"* include corrupt, fraudulent, collusive, and coercive practices. The World Bank may also sanction a firm or individual for having engaged in obstructive practices in connection with an investigation (World Bank Group, 2019).

The firms and individuals listed, are debarred and thus ineligible to participate in World Bank-financed contracts for the periods indicated. These entities are sanctioned as a result of prohibited conduct defined in the applicable Procurement or Consultant Guidelines or in the World Bank Procurement Regulations for Investment Project Financing Borrowers (World Bank Group, 2022). As of 08 April 2022, there were 1,212 firms and individuals that had been debarred by the World Bank due to various grounds illustrated and summarized in Figure 4.1.

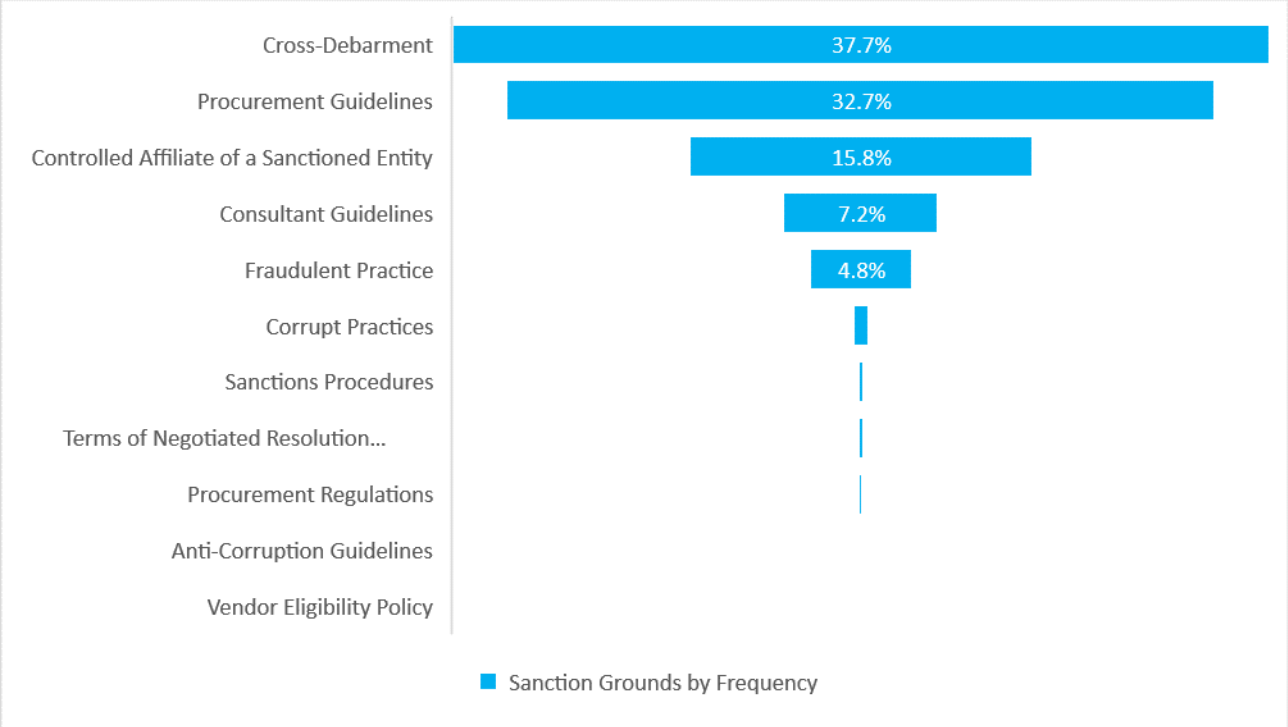


Figure 4.1: Summary of Ineligible Firms and Individuals by Grounds (World Bank Group, 2022)

4.2.3 Risk Profile of a Fraud Perpetrator

To determine the risk profile of a fraud perpetrator, secondary data from a study conducted by the ACFE that included perpetrator data from more than 2,000 fraud cases was analysed. This data identified the common characteristics and risk profiles of those who committed occupational fraud and aided the researcher to better understand red flags of fraud in organizations. The data was used by the researcher to develop an algorithm to calculate the risk profile of employees in the Bank thereby providing anti-fraud professionals with a training and awareness tool to understand the human face of procurement fraud. According to the ACFE (2020) study, the following common characteristics of fraud perpetrators were identified:

i. **Perpetrator's Tenure:** Employees who had worked for 1 – 5 years were more prone to commit fraud (i.e., 46%) and the longer a fraud perpetrator works for a company, the more damage that person's scheme is like to cause.

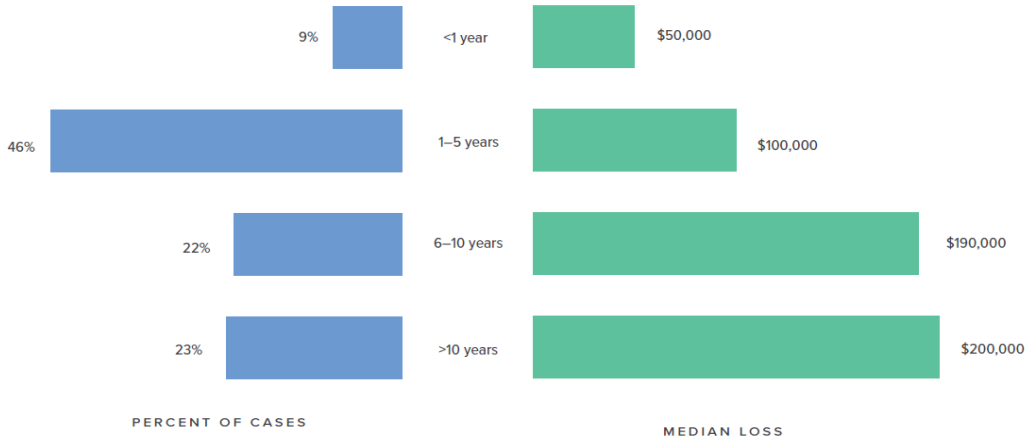


Figure 4.2: How does the perpetrator's tenure relate to occupational fraud (ACFE, 2020)

ii. **Perpetrator's Department:** The following eight departments (i.e., Operations, Accounting, Executive/upper management, Sales, Customer Service, Administrative Support, Finance and Purchasing) accounted for 76% of all occupational frauds in the study and were classified as high-risk departments. This illustrated the relative risks of occupational fraud in the different parts of a typical organization. For example, the operations (15%), Accounting (14%), and sales (11%) departments were both associated with high frequency and median loss, which indicates that fraud risks in these areas should be carefully addressed in any anti-fraud program.

iii. **Perpetrator's Gender:** More than 70% of the perpetrators in the study were males. Men also caused a significantly larger median loss (USD 150,000) than women (USD 85,000). However, in Sub-Saharan Africa, males accounted for 84% of fraud perpetrators.

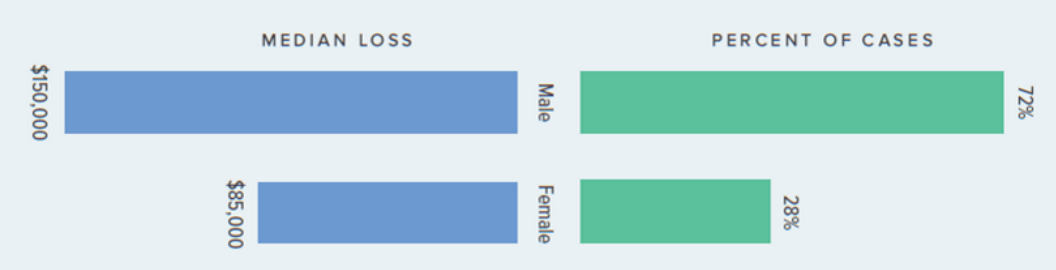


Figure 4.3: How perpetrators' gender relates to occupation fraud (ACFE, 2020)

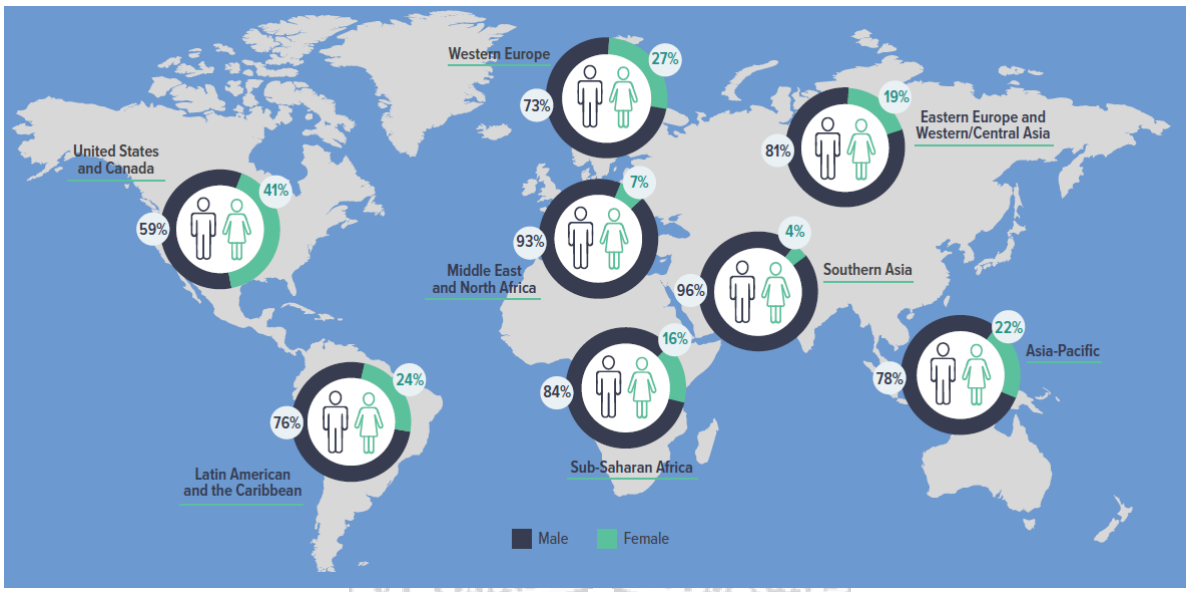


Figure 4.4: How gender distribution of perpetrators varies by region (ACFE, 2020)

iv. **Human Resources-Related Red Flags:** In some circumstances, negative events surrounding a person's conditions of employment (such as poor performance evaluations, loss of pay or benefits, fear of job loss, etc.) can cause financial stress or resentment toward the employer, which might play a role in the decision to commit fraud. 42% of fraudsters had experienced some form of HR-related red flags prior to or during the time of their frauds. The most common of these were negative performance evaluations (13% of cases), fear of job loss (12%), denied raise or promotion (10%), complained about inadequate pay (8%), refusal to take vacations (7%) and unwillingness to share duties (15%).

v. **Perpetrator's Age:** 53% of fraudsters in the study were between the ages of 31 and 45. Median losses, on the other hand, tended to rise along with the age of the perpetrator. Those in the 56 to 60 and 60+ age ranges together accounted for less than 10% of all cases but caused the highest median losses in any age range.

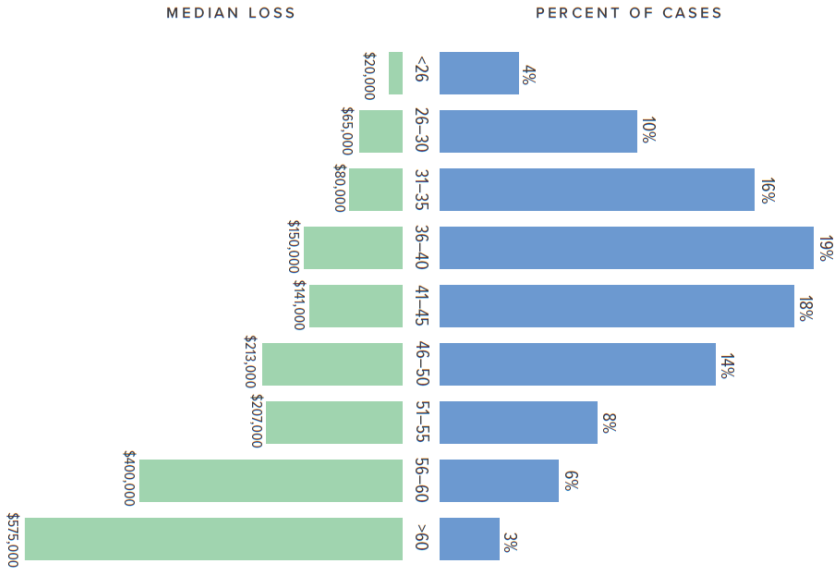


Figure 4.5: How the perpetrator's age relates to occupational fraud (ACFE, 2020)

vi. **Perpetrator's Education Level:** Fraudsters with a postgraduate degree caused a median loss of USD 200,000. Generally, those with higher levels of education tend to hold higher positions of authority and also have greater technical capabilities for committing fraud.

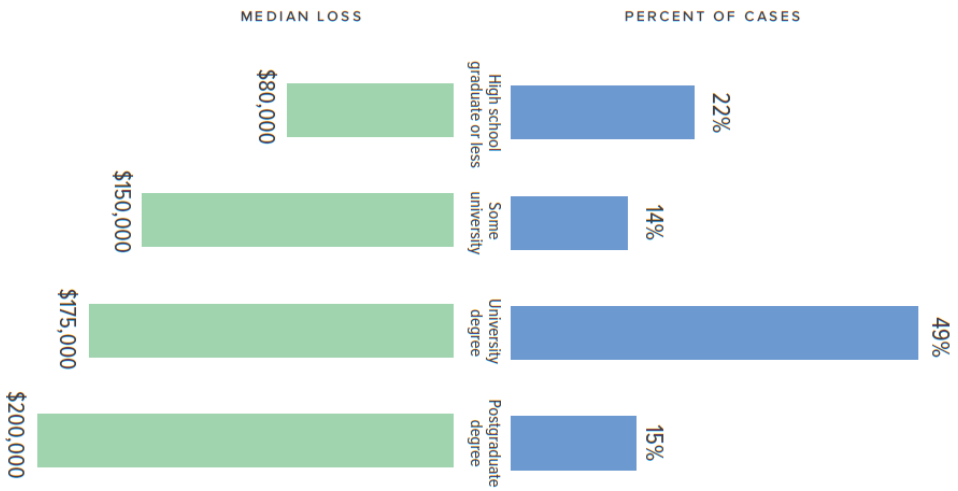


Figure 4.6: How the perpetrator's education level relates to occupational fraud (ACFE, 2020)

4.3. System Analysis

4.3.1 Functional Requirements

The system aims to bring various advantages for its users such as the ability to detect the occurrence of potential procurement related fraudulent activity. Through the interviews conducted and secondary data analysed, the research found out that the functional requirements of the system should include:

i. The system should be able to capture employee details from the HR system and calculate the fraud risk profile for the employee based on their attributes (i.e., age, gender, tenure, education level, performance rating, disciplinary record, etc). The computed risk profile will assist in the detection of potentially fraudulent activity committed by the employee.
ii. The system should be able to capture vendor data and procurement related transactions from the ERP system.
iii. The system should be able to capture additions, modifications and deletions of employee, vendor, and procurement related transaction data as events indicating the date/time of the event and the user responsible for the event.
iv. The system should be able to perform data mining algorithms on the events and flag events as either fraudulent, possible fraudulent, or non-fraudulent based on defined fraud rulesets.
v. The system should be able to upload the World Bank listing of ineligible firms and individuals to be used by data mining algorithms to perform reference check matches.
vi. The user (fraud analyst) should be able to view the events that have been flagged as fraudulent or possible fraudulent for investigation purposes. Thereafter, the user should be able to capture and update the outcome of their investigation on the events.
vii. The user (fraud analyst) should be able to update the fraud ruleset thresholds used by the data mining algorithms.
viii. The system should be able to automatically send notifications to the users (fraud analysts) email addresses on the events flagged as fraudulent.
ix. The user should be able to change their user account passwords according to predetermined password policies.
x. The system administrator should be able to create, update and disable users.

4.3.2 Non-Functional Requirements

The non-functional requirements identified for the prototype included:

i. **Availability:** The system should be easily accessible and available to all users whenever they require it to be so. In order to ensure high availability, the system should be created using reliable and robust software.

ii. **Security:** Due to the personal data that will be fed into the system, it is vital that the developed system is secure. The system should be resistant to any vulnerable states as well as putting its data at risk. The system should only be accessible by authorized users and assure that all data inside the system or its part are protected against malware attacks or unauthorized access.

iii. **Usability:** The system should be easy to use for all users, regardless of their knowledge, familiarity, or situation. It should be designed in a user-friendly way to ensure users are confident, and comfortable, when using the system. The user interface should be clear, concise, consistent, and attractive allowing users to easily interact with the system.

iv. **Robust:** The system should be robust, ensuring it can survive any errors during the execution process, whilst any errors have minimal impact on the user.

v. **Response Time:** In order to meet user needs, the developed system should reach quick response times, ensuring the users can be provided with the relevant output in a reasonable time frame. Using appropriate hardware and software will ensure that the system can perform to its expected level.

vi. **Maintainability:** The system should be easily maintained and be able to sustain any changes to increase performance or other qualities, fixes, or meet any new necessary requirements to adapt to a changing environment.

vii. **Portability:** The system should be accessible and operate satisfactorily from different environments. The system should be a web application that provides for a cross-platform, cross-browsing, and mobile-responsive solution.

viii. **Localization:** The system should be in line with the context of the local Kenyan market-to-be. The local system language should therefore be in English. The date format must follow: *"date-month-year"* format and should be based on the East Africa Time zone (EAT). The currency to be adopted by the system should be Kenya Shilling (KES).

4.4. Technical System Architecture

The system architecture is based on a **three-tier client server model** divided into three major components i.e., presentation, business logic and data management tiers. The **presentation tier** is accessible via a web browser, and it enables the end user to interact with the system. It sends HTTPS requests (i.e., GET, POST, PUT & DELETE) methods and receives their HTTPS response. The **business logic tier** contains the APIs that receive HTTPS requests from the presentation tier, interrogates the database to execute the data mining algorithms and communicates the results back to the presentation tier. It is based on Restful Web Services built on the REST (Representational State Transfer) framework to expose the system's APIs in a secure, uniform, and stateless manner to the presentation tier. The underlying protocol for REST is HTTPS. The **data management tier** hosts the relational database that stores the data consumed by the system. The tier handles the Create, Read, Update and Delete (CRUD) operations required for interacting with the database. Figure 4.7 illustrates the Technical System Architecture for the prototype.

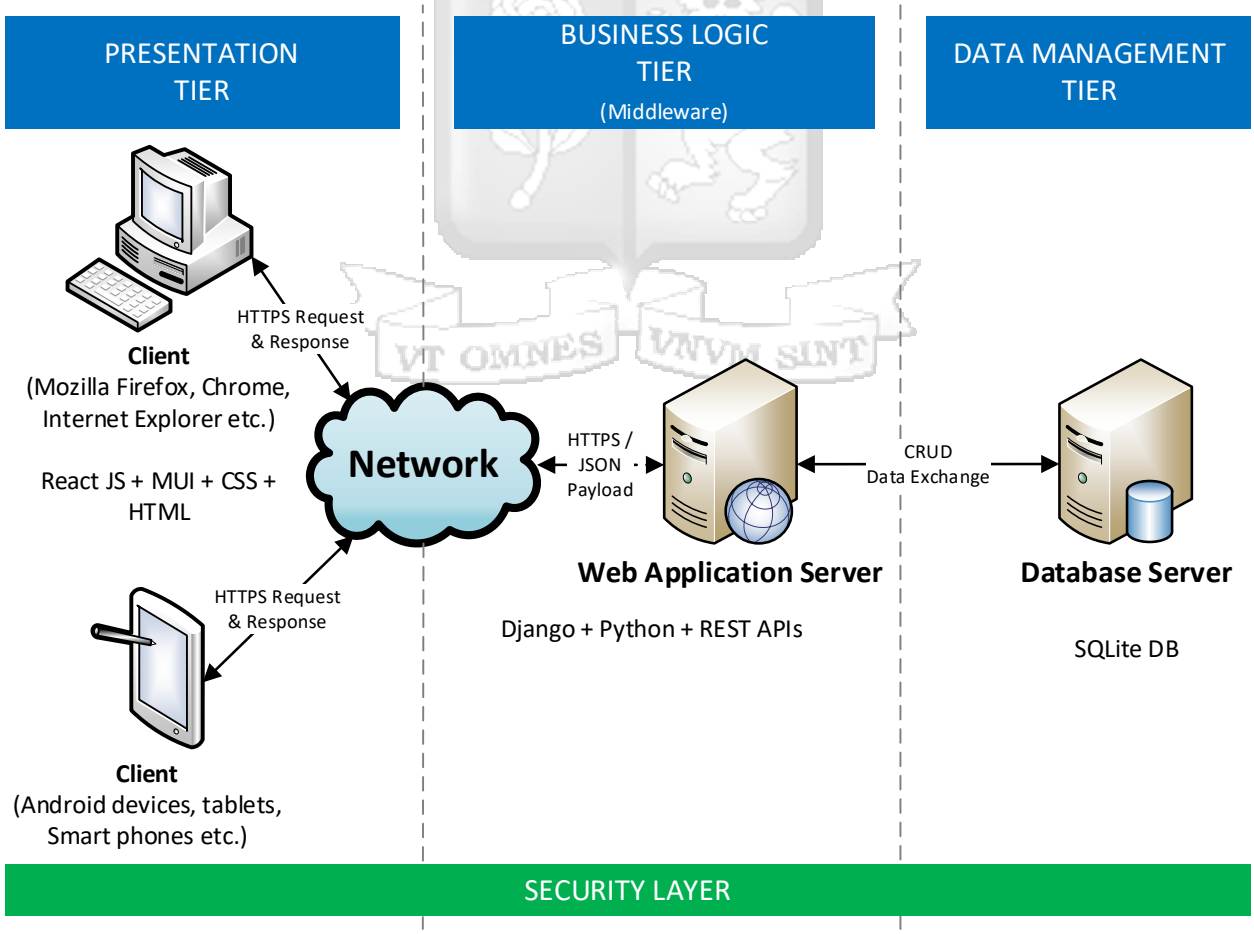


Figure 4.7: Technical System Architecture

4.5. System Process Modelling

4.5.1 Context Level Diagram

Phase II of the expert system development life cycle shown in Figure 3.1, requires the creation of a conceptual design and plan that will provide a general idea of what the system will look like and how it will solve the problem. The design will also show the general capabilities of the system, the interfaces with other computer-based information systems, the areas of risk, and the required resources (Turban & Aronson, 2001).

The Context Level Diagram in Figure 4.8 illustrates the data flowing across the boundaries of the system from processes within the system to external sources and recipients, and vice versa. The main process is to detect procurement fraud and four (4) external entities were identified namely: ERP System, HR System, Fraud Analyst and System Admin.

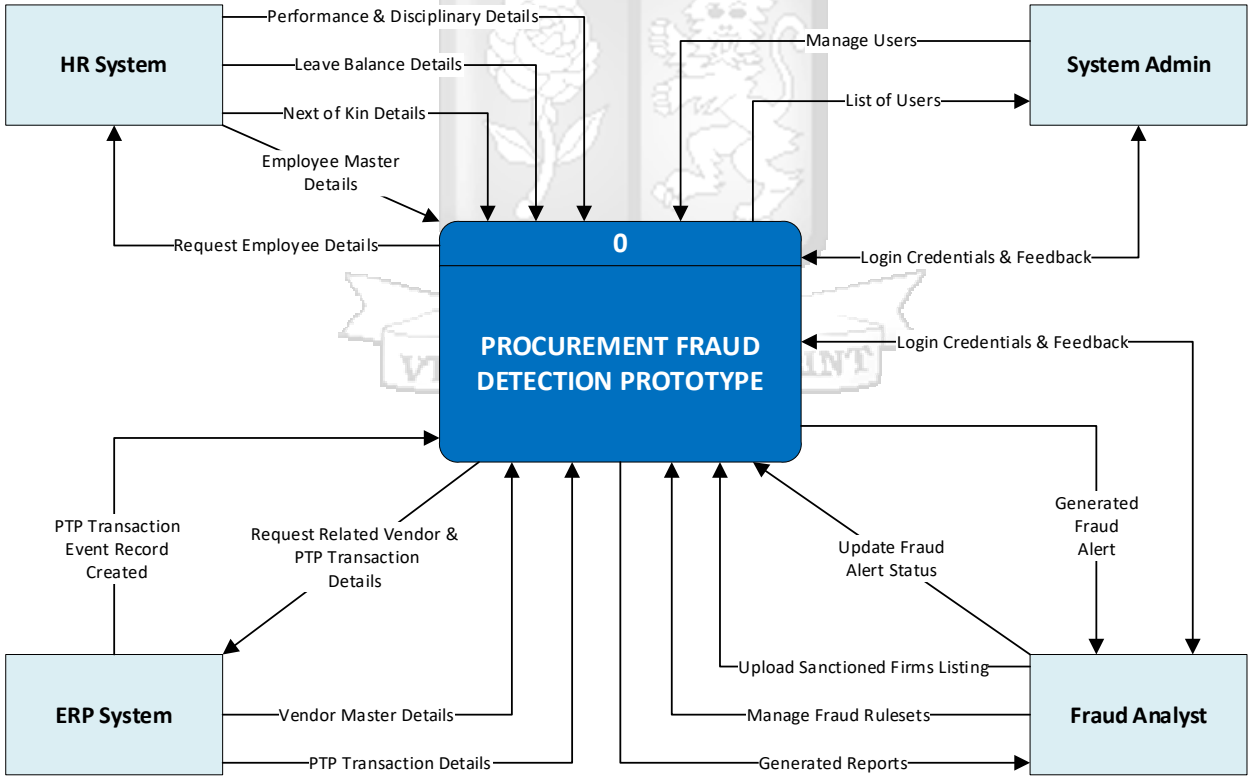


Figure 4.8: Context Level Diagram

4.5.2 Level 1 Data Flow Diagram

A data flow diagram (DFD) provides a logical graphical model of information flow, partitioning a system into modules that show manageable levels of detail and specifies the processes or transformations that occur within each module and the interfaces that exist between them (Laudon & Laudon, 2014). As part of the system's conceptual design, the data flow diagram (DFD) was used to represent the system's six (6) component processes and the flow of data between them. The major processes performed by the fraud detection prototype include:

i. **Capture PTP Transaction Event**

This process will capture the details of a PTP transaction event that the system will analyse to determine whether it is fraudulent. The PTP transaction event may include: creation, modification, or deletion of a vendor, purchase requisition, purchase order, good receipt note, vendor invoice, or vendor payment. Thereafter, this process will create a fraud review case which the fraud analysts will use to track and view the results (i.e., 'fraudulent event' or 'non-fraudulent event') of the data mining algorithms for each specific fraud review case.

ii. **Capture Vendor and PTP Transactions**

This process aims to retrieve vendor details of the vendor who is associated with the PTP transaction event that the system will analyse to determine whether it is fraudulent. It will also retrieve the related PTP transactions associated with the PTP transaction event being analysed. This information will be retrieved from the Bank's ERP system and will be consumed by the data mining algorithms to detect fraudulent activities.

iii. **Capture Employee Details**

This process aims to retrieve details of the staff who created or approved the PTP transaction event. The employee details will be retrieved from the Bank's HR system and will include details such as: employee's details, employee's leave balance, employee's next of kin details, employee's performance history, and employee's disciplinary history. Once retrieved, data mining algorithms will be applied to calculate the employee fraud risk profile and detect potentially fraudulent activities performed by the employees.

iv. **Manage Fraud Ruleset and Upload Sanctioned Firms**

This process provides an interface to the users to amend the thresholds to the fraud rulesets that are used by the data mining algorithms to detect fraudulent activities. It allows a user to disable an existing ruleset so that the data mining algorithm excludes it when executing the algorithms.

In addition, this process enables the fraud analyst to upload the World Bank listing of ineligible / sanctioned firms and individuals. This list will be used to by the data mining process to perform fuzzy string-matching to determine whether a vendor exists in this list.

v. **Data Mine and Generate Alerts**

This process executes data mining algorithms on the PTP events to check and determine whether it is fraudulent or not. Once complete, the process will flag/earmark the event as either fraudulent, possible fraudulent or non-fraudulent and will generate a fraud alert to the fraud analyst in case it determines that the event is fraudulent.

In addition, this process will calculate and determine the fraud risk profile for each employee in the Bank based on the employee's attributes and fraud rulesets. Together with the risk profile, the process will check whether PTP events initiated by the employee are suspicious enough to be flagged as fraudulent or not. Lastly, the process will also generate reports and alerts that will be consumed by the fraud analysts.

vi. **Login and Manage System Users**

This process handles the login of users into the prototype. In addition, the process manages the creation of new users and assigns them to their respective user roles such as 'administrator' or 'fraud analyst'. It also manages the modification and disabling of existing users in the system.

The required system process was defined and illustrated diagrammatically using the required data flow diagram in Figure 4.9. The data flow diagram summarized the processes, data stores, external entities, and the data flows.

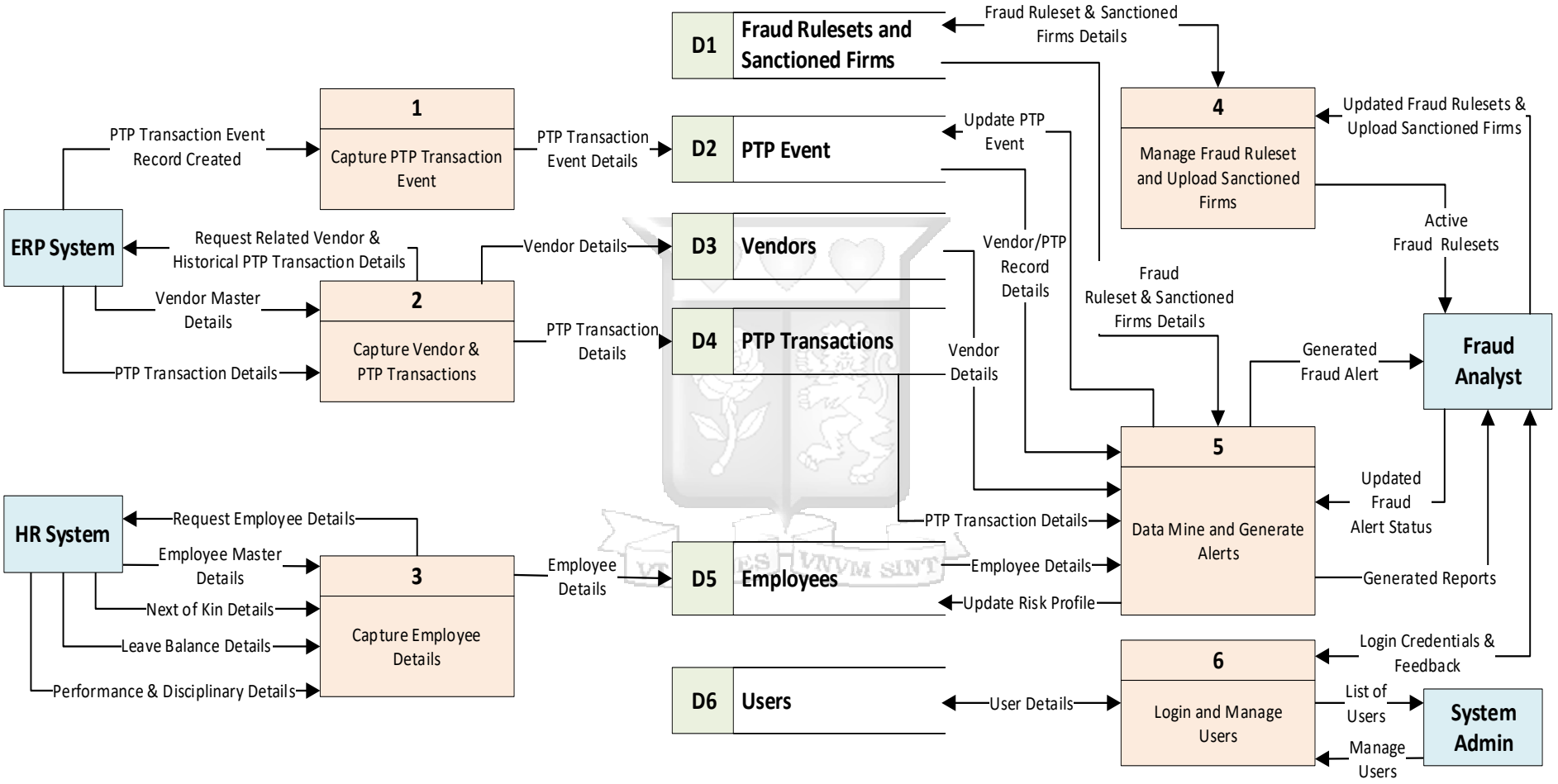


Figure 4.9: System Data Flow Diagram (Level 1)

4.6. Data Model

4.6.1 Entity Relationship Diagram

An entity relationship diagram is a methodology for documenting databases illustrating the relationship between various entities in the database (Laudon & Laudon, 2014). Based on the data flow diagram in Figure 4.9, Table 4.1 shows the list of identified entity types and Table 4.2 shows the relationships among the twelve (12) identified entity types.

Table 4.2: List of identified Entity Types

#	Entity Types	#	Entity Types
1	Fraud Rulesets & Sanctioned Firms	7	Vendor
2	Fraud Alerts	8	Vendor Directors
3	PTP Event	9	Employee
4	PTP Transaction Header	10	Employee Risk Profile Scores
5	PTP Transaction Detail	11	Fraud Analyst User
6	Inventory	12	User Role

Table 4.3: ER Relationship Types

#	Entity Type	Card.	Partic.	Relationship Type	Entity Type	Card.	Partic.
1	Fraud Rulesets & Sanctioned Firms	M	T	Applied In	PTP Event	N	P
2	Fraud Rulesets & Sanctioned Firms	M	P	Updated By	Fraud Analyst User	1	T
3	PTP Event	M	P	Assigned To	Fraud Analyst User	1	T
4	PTP Event	M	P	Performed By	Employee	1	T
5	PTP Event	M	T	Related To	PTP Transaction Header	1	P

#	Entity Type	Card.	Partic.	Relationship Type	Entity Type	Card.	Partic.
6	PTP Transaction Header	1	P	Contains	PTP Transaction Detail	M	T
7	PTP Transaction Detail	M	T	Utilizes	Inventory	1	P
8	PTP Event	M	P	Involves	Vendor	1	T
9	PTP Event	1	T	Generates	Fraud Alert	M	P
10	Vendor	M	T	Owned By	Vendor Directors	N	P
11	Employee	1	T	Has	Employee Risk Profile Scores	M	P
12	Fraud Analyst User	M	P	Assigned	User Role	1	T

KEY:

1	-	One	**CARDI.**	-	Cardinality
M/N	-	Many	**PARTIC.**	-	Participation
T	-	Total Participation	**P**	-	Partial Participation

Figure 4.10 shows the Entity Relationship Diagram based on the identified entities and relationship types.

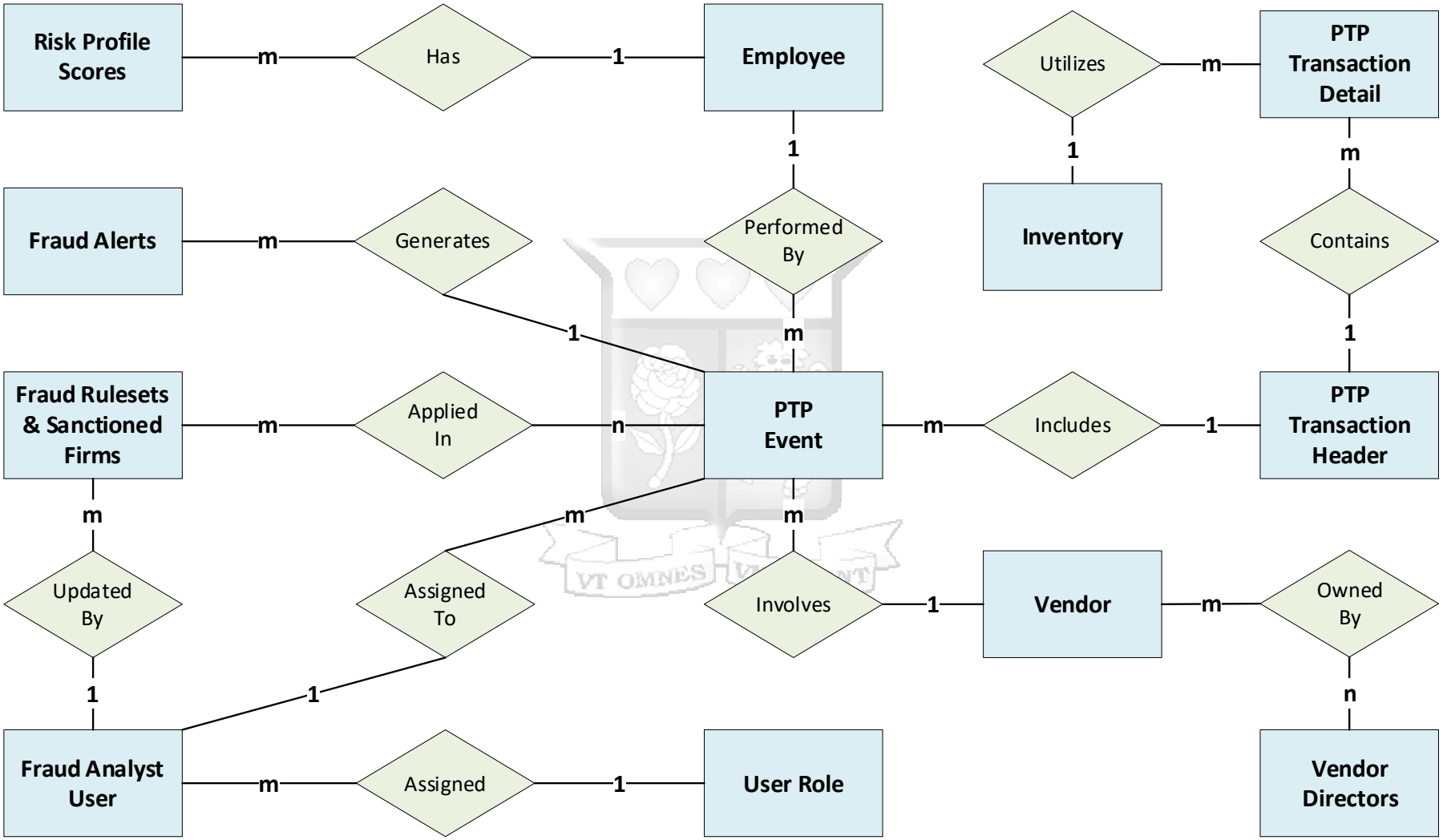


Figure 4.10: Entity Relationship Diagram (ERD)

4.6.2 Validating Relations Using Normalization

The relational mapping illustrated in Appendix F led to the creation of four (4) additional relations in scenarios where there was a many to many (M : N) relationship between the identified entity types in the Entity Relationship Diagram (ERD) illustrated in Figure 4.10. The additional relations created include:

i. **Vendor Ownership** – Captures the relationship between the directors and the vendors they own.
ii. **Fraud Review Case** – Stores the summary results of the positive and negative fraud tests results conducted on the PTP Events (i.e., creation, modification, and deletion events).
iii. **Fraud Results** – Stores the detailed results of the positive fraud test results with reference to the source record (e.g., employee, vendor, sanctioned firm, etc.) for ease of trace back.
iv. **Sanctioned Firms** – Stores the World Bank listing of ineligible firms and individuals.

This phase then aimed to establish whether the groupings of attributes in each identified relation in the relational mapping were compliant with the rules of normalization. This ensured that the set of relations identified had minimal and sufficient number of attributes that were necessary to support the data requirements for the procurement fraud detection prototype. The normalization process also ensured that the identified relations had minimal data redundancy so as to avoid problems with update, insertion, and deletion anomalies in the database during data manipulation tasks performed by the prototype.

The relational mapping and data catalogue were subjected through the process of normalization, and it was established that all of the 16 relations were in the **Third Normal Form** (3NF) because of the following:

i. There were **no repeating groups** (nested relations) identified in each of the relations.
ii. Every non primary key attributes in each of the relations were **not partially dependent** on their primary keys.
iii. In all the identified relations, there were no non primary key attributes that were **transitively dependent** on their respective primary key.

4.6.3 Database Indexes to aid in Data Mining

When relations are very large, it becomes expensive to scan all the tuples of a relation to find those (perhaps very few) tuples that match a given condition. The technology of implementing indexes on large relations is of central importance in the implementation of DBMS's. This is because the existence of an index on an attribute may speed up greatly the execution of those queries in which a value, or range of values, is specified for that attribute, and may speed up joins involving that attribute as well (Garcia-Molina, Ullman, & Widom, 2009).

Given that the prototype will greatly rely on data mining techniques to detect potential procurement fraud, the designed data model will at minimum include indexes on the primary key of each relation to improve the speed of data retrieval operations on the database table.

In addition to the primary key indexes, the data model design for the prototype includes the following additional indexes providing the basis for both rapid random lookups and efficient access of ordered records.

Table 4.4: List of Database Indexes

#	Relation	Indexes
1	PTPEvent	• EventID [Primary Key] • PTPTrxID • VendorID • EmployeeID • EventSource
2	Fraud Alerts	• AlertID [Primary Key] • EventID
3	Fraud Review Case	• FraudReviewCaseID [Primary Key] • EventID • RuleID
4	Fraud Result	• FraudResultID [Primary Key] • FraudReviewCaseID

4.6.4 *Database Schema*

The derived database schema which identified all the tables and attributes that were created in the database is shown in Figure 4.11.

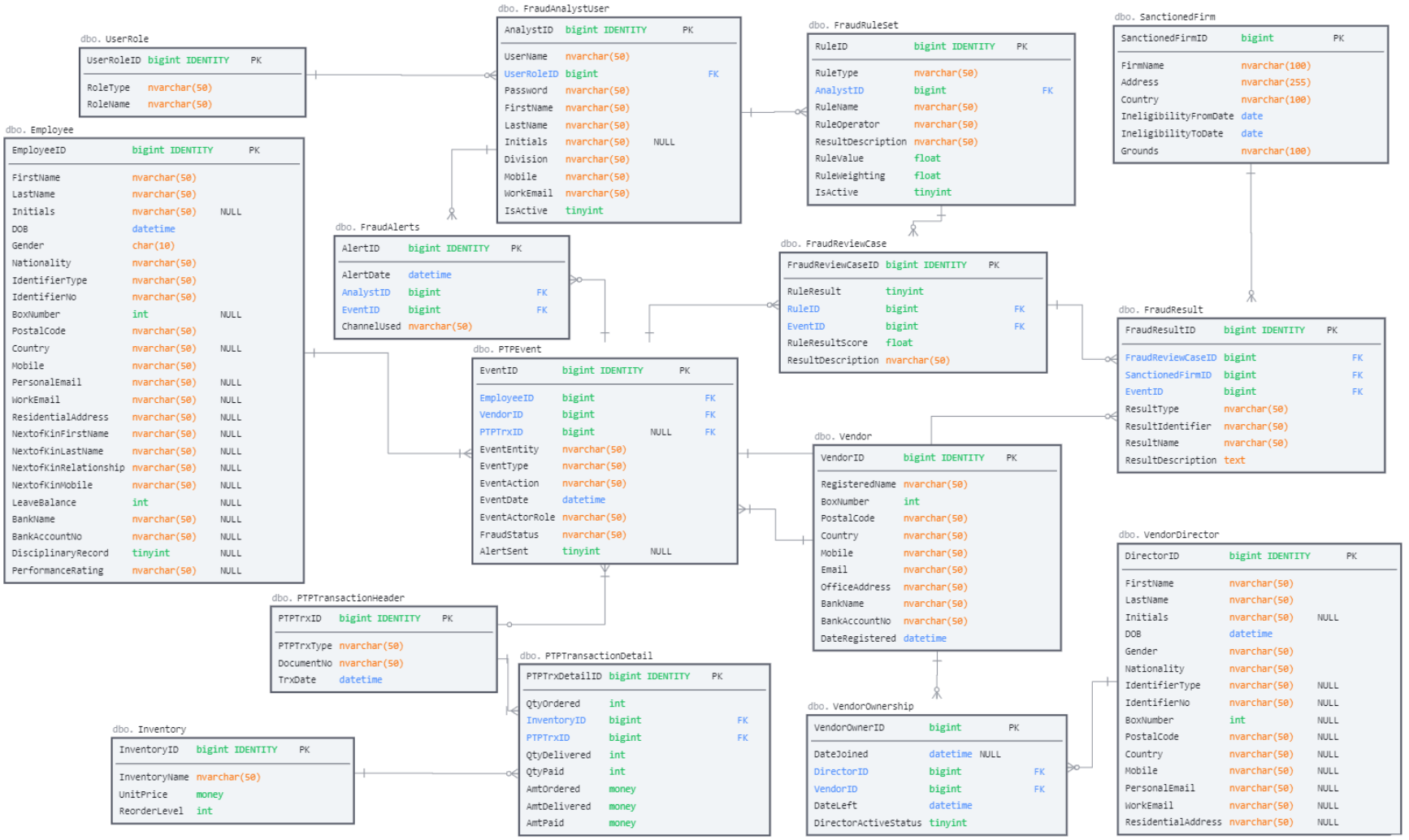


Figure 4.11: Database Schema

4.7. Security Design

Security design considerations were made to ensure that the personal data fed into the system remains secure and the system is resistant to any vulnerable states. The security design considerations involved the following aspects:

i. **Password Hashing:** User login credentials i.e., passwords were hashed using the Secure Hash Algorithm (SHA) and stores an irreversible SHA256 hash in the database to ensure even the database administrator cannot obtain the user's plaintext password. SHA is a government standard promoted by the National Institute of Standards and Technology.

ii. **Salting:** The password hashes also include a cryptographic salt (i.e., a random value) that is added to the end of the password before the system hashes the password in order to protect against rainbow table attacks The system utilizes a specialized password hashing function i.e., PBKDF2 algorithm that creates the SHA256 hashes using salts.

iii. **Password Policy:** The system required users to create strong login credentials based on a predetermined password policy. The password policy defined the type of characters to be used, minimum number of characters, duration of use of the password and prohibits the use of a password that has been used before.

iv. **Token Authentication:** To safeguard the backend APIs, the system implemented this HTTP authentication scheme that involves generating tokens by the server in response to a login request. The client must then send this token in the Authorization header when making requests to be able to access the protected API resources.

v. **Least Privilege:** The system utilized a Role Based Access Control (RBAC) to ensure that the authenticated users only access the modules they require to perform their jobs.

vi. **Whitelisting IP Addresses:** The system utilized Cross-Origin Resource Sharing (CORS) to whitelist only the domains that are allowed to access and consume the API resources.

vii. **Logging and Monitoring:** The system logged the date, time and user details that created or modified the data records to maintain an audit trail. The date and time were auto populated with the time on the database server to ensure unified timestamping.

4.8. User Interface Design

4.8.1 Sitemap

A sitemap was created to provide information about the web pages in the system and the relationship between them. The sitemap also illustrates how the end user will interact and navigate through the system and has been illustrated in Figure 4.12.

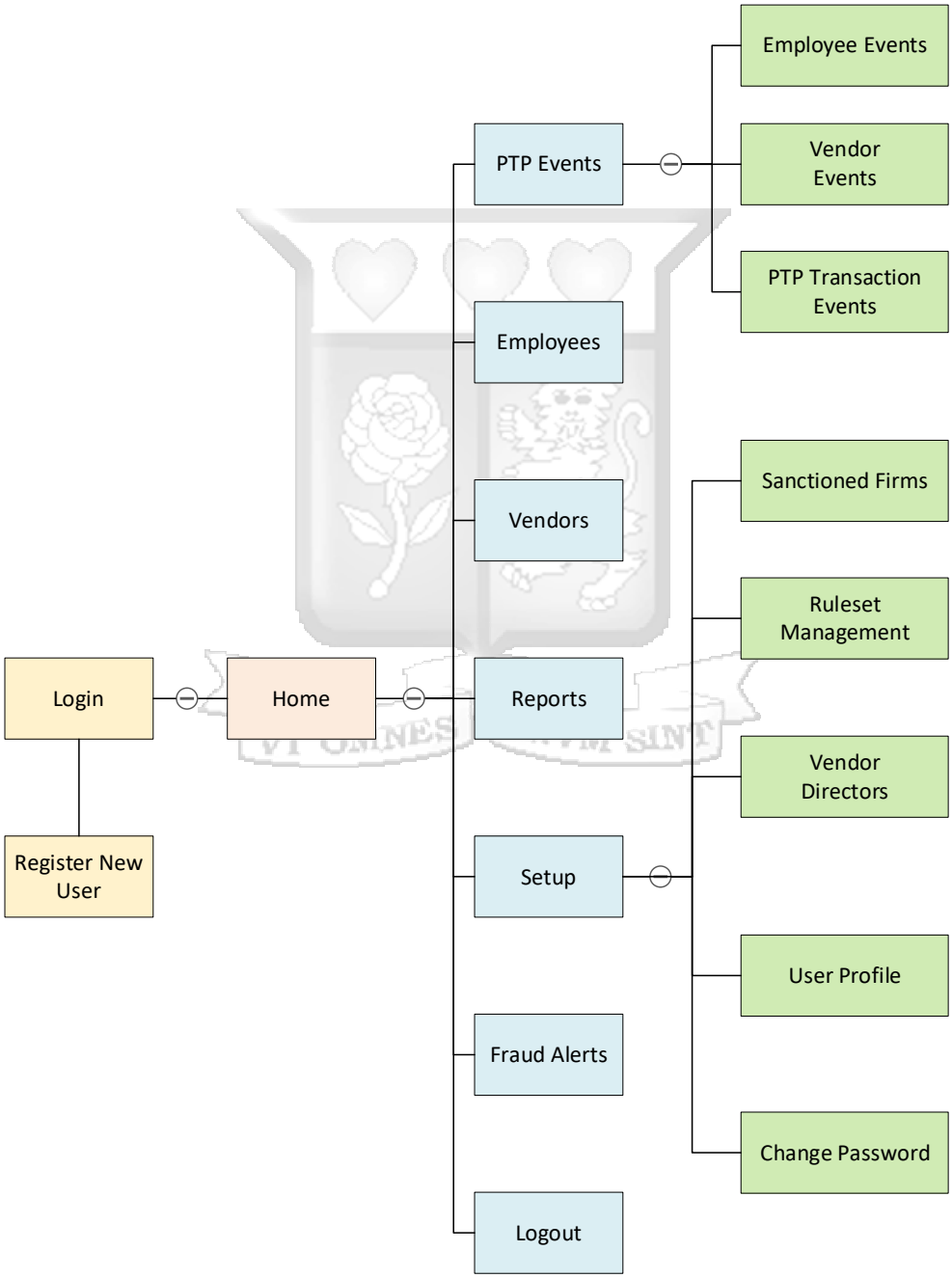


Figure 4.12: System sitemap

4.8.2 Wireframes

To provide an interaction between the end users and the prototype, front-end user interfaces were designed using wireframes. Based on the designed wireframes the front-end user interfaces were be built. Figures 4.13, 4.14 and 4.15 shows some wireframes that were designed for the user interfaces.

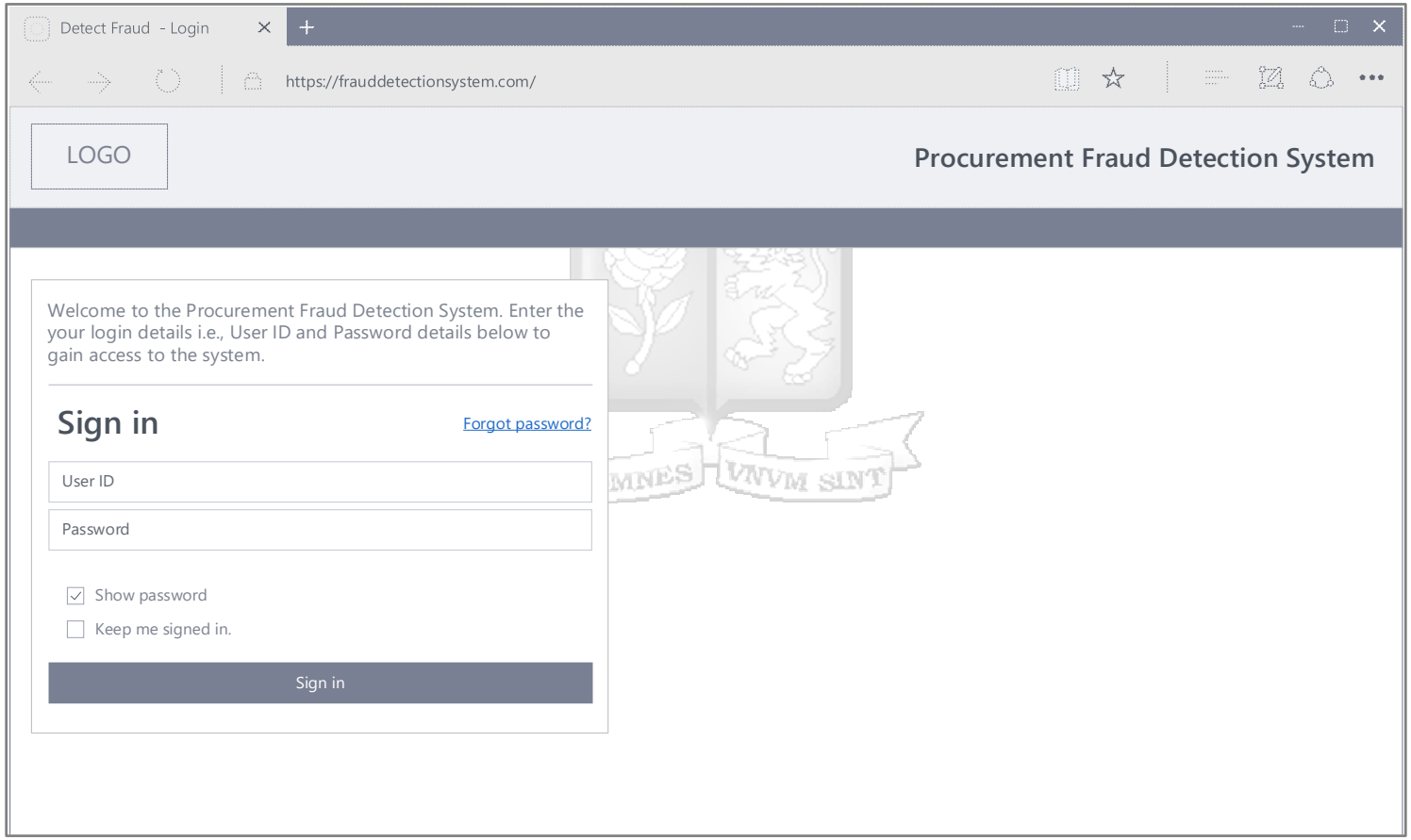


Figure 4.13: Wireframe for the Login Page

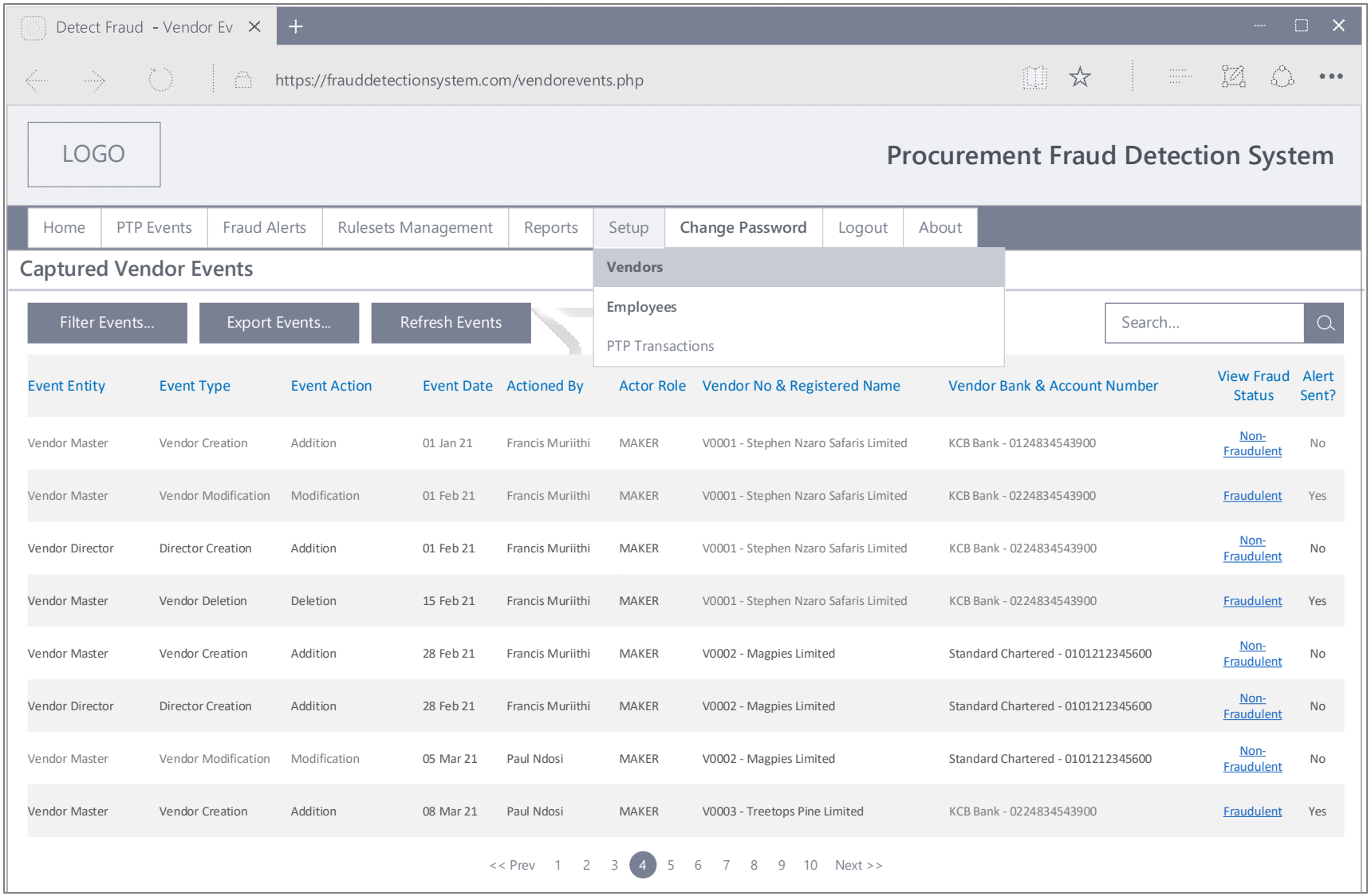


Figure 4.14: Wireframe for the Events Page

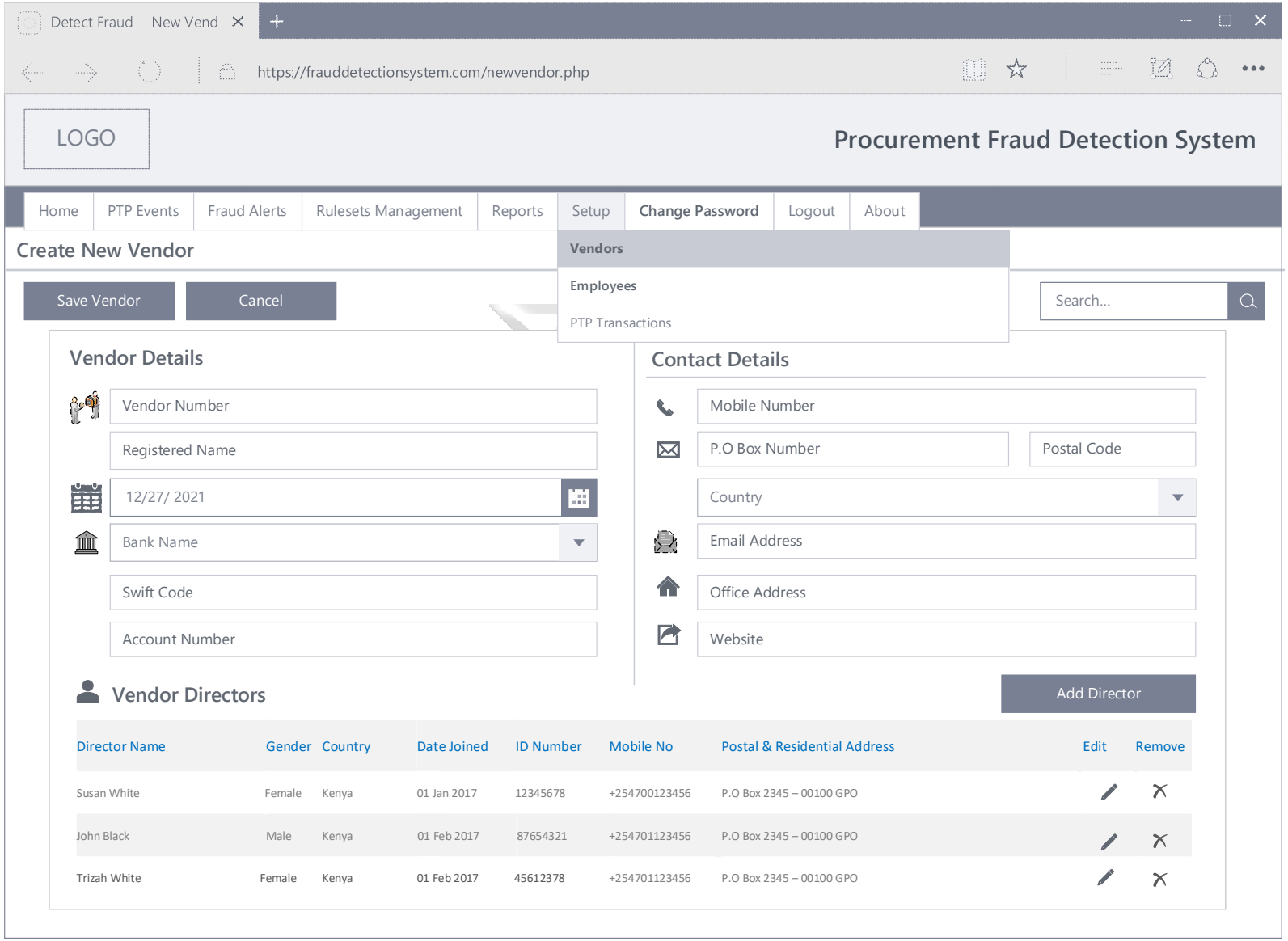


Figure 4.15: Wireframe for the Create New Vendor Page

Chapter 5: System Implementation and Testing

5.1. Introduction

Using the user requirements gathered and Design Document Specification (DDS) documented in Chapter 4, the prototype was developed, implemented, and tested. The prototype was developed in line with the **three-tier client server model** documented in the technical architecture using the following technologies: SQLite database, Django, Django REST framework, Pandas, React JS, and Material UI. Integration Testing, System Testing and User Acceptance Testing were performed to test the functionality and readiness of the prototype.

5.2. System Implementation

5.2.1 Implementation of the Data Management Tier

Using the Database Schema documented in Chapter 4, the database, data tables, field attributes and indexes were created and implemented in the SQLite database. This was achieved using **data dictionaries** and **SQL Data Definition Language (DDL).** The SQLite database that was used by the prototype to store employees, vendors, PTP transactions, PTP events, fraud results and fraud alerts data was hosted locally on a PC.

i. Data Dictionaries

Data dictionaries were documented for each data table in the database. The data dictionaries highlighted the attributes, data type, maximum length, null values indicator, and additional notes. It is from these data dictionaries that the table structures were implemented. Tables 5.1, and 5.2 show data dictionaries that were documented for some tables in the database.

Table 5.1: Data Dictionary for the PTP Event Table

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	EventID	bigint	8	False	Primary Key
FK	PTPTrxID	bigint	8	True	FK: PTPTransationHeader
FK	EmployeeID	bigint	8	True	FK: Employee

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
FK	VendorID	bigint	8	True	FK: Vendor
	EventEntity	Nvarchar (32)	64	False	
	EventType	Nvarchar (32)	64	False	Employee | Vendor | LPO |
	EventAction	Nvarchar (32)	64	False	Create | Modify | Delete
	EventDate	datetime	8	False	
	EventActorRole	Nvarchar (8)	16	False	Maker | Checker
	FraudStatus	Nvarchar (8)	16	False	Fraudulent | Non-Fraudulent
	AlertSent	tinyint	1	True	

Table 5.2: Data Dictionary for the Fraud Ruleset Table

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	RuleID	bigint	8	False	Primary Key
FK	AnalystID	bigint	8	False	FK: FraudAnalystUser
	RuleType	Nvarchar (50)	100	False	Risk Profile | Vendor & Employee Rules
	RuleName	Nvarchar (50)	100	False	
	RuleOperator	Nvarchar (50)	100	False	
	ResultDescription	Nvarchar (255)	510	False	
	RuleValue	float	8	False	
	RuleWeighting	float	8	False	
	IsActive	tinyint	1	False	

ii. Implementing Data Tables, Field Attributes, and Indexes

Once the data dictionaries were documented, the data tables, field attributes and indexes were implemented using **SQL Data Definition Language (DDL)** in the SQLite database. Appendix H shows the SQL data definition language that was used to create the data tables, their attributes, and indexes.

5.2.2 Implementation of the Business Logic (Middleware) Tier

To develop the business logic tier, **pseudo-code** and **flowcharts techniques** were used to document the program logic for the prototype. The pseudo-code and flowcharts were implemented using the **Python programming language** to develop the prototype.

Using the user requirements gathered, the processing specification was developed and implemented in the business logic (middleware) tier. The fraud rulesets and data mining algorithms that the prototype relied on to detect potential fraud were implemented in this tier as part of the prototype's **inference engine**. Over 80 fraud rules relating to employees and vendors were created and fed to the inference engine. Each rule was assigned a score that varied based on the interviews conducted. For example, the rule matching bank details between employees and vendors was given a higher priority compared to the rule checking for events posted after working hours.

Further, the prototype created events for the employee, vendor and PTP transaction records that were created, modified, and deleted directly in the HR & ERP source systems by the HR & Procurement Staff respectively. These events (i.e., creations, modifications, and deletions) were subjected to real-time data mining algorithms that determined whether to flag the event as either non-fraudulent, possible fraudulent or fraudulent. The algorithms applied included rule-based analytics, exact matches, fuzzy matches, and outlier analysis.

Python's FuzzyWuzzy library was used to develop the fuzzy string-matching fraud tests by calculating the Levenshtein Distance similarity ratio between two strings such as employee-vendor name matching test. In the fuzzy matching algorithm, strings were tokenized and pre-processed by converting to lower case and getting rid of punctuations. Thereafter, a set operation was performed to take out common tokens. As a result, extra or repeated words did not matter in the calculation. Figure 5.1 illustrates the pseudo-code and Figure 5.2 the flowchart program logic for an employee related event once it is captured into the prototype. Appendix I.1 shows the Python code snippet that implemented the fuzzy string matching and rule-based data mining algorithm and Appendix I.2 shows the Python code snippet that implemented z-score outlier analysis.

The **HTTP request handler** was also implemented in this tier, and it was responsible for every data exchange involving a request from the presentation tier and a response from the data

management tier. The request handler was then mapped to a specific URL so that HTTP requests and responses could be channelled through that URL.

Seq.	Pseudo-code	Notes
1.	START	
2.	CAPTURE Data from HR AND ERP Systems as Events	
3.	Determine the Event Source	E.g., Employee, Vendor, etc.
4.	IF Event Source is Employee:	
5.	Determine the Event Action	E.g., Create, Modify, Delete
6.	IF Event Action = 'Create' OR 'Modify':	
6a.	RETRIEVE Fraud Ruleset Values and Sanctioned Firms	
6b.	CALCULATE Employee Fraud Risk Profile	Use Employee Attributes (e.g., age, tenure, gender, etc.)
6c.	EXECUTE Rule-Based AND Fuzzy Fraud Tests	
6d.	IF Fraud Test Results = 'Negative':	
6di.	Flag Event as 'Non-Fraudulent'	
6e.	IF Fraud Test Results = 'Positive':	
6ei.	Determine the Fraud Result Score	
6f.	IF Fraud Result Score > 80%:	Levenshtein Distance greater than 80% represents a very close resemblance based on data tests performed. For example, Mobile No: 0721-852044 and Mobile No: 0777-852044 leads to a Levenshtein Distance of 80% which would warrant an investigation to reconfirm.
6fi.	Flag Event as 'Fraudulent'	
6fii.	UPDATE Employee Risk Profile Score to 99.00%	
6fiii.	SEND Fraud Alert to Fraud Analyst	
6fiv.	SAVE Fraud Alert Sent	
6g.	IF Fraud Result Score >= 50% AND < 80%:	
6gi.	Flag Event as 'Possible Fraudulent'	
6h.	IF Fraud Result Score < 50%:	
6hi.	Frag Event as 'Non-Fraudulent'	
7.	IF Event Action = 'Delete':	Delete Events are automatically flagged as 'Fraudulent' based on interviews conducted and business rules.
7a.	Flag Event as 'Fraudulent'	
7b.	SEND Fraud Alert to Fraud Analyst	
7c.	SAVE Fraud Alert Sent	
8.	SAVE the Event AND Fraud Status	
9.	SAVE the Positive AND Negative Test Results	
10.	END	

Figure 5.1: Program Flow Pseudo-code for an Employee Related Event

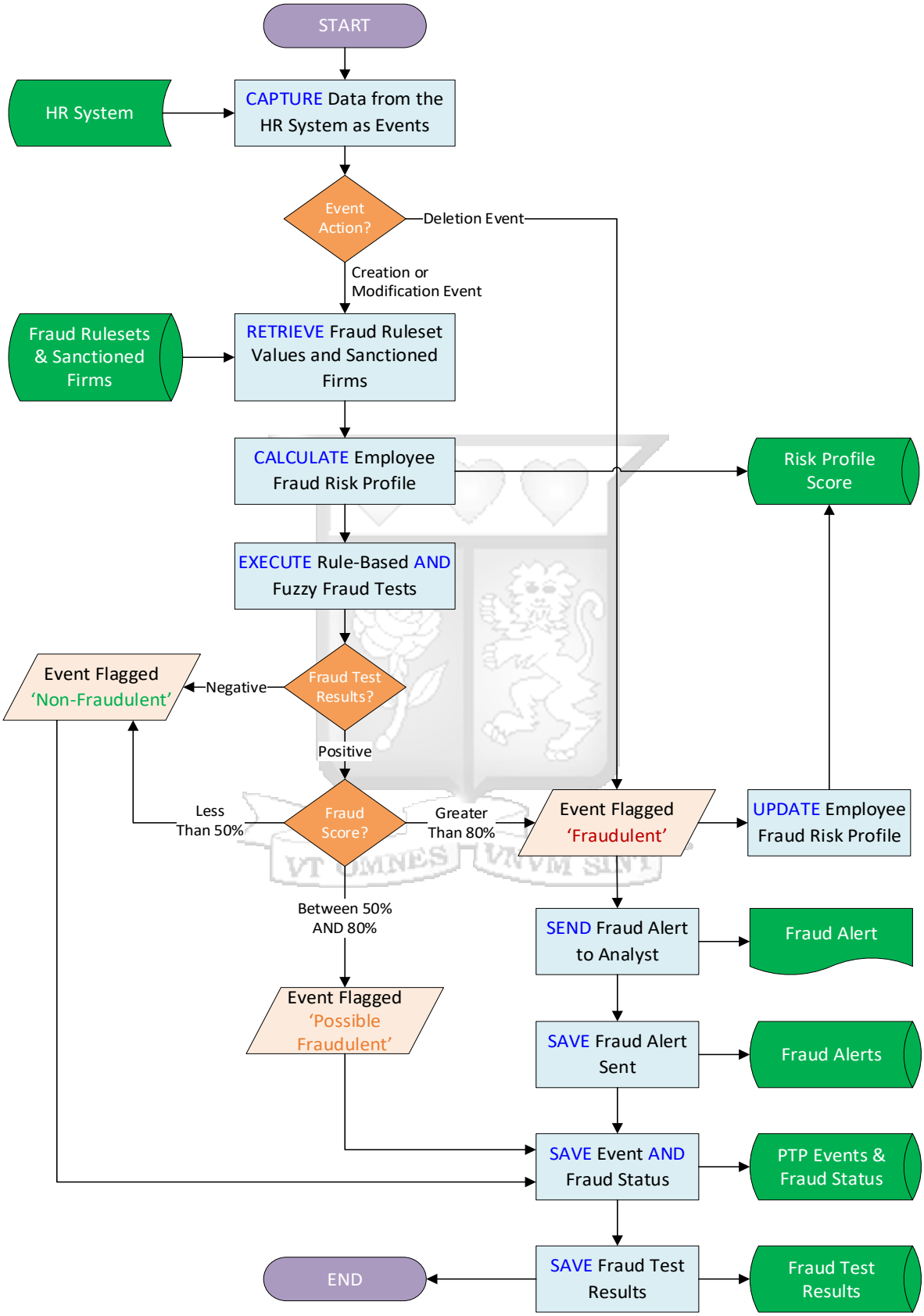


Figure 5.2: Program Flowchart for an Employee Related Event

The following are examples of the data mining fraud rules that were implemented in the inference engine within the Business Logic (Middleware) Tier:

i. **Vendor Bank Details Test:** This fraud test uses exact matching algorithm to check for scenarios where an employee's bank details match the bank details for a vendor.

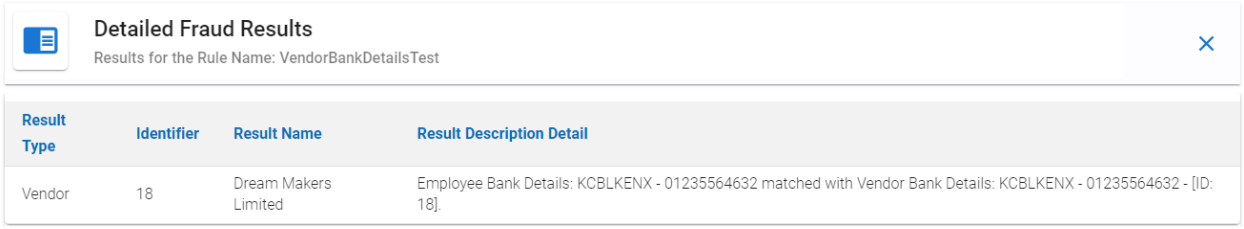


Figure 5.3: Vendor Bank Details Test

ii. **Vendor Director Mobile Number Fuzzy Test:** This fraud test uses fuzzy string-matching algorithm to check for scenarios where an employee's mobile number has a close similarity or resemblance with the mobile number for a vendor director.

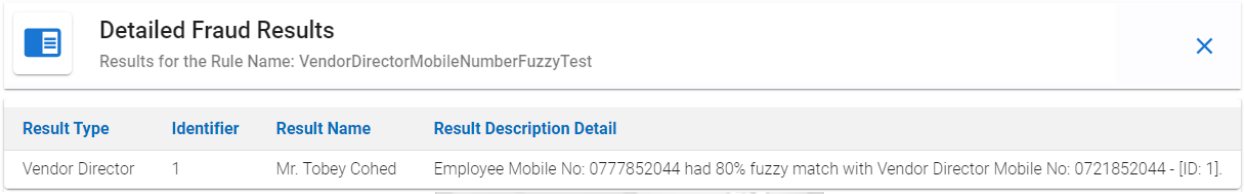


Figure 5.4: Vendor Director Mobile Number Fuzzy Test

iii. **Sanctioned Firm Name Fuzzy Test:** This fraud test uses fuzzy string-matching algorithm to check for scenarios where a registered vendor has a close similarity or resemblance with an entity in the World Bank list of ineligible firms and individuals.

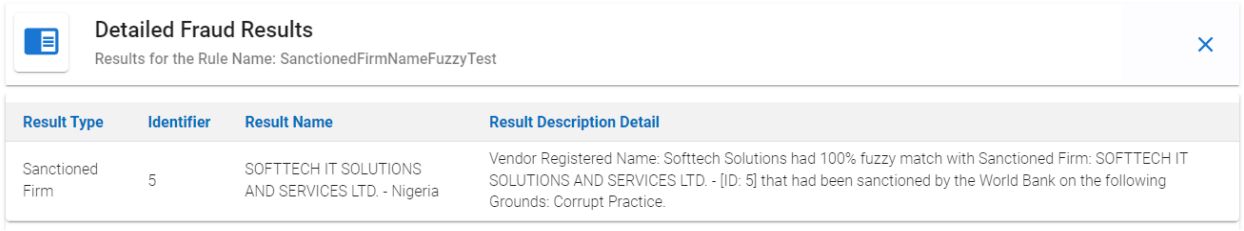
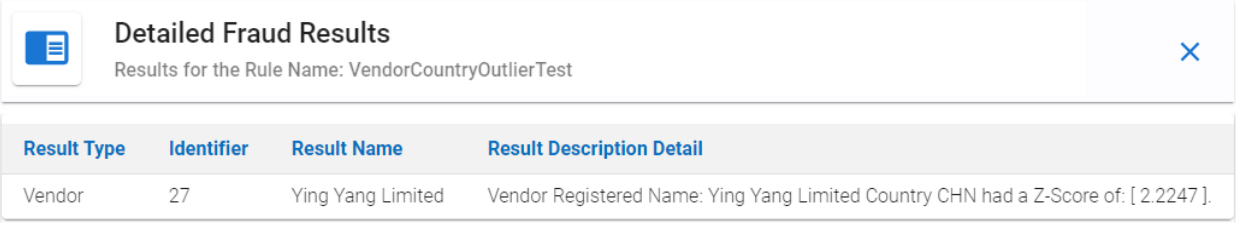


Figure 5.5: Sanctioned Firm Name Fuzzy Test

iv. **Vendor Country Outlier Test:** This fraud test uses outlier algorithm to calculate the z-score for the vendor's country and compare with the z-score for other vendors. The test then flags z-scores for countries that are greater than 2.0 (> 2.0) and less than -2.0 (< -2.0).



v. **Vendor Duplicate Bank Details Test:** This fraud test uses duplicate detection to check for scenarios where the bank details for a registered vendor is shared with other vendors.

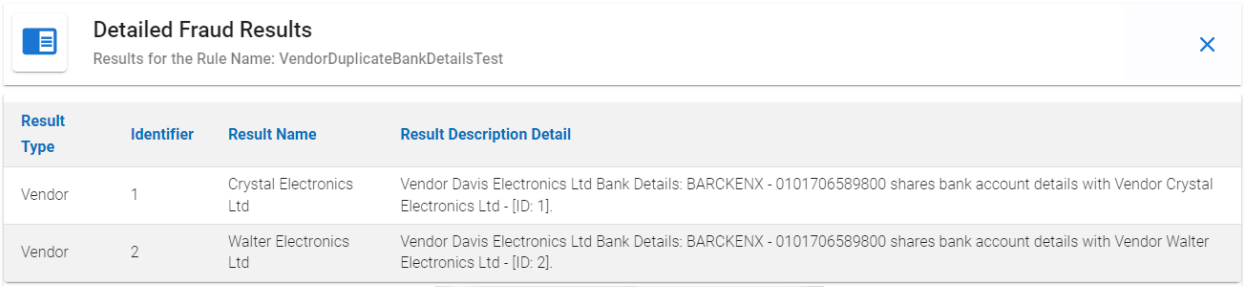


Figure 5.6: Vendor Duplicate Bank Details Test

vi. **Posted On Sunday Test:** This fraud test checks for scenarios where the related Employee Event was created or modified on a Sunday (i.e., a non-working day).

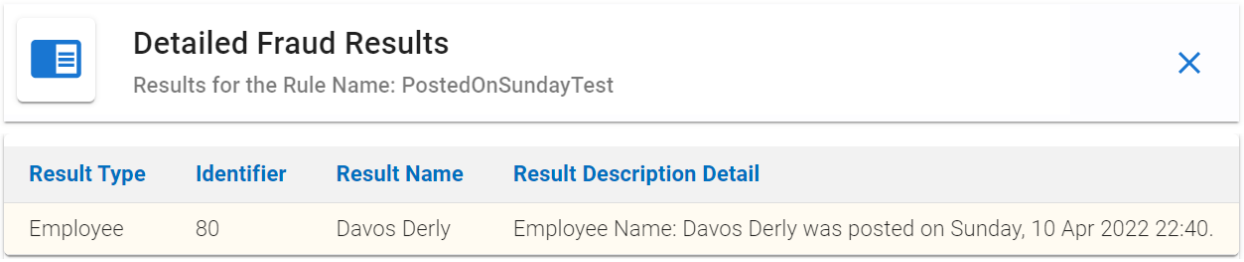


Figure 5.7: Posted On Sunday Test

vii. **Sanctioned Firm Postal Address Fuzzy Test:** This fraud test checks for scenarios where the vendor's postal address has a close similarity with a World Bank Sanctioned entity.

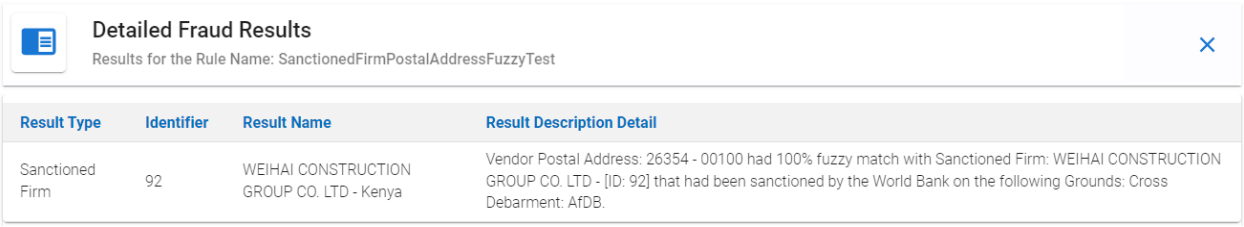


Figure 5.8: Sanctioned Firm Postal Address Fuzzy Test

5.2.3 Implementation of the Presentation Tier

Using the site map and wireframes documented in Chapter 4, user interfaces were developed using **React JS** and **Material UI**. These interfaces provided a user-friendly way for the user to access, navigate and interact with the prototype. Each interface presented a unique function provided by the prototype such as capturing, modifying, searching, and viewing of data records.

i. Employee Details User Interface

This interface displays the overall calculate fraud risk profile score for the employee based on their attributes. The overall risk profile scores are colour-coded using the Red-Amber-Green (RAG) colour notations for ease of interpretation by the end user. An overall risk profile score below 50% is highlighted in 'green', a score between 50% and 80% is highlighted in 'amber' and a score greater than 80% is highlighted in 'red'. In addition, this interface displays the Employee's Bio Details, Contact Details, Bank Details, HR Details and Next of Kin Details. Figure 5.9 shows the user interface that displays the employee details to the user.

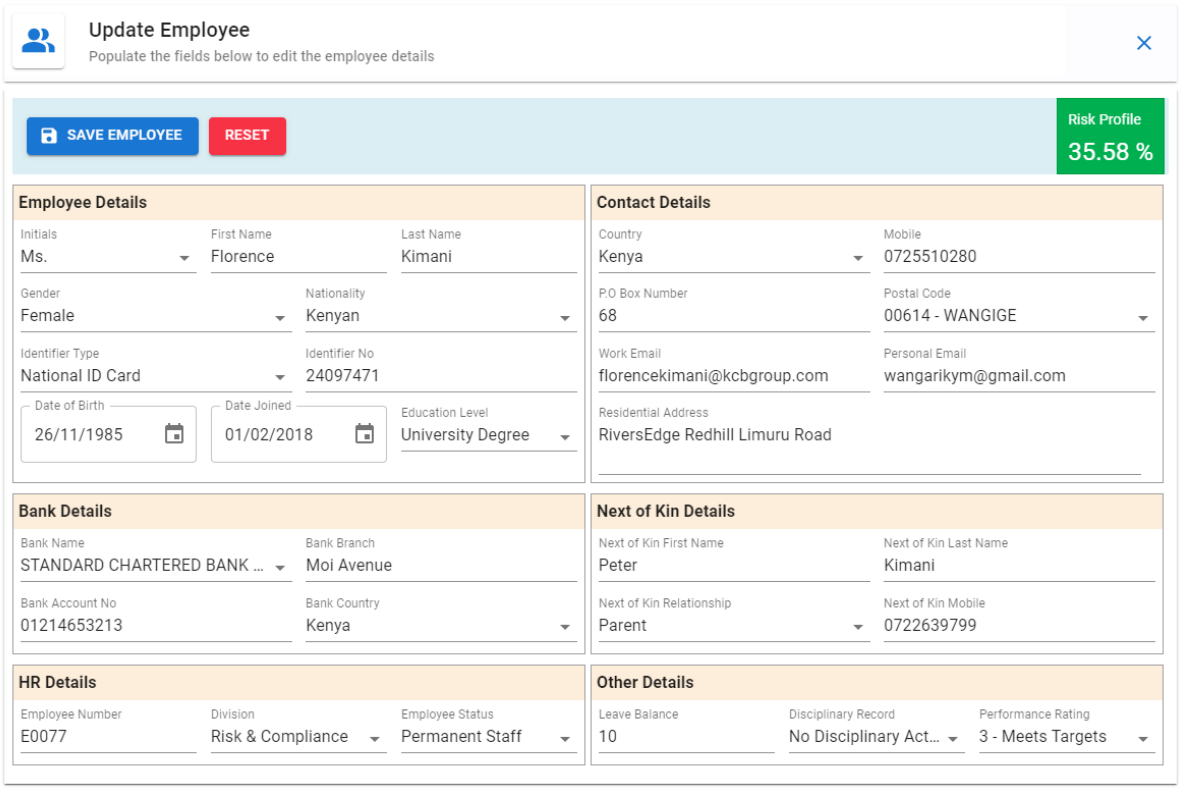
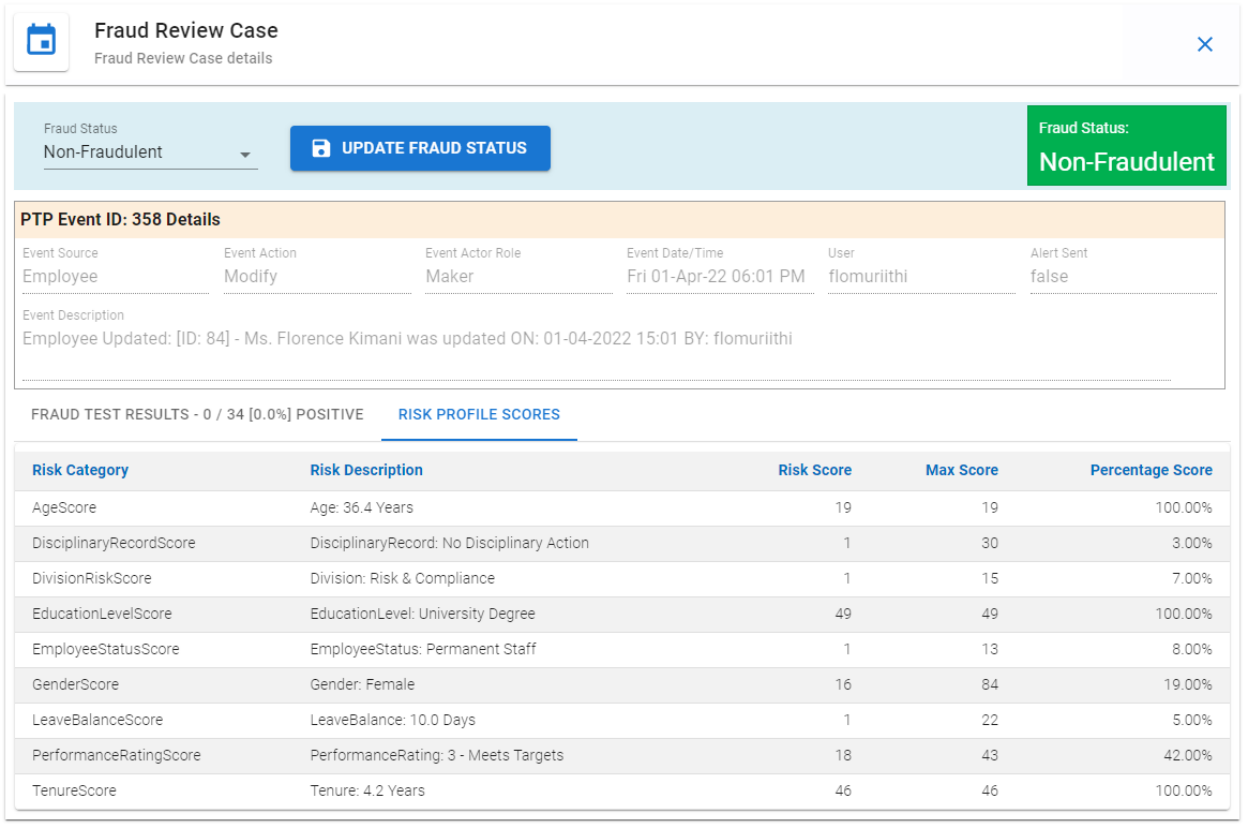


Figure 5.9: Interface Displaying Employee Details and Overall Risk Profile Score

ii. Employee Risk Profile Scores User Interface

This interface displays how the employee's risk profile score was calculated based on the various risk categories. For each employee, **nine (9) risk category scores** are calculated and totalled. These risk categories and scores are based on the secondary data collected in Chapter 4 and they include: age, disciplinary record, division/department, education level, employee status, gender, leave balance, performance rating and employment tenure factors. The risk profile information is available to the fraud analysts for awareness purposes as per their user requirements.

In addition, this interface displays the details of the specific PTP Event i.e., Event Source, Event Action, Event Actor Role, Event Date / Time, User, Event Description and Fraud Status for the particular Event. Figure 5.10 shows the user interface that displays how the employee's risk profile scores were calculated and the fraud status of the event.



Risk Category	Risk Description	Risk Score	Max Score	Percentage Score
AgeScore	Age: 36.4 Years	19	19	100.00%
DisciplinaryRecordScore	DisciplinaryRecord: No Disciplinary Action	1	30	3.00%
DivisionRiskScore	Division: Risk & Compliance	1	15	7.00%
EducationLevelScore	EducationLevel: University Degree	49	49	100.00%
EmployeeStatusScore	EmployeeStatus: Permanent Staff	1	13	8.00%
GenderScore	Gender: Female	16	84	19.00%
LeaveBalanceScore	LeaveBalance: 10.0 Days	1	22	5.00%
PerformanceRatingScore	PerformanceRating: 3 - Meets Targets	18	43	42.00%
TenureScore	Tenure: 4.2 Years	46	46	100.00%

Figure 5.10: Interface Displaying Employee Risk Profile Scores

iii. Employee Fraud Test Results User Interface

The interface in Figure 5.11, displays details of the specific PTP Event e.g., a modification of an employee event. For this modification event, the interface displays the fraud status of the event which is colour-coded using the Red-Amber-Green (RAG) notation for ease of interpretation by the end user. If an event is flagged, fraudulent, it is highlighted in 'red'. Events flagged as possible fraudulent are highlighted in 'orange' and events flagged non-fraudulent are highlighted in 'green'.

In addition, this interface displays to the end user the fraud test results for the employee related event by enumerating the fraud tests that were performed and their respective results i.e., Positive or Negative. Figure 5.11 shows that the employee modification event had 9 out of 34 (i.e., 26%) fraud tests with positive results resulting to the fraudulent fraud status.

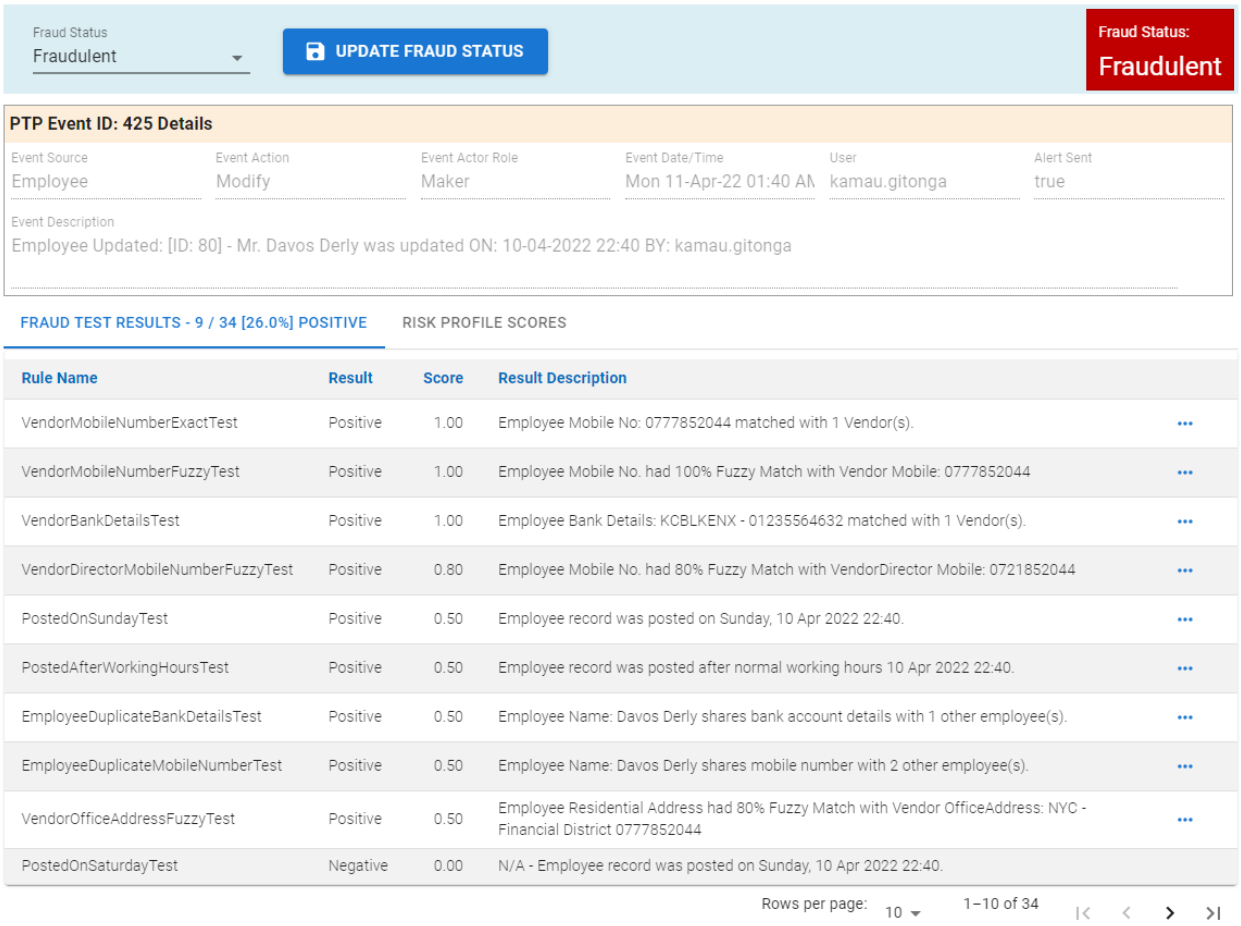


Figure 5.11: Interface Displaying Employee Fraud Tests Performed and Respective Results

iv. **Fraud Alerts User Interface**

Figure 5.12 shows the user interface that displays the fraud alerts that have been sent to the fraud analysts for their action and investigation. Figure 5.13 shows the actual email message notification that was received by the fraud analyst.

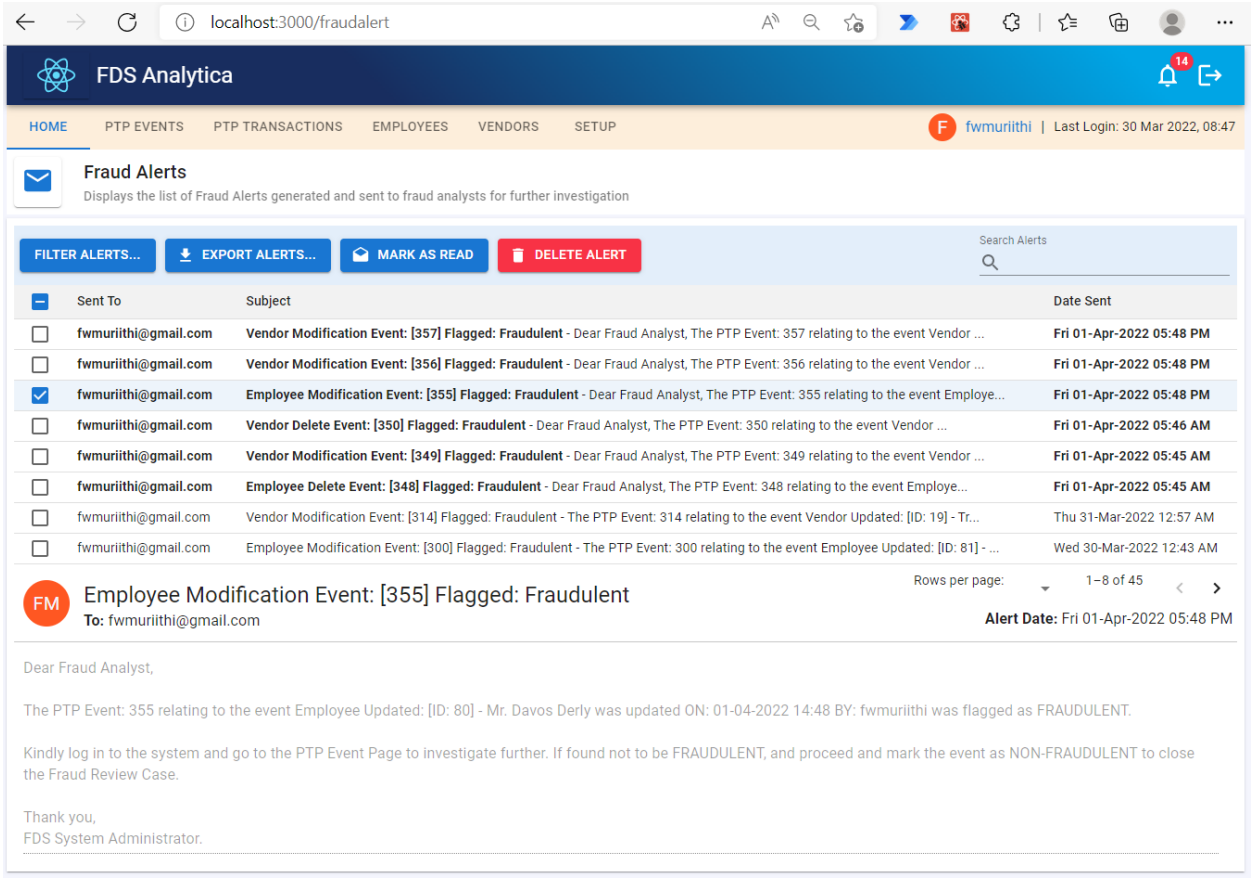


Figure 5.12: Interface Displaying Fraud Alerts Sent to Analysts

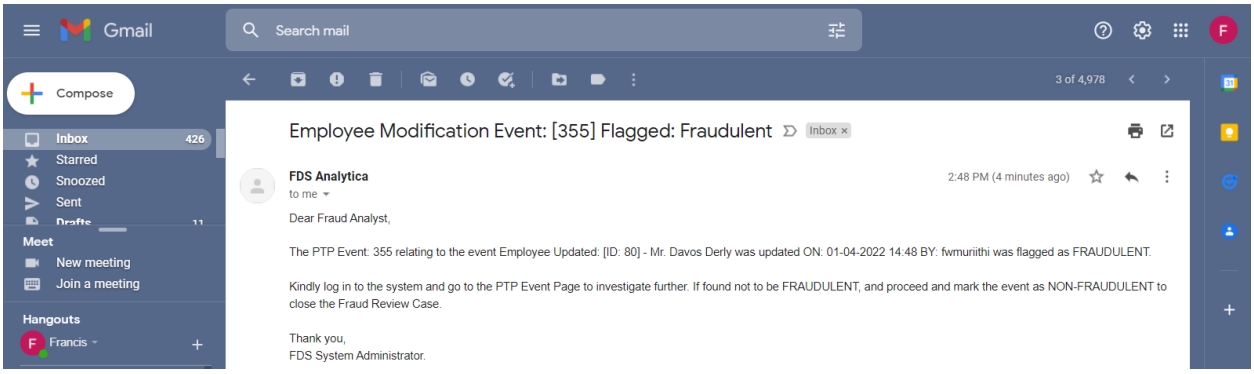


Figure 5.13: Interface Displaying Fraud Alert Email Received by the Analyst

v. Fraud Rulesets and Sanctioned Firms User Interface

This interface displays the fraud rulesets and sanctioned firms listing that the prototype is relying on to execute the fraud tests to determine whether an event is fraudulent, possible fraudulent or non-fraudulent. This interface provides the user the option to modify the weightings and thresholds for each rule according to changing business needs and environment. Figure 5.14 illustrates the fraud rulesets user interface and Figure 5.15 shows the sanctioned firms listing.

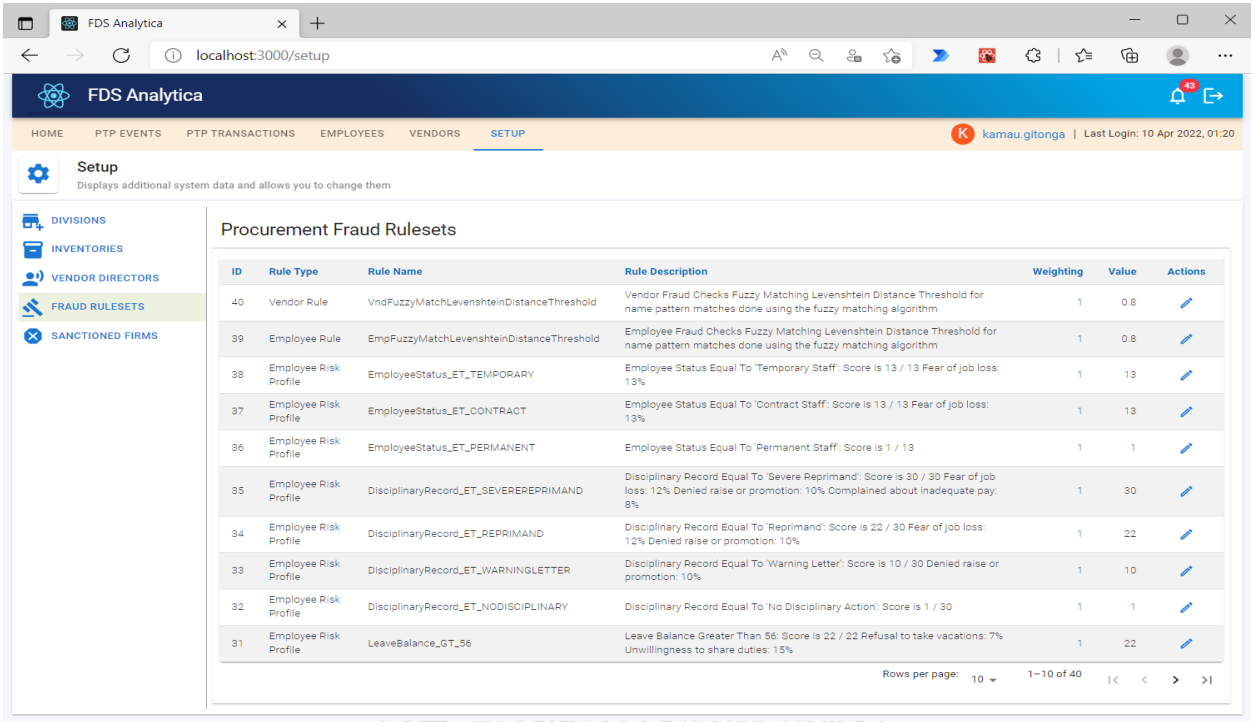


Figure 5.14: Interface Displaying Fraud Rulesets

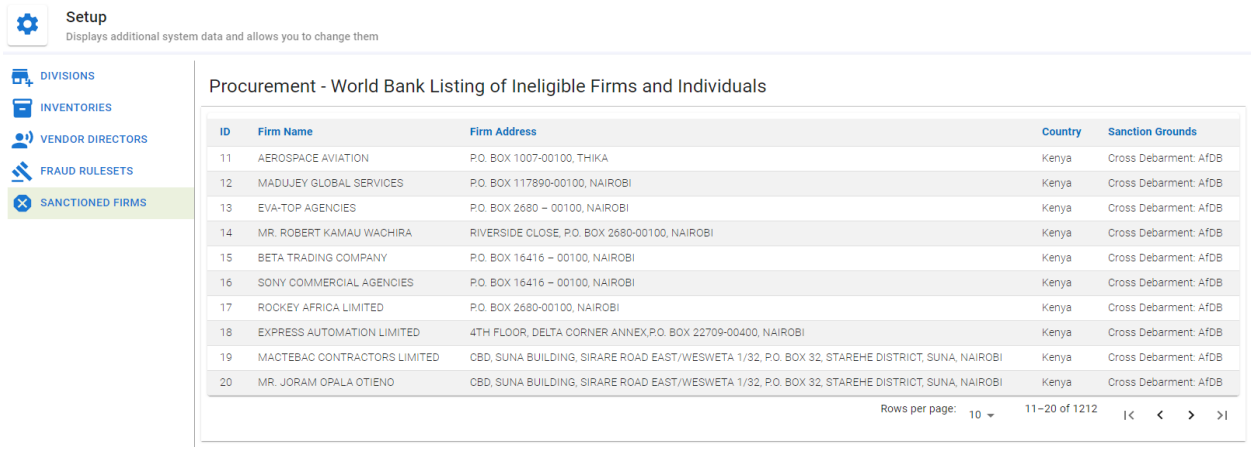


Figure 5.15: Interface Displaying the Sanctioned Firms and Individuals

5.2.4 Security Layer Implementation

Using the security design documented in Chapter 4, several security features were built and incorporated into the prototype across the presentation, business logic and data management tiers. These features aided in ensuring that the personal data fed into the prototype remained secure.

i. Login and URL Routes Authentication Security

The login capability was built in to allow only authenticated users had access to the system. This ensured that private URLs (e.g., Home Page, Employees Page, etc.) would only be accessible once the user had logged in. If a user tried to access any private URL, the system would automatically re-direct them back to the Login Page, (a public URL). Figure 5.16 shows the login page that the user is welcomed with when they first access the prototype. Appendix I.3 shows the React JS source code snippet that implemented the URL Routes authentication security feature.

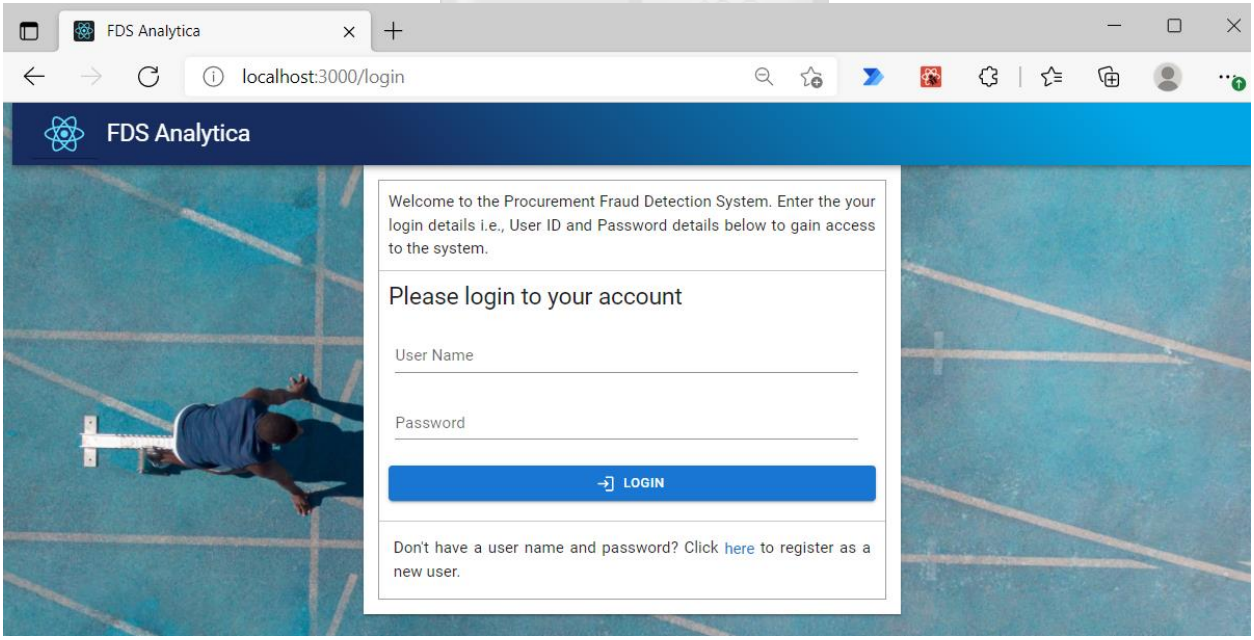


Figure 5.16: Login Page

ii. Password Hashing and Salting

The user login credentials i.e., passwords were hashed using the Secure Hash Algorithm (SHA) and stored in an irreversible SHA256 hash in the database. The system utilized a specialized password hashing function i.e., PBKDF2 algorithm that created the SHA256 hashes using salts. Figure 5.17 shows the how the hashed user passwords have been stored in the database.

id	password	last_login	is_superuser	username
1	pbkdf2_sha256$320000$hWvYnl7fdDxqlxfkK5uPa0$S84MDapVFBpk9TzkyUWiJnRee9ltg+ai5H0j4ifnDws=	2022-03-13 11:33:08.789603	1	admin
5	pbkdf2_sha256$320000$AeoobxcMIIIGkEZhfIHiAR$nx450Mq0LL8JQEpdA7Y1XSTZxdnguceTLkv7AMKTedM=	2022-03-06 09:37:16.605079	1	paco
6	pbkdf2_sha256$320000$qh4shImPFpeBKQu4DsCBkJ$Q/H5fx+U1QnQVdv+noN+4btHjWriz7FGTNe27aVpKng=	2022-03-13 17:27:17.948803	1	fwmuriithi
14	pbkdf2_sha256$320000$r8Si50Mqr9A3TX15904us6$2FIrnGzkEN/BFt5Hf15VKLR45pTK8615fCIkaTJHDJ8=	2022-02-26 11:26:21.120528	0	dkamau
22	pbkdf2_sha256$320000$5OvaUfQPVFAVFQQey7DlPW$mOU9XMH+xgaFWGJD20opS8vGNjQ7O6hdplsnRs/...	2022-02-26 20:54:31.203835	0	tracy
25	pbkdf2_sha256$320000$uj4lGxwNpG6X49BigBj2wz$T3DT9N7Xmr0DV918ul/gOJ1ZSOEH3DN5uCg1bn1Rjyg=	2022-03-08 17:39:49.856423	0	JTerry
26	pbkdf2_sha256$320000$ubrqWyLzpbFr8DnpvTQfUK$15RwWDJCf4ndmf2mrRNG3Znja7FysSjNixEVNCMvvto=	2022-03-10 12:44:59.394965	0	wangari

Figure 5.17: User Passwords Hashed in the Database

iii. Token Authentication

The system implemented Token Authentication to safeguard the backend APIs. The system generated a unique token for each user during user registration. The client appends the user token in the Authorization Header when making requests to access the protected API resources. Figure 5.18 shows the unique tokens generated by the prototype for each user for their authentication. Appendix I.4 shows the code snippet that implemented the token authentication capability.

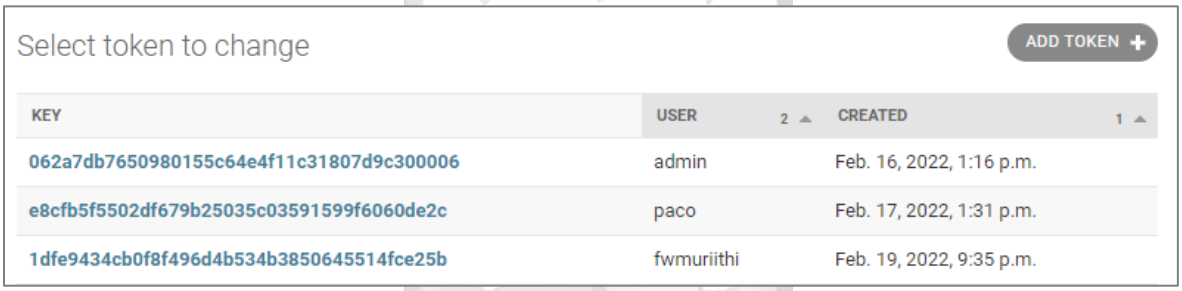


Figure 5.18: Unique Tokens Generated For Each User

iv. Password Policy

The system implemented a strong password policy. Figure 5.19 illustrates the password policy being enforced when a user is changing their password.

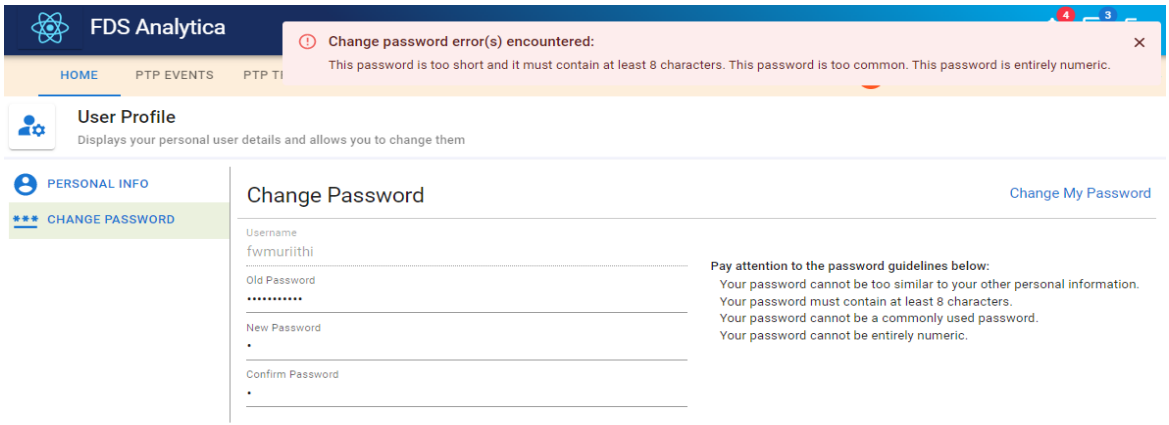


Figure 5.19: Password Policies Enforced by the System

5.3. System Testing

After developing the prototype, the prototype was tested to verify that the functional and non-functional requirements documented in Chapter 4 have been met. Testing was also done to find any defects and failures caused by bugs for remediation. The testing was carried out in two (2) testing levels i.e., **Integration Testing** and **System Testing** and their respective test cases were designed, developed, and executed. The test cases consisted of a set of input values, execution preconditions, expected results, and the actual test results.

5.3.1 Integration Testing

The main objective of the integration testing was to test interfaces and interactions with different parts of the prototype such as the interface between the presentation, business logic and data management tiers. The integration tests verifying communication of the Restful APIs with all the tiers documented in the Technical Architecture were performed using the Postman API Tool.

Postman is an API platform that includes a comprehensive set of tools that help accelerate the API lifecycle—from design, testing, documentation, mocking and discoverability of the APIs. It enables the exploring, debugging, and testing of APIs by allowing the tester to define API requests for HTTP, REST, SOAP, GraphQL, and WebSockets (Postman, Inc., 2022). Some of the API integration tests that were executed via Postman are shown in Table 5.3.

Table 5.3: Integration Testing Test Cases

ID	Test Scenario	Expected Results	Test Results
1.	**API Request Method:** GET **API Request URL:** http://127.0.0.1:8000/fraudalert/ **API Header Authorization:** Token [Valid user token key]	**Response Status:** 200 OK **Response Body:** Retrieve all 28 fraud alerts in the database.	**Pass** Refer to Figure 5.20
2.	**API Request Method:** GET **API Request URL:** http://127.0.0.1:8000/employee/ **API Header Authorization:** Token [Invalid OR No user token key provided]	**Response Status:** 401 Unauthorized **Response Body:** Detail: Invalid token. OR Detail: Authentication credentials were not provided.	**Pass** Refer to Figure 5.21

ID	Test Scenario	Expected Results	Test Results

3.	**API Request Method:** POST
	API Request URL:
	http://127.0.0.1:8000/vendor/
	API Header Authorization:
	Token [Valid user token key]
	API Body Request:
	{
	"RegisteredName": Davis Ltd",
	"DateRegistered":"2010-11-01",
	"SingleSourced":"Yes",
	"BoxNumber":"56871",
	"PostalCode":"00200",
	"Country":"KEN",
	"Mobile":"+254721585858",
	"Email":"info@daelec.com",
	"OfficeAddress":"Westlands",
	"BankName":"BARCKENX",
	"BankAccountNo":"01019800",
	"BankCountry":"KEN",
	"RiskProfile":"0"
	}
		Response Status:
		201 Created
		Response Body:
		Success: Vendor Record Created in Database
			Pass
			Refer to Figure 5.22

4.	**API Request Method:** PUT
	API Request URL:
	http://127.0.0.1:8000/vendor/
	API Header Authorization:
	Token [Valid user token key]
	API Body Request:
	{
	"VendorID":22,
	"RegisteredName": Travis Ltd",
	"VendorNumber":"V0012"
	}
		Response Status:
		201 Created
		Response Body:
		Success: Vendor Record Updated in Database
			Pass
			Refer to Figure 5.23

5.	**API Request Method:** DELETE **API Request URL:** http://127.0.0.1:8000/vendor/22 **API Header Authorization:** Token [Valid user token key]	**Response Status:** 200 OK **Response Body:** Success: Vendor ID 22 deleted in the database.	**Pass** Refer to Figure 5.24

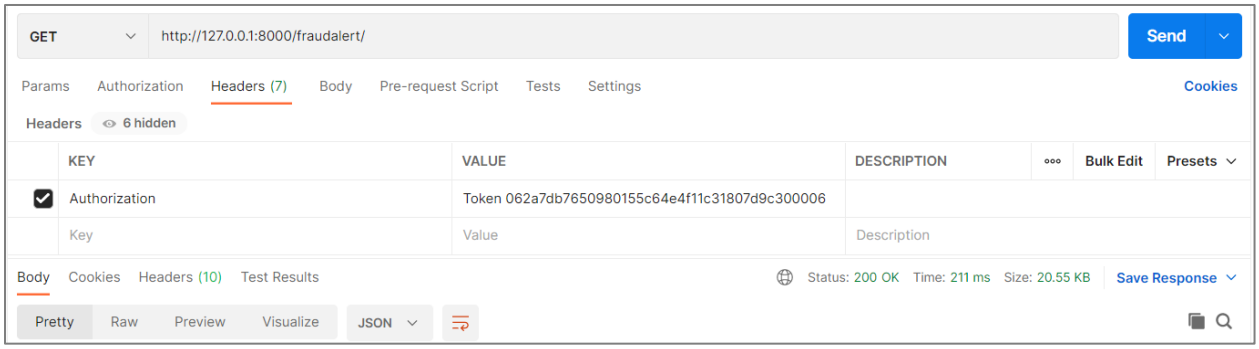


Figure 5.20: Postman GET Request Test Results for Fraud Alerts

Figure 5.21: Postman GET Request Without Token Provided

Figure 5.22: Postman POST Request Test Results for Creating Vendor

Figure 5.23: Postman PUT Request Test Results for Updating Vendor

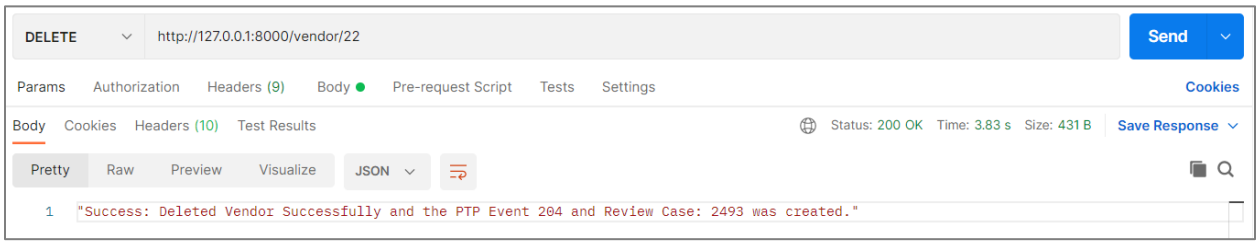


Figure 5.24: Postman DELETE Request Test Results for Deleting Vendor

5.3.2 System Testing

The main objective of the system testing was to test the behaviour of the whole prototype. This test investigated the functional requirements, non-functional requirements, and data quality characteristics of the prototype. The system testing of the requirements was performed using black-box techniques and the respective test results were documented in test cases. Some of the system tests that were executed are shown in Table 5.4.

Table 5.4: System Testing Test Cases

ID	Test Scenario	Expected Results	Test Results
1.0	**Login →** http://localhost:3000/login		
1.1	Username or Password not entered	Error message	**Pass**
1.2	Incorrect username or password	Error message	**Pass**
1.3	Navigate to Private URL Route without logging into the system	Redirect user to Login URL page	**Pass**
1.4	Valid username and password entered	Redirect user to Home URL	**Pass**
2.0	**Register New User Account →** http://localhost:3000/registeruser		
2.1	Fields with no details entered	Error message	**Pass**
2.2	Password and Confirm Password are different	Error message	**Pass**
2.3	Enter an existing username or email	Error message	**Pass**
2.4	Enter valid details in all fields (Name, Email, Username, Password and Confirm Password)	Create user and success message	**Pass**
3.0	**Change Password →** http://localhost:3000/userprofile		
3.1	Fields with no details entered	Error message	**Pass**
3.2	Incorrect old password	Error message	**Pass**
3.3	New Password and Confirm Password are different	Error message	**Pass**
4.0	**Fraud Review Case →** http://localhost:3000/ptpevent		
4.1	Delete Vendor	Flag Event as Fraudulent	**Pass** Refer to Figure 5.25
5.0	**Employee Risk Profile Calculation →** http://localhost:3000/employee		
5.1	Modify Employee with the following attributes: - **Gender:** Male	Employee successfully updated with risk	**Pass** Refer to Figure 5.26

ID	Test Scenario	Expected Results	Test Results
	- **Date of Birth:** 04 Sep 1983 - **Date Joined:** 04 Sep 2019 - **Education Level:** University Degree - **Division:** Shared Services - **Employment Status:** Contract Staff - **Leave Balance:** 60 days - **Disciplinary Record:** Severe Reprimand - **Performance Rating:** 1 – Does Not Meet	profile scores of 100% for each of the nine (9) risk factors.	

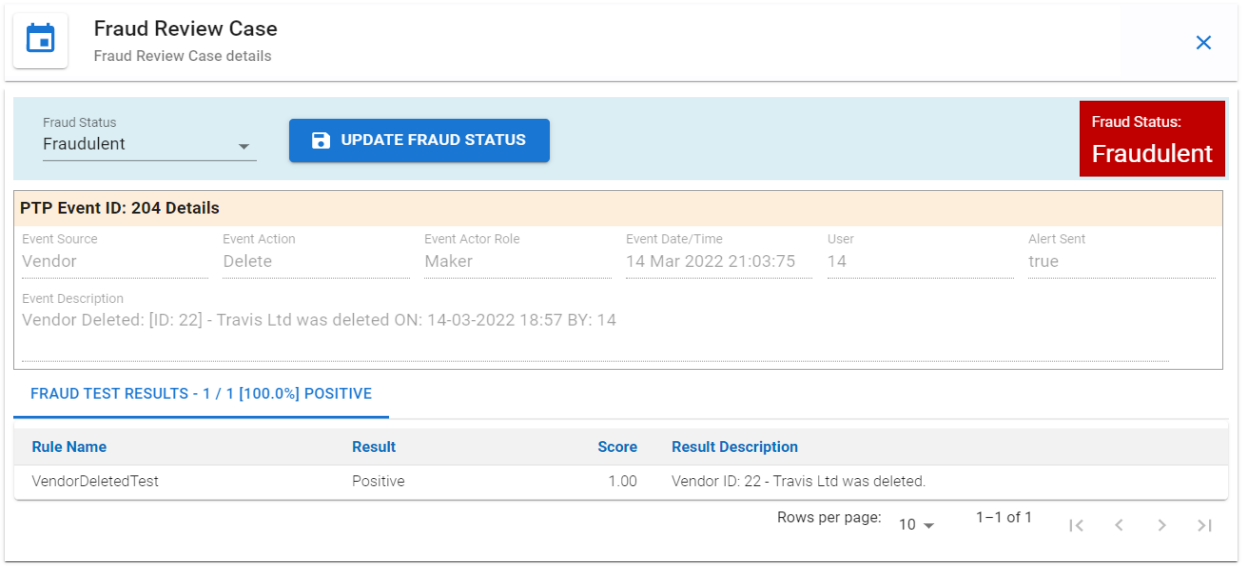


Figure 5.25: Deleted Vendor Event is Flagged Fraudulent

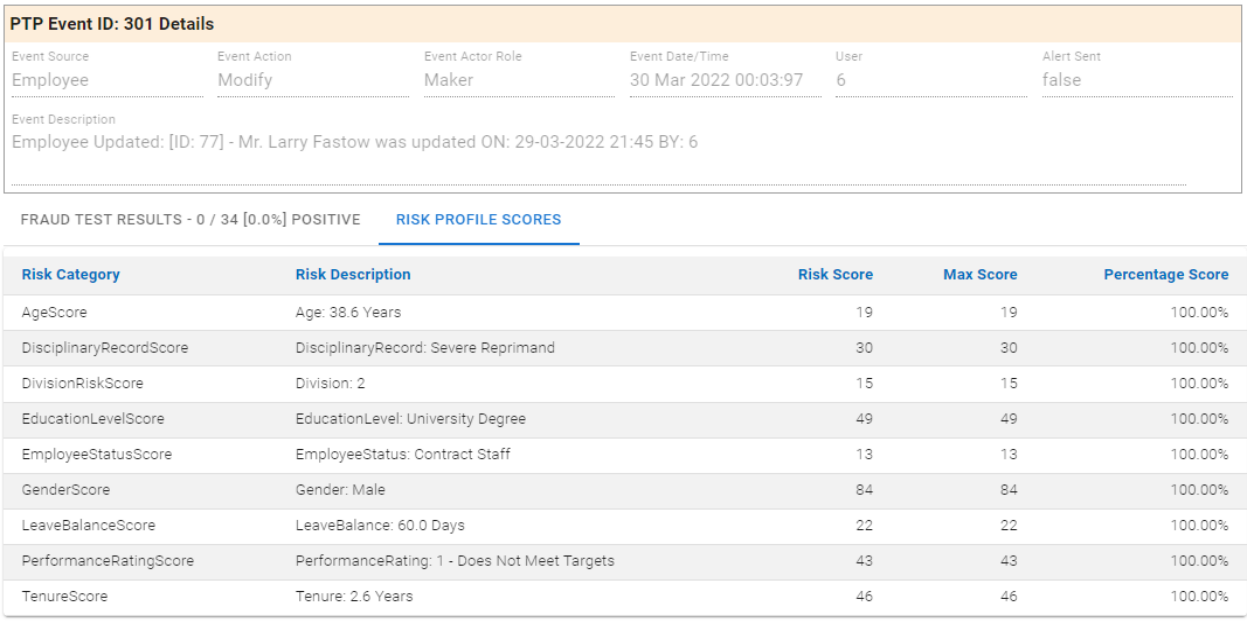


Figure 5.26: Employee with 100% on all risk factors

5.4. System Validation and Deployment

After performing integration and systems tests on the prototype, **User Acceptance Tests (UAT)** were conducted to validate and assess the prototype's readiness for deployment. The UATs were performed with the aid of an acceptance testing questionnaire (Appendix E) that was administered to same group of respondents that provided the user requirements. The questionnaire required the respondents to rate the prototype in the following areas: **functionality**, **usability**, **performance**, and **data security** by providing a score on a scale of 1 – 5, (where 5 represented strongly agree and 1 strongly disagree). The respondent scores were summarized using graphs.

5.4.1 Functionality and Usability Readiness

With regards to **functionality readiness**, respondents were required to assess the whether the prototype adequately met their user requirements and whether it generated accurate results with minimal false positives for the events analysed. The test revealed that 13% of the respondents strongly agreed, 63% agreed, and 25% were neutral. This is illustrated in Figure 5.27.

On **usability readiness,** the test revealed that 63% of the respondents strongly agreed that the user interface was user friendly, and that they were capable of using the prototype with minimum training given its intuitive and graphical user experience. 25% agreed and 13% were neutral regarding usability readiness of the prototype. This is illustrated in Figure 5.28.

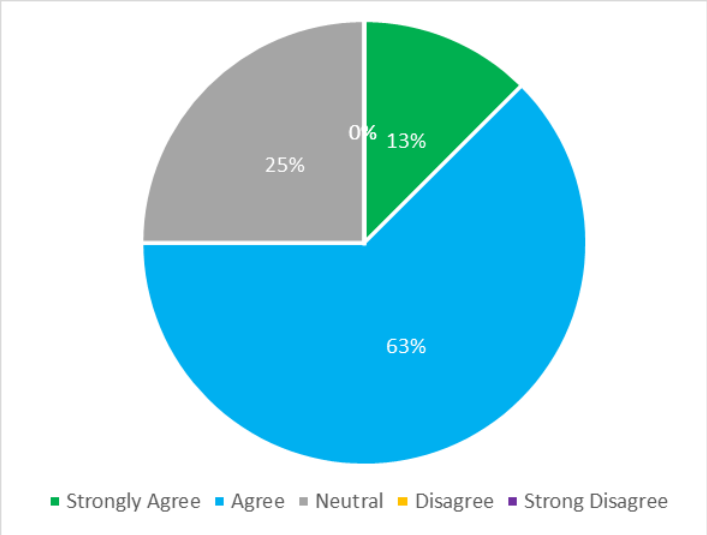


Figure 5.27: Validation of Functionality Readiness

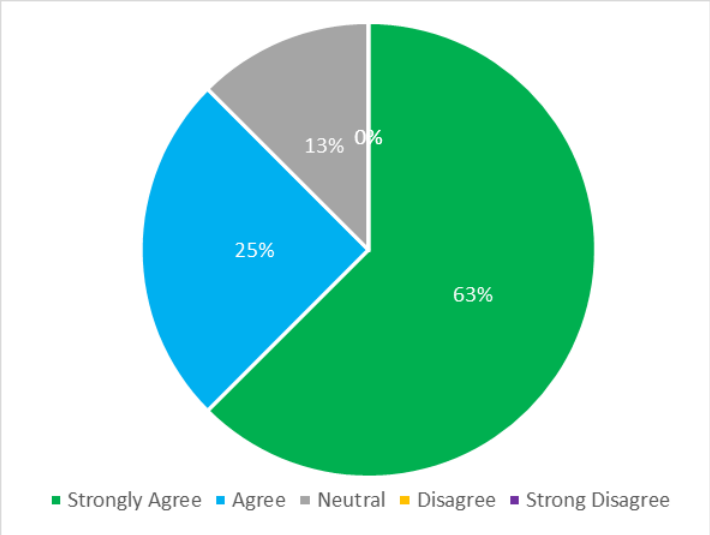


Figure 5.28: Validation of Usability Readiness

5.4.2 Performance and Data Security

With regards to **performance readiness**, respondents were required to assess the responsiveness of the prototype's user interface during their interactions with the prototype. In addition, the respondents were required to assess the robustness of the prototype by validating whether it handled system and data errors without crashing, hanging, or shutting down unexpectedly. The test revealed that 50% of the respondents agreed that the prototype was robust and its performance adequate, while 38% were neutral, and 13% disagreed. This is illustrated in Figure 5.29.

On **data security readiness,** respondents were required to assess whether confidentiality of data was maintained by restricting access to authorized users only. In addition, the respondents assessed whether the integrity of the data was preserved by the prototype during processing. The test revealed that 63% of the respondents agreed that the data security readiness was adequate and 38% were neutral. This is illustrated in Figure 5.30.

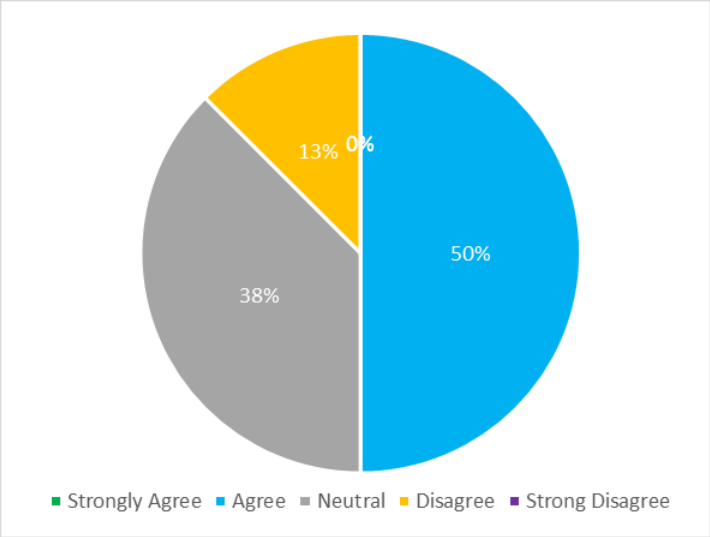


Figure 5.29: Validation of Performance and Scalability Readiness

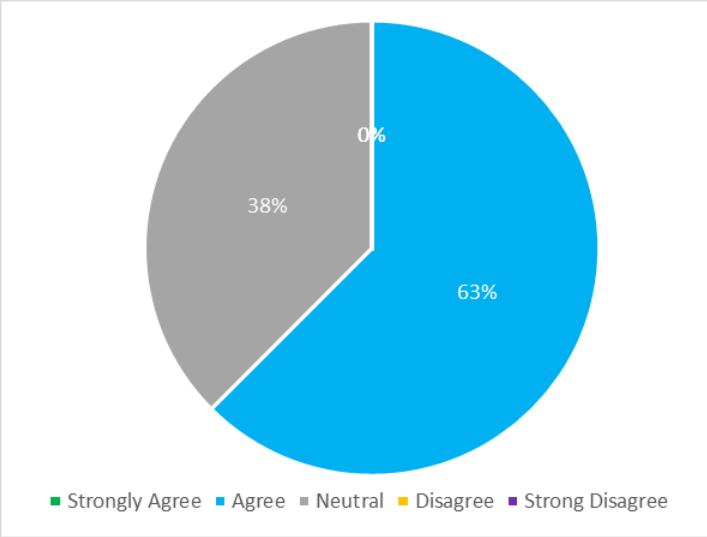


Figure 5.30: Validation of Data Security Readiness

Chapter 6: Discussion

6.1. Introduction

As fraud schemes become more sophisticated and the amount of data organizations produce continues to grow, it's becoming increasingly difficult to uncover procurement fraud using manual checks and balances. This research investigated procurement fraud in Kenya's banking industry and developed a prototype to proactively detect the fraud using data mining algorithms such as rule-based analytics, fuzzy string-matching, and z-score outlier analysis. The aim was to assist organizations reduce the time taken to detect the fraud, ultimately reducing the cost of the fraud.

To develop the prototype, the research sought to understand who was most likely to perpetrate procurement fraud, the fraud schemes used and what were the most common red flags indicators. The research also investigated the challenges hampering earlier detection of the fraud and the effectiveness of the current fraud detection methods. This was achieved through interviews where respondents provided responses which were analysed and summarized using charts.

6.2. Findings

6.2.1 Procurement Fraud Perpetrators and Schemes

The first objective of the study was to identify schemes and red flag indicators exist in procurement fraud. According to the findings, the research revealed that 87.5% of procurement fraud was perpetrated through collusion of both the employee and vendor and 12.5% was perpetrated by employees. These results build on existing evidence discussed in section 2.3.4 in the literature review which also noted that collusion frauds were on the rise with more than half of the frauds being committed by two or more fraudsters working in collusion. These results indicate that for the prototype to be effective, it should detect any close associations between employees and vendors using data mining algorithms on available data.

The research further revealed that Phantom Vendor Scheme at 87.5% was most commonly used by employees, Inflated Claims Scheme at 54.5% was most commonly used by vendors, and Bid Manipulation Scheme at 50%, was most commonly used by both employees and vendors to perpetrate procurement fraud. This is illustrated in Figure 6.1.

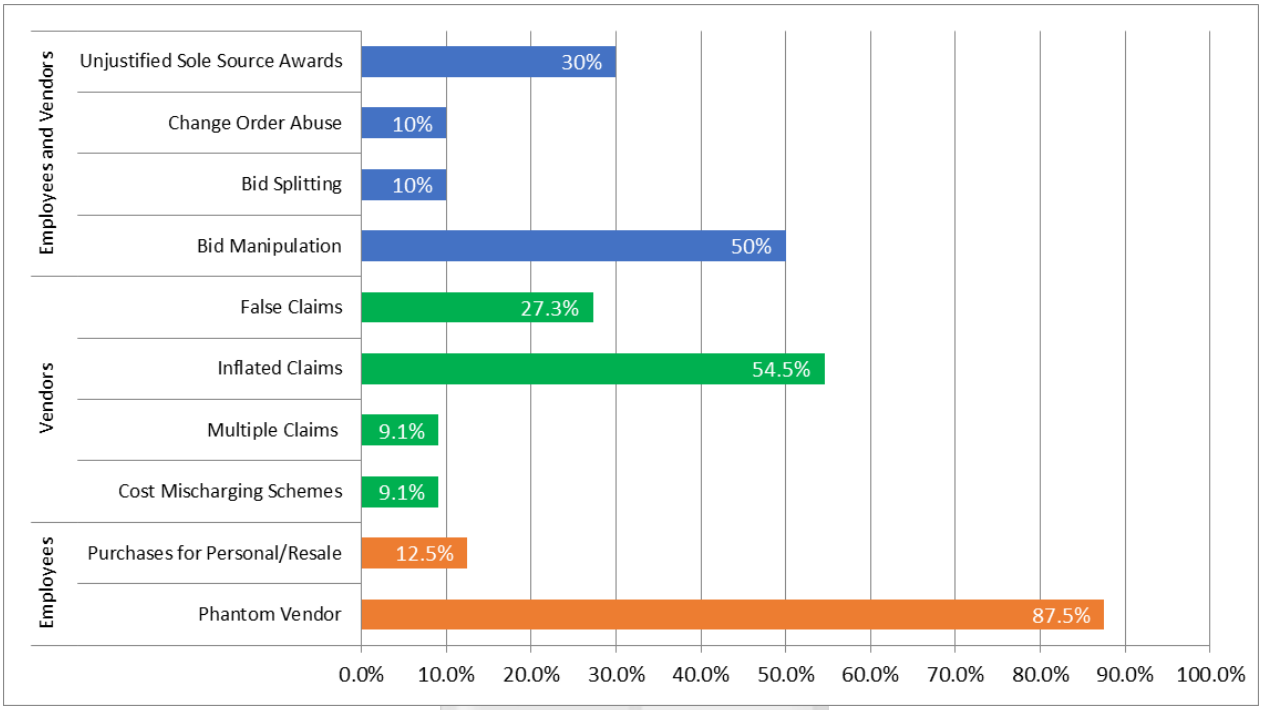


Figure 6.1: Fraud schemes used by both employees and vendors

6.2.2 Procurement Fraud Red Flag Indicators

Further to identifying the fraud schemes, the research sought to also identify the red flag indicators in procurement fraud in order to create fraud rulesets to detect the fraud. Section 2.3.3 of the literature review explored the different red flag indicators present in procurement fraud categorized into six groups: accounting anomalies, internal control weaknesses, analytic anomalies, extravagant lifestyles, unusual behaviour, and tips and complaints.

The research results established that over 90% of the most common red flag indicators in procurement fraud were analytical anomalies, internal control weaknesses, accounting anomalies and human resource-related. This is summarized in Figure 6.2. These findings significantly contributed to determine the fraud rulesets to be incorporated in the prototype's inference engine with over 80 rules developed into the prototype to aid detect these red flag indicators.

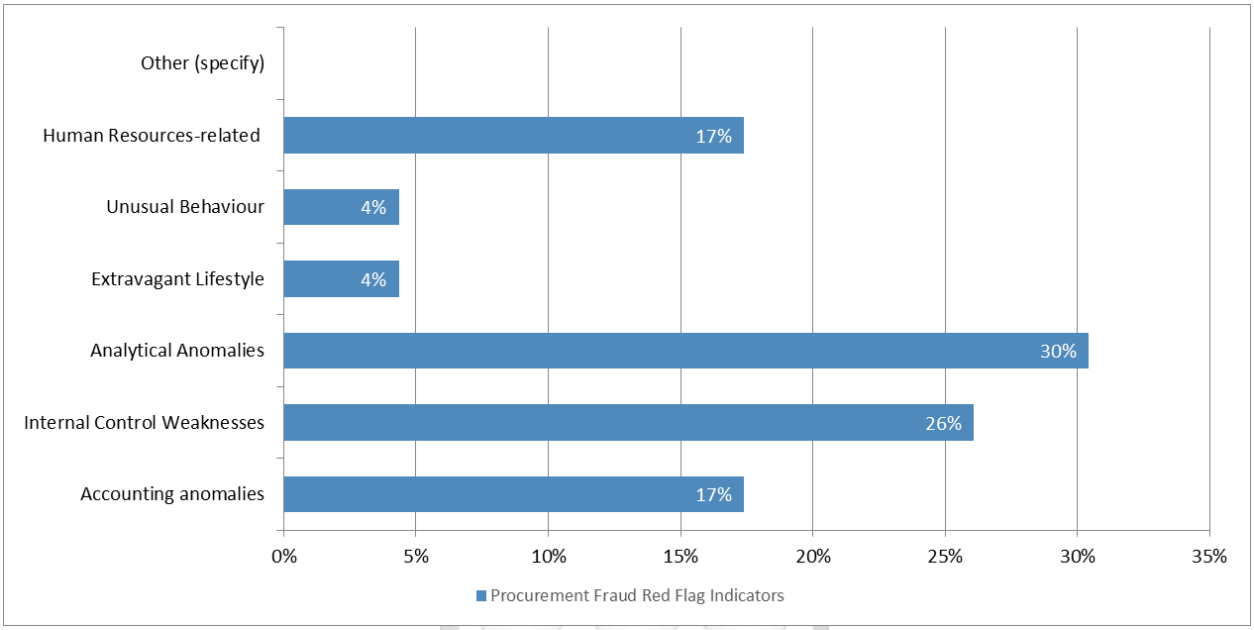


Figure 6.2: Most common red flag indicators for procurement fraud

6.2.3 Challenges in detecting procurement fraud

The second objective was to explore the challenges in detecting procurement fraud and section 2.3.4 of the literature review explored the different challenges encountered when detecting procurement fraud. The research results indicate that the manual nature of the procurement process was the main obstacle that hindered the earlier detection of procurement fraud at 35%. This was followed by the flexibility of procurement fraud schemes at 20% and unrealistic expectations on external auditors to detect fraud at 15%. These results are illustrated in Figure 6.3.

Although the results suggest that the manual nature of the procurement process posed the greatest challenge, the research revealed that employee, vendor and PTP transactional data in 100% of the Banking institutions sampled, was stored in electronic form. This data was also readily accessible and available from the HR and ERP systems. This was a critical finding as it provided an opportunity for the prototype to apply data mining algorithms on the electronic data thereby overcoming the main challenge (i.e., manual nature of procurement process) hindering earlier detection of procurement fraud.

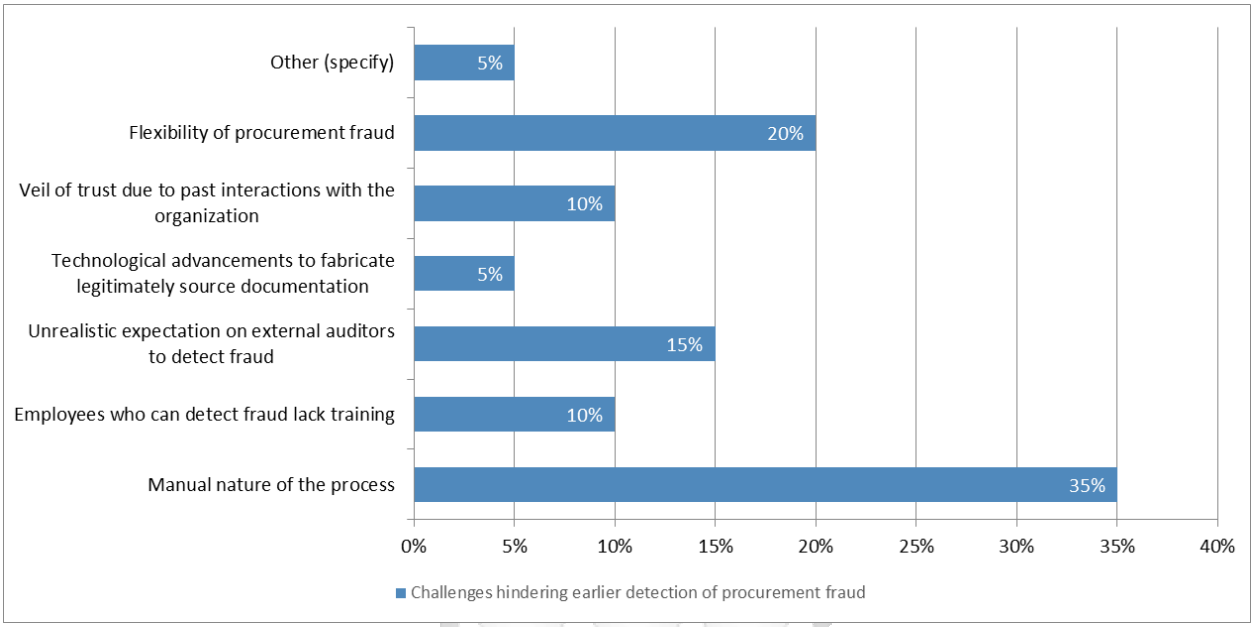


Figure 6.3: Challenges hindering earlier detection of procurement fraud

6.2.4 Methods used to detect procurement fraud

The third objective was to review the methods and architectures used to detect fraud. To develop a suitable solution, the research findings were used to help establish the most appropriate method and approach. The literature review discussed the different methods and architectures that are used to detect procurement fraud including external audits, management reviews, internal audits, whistle-blower tips, job rotation with mandatory leave, and proactive data monitoring/data analysis. Using a 5-point Likert scale (with 5 being most effective and 1 the least effective) the research results revealed that proactive data monitoring / data analysis at 50%, was the most effective method in detecting procurement fraud especially when the timeliness factor was considered. These results build on existing evidence that when fraud is detected proactively, it tends to be detected more quickly and thus causes lower losses. External audit of internal controls over financial reporting was identified as the least effective method at 50%, followed by hotline / whistle-blowers tips at 31%. This is illustrated in Figure 6.4.

Although the research results revealed that proactive data monitoring / data analysis was the most effective fraud detection method, it was not commonly used in organizations. The literature review, notes that whistle-blower tips were the most common way occupational frauds were discovered despite it being a passive fraud detection method. The research results collaborate the literature

review regarding the effectiveness of proactive data monitoring / data analysis as an anti-fraud control to reduce the duration taken to detect fraud.

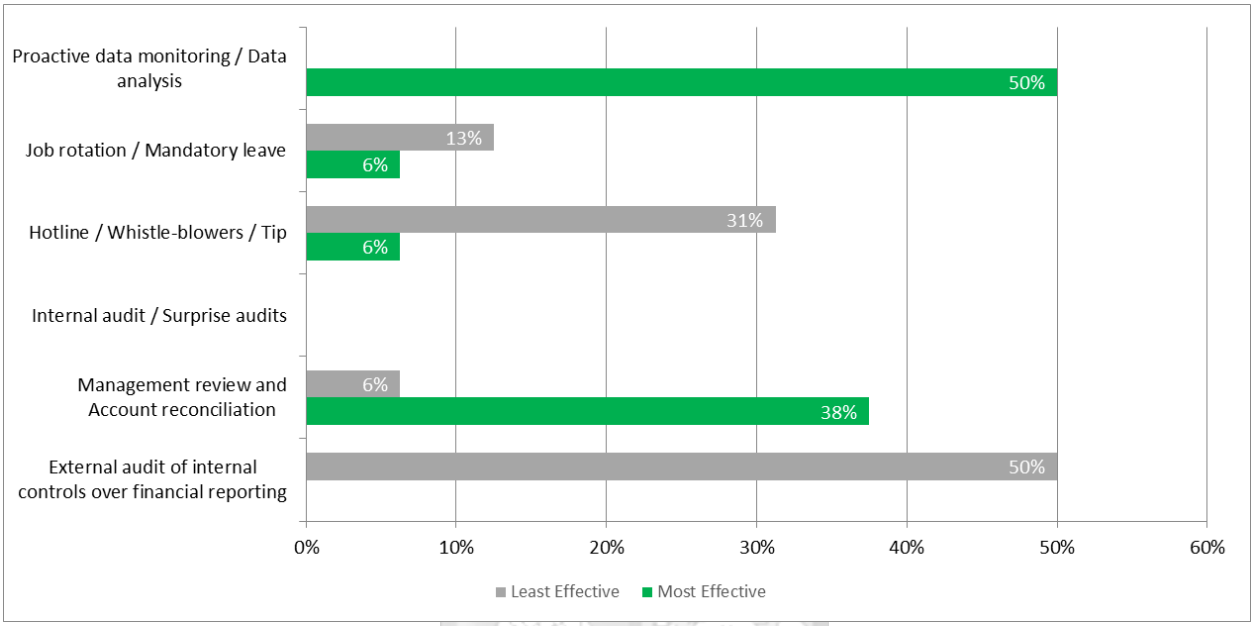


Figure 6.4: Effectiveness of Current Procurement Fraud Detection Methods

6.2.5 Develop and Test the Prototype

The fourth objective was to develop a prototype using data mining techniques to detect procurement fraud. The research established that data mining algorithms such as rule-based analytics, fuzzy string matching, and outlier analysis play a key role in fraud detection. The data mining algorithms used to develop the prototype were in line with the algorithms discussed in section 2.5 of the literature review. The prototype developed executes fraud rules on existing datasets to detect suspicious events and sends fraud alerts to the fraud analysts.

The fifth objective was to test the ability and readiness of the prototype. Integration and System test cases were designed and executed to test the functionality of the prototype. The acceptance testing questionnaire in Appendix E was used to test the readiness of the prototype covering four areas namely: functionality & accuracy, usability, performance & robustness, and data security. Based on the results of the acceptance tests, the functionality readiness of the prototype was strongly rated at 88%. Performance readiness was rated at 50% and data security readiness at 63%.

6.3. Advantages of the Prototype

The manual nature procurement process is seen has a major deterrent in detecting procurement fraud. However, with the existence of electronic procurement related data in Banking Institutions, the developed prototype was able to leverage on data mining algorithms to detect procurement fraud. Consequently, the developed prototype helps fraud analysts to detect any close associations between employees and vendors using data mining algorithms on the available data. The prototype achieves this by applying real-time data analysis on employee and vendor data from the HR and ERP source systems seeking to detect pattern matches between the two datasets using predetermined fraud rulesets. The prototype also relies on data of ineligible firms and individuals from the World Bank. The prototype uses over 80 fraud rulesets using rule-based, fuzzy string-matching, and z-score outlier analysis algorithms to detect suspicious procurement related-events. When positive matches are identified, the prototype flags the respective event as fraudulent.

Once an event is flagged fraudulent, the prototype immediately sends an email notification alert to the fraud analyst. This alert prompts the fraud analyst to investigate the event flagged fraudulent and captures their final finding in the prototype. This greatly reduces the time taken to detect fraudulent activities thereby reducing the cost of the fraud.

6.4. Limitations of the Prototype

The prototype does not consider external data sources such as Credit Reference Bureau Data, Company Registry Data, Land & Securities Registry Records to detect procurement fraud. Other than the World Bank listing of ineligible firms and individuals (which has to be manually uploaded due to the lack of an interface with the World Bank), the prototype mainly relies on data that is internal to the organization which limits its ability to detect more employee-vendor associations.

Although the inclusion of new rules is comparably simple, the prototype uses a finite number of fraud rules to detect fraudulent events. Therefore, any fraud scheme scenario perpetrated outside these rules would not be detected until new rules related to that fraud scheme are added to the prototype. The prototype also relies on structural attributes such as age, disciplinary record, division/department, education level, employee status, gender, leave balance, performance rating and employment tenure factors to calculate the employee risk profile that is used for training and awareness and to provide context and perspective to the fraud analysts.

Chapter 7: Conclusion and Recommendation

7.1. Conclusion

Early detection of procurement fraud schemes was hampered by the manual aspects of the procurement process especially the sourcing stage, and the flexible nature of the fraud. As a result, common traditional fraud detection processes such as whistle-blowing and periodic external audits faced inherent obstacles that hampered their ability to promptly detect the fraud. This resulted in situations where the fraud would go undetected for many months or even years.

Due to its proactive nature, the presence of data monitoring and data analysis as an anti-fraud control in an organization reduces the duration taken to detect fraud by over 30%. Despite this, only 31% of organizations had implemented proactive data monitoring/analysis with the rest exposed to fraud schemes going undetected for long periods of time. This research filled this gap by developing a prototype that utilizes rule based, fuzzy string-matching, and z-score outlier analysis data mining algorithms to detect fraud schemes related to the procurement process. The data mining algorithms were applied in real-time on employee and vendor data created, modified, or deleted in the prototype thereby facilitating prompt detection of any suspicious events.

Although there are numerous schemes that can be used to perpetrate procurement fraud, the research revealed that creation of phantom vendors and collusion of both employees and vendors remained the most prevalent procurement fraud schemes used in Kenyan Banking institutions. This was a significant finding in the research as it enabled the identification of red flag indicators to create the necessary fraud rulesets that were used by the prototype to proactively detect these fraud schemes. The prototype implemented over 80 fraud rules that were developed based on user requirements gathered from the in-depth interviews conducted during the research. These fraud rules were applied in real-time on the employees, vendors, procurement related data, and the World Bank listing of ineligible firms and individuals. This was made possible as the research established that Tier 1 banking institutions had implemented ERP and HR systems to manage their PTP and HR processes respectively which meant this data was readily available and electronically accessible. Further, the World Bank listing of ineligible firms and individuals consisting of 1,212 names was readily available from the World Bank website and was uploaded to the prototype's database for data mining purposes.

Secondary data from over 2,000 fraud cases was also analysed to calculate the risk profile for employees using structural attributes such as age, disciplinary record, division/department, education level, employee status, gender, leave balance, performance rating and employment tenure factors. The risk profile was a requirement by the fraud analysts for training and awareness purposes rather than stereotyping the organization's employees.

Using the defined fraud rulesets and data mining algorithms, the prototype determined whether to flag a procurement related event as non-fraudulent, possible fraudulent or fraudulent as soon as the event occurs. Fraud analysts are immediately notified of events flagged fraudulent via email notifications facilitating earlier detection of suspicious activities which would otherwise have gone unnoticed for a long period of time. The prompt notification of suspicious activities by the prototype reduces the average number of days it takes organizations to detect procurement fraud which also assists the prompt recovery of fraud proceeds effectively reducing the cost of the fraud.

7.2. Recommendations

For the prototype to work effectively in any organization, it will need to be directly integrated with the HR and ERP source systems to allow seamless transfer of data from these systems into the prototype in real-time. The researcher recommends that a data integration and mapping exercise should be conducted to map the data fields from the source systems. The data integration could also be achieved using APIs, where the source systems will expose their APIs for consumption by the prototype.

Further, to improve the prototype's accuracy, there will be a need to incorporate additional external data sources, and fast computing methods in determining fraudulent activities and reduce false alarms. Therefore, the research recommends the following external data sources to be incorporated and utilized by the prototype in order to improve its overall accuracy:

i. **Credit Reference Bureau Data:** This data source will provide access to negative credit information to detect financial pressures experienced by employees or vendors. Information such as defaulted loans, unpaid credit card debts, or issuance of bounced cheques will also be useful in determining the fraud risk profile of an employee and vendor.
ii. **OFAC SDN Data:** The OFAC's Specially Designated Nationals and Blocked Persons List ("SDN List") has approximately 6,300 names connected with sanctions (U.S. Department

of the Treasury, 2021). The SDN list would determine whether a vendor needs to be flagged fraudulent on the basis that its directors or senior management appear on this list.

iii. **Company Registry Data:** Based on this data source, changes of vendor directorship at the company registry will be detected in real time and flagged for investigation.

iv. **Land and Securities Registry Records:** Most proceeds of fraud schemes are usually channelled to real estate properties and acquisition of equities. Integrating records from these registries will aid in detecting whether an employee is living beyond their means.

v. **Bank Account and M-Pesa Deposit Transaction Records:** Banking and M-Pesa transactions for employees would be useful in detecting and monitoring cash and EFT deposit transactions that are above a certain predetermined limit. This will detect whether an employee has an unexplained additional income generated from sources other than employment thereby determining whether an employee is living beyond their means.

7.3. Future Work

Given the flexibility of procurement fraud, no number of rules can be developed to cover all possible scenarios of fraud schemes. Therefore, in addition to the rule-based, fuzzy matching, and outlier analysis data mining algorithms used, further research is needed to incorporate Artificial Neural Network (ANN) and big data analysis in the prototype to extend its procurement fraud detection capabilities.

Future research should also consider situational attributes to calculate the employee risk profile to detect behavioural red flags such as living beyond their means; financial difficulties; irritability, suspiciousness, or defensiveness; addiction problems; divorce/family problems; excessive pressure from within the organization etc.

References

ACFE. (2014). *Fraud Examiners Manual* (2014 International ed., Vol. I). Austin, Texas: Asscoiation of Certified Fraud Examiners (ACFE).

ACFE. (2020). *2020 Global Fraud Study: Report to the Nations on Occupational Fraud and Abuse.* Austin, TX. Retrieved October 3, 2021, from https://acfepublic.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf

Akerkar, R. (2013). Advanced data analytics for business. In R. Akerkar, & R. Akerkar (Ed.), *Big data computing* (pp. 373-397). Boca Raton: Chapman and Hall/CRC.

Albrecht, S. W., Albrecht, C. C., & Albrecht, C. O. (2006). *Fraud Examination* (2nd ed.). Toronto, Canada: Thomson South-Western.

Albrecht, S. W., Albrecht, C. O., Albrecht, C. C., & Zimbelman, M. F. (2012). *Fraud Examination* (4th ed.). Mason, OH, USA: Cengage Learning.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Freeman, J., & Eddie, S. (2010). *Statistics For Business and Economics* (2nd ed.). Hampshire, United Kingdom: Cengage Learning EMEA.

Baesens, B., Vlasselaer, V., & Verbeke, W. (2015). *Fraud Analytics Using Descriptive, Predictive and Social Network Techniques: A Guide to Data Science for Fraud Detection.* North Carolina, United States of America: SAS Institute Inc.

Caulfield, T. (2010, September). The Anatomy & Illusiveness of Procurement Fraud. *Journal of the Association of Inspectors General, VII*(3). Retrieved November 17, 2021, from http://www.theiia.org/chapters/pubdocs/27/AIG_Article.pdf

Central Bank of Kenya. (2020). *Bank Supervision Annual Report 2020.* Nairobi: Central Bank of Kenya. Retrieved November 23, 2021, from Central Bank of Kenya: https://www.centralbank.go.ke/uploads/banking_sector_annual_reports/1375903848_Bank%20Supervision%20Annual%20Report%202020.pdf

Cervo, D., & Allen, M. (2011). *Master Data Management in Practice.* Hoboken, New Jersey, United States of America: John Wiley & Sons, Inc.

Cronie, G. (2008). Purchase-to-Pay Processes: ING Guide to Financial Supply Chain Optimisation. *TMI*. Retrieved September 3, 2014, from http://web.utk.edu/~jwachowi/INGpart3.pdf

Davenport, T., & Harris, J. (2007). Competing on analytics: The new science of winning. *Harvard Business Review*.

Deloitte. (2021, December 02). *Profiling the fraudster: understanding the threat of insider fraud*. Retrieved December 02, 2021, from Deloitte: https://www2.deloitte.com/mt/en/pages/risk/articles/mt-risk-article-profiling-the-fraudster.html

DiNapoli, T. P. (2008). *Red Flags for Fraud.* State of New York Office of the State Comptroller. Retrieved September 26, 2014, from http://web4.osc.state.ny.us/localgov/pubs/red_flags_fraud.pdf

Django Software Foundation. (2021, December 21). *Meet Django*. Retrieved from Django: https://www.djangoproject.com/

Doig, A. (2012). *Fraud: The Counter Fraud Practitioner's Handbook.* (A. Doig, Ed.) Gower Publishing.

EACC. (2015). *Report of Activities and Financial Statements for the Financial Year 2104/2015 for the EACC.* Nairobi: Ethics and Anti-Corruption Commission.

Encode OSS Ltd. (2021, December 21). *Django Rest Framework: Home*. Retrieved from Django Rest Framework: https://www.django-rest-framework.org/

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2009). *Database Systems The Complete Book* (2nd ed.). New Jersey, United States of America: Pearson Prentice Hall.

Gelinas, U. J., & Sutton, S. G. (2002). *Accounting Information Systems.* Thomson Learning.

Green, W. (2014). Procurement fraud second most common economic crime globally. *CIPS Supply Management*.

GSA OIG. (2012). *Procurment Fraud Handbook.* Retrieved September 25, 2014, from http://www.gsaig.gov/?LinkServID=6486B647-A5DF-C154-010A408470CAE0B8

Guttag, J. V. (2017). *Introduction to Computation and Programming Using Python with Application to Understanding Data* (2nd ed.). Massachusetts: The MIT Press.

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques.* San Francisco: Morgan Kaufmann Publishers.

Hollander, A. S., Denna, E. L., & Cherrington, J. O. (2000). *Accounting, Information Technology, and Business Solutions* (2nd ed.). New York: McGraw-Hill/Irwin.

Hurt, D. L. (2010). *Accounting Information Systems: Basic Concepts & Current Issues* (2nd ed.). New York: McGraw-Hill/Irwin.

IBM Cloud Education. (2020, August 19). *Supervised Learning*. Retrieved November 30, 2021, from IBM Cloud Web site: https://www.ibm.com/cloud/learn/supervised-learning

Kearney, D. (2013, November 07). *Applying data analytics logic to supplier management*. Retrieved February 05, 2017, from Procurement Leaders: https://www.procurementleaders.com/blog/my-blog--guest-blog/applying-data-analytics-logic-to-supplier-management-369396

KPMG. (2010). *Procurement Fraud in Consumer Companies: Preventing, Detecting and Taking Action.*

KPMG. (2013). *A survey of fraud, bribery and corruption in Australia & New Zealand 2012.*

Laudon, K. C., & Laudon, J. P. (2014). *Management Information Systems: Managing the Digital Firm* (13th ed.). Essex, England: Pearson Education Limited.

Lemon, J. (2017). *How a hybrid anti-fraud approach could have detected and prevented fraud in government acquisition programs (White paper).* Retrieved February 05, 2017, from SAS: http://www.sas.com/en_ca/whitepapers/how-a-hybrid-anti-fraud-approach-105793.html

Marr, B., & Ward, M. (2019). *Artificial Intelligence in Practice.* John Wiley & Sons, Inc.

Material-UI SAS. (2021, December 21). *MUI: The React component library you always wanted*. Retrieved from The React UI Library: https://mui.com/

Maurno, D. (2013). The latest fraud-finding tools. *Compliance Week*, 38– 39.

McNeal, A. (2014). The statistical profile of a fraudster. *Fraud Magazine*. Retrieved December 01, 2021, from https://www.fraud-magazine.com/article.aspx?id=4294984634

Meta Platforms, Inc. (2021, December 21). *React – A JavaScript library for building user interfaces*. Retrieved from React: https://reactjs.org/

Microsoft Corp. (2021, December 21). *Code Editing Redefined*. Retrieved from Visual Studio Code: https://code.visualstudio.com/

Nation Media Group. (2017, February 6). *11 mega scandals that hit the Jubilee Government*. Retrieved March 7, 2017, from Daily Nation: http://www.nation.co.ke/newsplex/jubilee-scandals/2718262-3802442-m5abehz/index.html

National Fraud Authority. (2011). *Procurement Fraud in the Public Sector.* Retrieved September 16, 2014, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/118460/procurement-fraud-public-sector.pdf

Ng, J., & Shah, S. (2020). *Hands-On Artificial Intelligence for Banking.* Birmingham: Packt Publishing Ltd.

Oracle Applications Users Group (OAUG). (2011). *Strategies for Managing Risky Business Process: 2011 OAUG Enterprise Governance, Risk and Compliance Survey.* Oracle. NJ: Unisphere Reseach. Retrieved September 1, 2014, from http://www.oracle.com/us/solutions/corporate-governance/strategie-mana-riskybus-proc-1368906.pdf

Padgett, S. (2015). *Profiling the Fraudster: Removing the Mask to Prevent and Detect Fraud.* Hoboken,, New Jersey, United States of America: JohnWiley & Sons, Inc.

Passas, P. (2007). *Corruption in the procurement process/outsourcing Government functions.* Boston: Institute for Fraud Prevention.

Postman, Inc. (2022, March 14). *Postman API Tools*. Retrieved from Postman: https://www.postman.com/product/tools/

Purcell, J. (2016, January 31). *Employee Management Security Controls.* Retrieved March 2017, 17, from SANS Software Security: https://software-security.sans.org/resources/paper/cissp/employee-management-security-controls

PwC. (2012). *Optimize procure-to pay processes for profitability, efficiency, and compliance.*

PwC. (2014). *Global Economic Crime Survey: Kenya Report.*

Rose, D. (2020). *Artificial Intelligence for Business* (2nd ed.). Pearson FT Press.

Santhi, V., Acharjya, D., & Ezhilarasan, M. (2016). *Emerging Technologies in Intelligent Application for Image and Video Processing.* Hershey, United States of America: IGI Global.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research Methods for Business Students* (6th ed.). Essex, England: Pearson Education Limited.

SEC. (2002, June 24). *All About Auditors: What Investors Need to Know*. Retrieved March 23, 2017, from U.S Securities and Exchange Commission: https://www.sec.gov/reportspubs/investor-publications/investorpubsaboutauditorshtm.html

Singh, N., Lai, K.-h., Vejvar, M., & Cheng, E. T. (2019, July 08). Data-driven auditing: A predictive modeling approach to fraud detection and classification. *Journal of Corporate Accounting & Finance, 30*(3), 64-82. Retrieved from https://doi.org/10.1002/jcaf.22389

Stiner, S. (2016, August 24). *Rapid Application Development (RAD): A Smart, Quick And Valuable Process For Software Developers*. Retrieved March 28, 2017, from Forbes: https://www.forbes.com/sites/forbestechcouncil/2016/08/24/rapid-application-development-rad-a-smart-quick-and-valuable-process-for-software-developers/#1e9ad20619e8

Taulli, T. (2019). *Artificial Intelligence Basics: A Non-Technical Introduction.* Monrovia, USA: Apress.

The Pandas Development Team. (2021, December 21). *Pandas - Python Data Analysis Library*. Retrieved from Pandas: https://pandas.pydata.org/

The SQLite Consortium. (2021, December 2021). *What Is SQLite?* Retrieved from SQLite: https://www.sqlite.org/index.html

Turban, E., & Aronson, J. E. (2001). *Decision Support Systems and Intelligent Systems* (6th ed.). New Jersey: Prentice Hall International, Inc.

U.S. Department of the Treasury. (2021, October 12). *Specially Designated Nationals And Blocked Persons List (SDN) Human Readable Lists*. Retrieved December 13, 2021, from U.S. Department of the Treasury: https://home.treasury.gov/policy-issues/financial-sanctions/specially-designated-nationals-and-blocked-persons-list-sdn-human-readable-lists

Vona, L. (2011). *The Fraud Audit: Responding to the Risk of Fraud in Core Business Systems.* John Wiley & Sons, Inc.

Vona, L. (2017). *Fraud Data Analytics Methodology.* New Jersey: John Wiley & Sons, Inc.

World Bank Group. (2019, April 01). *World Bank Sanctions System: Tackling Fraud & Corruption through a Two-Tier Administrative Process.* Retrieved April 11, 2022, from The World Bank: https://www.worldbank.org/content/dam/documents/sanctions/office-of-suspension-and-debarment/2019/may/OSDFactSheetApril2019.pdf

World Bank Group. (2022, April 08). *Procurement - World Bank Listing of Ineligible Firms and Individuals.* Retrieved April 08, 2022, from The World Bank: https://www.worldbank.org/en/projects-operations/procurement/debarred-firms

Appendices

Appendix A: Similarity Report

Curiginal

Document Information

Analyzed document	2022 - A Prototype for Detecting Procurement Fraud - Francis Muriithi - 008096.pdf (D133644093)
Submitted	2022-04-14T17:16:00.0000000
Submitted by	
Submitter email	Fwmuriithi@gmail.com
Similarity	8%
Analysis address	library.strath@analysis.urkund.com

Sources included in the report

SA	**Strathmore University / Thesis WIP 221121.docx** Document Thesis WIP 221121.docx (D119507624) Submitted by: brian.onyango@strathmore.edu Receiver: mwachira.strath@analysis.urkund.com	2
SA	**Strathmore University / Amon - Research Proposal; Chapter 1-3.docx** Document Amon - Research Proposal; Chapter 1-3.docx (D123284860) Submitted by: Amon.nyansera@strathmore.edu Receiver: mwachira.strath@analysis.urkund.com	1
W	URL: https://acfepublic.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf Fetched: 2022-04-14T17:28:00.0000000	23
SA	**ACFE_2018-report-to-the-nations.pdf** Document ACFE_2018-report-to-the-nations.pdf (D54089959)	5
W	URL: https://www2.deloitte.com/content/dam/Deloitte/xe/Documents/About-Deloitte/mepovdocuments/mepov13/dtme_mepov13_Procurement%20fraud.pdf Fetched: 2020-07-21T10:38:31.8700000	1
SA	**Strathmore University / Seminar paper Final.edited.doc** Document Seminar paper Final.edited.doc (D123291030) Submitted by: Franklin.Chebii@strathmore.edu Receiver: mwachira.strath@analysis.urkund.com	1
W	URL: https://www.withum.com/resources/2020-report-to-the-nations-insights-and-lessons-on-occupational-fraud/ Fetched: 2021-07-27T01:05:22.3070000	1
W	URL: https://www.sec.gov/reportspubs/investor-publications/investorpubsaboutauditorshtm.html Fetched: 2022-04-14T17:31:00.0000000	2
W	URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/118460/procurement-fraud-public-sector.pdf Fetched: 2020-05-23T11:23:19.6030000	1
W	URL: http://www.fraudconference.com/uploadedFiles/Fraud_Conference/Content/Course-Materials/presentations/23rd/ppt/post-07-Performance-Schemes.pdf Fetched: 2022-03-28T08:42:01.0530000	1

Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

23rd March 2022

Mr Muriithi Francis,
FMuriithi@strathmore.edu

Dear Mr Muriithi,

RE: A Prototype for Detecting Procurement Fraud Using Data Mining Techniques: Case of Banking Industry in Kenya

This is to inform you that SU-IERC has reviewed and **approved** your above **SU masters'** research proposal. Your application reference number is **SU-IERC1278/22.** The approval period is **23rd March 2022 to 24th March 2023.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-IERC

Cc: Prof Fred Were,
Chairperson; SU-IERC

STRATHMORE UNIVERSITY INSTITUTIONAL ETHICS REVIEW COMMITTEE
(SU-IERC)
23 Mar 2022
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Appendix C: NACOSTI Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **524705**

Date of Issue: **30/March/2022**

RESEARCH LICENSE

This is to Certify that Mr.. Francis W. Muriithi of Strathmore University, has been licensed to conduct research in Nairobi on the topic: A Prototype for Detecting Procurement Fraud Using Data Mining Techniques: Case of Banking Industry in Kenya for the period ending : 30/March/2023.

License No: **NACOSTI/P/22/16583**

524705

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix D: Interview Questions

Strathmore UNIVERSITY

INTERVIEW GUIDE

Researcher: Francis W. Muriithi

MSc. IT, Strathmore University

This research will be used for academic purpose only. Its main objective is to find out users' experience in detecting procurement fraud in Kenyan Banks. Kindly provide your honest opinion on the same. Please note that your responses will be treated as private and confidential.

Interviewee:		**Date:**	

SECTION 1: BACKGROUND

1.1. What is your primary occupation and professional background? Select one option

- ☐ Forensic Examiner/Investigator
- ☐ Internal Auditor
- ☐ Risk and Controls Professional
- ☐ External/Independent Audit
- ☐ Other (Specify) .....

1.2. How many years of experience do you currently have in fraud examination, risk and controls or audit profession? Select one option

- ☐ ≤ 5 years
- ☐ 6 – 10 years
- ☐ 11 – 15 years
- ☐ > 15 years

SECTION 2: PROCUREMENT FRAUD SCHEMES AND RED FLAG INDICATORS

2.1 In your view, who is <u>most likely to perpetrate</u> procurement fraud?

Select one option

- ☐ Employees
- ☐ Vendors
- ☐ Collusion of both employee and vendor
- ☐ Other Party (Specify) .....

2.2 In your view, what are the common ways in which procurement frauds are perpetrated by <u>employees</u>?

Select all that apply

- ☐ Phantom Vendor
- ☐ Purchases for Personal/Resale
- ☐ Other (specify) .....

2.3 In your view, what are the common ways in which procurement frauds are perpetrated by <u>vendors</u>?

Select all that apply

- ☐ Cost Mischarging Schemes
- ☐ Multiple Claims
- ☐ Inflated Claims
- ☐ False Claims
- ☐ Other (specify) .....

2.4 In your view, what are the common ways in which procurement frauds are perpetrated by <u>both employees and vendors</u>?

Select all that apply

- ☐ Bid Manipulation
- ☐ Bid Splitting
- ☐ Change Order Abuse
- ☐ Unjustified Sole Source Awards
- ☐ Other (specify) .....

2.5 In your view, what are the <u>most common red flag indicators</u> for procurement related frauds?

Select all that apply

- ☐ **Accounting anomalies** (e.g., purchase documents alterations, common names or addresses of payees, duplicate payments, journal entries made by unauthorized individuals etc.)
- ☐ **Internal Control Weaknesses** (e.g., lack of segregation of duties, lack of proper authorization, overriding of existing controls etc.)
- ☐ **Analytical Anomalies** (e.g., excess purchases; unexplained inventory shortages or adjustments; cash shortages or overages; excessive late charges; and significant changes in account balances or ratios etc.)
- ☐ **Extravagant Lifestyle** (e.g., live beyond their means since their income does not support their lifestyle)
- ☐ **Unusual Behaviour** (e.g., unusually close association with vendors, refusal to take vacations, financial difficulties, unwillingness to share duties etc.)
- ☐ **Human Resources-related** (e.g., poor performance evaluations, fear of job loss, denied raise or promotion, disciplinary actions received etc.)
- ☐ Other (specify) .....

2.6 Does your organization have <u>electronic</u> employee, vendor and PTP transactional <u>data</u>?

Select one option

- ☐ Yes
- ☐ No

SECTION 3: CHALLENGES & DETECTION OF PROCUREMENT FRAUD

3.1 In your view, what are the <u>main challenges</u> that hinder the earlier detection of procurement fraud in organizations?

Select all that apply

	Manual nature of the process
	Employees who can detect fraud lack training
	Unrealistic expectation on external auditors to detect fraud
	Technological advancements to fabricate legitimate source documentation
	Veil of trust due to past interactions with the organization
	Flexibility of procurement fraud
	Other (specify) .....

3.2 On a scale of 1 – 5, (where 5 is the most effective and 1 the least effective), rate the effectiveness and timeliness of the current methods used in <u>detecting procurement fraud</u>.

#	Method of Detecting Fraud	1	2	3	4	5
1	External audit of internal controls over financial reporting					
2	Management review and Account reconciliation					
3	Internal audit / Surprise audits					
4	Hotline / Whistle-blowers / Tip					
5	Job rotation / Mandatory leave					
6	Proactive data monitoring / Data analysis					

Appendix E: Acceptance Testing Questionnaire

Strathmore UNIVERSITY

ACCEPTANCE TESTING QUESTIONNAIRE

Researcher: Francis W. Muriithi

MSc. IT, Strathmore University

This research will be used for academic purpose only. Its main objective is to find out users' experience in using the procurement fraud detection prototype. Please note that your responses will be treated as private and confidential.

Kindly rate the readiness of the prototype on a scale of 1 – 5, (where 5 is strongly agree, 4 is agree, 3 is neutral, 2 is disagree, and 1 is strongly disagree) in the sections below.

Respondent:	

SECTION 1: FUNCTIONALITY & ACCURACY					
Factors	**1**	**2**	**3**	**4**	**5**
Satisfactorily meets your stated user requirements					
Generates accurate results in the events analysed					

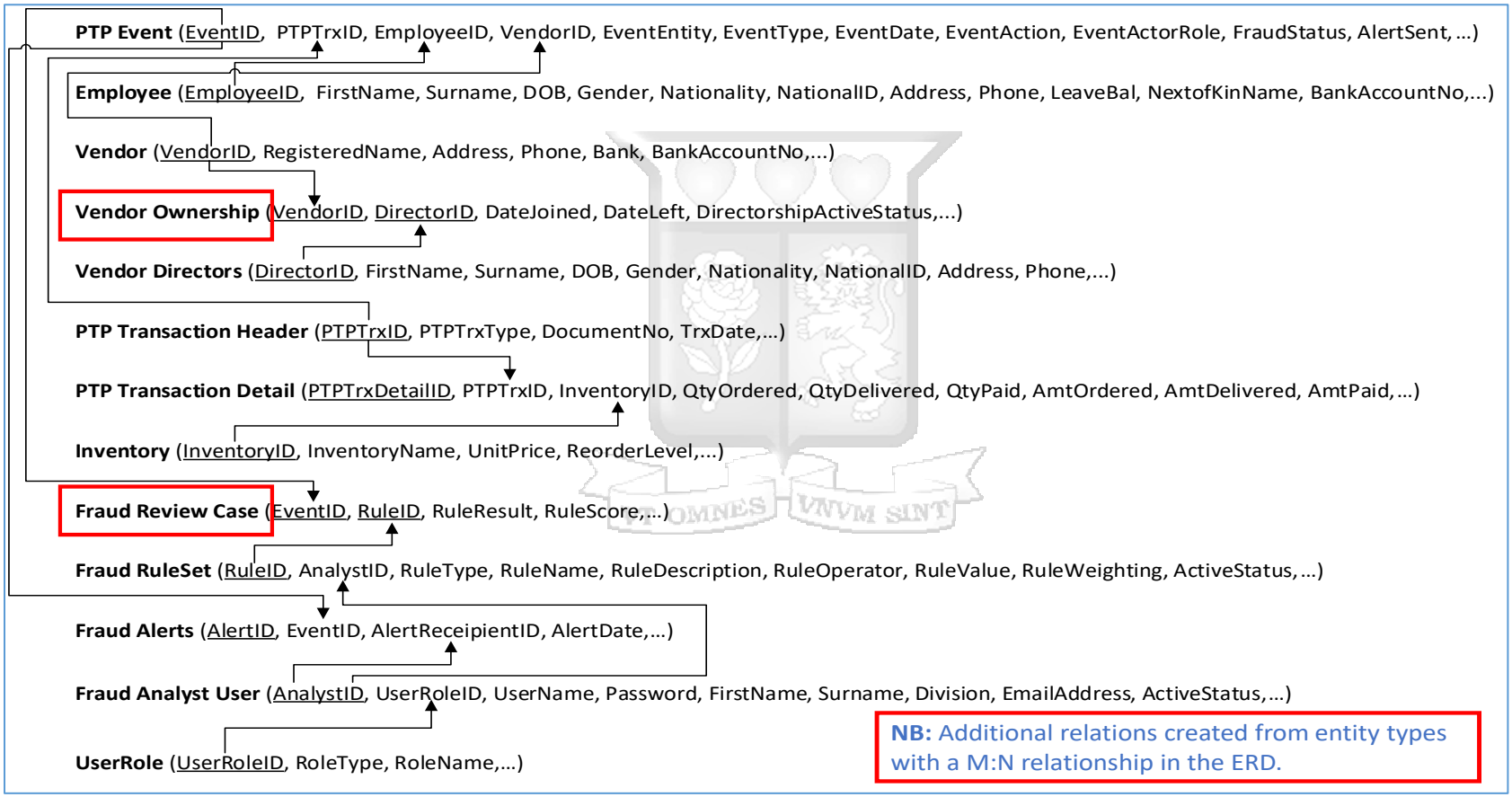
SECTION 2: USABILITY					
Factors	**1**	**2**	**3**	**4**	**5**
Easy to use and intuitive user interface					
Minimum training needed to learn and use					

SECTION 3: PERFORMANCE & ROBUSTNESS					
Factors	**1**	**2**	**3**	**4**	**5**
User interface is responsive in nature					
Robust and handles errors without crashing					

SECTION 4: DATA SECURITY					
Factors	**1**	**2**	**3**	**4**	**5**
Confidentiality of data is restricted to authorized users					
Integrity of data is maintained					

Appendix F: Relational Mapping of the ERD into Relations

From the Entity Relationship Diagram in Figure 4.9, the database relations were identified using the relational mapping process and attributes of the relations are captured in greater detail in Appendix G: Data Dictionary.



Appendix G: Sample Data Dictionaries

i. Entity Type: Employee

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	EmployeeID	bigint	8	False	Primary Key
	FirstName	Nvarchar (50)	100	False	
	LastName	Nvarchar (50)	100	False	
	Initials	Nvarchar (8)	16	True	
	DOB	datetime	8	False	Date of Birth
	DOJ	datetime	8	False	Date of Joining
	EducationLevel	Nvarchar (32)	64	False	Primary School | High School | Diploma | University Degree | Postgraduate Degree
	Gender	Char (10)	10	False	
	Nationality	Nvarchar (32)	64	False	
	IdentifierType	Nvarchar (32)	64	False	Passport | ID Card | Driving License
	IdentifierNo	Nvarchar (32)	64	False	
	BoxNumber	int	4	True	
	PostalCode	Nvarchar (8)	16	True	
	Country	Nvarchar (64)	128	True	
	Mobile	Nvarchar (32)	64	False	
	PersonalEmail	Nvarchar (255)	510	True	
	WorkEmail	Nvarchar (255)	510	False	
	ResidentialAddress	Nvarchar (255)	510	True	
	NextofKinFirstName	Nvarchar (50)	100	False	
	NextofKinLastName	Nvarchar (50)	100	False	
	NextofKinRelationship	Nvarchar (32)	64	False	
	NextofKinMobile	Nvarchar (32)	64	False	
	LeaveBalance	int	4	False	
	BankName	Nvarchar (50)	100	False	
	BankBranch	Nvarchar (50)	100	False	
	BankAccountNo	Nvarchar (50)	100	False	
	BankCountry	Nvarchar (64)	128	False	
	DisciplinaryRecord	Nvarchar (32)	64	False	
	PerformanceRating	Nvarchar (32)	64	False	
	RiskProfile	int	4	False	
	RiskScore	int	4	False	

ii. Entity Type: RiskProfile

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	RiskProfileID	bigint	8	False	Primary Key
FK	EmployeeID	bigint	8	False	FK: Employee
	RiskCategory	Nvarchar (50)	100	False	Tenure | Gender | Age | Education Level | Performance Rating | Leave Balance
	RiskDescription	Nvarchar (255)	510	False	
	RiskScore	float	8	False	
	RiskMaxScore	float	8	False	

iii. Entity Type: FraudAlerts

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	AlertID	bigint	8	False	Primary Key
FK	EventtID	bigint	8	False	FK: PTPEvent
FK	AnalystID	bigint	8	False	FK: FraudAnalystUser – Alert Recipient
	AlertDate	datetime	8	False	
	ChannelUsed	Nvarchar (8)	16	False	Email | SMS | Notification

iv. Entity Type: Vendor

Key	Attribute Name	Data Type	Max Length (Bytes)	Allow Nulls	Notes
PK	VendorID	bigint	8	False	Primary Key
	RegisteredName	Nvarchar (255)	510	False	
	VendorNo	Nvarchar (50)	100	False	
	SingleSourced	Char (10)	10	False	Yes | No
	DateRegistered	datetime	8	False	
	BoxNumber	int	4	False	
	PostalCode	Nvarchar (8)	16	False	
	Country	Nvarchar (32)	64	False	
	Mobile	Nvarchar (32)	64	False	
	Email	Nvarchar (255)	510	False	
	Website	Nvarchar (50)	100	True	
	OfficeAddress	Nvarchar (255)	510	True	
	BankName	Nvarchar (50)	100	False	
	BankBranch	Nvarchar (50)	100	False	
	BankAccountNo	Nvarchar (50)	100	False	
	BankCountry	Nvarchar (64)	128	False	
	RiskProfile	int	4	False	

Appendix H: Sample SQL Data Definition Language Implementing Data Tables

H.1: SQL DDL Creating the PTP Event Table and Indexes

```
CREATE TABLE FraudDetectApp_PTPEvent (
      EventID          INTEGER          NOT NULL, PRIMARY KEYAUTOINCREMENT,
      EventSource      VARCHAR (50)     NOT NULL,
      EventType        VARCHAR (50)     NOT NULL,
      EventAction      VARCHAR (50)     NOT NULL,
      EventDate        DATETIME         NOT NULL,
      EventActorRole   VARCHAR (50)     NOT NULL,
      EventTypeDesc    TEXT             NOT NULL,
      FraudStatus      VARCHAR (50)     NOT NULL,
      AlertSent        BOOL             NOT NULL,
      DateCreated      DATETIME         NOT NULL,
      CreatedBy_id     INTEGER          REFERENCES auth_user (id) DEFERRABLE INITIALLY
                                        DEFERRED,
      EmployeeID_id    INTEGER          NOT NULL
                                        REFERENCES EmployeeApp_Employee (EmployeeID)
                                        DEFERRABLE INITIALLY DEFERRED,
      PTPTrxID_id      INTEGER          REFERENCES PTPTransaction_PTPTransactionHeader
                                        (PTPTrxID) DEFERRABLE INITIALLY DEFERRED,
      VendorID_id      INTEGER          NOT NULL
                                        REFERENCES VendorApp_Vendor (VendorID) DEFERRABLE
                                        INITIALLY DEFERRED
);
CREATE INDEX FraudDetectApp_PTPEvent_EmployeeID_id_Index ON FraudDetectApp_PTPEvent
(EmployeeID_id);

CREATE INDEX FraudDetectApp_PTPEvent_PTPTrxID_id_Index ON FraudDetectApp_PTPEvent
(PTPTrxID_id);

CREATE INDEX FraudDetectApp_PTPEvent_VendorID_id_Index ON FraudDetectApp_PTPEvent
(VendorID_id);

CREATE INDEX FraudDetect_EventSo_Index ON FraudDetectApp_PTPEvent (EventSource);
```

H.2: SQL DDL Creating the FraudRuleset Table and Indexes

```
CREATE TABLE FraudDetectApp_FraudRuleSet (
      RuleID           INTEGER          NOT NULL PRIMARY KEY AUTOINCREMENT,
      RuleType         VARCHAR (50)     NOT NULL,
      RuleName         VARCHAR (50)     NOT NULL,
      RuleOperator     VARCHAR (50)     NOT NULL,
      RuleValue        REAL             NOT NULL,
      RuleWeighting    REAL             NOT NULL,
      RuleDescription  TEXT             NOT NULL,
      IsActive         BOOL             NOT NULL,
      DateCreated      DATETIME         NOT NULL,
      DateModified     DATETIME         NOT NULL,
      CreatedBy_id     INTEGER          REFERENCES auth_user (id) DEFERRABLE INITIALLY
                                        DEFERRED,
      ModifiedBy_id    INTEGER          REFERENCES auth_user (id) DEFERRABLE INITIALLY
                                        DEFERRED
);
CREATE INDEX FraudDetect_RuleName_Index ON FraudDetectApp_FraudRuleSet (RuleName);

CREATE INDEX FraudDetect_RuleType_Index ON FraudDetectApp_FraudRuleSet (RuleType);
```

H.3: SQL DDL Creating the RiskProfile Table and Indexes

```
CREATE TABLE EmployeeApp_RiskProfile (
      RiskProfileID        INTEGER          NOT NULL PRIMARY KEY AUTOINCREMENT,
      RiskCategory         VARCHAR (50)     NOT NULL,
      RiskDescription      TEXT             NOT NULL,
      RiskScore            REAL             NOT NULL,
      RiskMaxScore         REAL             NOT NULL,
      RiskScorePercentage  DECIMAL          NOT NULL,
      IsActive             BOOL             NOT NULL,
      DateCreated          DATETIME         NOT NULL,
      DateModified         DATETIME         NOT NULL,
      CreatedBy_id         INTEGER          REFERENCES auth_user (id) DEFERRABLE
                                            INITIALLY DEFERRED,
      ModifiedBy_id        INTEGER          REFERENCES auth_user (id) DEFERRABLE
                                            INITIALLY DEFERRED
      EmployeeID_id        INTEGER          NOT NULL
                                            REFERENCES EmployeeApp_Employee
                                            (EmployeeID) DEFERRABLE INITIALLY DEFERRED,
);
CREATE INDEX FraudDetect_RuleName_Index ON FraudDetectApp_FraudRuleSet (RuleName);

CREATE INDEX FraudDetect_RuleType_Index ON FraudDetectApp_FraudRuleSet (RuleType);
```

Appendix I: Sample Source Code Snippets

I.1: Code Snippet Implementing Fuzzy Matching and Rule-Based Analytics

```
import sqlite3, import pandas as pd
from fuzzywuzzy import fuzz, process

#Execute Fuzzy Match Checker for name/string pattern matches
def FuzzyMatchChecker (SearchField,SearchTable,TextToSearch):
    # Read sqlite query results into a pandas DataFrame
    con = sqlite3.connect("D:/SystemImplementation/FDSWebServerProject/db.sqlite3")
    SearchTableName = SearchTable.split("_")[1]
    # Return all results of query in pandas dataframe
    dataFrame = pd.read_sql_query('SELECT ' + SearchField + ' FROM ' + SearchTable, con)
    #Convert pandas data frame to list
    ListToSearchIn = dataFrame[SearchField].tolist()
    #Fuzzy search results in list
    SearchResults=process.extract(TextToSearch, ListToSearchIn, scorer=fuzz.token_set_ratio)
    FuzzySearchScore = SearchResults[0][1]/100
    FuzzySearchScoreDescription = str(SearchResults[0][1]) + '% Fuzzy Match with ' +
    SearchTableName + ' ' + SearchField + ': ' + str(SearchResults[0][0])
    #Close the DB connection
    con.close()
    return FuzzySearchScore, FuzzySearchScoreDescription

#Execute Rule-based Analytics to detect potential fraudulent employees
def EmployeeFraudChecks (employee_data):
    #Variables to store the results of fraud tests performed
    EmpFullName = str(employee_data['FirstName']) + " " + str(employee_data['LastName'])
    #Get fuzzy match Levenshtein distance threshold value
    LevenshteinDistanceThreshold = GetFraudRulesetValue('EmpFuzzyMatchLevenshteinDistanceThreshold')[0]
    #Check for fuzzy matches in employee name and vendor name
    VendorRegisteredNameFuzzyTest = FuzzyMatchChecker('RegisteredName','VendorApp_Vendor',EmpFullName)
    if VendorRegisteredNameFuzzyTest[0] >= LevenshteinDistanceThreshold :
        EmployeeFraudTestResult["VendorRegisteredNameFuzzyTest"] =
        [VendorRegisteredNameFuzzyTest[0],"Positive",VendorRegisteredNameFuzzyTest[0],
        "Employee Name had " + VendorRegisteredNameFuzzyTest[1]]
    else:
        EmployeeFraudTestResult["VendorRegisteredNameFuzzyTest"] =
        [VendorRegisteredNameFuzzyTest[0],"Negative",VendorRegisteredNameFuzzyTest[0],"N/A"]
    return EmployeeFraudResultScore
```

I.2: Code Snippet Implementing Outlier Analytics by Calculating the Z-Score

```
#Calculate the Z-Score to determine outlier in the Vendor Country Field
def VendorCountryOutlierChecker (VendorRecord, isNewVendor):
    # Read sqlite query results into a pandas DataFrame
    con = sqlite3.connect("D:/System Implementation/FDSWebServerProject/db.sqlite3")
    CountryNameToSearch =str(VendorRecord['Country'])
    ZScore = 0
    #Retrieve Country Profile for Vendor Being Added/Updated
    dfCountryProfile = pd.read_sql_query("SELECT ... FROM LookupApp_Country, con)
    CountryProfile = dfCountryProfile['CountryProfile'][0]
    #Retrieve List of Vendors with respective Country Profile
    dataFrame = pd.read_sql_query(SELECT ... FROM VendorApp_Vendor, con)

    if isNewVendor == True :
        #Add new record in dataframe - For Creation Events
        dataFrame.loc[len(dataFrame.index)] = [0, CountryNameToSearch, CountryProfile]
        #Calculate the Z-Score based on Country Profile
        z = stats.zscore(dataFrame['CountryProfile'])
        #Add the z-score as a column in the dataframe
        dataFrame['ZScore'] = z
        #Retrieve and save the calculated ZScore
        ZScore =dataFrame.iloc[-1:]['ZScore'].values[0]

    else :
        #Calculate the Z-Score based on Country Profile
        z = stats.zscore(dataFrame['CountryProfile'])
        #Add the z-score as a column in the dataframe
        dataFrame['ZScore'] = z
        #Retrieve and save the calculated ZScore
        ZScore =dataFrame.at[editVendorRecordIndex,'ZScore']

    #Close the DB connection
    con.close()
    return ZScore
```

I.3: Code snippet Protecting Private URL Routes

```
const authenticate = (component) => {
    //Route to private route only if authenticated
    return token ? component : <Navigate to="/login" />;
}
<Header isToken ={token}/>
<Routes>
     <Route element={<WithoutNavigation/>}>
        <Route path='/login' element= {token ? <Home/>:<Login setToken={setToken} />}/>
        <Route exact path='/registeruser' element={<RegisterUser/>}/>
     </Route>
    <Route element={<WithNavigation/>}>
        <Route exact path='/' element={authenticate(<Home/>)} />
        <Route exact path='/employee' element={authenticate(<EmployeePage/>)}/>
        <Route exact path='/newemployee' element={authenticate(<AddEditEmployeeModal/>)}/>
        <Route exact path='/vendor' element={authenticate(<VendorPage/>)} />
        <Route exact path='/fraudalert' element={authenticate(<FraudAlertPage/>)} />
        <Route path ="*" element={authenticate(<NotFound/>)}/>
     </Route>
</Routes>
```

I.4: Code Snippet Creating Tokens and Embedding Tokens in Header Requests

```
#Create hashed passwords & tokens for newly registered users
def create(self, validated_data):
     user= User.objects.create_user(**validated_data)
     Token.objects.create(user=user)
     return user

//Get Employee Data from Back-end API
const fetchEmployeeData = () => {
     axios.get(process.env.REACT_APP_API+'employee', {
         headers: {
              'Authorization': `Token ${token}`
         }
     })
     .then((response) => {
         setEmployees(response.data);
     })
      .catch((error) => {
          console.log(error)
     })
}
```