International Journal of Auditing and Accounting Studies
ISSN: 2582-3272 • Volume 6; Number 2; 2024 : pp. 159-186
https://DOI:10.47509/IJAAS.2024.v06i02.02

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

MACHINE LEARNING ON TRIAL: ASSESSING ITS EFFICACY IN DETECTING FINANCIAL STATEMENT FRAUD

Sara H. Sabry[1*] and Yara Ibrahim[2]

[1]Faculty of International Business and Humanities, Egypt Japan University of Science and Technology, Alexandria, Egypt. E-mail: sara.hussein@ejust.edu.eg
[2]Faculty of International Business and Humanities, Egypt Japan University of Science and Technology, Alexandria, Egypt. E-mail: yara.magdy@ejust.edu.eg
**Corresponding author: sara.hussein@ejust.edu.eg*

Received: 10 March 2024; Revised: 1 April 2024; Accepted: 16 May 2024 : 30 June 2024

***Abstract**:* Machine learning offers the potential to revolutionize financial fraud detection, providing a powerful alternative to the limitations of traditional financial ratio analysis. By employing panel data derived from financial statements spanning from 2017 to 2022, this study investigates the feasibility of implementing these methodologies to augment the efficacy of fraud detection systems specifically tailored to the Egyptian stock market. In contrast to conventional financial ratio analysis, the applied techniques evaluate and contrast three sophisticated machine learning algorithms—namely logistic regression (LR), support vector machine (SVM), and XGBoost. The findings demonstrate a notable disparity in performance when comparing machine learning methods to conventional methodologies using performance metrics indicators, specifically recall, Fi-score, and precision, implying that machine learning can provide more precision. Furthermore, this research indicates that XGBoost routinely exhibits superior performance compared to the alternative approaches in critical fraud detection measures. In brief, the researchers analyzed the ramifications of these findings for accountants and auditors in Egypt, underscoring the importance of employing a sophisticated methodology that integrates machine learning with expert opinion and ample comprehension of the financial reporting environment in Egypt in order to optimize the efficacy of fraud detection. Ultimately, it is the responsibility of auditors and accountants to thoroughly evaluate and select machine learning methods that are ideal in alignment with their specific data characteristics, risk tolerances, and transparency requirements. Furthermore, it is crucial to underline the significance of audit quality: auditors must be aware of how audit quality indicators, such as audit tenure and productivity, might influence the identification of fraudulent activities. Higher risk locations may be acknowledged

To cite this paper:
Sara H. Sabry & Yara Ibrahim (2024). Machine Learning on Trial: Assessing ITS Efficacy in Detecting Financial Statement Fraud. *International Journal of Auditing and Accounting Studies*. 6(2), 159-186. *https://DOI:10.47509/IJAAS.2024.v06i02.02*

by these indications, necessitating a more thorough analysis and the use of more sophisticated analytical tools. This research explores how Machine Learning (ML) can empower auditors in fraud detection. The paper recommends utilizing unsupervised ML to identify anomalies in vast datasets, flagging unusual patterns for investigation. By embracing these ML techniques, auditors can pave the way for a more data-driven and efficient approach to uncovering fraudulent activities.

Keywords: Fraud detection- machine learning- financial ratio detection- logistic regression- support vector- XGBoost.

1. INTRODUCTION

A party or individual that commits fraud does so with the intention of deceiving another party, evading accountability, or causing non-financial or financial harm. Moreover, fraud is the act of a person or group of people gaining an unfair advantage within a company. Associates of the firm, both internal and external, are capable of committing fraud. This notion usually entails fabricating a financial statement to persuade potential investors to put money into the business (Rashid *et al.*, 2022). In a time when financial crimes and regulatory pitfalls are on the rise, corporate fraud is considered a serious risk to the company and its stakeholders.

Statement on Auditing Standards 99 (Auditing and Assurance Standards Board, 2002) defined fraud as an intentional act intended to cause a significant deception in financial reporting with regard to Fraud in a Financial Statement Audit. Theft, fraudulent expenditure, and the falsification of financial documents are a few examples of such crimes. Honest people are susceptible to pressure from the corporate world, which often results in them acting in a way that misrepresents financial statements. Everywhere fraud occurs, it undermines a company's trust, revenue, and reputation. When stakeholders, such as members of the audit committee and board, senior management, employees, auditors, creditors, shareholders, and pensioners, commit fraud, the performance of several firms suffers (Rashid *et al.*, 2022).

Hamal and Senvar (2021) postulate that auditors and decision-makers need advanced analytical tools and procedures, not traditional methods, to detect fake financial statements. There is no established method in the literature for identifying financial accounting fraud. However, financial ratios, such as the Altman Z-score and the Beneish model, were formerly employed as statistical models. Subsequently, researchers utilized data mining methods to identify instances of financial statement fraud. Data mining techniques such as decision trees, logistic regression (LR), and artificial neural networks are utilized.

Additionally, recent research has focused on using hybrid systems that incorporate data mining methods to uncover financial accounting fraud. The research indicates that financial accounting fraud can be identified through techniques for extracting information from financial statements. An in-depth analysis of financial statistics can uncover indications of fraudulent behavior. It is vital to understand the most widely utilized financial metrics to detect financial accounting fraud. Moreover, using numerous financial ratios to identify financial accounting fraud may lead to mistakes in detection.

Business fraud is not a new phenomenon; in fact, it was first made public in 2001 when Enron, one of the largest business bankruptcies ever, failed (Mangala & Kumari, 2017). (Price Waterhouse Coopers (PWC) (2022), out of organizations generating global annual revenues exceeding US$10 billion, 52% encountered fraud in the last 24 months. Among this subset, almost 20% stated that their most disruptive incidence resulted in a financial effect surpassing US$50 million. Less than 100 million dollars in revenue, 38% of smaller organizations were impacted by fraud, with around 25% of them experiencing losses exceeding 1 million dollars.

According to estimates from the (Chartered Institute of Management Accountants, 2009), even while larger organizations are more likely to be affected by economic crime, fraud may nonetheless be more costly for small businesses. Small businesses witnessed an average fraud occurrence of $98,000, compared to major corporations that saw an average fraud event of $105,500. Compared to large organizations, small businesses may experience losses from fraud that are up to 100 times higher per employee. Fraud also has negative effects that go beyond only the immediate loss of money. Collateral damage can include harm to a company's reputation, branding, personnel morale, and external business relationships (Bierstaker *et al.*, 2006).

Financial transactions lay the basis of contemporary society. Unfortunately, there is widespread exploitation of the illicit financial system. Fraud controls seek to identify these questionable behaviors, but in order to fully understand their worth and effectiveness, a thorough examination is necessary. This research is frequently carried out in retrospect because of the size and confidential nature of these financial transactions. Because of the concealed fraud problem, financial institutions lack the information necessary to configure and fine-tune their fraud management systems. New techniques for fraud detection systems that can handle big datasets more effectively and efficiently are of interest to many enterprises (Xu *et al.*, 2023). Accordingly, the purpose of this paper is to investigate new methods to detect fraud beyond financial ratios and overcome the obstacles accountants and auditors face with such voluminous datasets. This research focuses only on three of those methods: LR and simple vector method using machine learning tools.

The remaining research is organized as follows: The initial part consists of a literature study on machine learning methods for fraud detection. The study's research approach is outlined in the second section. The results related to the three research topics are reported in the third part. The fourth section provides a conclusion and recommendations for practitioners on detecting fraud in financial reports.

2. LITERATURE REVIEW

The next section is divided into three sub-sections, the first of which displays prior literature concerning fraud detection in the context of financial ratios. The second sub-section displays the machine learning tools used in the accounting context. In contrast, the final and third sub-sections reveal the use of machine learning tools in fraud detection with regard to financial statements.

2.1. Fraud Detection in the Financial Context

The purpose of financial statements is to fairly report the company's cash flows, operating performance, and financial status. The rationale behind this is that the data shown in financial statements serves as the foundation for choices made by governmental organizations, investors, and other stakeholders regarding the course of a company. The worldwide rules on auditing, however, state that management's ability to falsify financial statements and modify accounting records by subverting safeguards that otherwise seem to be working well puts it in a unique position to commit fraud. Consequently, it is imperative to examine the various techniques for identifying fraudulent activity in financial accounts (Kanapickienė & Grundienė, 2015).

Some earlier research studies have developed models for the identification of fraud; almost all of them are based on the analysis of financial ratios for fraud discovery. According to these research ratios, it is a useful technique for determining market loss and evaluating performance. Some academics also put out a number of financial ratios, including financial leverage, profitability, asset composition, liquidity, and examples of such measurements that may subtly include fraud. Some empirical researchers have recently developed several models that analyze financial loss and identify fraud using financial ratios; many of these models tend to predict various market occurrences, such as fraud, manipulation of earnings, control of earnings, and bankruptcy

(Dalwai *et al.*, 2021). Examples of these models include the Dechow F-score model, the Altman Z-score model, the Beneish model, and the Jones (Huerta & Jensen, 2017). From the above discussion, financial statements provide crucial information for investors and other stakeholders to make decisions, but traditional methods of detecting fraud in these statements may not be enough. The following subsection explores alternative techniques to identify fraudulent activity in financial statements.

2.2. Fraud Detection with Machine Learning Tools

Financial data is becoming increasingly complicated, requiring advanced detection technologies to keep up with the growing strategies of fraudsters. Researchers have utilized machine learning (ML) methods to automate and potentially objectively detect fraudulent activities in financial accounts. This literature review investigates how previous studies have examined the use of different machine learning algorithms in fraud detection. We analyze the algorithms used and their efficacy in detecting fraudulent patterns and compare the performance of these machine-learning models with older approaches. We examine the aspects that impact the success of these methods in the specific area of detecting financial statement fraud.

Hamal and Senvar (2021) studied Turkish SMEs at risk of fraud and their creditor banks to assess the effectiveness of machine learning classifiers in detecting financial accounting fraud. Analyzed financial statements of 341 Turkish SMEs from 2013 to 2017 through data pre-processing and feature selection methods to identify key financial ratios impacting fraudulent financial statements. Evaluated and compared seven classifiers: SVM, Naive Bayes, artificial neural network, K-nearest neighbor, random forest, LR, and bagging) using performance metrics. Researchers found that the random forest oversampling model, without feature selection, outperforms all other models.

Moreover, Kaminski *et al.* (2004) used univariate analysis (paired sample T-test) to compare 21 financial ratios of two groups over seven years to examine the usage of ratios for detecting financial accounting fraud. According to the study, 16 out of 21 financial ratios were found to be statistically significant across the seven years. Four ratios were significant for both periods, while nine ratios were significant for only one time period. Three ratios were significant for three distinct periods. Discriminant analysis identified financial accounting fraud, with misclassifications for fraud firms ranging from 58 percent to 98 percent. It was concluded that certain financial statistics have little ability to detect financial accounting fraud.

Current accounting research focuses on assessing the efficiency of various statistical and machine learning methods, like LR and artificial neural networks (ANN), to enhance the identification of financial statement fraud. This research is essential due to the distinctiveness of financial statement fraud for every organization. The domain in question is characterized by a low ratio of fraud to non-fraud firms (high-class imbalance), a low ratio of false positive to false negative misclassification costs (high-cost imbalance), attributes that are noisy and can indicate both fraudulent and non-fraudulent activities and intentional actions by fraudsters to hide fraud by making fraud firm attributes resemble non-fraud firm attributes. It is uncertain if statistical and machine learning techniques, often known as classification algorithms, that perform well in other areas can do well in detecting financial statement fraud due to specific characteristics. Therefore, there is a need for research focused on detecting financial statement fraud (Bussmann *et al.*, 2021; Perols, 2011; Rashid *et al.*, 2022).

From the above discussion, prior research emphasizes the continuous investigation of several methodologies to identify fraudulent financial statements. These methodologies encompass conventional statistical techniques as well as more contemporary machine-learning approaches. Although both methodologies exhibit promise, scholarly investigations indicate that financial fraud data possesses distinct attributes—including cost asymmetry and class imbalance—that require specific procedures to ensure maximum performance. Even though several techniques have shown encouraging outcomes, it is essential to acknowledge the particularities of the Egyptian environment and, by extension, the financial reporting context.

In light of the numerous obstacles identified in the prior research, this study emphasizes the need for specialized research on the detection of financial statement fraud in Egypt that employs more sophisticated techniques and moves beyond conventional tools in order to alleviate the complexities of financial reporting in the Egyptian culture. The objective of this study is to fill this void through the development of precise research questions that are tailored to the complexities of detecting financial statement fraud in the Egyptian environment. Thus, the research questions are formulated as follows:

RQ1: To what extent are machine learning techniques well-suited for identifying fraud in the context of financial statements in terms of performance evaluation?

RQ2: How do the three machine learning techniques provide differing outcomes?

RQ3: What are the financial ratio determinants that significantly predict fraud detection in each of the three machine learning methods?

While traditional financial ratio analysis has been a cornerstone of fraud detection, its limitations are increasingly recognized. Existing research suggests promise in employing machine learning (ML) techniques for more comprehensive and nuanced analysis of financial data. This study builds upon this foundation by exploring the efficacy of specific ML algorithms in identifying fraudulent financial statements. The following section details the research methodology employed to evaluate the effectiveness of these ML models in a real-world context.

3. METHODOLOGY

3.1. Sample and Data Collection

To test the research hypotheses and answer the research questions, this research relies on an empirical investigation of panel data for firms that are listed in EGX 100 from 2017 to 2022. The Egyptian Exchange (EGX) is a vital part of Egypt's financial landscape. The EGX 100 is a key index tracking the performance of the 100 most active listed companies. By focusing on the EGX 100, this research delves into the financial health and activities of some of Egypt's most influential corporations (Appendix A: sample by sector). Examining these leading companies provides valuable insights into the broader Egyptian economy and its key players. The study has two stages. The first stage involves the pre-processing of the data under investigation; the initial step involves conducting feature selection procedures and financial ratios calculations for the research sample based on certain selection criteria highlighted below (Ezat, 2019; Salehi *et al.*, 2020):

1. Firms should be at least for one year in the period of investigation from 2017 to 2022.
2. Financial and non-bank financial firms are excluded from the sample due to their special nature, which differs from that of non-financial firms in Egypt.
3. Firms that prepare their financial statements in U.S. dollars are also excluded; thus, only Egyptian currency financial statements are included in the sample.
4. Firms with missing financial statements are excluded from the sample.

The initial sample consisted of 102 firms for six years = 612, excluding banks (10), non-bank financial firms (14), and firms whose currency is U.S. dollars (1), as shown in the table below (1). Also, table (2, Appendix A) represents the firms included in the study by the industrial sector. The second stage involves assessing and comparing the performance of classifiers. Three classifiers - SVM, LR, and XGboost - are compared in the second stage using performance measures.

Table 1: Sample Selection

	Firm	*Firm year observation*
Initial Sample	240	1440
Exclude banks	15	90
Exclude non-bank financial firms	34	204
Exclude firms operating in U.S dollars	1	6
Exclude unavailable reports	6*	360
Final Sample	184	780

*The number of excluded reports is 6 (6 * 6 years= 36 only from 360 observations) firms with no reports available, while the remaining 324 observations stem from the fact that firms have some reports available for some years while the others are not; thus, not all the firm's reports were excluded.

The research methodology employed in this study consists of three distinct phases:

Phase one: Performance evaluation of Machine Learning. In this study, we conduct a comprehensive evaluation of the fraud detection capabilities of three machine learning algorithms that were utilized: LR, SVM, and XGBoost. The focus of this review is to assess the strengths and limitations of each approach in detecting fraudulent financial activity through the comparison of performance measures. During this phase, we utilized recall, accuracy, and F-score measures to effectively assess the performance of the three machine learning algorithms (SVM, XGBoost, and LR) in detecting false financial statements. The metrics will offer valuable information into the capacity of each approach to effectively detect fraudulent activity (recall), prevent the misclassification of legitimate transactions (precision), and attain a harmonious integration of both (recall and precision) (F-score).

Phase two: Financial ratios against Machine Learning: This stage involves a comprehensive evaluation of the limitations of conventional financial ratio analysis in comparison with the three machine-learning technologies utilized for the purpose of detecting fraud. Our primary objective is to ascertain if machine learning provides more effective predictive capabilities than conventional ratio analysis or if a smart integration of both approaches might yield improved outcomes. The three comparisons were as follows: Comparison one was conducted between the traditional method results (fraud versus non–fraud) and results (fraud versus non–fraud) of LR. Comparison two was conducted between the results of the traditional method (fraud versus non–fraud) and the results of the SVM method (fraud versus non–fraud). Comparison three was conducted between the traditional method results (fraud versus non–fraud) and results (fraud versus non–fraud) of the XGBoost method.

Phase three: Identification of key determinants The ultimate stage aims to shed light on the fundamental factors that exhibit a robust correlation with fraudulent conduct within businesses. Our objective is to enhance the accuracy of preventative measures and refine fraud detection algorithms with a comprehensive understanding of these foundational drivers.

3.2. Measurement of Research Indicators and Measurement

The next section discusses the research indicators and their measurement.

Fraud Detection Method: The following section presents the three methods utilized in this research to detect fraudulent financial statements:

The first stage of this research relies on five financial statement indicators calculated based on the 47 financial sub-indicators adapted from previous literature (Perols, 2011). The five financial ratios are calculated as follows:

Table 2: Indicators Measurement

Indicator	*Measurement*	*Source*
Audit Turnover	Measured as a dummy variable, taking 0 if the auditor did not change from the previous year and 1 if the auditor changed	(Perols, 2011)
Big 4 audit firms	This is measured as a dummy variable, taking 1 if the firm was audited by one of the Big 4 audit firms (PWC, KPMG, Deloitte and Touch, EY, and Accountability State Authority).	(Perols, 2011)
Total Accruals	$1.2*\left(\frac{Working\ Capital}{Total\ Assets}\right)+1.4*\left(\frac{Net\ Income}{Total\ Assets}\right)+3.*\left(\frac{EBIT}{Total\ Assets}\right)+0.6 *\left(\frac{Equity}{Total\ Assets}\right)+1.0*\left(\frac{Retained\ Earnings}{Total\ Assets}\right)$ If total accruals are above 1.81, this is an indicator of potential fraud, thus taking the value of 1, and less than 1.81 will take the value of 0 for a non-fraud indicator.	(Mahama, 2015)

Indicator	*Measurement*	*Source*
Unexpected employee productivity	$\frac{Firm\ ROE_t - Firm\ ROE_{t-1}}{Firm\ ROE_{t-1}} - \frac{Industry\ ROE_t - Industry\ ROE_{t-1}}{Industry\ ROE_{t-1}}$ If unexpected employee productivity calculated above is positive, it is an indicator of fraud; thus, it would be given the value of 1. If the value is below 0, it is an indicator of non–fraud; thus, it would be given the value of 0.	(Perols, 2011)
Change in A/R	$\frac{Accounts\ Receivables - Accounts\ Receivables_{-1}}{Accounts\ Receivables_{-1}}$ This is interchanged with a dummy variable based on the following scale: 1, which indicates fraud when the change in accounts receivables is above 1.1, and no fraud when it is less than 1.1.	(Perols, 2011)

Machine learning is a branch of artificial intelligence that enables systems to gain knowledge from data and gradually become more efficient at a given activity. It entails the creation of algorithms that provide computers with the ability to see patterns in data, anticipate outcomes, and make judgments. Machine learning algorithms can identify complex patterns and anomalies that conventional methods may miss by utilizing enormous volumes of past transaction data (Noviandy *et al.*, 2023). In the subsections that follow, the theoretical foundation of the classifiers employed in this work is outlined.

(A) Logistic Regression (LR) Machine Learning

LR is a common algorithm for classifying data into two categories. It estimates the probability of a particular outcome using the entered variables. The model presents the relationship between the input variables and the log odds of the target variable using a logistic function. The output is restricted between values 0 and 1, allowing it to be interpreted as probability. LR is simple, efficient, and provides interpretable results, which makes it valuable for classification problems and has proven its performance with regard to financial statements fraud (Chen, 2016; Lin *et al.*, 2003; Perols, 2011).

(B) Support Vector Machine Learning (SVM)

SVM, introduced by Vapnik in 1995, is a machine learning method that relies on structural risk minimization and statistical learning theory. SVM determines the best-separating hyperplane to categorize several classes of data through a learning process. By reducing classification errors and increasing the geometric margin of the decision border to a set of points, SVM typically learns a binary linear decision function (BAO *et al.*, 2020). Additionally, the goal of SVM is to locate the optimal separation hyperplane in the feature space in order to maximize the positive and negative sample intervals on the training set. SVM can also be utilized to address non-linear problems (Wang *et al.*, 2018). Hajek *et al.* (2023) highlighted that SVM is a classifier that works especially well for detecting financial fraud since it can handle high-dimensional data.

(C) XGBoost Machine Learning

A robust and popular machine learning algorithm, XGBoost is well-known for its accuracy and efficiency in managing a wide range of data kinds and complexity. It belongs to the class of ensemble learning techniques known as gradient boosting frameworks, which pool the predictive strength of several models to produce a more powerful forecast. Gradient boosting is the foundation of XGBoost, which refines the technique by emphasizing regularization and iteratively fixing the mistakes created by the previous models. It adds decision trees to an ensemble one after the other. Through this iterative approach, XGBoost's prediction performance can be enhanced over time. The goal of XGBoost is to identify the best group of weak learners that, when added together, create a powerful prediction model. This is accomplished by measuring the discrepancy between expected and actual data and minimizing a loss function (Noviandy *et al.*, 2023; Wang *et al.*, 2018).

4. EMPIRICAL FINDINGS

The following analysis unfolds in three distinct phases: Phase One includes performance evaluation of machine learning. We commenced by conducting a thorough comparison of the fraud detection functionalities of the three machine learning technologies that were utilized: XGBoost, LR, and SVM (Alsuwailem *et al.*, 2023; Bussmann *et al.*, 2021). This assessment reveals the comparative merits and drawbacks of each in terms of detecting fraudulent financial transactions. Phase two then follows, which entails financial ratios against machine learning. This stage evaluates the performance of conventional financial ratios in comparison with our three machine-learning technologies designed for fraud detection. Our objective is to determine whether machine learning technologies provide more accurate predictions or whether they may be utilized in conjunction with one another. Ultimately, we reach phase three: identifying crucial determinants. The ultimate phase is to shed light on the fundamental indicators that indicate fraudulent behavior inside an organization. By comprehending these fundamental drivers, it is possible to refine models for detecting fraud and customize preventative actions with greater accuracy.

4.1. Phase One: Performance Evaluation

Table 3: Confusion Matrix

Prediction/ Target	*Positive*	*Negative*
Positive (fraudulent financial report)	True-Positive (*TP*)[a]	False-positive *(FP)*[b]
Negative (non-fraudulent financial report)	False-negative *(FN)*[c]	True-negative *(TN)*[d]
a True positive (TP): (actual positives) fraudulent financial statements are correctly identified by the model. b False positive (FP): The model incorrectly classifies the proportion of actual negatives. Non-fraudulent financial statements are identified as positives (fraudulent financial statements). c False negative (FN): The model incorrectly classifies the proportion of actual positive fraudulent financial statements as negative non-fraudulent financial statements. d True negative (TN) is the proportion of actual negatives: non-fraudulent financial statements correctly identified by the model.		

The uncertain distribution and impact of all fraudulent transactions is an inherent difficulty in financial fraud detection that needs to be addressed, as previously mentioned in the study (Lopez-Rojas & Barneaud, 2019). Existing fraud detection techniques rely on conventional measurements of classification performance when there is not a sufficient way to measure fraud detection performance. The capacity to recognize fraudulent transactions with accuracy is the most desired performance metric (true positive rate). Furthermore, one important feature of fraud detection systems is minimizing false positive and false negative transaction rates (refer to the confusion matrix in Table 3), particularly in an evolving fraudulent context (Lopez-Rojas & Barneaud, 2019).

Table 4: Performance Evaluation for the 3 methods

Logistic Regression					*Support Vector*					*XGBoost*				
	P	*R*	*Fi*	*S*		*P*	*R*	*Fi*	*S*		***P***	***R***	***Fi***	***S***
0	.94	.95	.94	179	**0**	.92	1	.96	179	**0**	.96	1	.98	181
1	.36	.31	.34	16	**1**	.0	.0	.0	16	**1**	1	.50	.67	14
Acc			.90	195	**Acc**			.90	195	**Acc**			.96	195
WA	.89	.9	.89	195	**WA**	.84	.92	.88	195	**WA**	.97	.96	.96	195
P = precision, R= recall, Fi= Fi score, S= Support, WA = weighted average fi-score														

Source: Python output results

In this research, fraud detection models were assessed using standard classification metrics. The number of financial statements accurately classified as fraudulent as a percentage of all fraudulent financial statements is known as the true positive rate or recall.

$$Recall = \frac{TP}{TP + FN}$$

Referring to (Table 4), with regards to the first performance evaluation indicator, recall for LR= .95, which indicates that LR was able to detect 95% of the non-fraudulent financial statements out of the total non-fraudulent financial, where there is a 5% of non-fraudulent financial statements were not identified as non-fraudulent reports. In the same vein, this method was able to detect only 31% of fraudulent financial statements. These results are consistent with prior literature (Alsuwailem *et al.*, 2023; Bussmann *et al.*, 2021; Zhou *et al.*, 2018). As for SVM and XGBoost, both = 1.00, they indicated that both methods were able to detect 100% of non-fraudulent financial statements and 0% of fraudulent reports identified as non-fraudulent reports. However, the SVM was able to detect 0% of fraudulent reports, while XGBoost detected 50% of fraudulent reports.

With regards to the second performance evaluation indicator, precision is the number of financial statements correctly identified as fraudulent, which is a percentage of all financial statements that are expected to be fraudulent. Financial institutions work to both comply with regulations and lower the risk of fraud; nevertheless, since FN is uncertain, it is challenging to assess Recall in the real world (hidden fraud). Consequently, financial organizations can only compute Precision:

$$Precision = \frac{TP}{TP + FP}$$

Referring to the precision of the three methods, with regard to LR precision, it is reported to be .94 with regards to detecting non-fraudulent financial statements, indicating that 94% of fraud incidents and non-fraudulent reports were correctly detected. In contrast, only 36% of the fraudulent reports were detected. As for the support vector, non-fraudulent reports were detected with 92% precision and 0% for fraudulent reports. Finally, XGBoost was able to detect 96% of non-fraudulent reports, and 100% of fraudulent reports were detected (Alsuwailem *et al.*, 2023; Bussmann *et al.*, 2021; Zhou *et al.*, 2018).

Finally, with regards to the third performance evaluation indicator, The Fi-score, which is the harmonic mean of recall and precision (Alsuwailem *et al.*, 2023; Bussmann *et al.*, 2021; Schlör *et al.*, 2021), has also been considered in previous research because it aims to solve the binary classification problem (fraudulent reports and non-fraudulent reports) by combining recall and precision. After all, developers typically have to choose between recall and precision. Fi-score ranges from 0 to 100%, whereas higher Fi-score results in better model performance.

$$Fi-score=2*\frac{Recall*Percision}{Recall+Percision}$$

In terms of LR, the model's fi-score was 94%, which indicates that it performed adequately in identifying financial statements that were not fraudulent but only 34% in identifying those that were fraudulent. Regarding the support vector, its fi-score of 96% indicates an effective performance when identifying reports that are not fraudulent. In comparison, its fi-score of 0% indicates a low performance when identifying fraudulent reports. Lastly, when it comes to the XGBoost technique, The fi-score for identifying non-fraudulent reports is stated to be 98%, suggesting a strong performance in this regard. For fraudulent reports, the fi-score is reported to be 67%, indicating a moderate performance that is greater than both SVM and LR—taking a closer look at all scores using weighted average fi-score to overcome the problem of imbalance samples (fraudulent versus non-fraudulent samples).

$$Macro\ \ Fi-score=\sum_{i=1}^{n}w_i*Fi-score$$

Where:

$$w_i=\frac{Number\ of\ samples\ in\ class}{total\ number\ of\ samples}$$

According to the research results and considering the imbalance in the research sample, the LR weighted average fi-score was 89%, while 88% for the SVM and finally for XGBoost = 96%. Those results signify that XGBoost reported a superior performance compared to both the other two methods. Those results are similar to results reported by (Hajek *et al.*, 2023; Noviandy *et al.*, 2023; Wang *et al.*, 2018; Zhou *et al.*, 2018).

From the above findings, it is revealed that some machine-learning tools outperform others. In this research and with regard to financial statement fraud detection, XGBoost outperforms both LR and SCV. Logistic Regression, while interpretable and efficient, might need help with complex non-linear relationships within financial data. SVM excels at identifying clear boundaries between fraudulent and legitimate statements but can be less effective with intricate data patterns. XGBoost's strength lies in its ensemble approach, combining multiple weaker decision trees to create a robust model. This feature allows XGBoost to capture the nuances and non-linearities present in financial data, potentially leading to superior performance in detecting fraudulent activity compared to Logistic Regression and SVM.

4.2. Phase two: Comparative Analysis

Subsequently, the research expands its scope by comparing conventional financial ratio analysis with the three machine learning technologies that were implemented: XGBoost, LR, and SVM. The purpose of this comparison study is to underscore the merits and drawbacks of each methodology in the realm of financial statement fraud detection (FSFD). By analyzing the outcomes, it is possible to determine which approach demonstrates the greatest overall precision. This comparison aids in identifying the method that yields the most dependable outcomes in the identification of fraudulent actions occurring inside financial statements. The possible complementarity of these techniques is analyzed similarly. By understanding the distinct characteristics and strengths of each method, we can explore the possibility of combining them strategically to create a more robust and comprehensive fraud detection system.

4.2.1. Descriptive Statistics Results

Table 5: Descriptive Analysis

	Tot.Acc	*UNEXEMPR*	*Δ in A/R*	*Big-4*	*Aud. Turn*	*Fin.Rat*	*Log. Reg*	*SVM*	*XGBoost*
count	780	780	780	780	780	780	780	780	780
mean	8.10	-2.587	380263	0.708	0.201	0.079	0.047	0.001	0.044
std	98.7	86.950	36411391	0.455	0.401	0.271	0.213	0.036	0.204
min	-12.5	-1013.96	-555044030	0.000	0.000	0.000	0.000	0.000	0.000
25%	0.64	-0.976	-0.262	0.000	0.000	0.000	0.000	0.000	0.000
50%	1.21	-0.082	0.000	1.000	0.000	0.000	0.000	0.000	0.000
75%	1.87	0.610	0.298	1.000	0.000	0.000	0.000	0.000	0.000
max	2320	1659	851368787	1.000	1.000	1.000	1.000	1.000	1.000

Source: Python output results

From the above table (5), it is clear that the four methods used in fraud detection were measured using a dummy variable, thus having minimum values of 0 and a maximum of 1. The mean of the LR method was 0.047, while that of the SVM was 0.001, and that of XGBoost was 0.044. Finally, the financial ratios have a mean of 0.079. On the other hand, the standard deviation of the four methods ranged from 0.036 to 0.271. Firms included in this research have a mean for total accruals of 8.101 and a standard deviation of 98.729.

Similarly, unexpected employee productivity has a mean of -2.587 and a standard deviation of 86.950. As for the change in accounts receivable, it has a mean of 380263 and a standard deviation of 36411391. As such, the firms that Big audits- 4 audit firms have a mean of 0.708 and a standard deviation of 0.455. Finally, the mean of the audit turnover is 0.201, and its standard deviation is 0.401.

4.2.2. Results of Comparative Analysis

Following prior research, financial ratios have proven their accuracy in detecting potential fraud in financial statements (BAO *et al.*, 2020b; Chen, 2016; Hamal & Senvar, 2021; Mahama, 2015; Perols, 2011). Thus, in the next phase of this research, after assessing the performance of each of the methods and following (Hamal & Senvar, 2021), a comparative investigation was conducted to examine whether there is a significant difference between the fraud detection methods as follows: the first step involved a comparison between the four research samples and to highlight whether there is a significant difference between three methods of fraud detection or not, the Kruskal Wallis non-parametric[1] The test is utilized, which fulfilled the statistical tool assumption from the number of samples being tested and normality as well. The purpose of the second analysis was to evaluate and further investigate the differences between each of the two samples and thus involved four sub-comparisons using financial ratios methods as a benchmark for comparison, comparison (1): financial Ratios methods versus LR, comparison 2: Financial Ratios methods versus support victor and comparison (3) Financial Ratios methods versus XGBoost method) Moreover, a non-parametric Mann-Whitney test was applied using Python programming language version 3.12.

4.2.2.1. Results of Kruskal Wallis

The researcher in this section performed a preliminary analysis to investigate whether there is a significant difference between the machine learning fraud detection methods used in this study.

Table 6: Kruskal Wallis Analysis

	Source	*ddof1*	*H*	*p-unc*
Kruskal	Method	3	58.65021	0.000

Source: Python output results

From the above table (6), the Kruskal-Wallis test[2] revealed statistically significant differences (H = 58.65, p < .000) in FSFD across the three groups defined by the independent variable fraud detection methods. This finding supports our research hypothesis, which predicted that the three machine-learning fraud detection methods would yield different results. This result aligns with previous researchers (BAO *et al.*, 2020a; Hamal & Senvar, 2021), who also found that there is a significant difference between machine learning fraud detection methods in the detection of financial statement fraud, suggesting that each machine learning tool employed in fraud detection vary in their underlying algorithms, strengths, weaknesses, and suitability for specific types of fraud patterns and datasets. These results highlight the potential influence of machine learning methods in financial statement fraud and justify more research into the fundamental processes that are responsible for these observed variations.

4.2.2.2. Results of Mann Whitney

The current section presented the results of the research's main analysis conducted to compare the three machine learning methods employed and the financial ratios method as a benchmark, where comparison one included an investigation of the difference between the financial ratios method and the LR method. Comparison two, however, examined the difference between the financial ratios and the SVM method, while comparison three tested the difference between the financial ratios method and the XGBoost method.

Table 7: Mann Whitney Results

Statistics	*Comparison 1*		*Comparison 2*		*Comparison 3*	
	Financial Ratios	*Logistic Regression*	*Financial Ratios*	*Support Vector*	*Financial Ratios*	*XGBoost*
Mann Whitney Result (statistic)	313950.0		327990.0		315120.0	
p-value	0.009		0.000		0.003	

Source: Python output results

As for comparison one, the researcher relied on the Whitney test[3] to investigate whether there is a significant difference between the financial ratios method and the LR method. Table (7) reveals that there is a significant difference between the financial ratios method and the LR (z = 313950.0, Sig. = 0.009). As for comparison two, the researcher continued in the same line stream to investigate the difference between the financial ratios method and the SVM method. Results reveal that there is a significant difference between the financial ratios method and the SVM method (z = 327990.0, Sig. = 0.000). As for comparison three, the researcher continued in the same line stream to investigate the difference between the financial ratios method and the XGBoost method. Results reveal that there is a significant difference between the financial ratios method and the XGBoost method (z = 315120.0, Sig. = 0.003). The Mann-Whitney tests identified statistically significant variations in the performance of fraud detection across all three machine-learning approaches. In light of these findings, it is critical to emphasize that no single strategy is universally superior in all fraud detection cases. The ideal selection is contingent upon several considerations, including the particular attributes of the dataset, limitations in processing resources, and the requirement for findings that are easily interpretable.

4.2. Phase three: Additional Analysis

4.3.1. Regression Analysis Results

4.3.1.1. Model One

Machine learning methods, namely LR, provide a significant asset in the continuous effort to combat financial fraud. Regarding their capacity to improve the fraud detection capabilities of a model, however, financial ratios vary considerably. In the LR framework for identifying fraudulent behavior in financial statements, this study seeks to shed light on the particular financial measures that have the most significant effect. We can enhance the performance of machine learning models by identifying these critical factors, which will allow them to differentiate between authentic and illegitimate financial statements more accurately.

Model 1 specification is as follows:
$Log.Reg_{it} = \beta_0 + \beta_1\ Tot.Acc_{it} + \beta_2\ UNEXEMPR_{it} + \beta_3\ \Delta \text{ in } A/R_{it} + \beta_4\ Big-4_{it} + \beta_5\ Audit.Turn_{it} + \varepsilon$

A notable finding is displayed in 9 Table 8, Appendix E). Among the five financial measures that were analyzed, only two have LR model-based influence on fraud detection that is statistically significant. Furthermore, the LR model indicates that companies that have had audits by the Big-4 audit firms are more readily identifiable as fraudulent (significance = 0.003). This observation implies that there may be a link between the detectability of fraudulent financial reporting and the reputation, and stringent processes associated with Big-4 corporations. The impact of audit turnover, which refers to the frequent replacement of auditors, on the model's fraud detection capability is similarly substantial (significance = 0.000). This observation suggests that during auditor changes, inconsistencies or anomalies can arise, which might increase the visibility of fraudulent practices. Comparing Impact: The standard coefficients for the two key factors, audit company size, and audit turnover, are very close (0.015 and 0.017), suggesting that they both have a comparable level of impact on the fraud detection capabilities of the LR model.

4.3.1.2. Model Two

In an ongoing attempt to identify financial ratios' determinants, machine learning techniques, viz. the SVM method, constitute an indispensable method. However, financial ratios exhibit significant variation with respect to their ability to enhance the fraud detection skills of a model. This study endeavors to elucidate the specific financial metrics that have the most substantial influence inside an SVM framework designed to detect fraudulent activity in financial statements. A more precise differentiation between financial ratios may be achieved by identifying these important elements, hence improving the performance of machine learning models.

Model 2 specification is as follows:

$$SVM_{it} = \beta_0 + \beta_1\ Tot.Acc_{it} + \beta_2\ UNEXEMPR_{it} + \beta_3\ \Delta \text{ in } A/R_{it} + \beta_4\ Big-4_{it} + \beta_5\ Audit.Turn_{it} + \varepsilon$$

Breaking down the key findings highlighted in (Table 9, Appendix E) and expanding them for clarity: Limited impact: Out of the five financial ratios analyzed, only one demonstrates a statistically significant impact on fraud detection using an SVM model. This result suggests that traditional financial ratios have less predictive power when employing this specific machine-learning method. Furthermore, the productivity factor unexpected changes in employee productivity emerge as the key predictor of fraud within the SVM model (significance = 0.000). This finding indicates that anomalies in productivity levels could be an early warning sign of fraudulent activity, warranting closer investigation. As for the coefficient, while the standardized coefficient of 0.0001 may appear minor, it is important to remember that in the context of statistical modeling, even slight variations can hold significance. In this case, the coefficient confirms the relationship between unexpected employee productivity and fraud detection within the SVM framework.

4.3.1.3. Model three

The next section delves deeper into the XGBoost method, aiming to identify the specific financial ratios that exert the most significant influence on its ability to detect fraud within financial statements. By identifying these pivotal factors, we can gain significant knowledge on the distinctive characteristics of this machine learning methodology in contrast to other techniques, such as LR and SVM. By conducting this comparison, we can underscore the subtleties and possible benefits of each approach within the intricate domain of financial fraud detection.

Model 3 specification is as follows:

$$XGBoost_{it} = \beta_0 + \beta_1\ Tot.Acc_{it} + \beta_2\ UNEXEMPR_{it} + \beta_3\ \Delta \text{ in } A/R_{it} + \beta_4\ Big-4_{it} + \beta_5\ Audit.Turn_{it} + \varepsilon$$

Based on the above (Table 10, appendix E), out of the five financial ratio determinants, only one significantly impacts the XGBoost fraud method detection. As such, audit turnover is significant in detecting fraud using the XGBoost method, with significance = 0.000 and a standardized coefficient of 0.2147.

In summary, this research offers valuable insights for real-world financial statement fraud detection. While XGBoost demonstrated superior performance in this study, the effectiveness of each model (Logistic *et al.*) is significantly impacted by specific factors within a company's environment. These findings translate to a more strategic approach to fraud detection. Instead of a one-size-fits-all solution, practitioners should consider the company's unique characteristics. For instance, if a Big 4 audit firm is involved, Logistic Regression might be a strong choice.

Conversely, companies experiencing unexpected fluctuations in employee productivity might benefit more from SVM. This research underscores the importance of understanding these contextual factors when selecting a fraud detection tool. By tailoring the approach to a company's specific environment, practitioners can leverage machine learning to achieve a more robust and effective defense against financial statement fraud.

5. CONCLUSION AND IMPLICATIONS

The findings of our inquiry about the utilization of machine learning to detect instances of financial fraud were quite persuasive. Although there were notable variations in performance among different algorithms, LR demonstrated encouraging prospects. Furthermore, the research unveiled that particular attributes of organizations, such as the size of the audit firms (Big-4), turnover rates (turnover), and unanticipated staff productivity, significantly impact the efficacy of diverse machine learning approaches.

This study provides valuable insights into the efficacy of different machine-learning methods for fraud detection and the importance of specific company attributes in influencing model performance. Results indicate that there is no one optimal algorithm: Performance in detecting fraud differs substantially among machine learning techniques. LR, SVM, and XGBoost demonstrate the potential, yet their efficacy is contingent upon the particular context and dataset at hand. Furthermore, the importance of audit-related factors: Significant explanatory power is possessed by audit firm size (Big-4 companies), audit turnover, and unanticipated changes in staff productivity across many machine learning models; therefore, auditors must carefully analyze these variables.

Concerning the ramifications of the research, algorithm selection is significant: organizations and auditors have to meticulously assess and choose machine-learning techniques that are in optimal accordance with their particular data attributes, risk appetites, and requirements for openness. Additionally, auditors must be aware of the impact that audit quality indicators, such as the size of the audit firm and audit turnover, have on the discovery of fraudulent activities. These indicators may point to heightened regions of risk that require more rigorous examination and more advanced analytical instruments. In conclusion, integrating machine-learning methodologies with conventional auditing approaches and taking into account various audit-related elements provides a more resilient strategy for identifying fraudulent activities. Furthermore, investigations into machine learning to detect fraud in financial statements have yielded noteworthy findings, underscoring the criticality of data quality and the necessity for prudence in the application of such technologies. Research undertaken in Egypt, for example, has provided evidence that the efficacy of different machine-learning approaches may vary depending on the quality of data contained inside financial statements. This research accentuates the significance of data quality in attaining precise and dependable outcomes. Although machine learning has much promise in the realm of fraud detection, accountants and auditors must maintain a watchful awareness of the constraints associated with these instruments. Possible biases in the data used to train the models or the inherent complexity of financial statement analysis, both of which these algorithms may not fully reflect, are examples of such limitations. Hence, professionals must exercise prudence and integrate the capabilities of machine learning with their financial acumen to guarantee thorough and precise fraud detection.

In relation to future investigations, a more comprehensive examination of the dynamic relationship among audit quality, financial ratios, and machine learning would significantly enhance our comprehension of fraud prevention. Additionally, hybrid approaches and ensemble strategies that capitalize on the respective capabilities of distinct algorithms may provide better predictive capability. Further research can determine the origins of model performance fluctuation. This may entail the identification of supplementary financial ratios or non-financial variables that exhibit a correlation with fraudulent operations. Ultimately, more studies may explore the potential benefits of integrating machine learning models (such as LR, SVM, and XGBoost) into ensemble models to improve the accuracy and resilience of fraud detection.

Acknowledgment

The authors are grateful to the anonymous reviewers for their helpful comments and to the editor of the journal for thoroughly editing this paper. However, for any errors, we owe the responsibility.

Declaration of conflict of interest

There exists no ethical issues bothering the study and sponsorship regarding funding and related issues of contradictions.

Notes

1. A non-parametric test was used after testing the research data for normality, and data was not normally distributed where Kolmogorov-Smirnov Z = .633 at significance = .818 (Appendix B).
2. Python output for the Kruskal Wallis test appendix C
3. Python output for the Mann Whitney test Appendix D

Reference

Alsuwailem, A. A. S., Salem, E., & Saudagar, A. K. J. (2023). Performance of Different Machine Learning Algorithms in Detecting Financial Fraud. *Computational Economics*, *62*(4), pp.1631–1667.

Auditing and Assurance Standards Board. (2002). *Statements on Auditing Standards 99.* Retrieved from:((https://us.aicpa.org/content/dam/aicpa/research/standards/auditattest/downloadabledocuments/au-00316.pdf).

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research*, *58*(1), pp.199–235.

Bierstaker, J. L., Brody, R. G., & Pacini, C. (2006). Accountants' perceptions regarding fraud detection and prevention methods. *Managerial Auditing Journal*, *21*(5), pp. 520–535.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics*, *57*(1), pp. 203–216.

Chartered Institute of Management Accountants. (2009). *Report to the Nation: Occupational Fraud and Abuse*. Retrieved from: (https://www.imanet.org/podcast/198

Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus*, *5*(1), 89.

Dalwai, T., Chinnasamy, G., & Mohammadi, S. S. (2021). Annual report readability, agency costs, firm performance: an investigation of Oman's financial sector. *Journal of Accounting in Emerging Economies*, *11*(2), pp. 247–277.

Ezat, A. N. (2019). The impact of earnings quality on the association between readability and cost of capital. *Journal of Accounting in Emerging Economies*, *9*(3), pp. 366–385.

Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework. *Information Systems Frontiers*, *25*(5), pp.1985–2003.

Hamal, S., & Senvar, O. (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. *International Journal of Computational Intelligence Systems*, *14*(1), pp. 769-782.

Huerta, E., & Jensen, S. (2017). An accounting information systems perspective on data analytics and big data. *Journal of Information Systems*, *31*(3), pp. 101–114.

Kaminski, K. A., Sterling Wetzel, T., & Guan, L. (2004). Can financial ratios detect fraudulent financial reporting? *Managerial Auditing Journal*, *19*(1), pp. 15–28.

Kanapickienė, R., & Grundienė, Ž. (2015). The Model of Fraud Detection in Financial Statements by Means of Financial Ratios. *Procedia - Social and Behavioral Sciences*, *213*, pp.321–327.

Lin, J. W., Hwang, M. I., & Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, *18*(8), pp. 657–665.

Lopez-Rojas, E. A., & Barneaud, C. (2019). *Advantages of the PaySim Simulator for Improving Financial Fraud Controls.* In Intelligent Computing(edited: Edgar A. Lopez-Rojas, Camille Barneaud). Volume 998 (pp. 727–736).

Mahama, M. (2015). Detecting corporate fraud and financial distress using the Altman and Beneish models.. *International Journal of Economics, Commerce and Management*, *3*(1), pp. 1–18.

Mangala, D., & Kumari, P. (2017). Auditors' perceptions of the effectiveness of fraud prevention and detection methods. *Indian Journal of Corporate Governance*, *10*(2), pp.118–142.

Noviandy, T. R., Idroes, G. M., Maulana, A., Hardi, I., Ringga, E. S., & Idroes, R. (2023). Credit card Fraud detection for contemporary financial management using XGBoost-driven machine learning and data augmentation techniques. *Indatu Journal of Management and Accounting*, *1*(1), pp. 29–35.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *AUDITING: A Journal of Practice & Theory*, *30*(2), pp. 19–50.

Price Waterhouse Coopers (PWC). (2022). *Global Economic Crime and Fraud Survey. Retrieved from: (*https://www.pwc.com/gx/en/forensics/gecsm-2022/pdf/PwC%E2%80%99s-Global-Economic-Crime-and-Fraud-Survey-2022.pdf) .

Rashid, M., Al-Mamun, A., Roudaki, H., & Yasser, Q. R. (2022). An overview of corporate fraud and its prevention approach. *Australasian Business, Accounting and Finance Journal*, *16*(1), pp.101–118.

Salehi, M., Lari Dasht Bayaz, M., Mohammadi, S., Adibian, M. S., & Fahimifard, S. H. (2020). Auditors' response to readability of financial statement notes. *Asian Review of Accounting*, *28*(3), pp.463–480.

Schlör, D., Ring, M., Krause, A., Hotho, A. (2021). *Financial Fraud Detection with Improved Neural Arithmetic Logic Units*. In: Bitetta, V., Bordino, I., Ferretti, A., Gullo, F., Ponti, G., Severini, L. (eds) Mining Data for Financial Applications. MIDAS 2020. Lecture Notes in Computer Science, vol 12591. Springer.

Wang, M., Yu, J., & Ji, Z. (2018). Credit fraud risk detection based on XGBoost-LR hybrid model. *In Proceedings of The 18th International Conference on Electronic Business* (pp. 336-343). ICEB, Guilin, China, December 2-6.

Xu, B., Wang, Y., Liao, X., & Wang, K. (2023). Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, *175*, 114037.

Zhou, H., Chai, H., & Qiu, M. (2018). Fraud detection within bankcard enrollment on mobile device based payment using machine learning. *Frontiers of Information Technology & Electronic Engineering*, *19*(12), pp.1537–1545.

APPENDICES

Appendix A: Sample by Sector

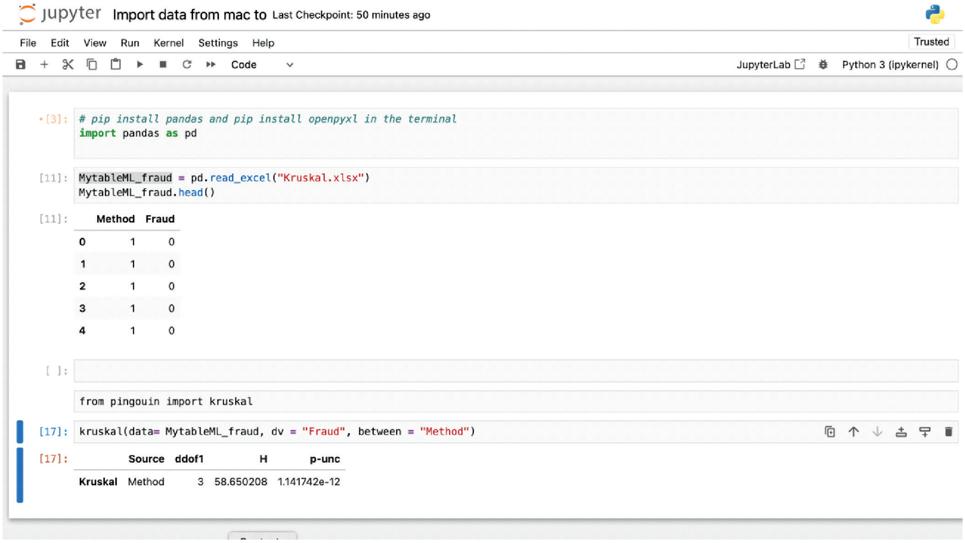
Sample by Sector

Sector	*Firm year observation*
Real estate	145
Industrial Goods, Services, and Automobiles	132
Food, Beverages and Tobacco	111
Health Care & Pharmaceuticals	90
IT, Media & Communication Services	69
Trade & Distributors	55
basic resources	50
Contracting & Construction Engineering	31
Travel & Leisure	25
Textile & Durables	18
Materials	16
Education Services	12
Shipping & Transportation Services	10
Paper & Packaging	10
Utilities	6
Total	780

Appendix B: Kolmogorov-Smirnov Z results

Test Statistics[a]		
		Fraud
Most Extreme Differences	Absolute	.032
	Positive	.000
	Negative	-.032
Kolmogorov-Smirnov Z		.633
Asymp. Sig. (2-tailed)		.818

Appendix C: Python output for Kruskal Wallas test



jupyter Import data from mac to Last Checkpoint: 50 minutes ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

```
•[3]: # pip install pandas and pip install openpyxl in the terminal
      import pandas as pd

[11]: MytableML_fraud = pd.read_excel("Kruskal.xlsx")
      MytableML_fraud.head()
```

[11]:

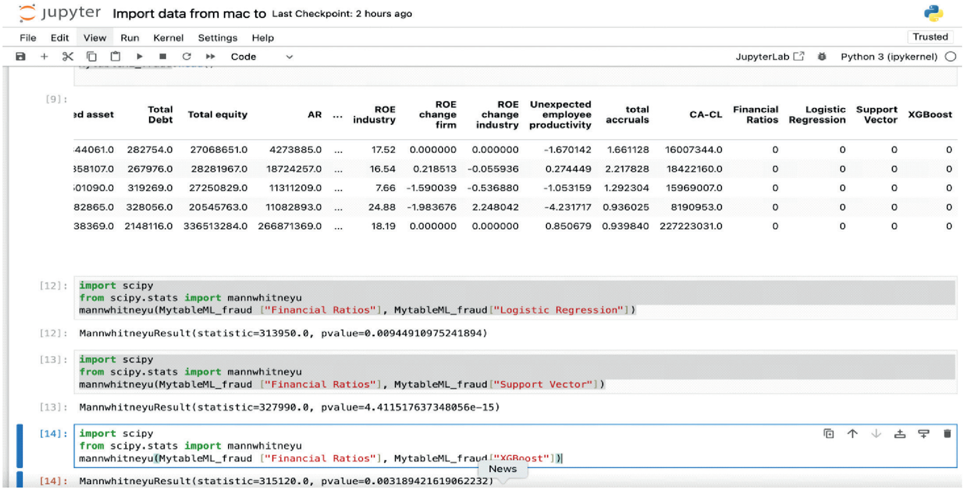
	Method	Fraud
0	1	0
1	1	0
2	1	0
3	1	0
4	1	0

```
[ ]:

      from pingouin import kruskal

[17]: kruskal(data= MytableML_fraud, dv = "Fraud", between = "Method")
```

[17]:

	Source	ddof1	H	p-unc
Kruskal	Method	3	58.650208	1.141742e-12

Appendix D: Python output for Mann Whitney test



jupyter Import data from mac to Last Checkpoint: 2 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
News

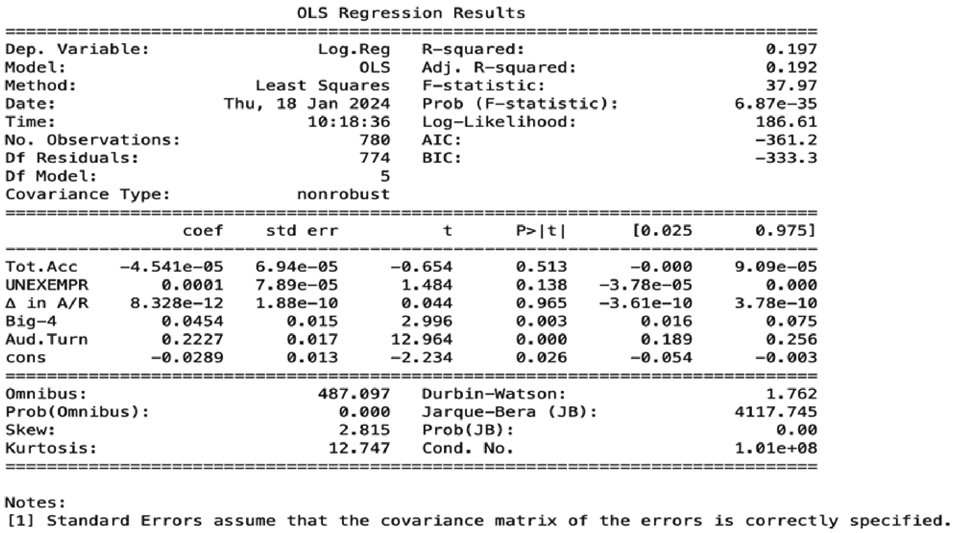
[9]:

ed asset	Total Debt	Total equity	AR	...	ROE industry	ROE change firm	ROE change industry	Unexpected employee productivity	total accruals	CA-CL	Financial Ratios	Logistic Regression	Support Vector	XGBoost
44061.0	282754.0	27068651.0	4273885.0	...	17.52	0.000000	0.000000	-1.670142	1.661128	16007344.0	0	0	0	0
358107.0	267976.0	28281967.0	18724257.0	...	16.54	0.218513	-0.055936	0.274449	2.217828	18422160.0	0	0	0	0
501090.0	319269.0	27250829.0	11311209.0	...	7.66	-1.590039	-0.536880	-1.053159	1.292304	15969007.0	0	0	0	0
82865.0	328056.0	20545763.0	11082893.0	...	24.88	-1.983676	2.248042	-4.231717	0.936025	8190953.0	0	0	0	0
38369.0	2148116.0	336513284.0	266871369.0	...	18.19	0.000000	0.000000	0.850679	0.939840	227223031.0	0	0	0	0

```
[12]: import scipy
      from scipy.stats import mannwhitneyu
      mannwhitneyu(MytableML_fraud ["Financial Ratios"], MytableML_fraud["Logistic Regression"])

[12]: MannwhitneyuResult(statistic=313950.0, pvalue=0.00944910975241894)

[13]: import scipy
      from scipy.stats import mannwhitneyu
      mannwhitneyu(MytableML_fraud ["Financial Ratios"], MytableML_fraud["Support Vector"])

[13]: MannwhitneyuResult(statistic=327990.0, pvalue=4.411517637348056e-15)

[14]: import scipy
      from scipy.stats import mannwhitneyu
      mannwhitneyu(MytableML_fraud ["Financial Ratios"], MytableML_fraud["XGBoost"])

[14]: MannwhitneyuResult(statistic=315120.0, pvalue=0.003189421619062232)
```

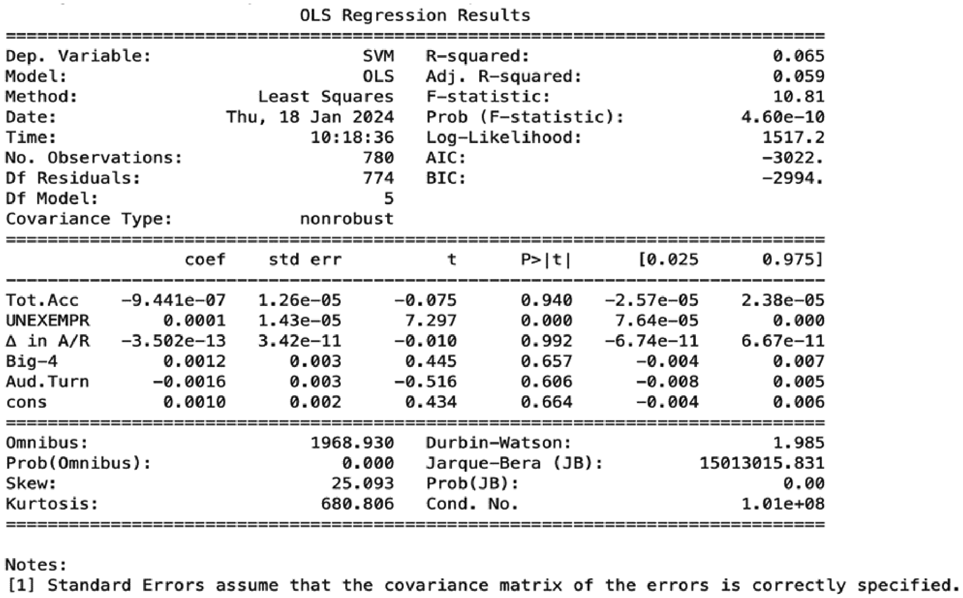
Appendix E: Python output for Regression Analysis

Regression Analysis Model 1



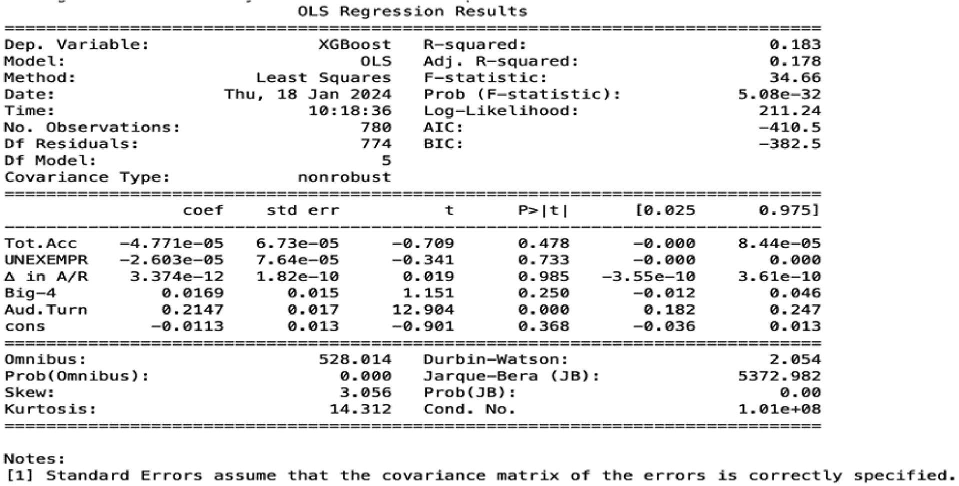
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Log.Reg   R-squared:                       0.197
Model:                            OLS   Adj. R-squared:                  0.192
Method:                 Least Squares   F-statistic:                     37.97
Date:                Thu, 18 Jan 2024   Prob (F-statistic):           6.87e-35
Time:                        10:18:36   Log-Likelihood:                 186.61
No. Observations:                 780   AIC:                            -361.2
Df Residuals:                     774   BIC:                            -333.3
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Tot.Acc    -4.541e-05   6.94e-05     -0.654      0.513      -0.000    9.09e-05
UNEXEMPR       0.0001   7.89e-05      1.484      0.138   -3.78e-05       0.000
Δ in A/R    8.328e-12   1.88e-10      0.044      0.965   -3.61e-10    3.78e-10
Big-4          0.0454      0.015      2.996      0.003       0.016       0.075
Aud.Turn       0.2227      0.017     12.964      0.000       0.189       0.256
cons          -0.0289      0.013     -2.234      0.026      -0.054      -0.003
==============================================================================
Omnibus:                      487.097   Durbin-Watson:                   1.762
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4117.745
Skew:                           2.815   Prob(JB):                         0.00
Kurtosis:                      12.747   Cond. No.                     1.01e+08
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Regression Analysis Model 2



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    SVM   R-squared:                       0.065
Model:                            OLS   Adj. R-squared:                  0.059
Method:                 Least Squares   F-statistic:                     10.81
Date:                Thu, 18 Jan 2024   Prob (F-statistic):           4.60e-10
Time:                        10:18:36   Log-Likelihood:                 1517.2
No. Observations:                 780   AIC:                            -3022.
Df Residuals:                     774   BIC:                            -2994.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Tot.Acc    -9.441e-07   1.26e-05     -0.075      0.940   -2.57e-05    2.38e-05
UNEXEMPR       0.0001   1.43e-05      7.297      0.000    7.64e-05       0.000
Δ in A/R   -3.502e-13   3.42e-11     -0.010      0.992   -6.74e-11    6.67e-11
Big-4          0.0012      0.003      0.445      0.657      -0.004       0.007
Aud.Turn      -0.0016      0.003     -0.516      0.606      -0.008       0.005
cons           0.0010      0.002      0.434      0.664      -0.004       0.006
==============================================================================
Omnibus:                     1968.930   Durbin-Watson:                   1.985
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         15013015.831
Skew:                          25.093   Prob(JB):                         0.00
Kurtosis:                     680.806   Cond. No.                     1.01e+08
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Regression Analysis Model 3



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                XGBoost   R-squared:                       0.183
Model:                            OLS   Adj. R-squared:                  0.178
Method:                 Least Squares   F-statistic:                     34.66
Date:                Thu, 18 Jan 2024   Prob (F-statistic):           5.08e-32
Time:                        10:18:36   Log-Likelihood:                 211.24
No. Observations:                 780   AIC:                            -410.5
Df Residuals:                     774   BIC:                            -382.5
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Tot.Acc    -4.771e-05   6.73e-05     -0.709      0.478      -0.000    8.44e-05
UNEXEMPR   -2.603e-05   7.64e-05     -0.341      0.733      -0.000       0.000
Δ in A/R    3.374e-12   1.82e-10      0.019      0.985   -3.55e-10    3.61e-10
Big-4          0.0169      0.015      1.151      0.250      -0.012       0.046
Aud.Turn       0.2147      0.017     12.904      0.000       0.182       0.247
cons          -0.0113      0.013     -0.901      0.368      -0.036       0.013
==============================================================================
Omnibus:                      528.014   Durbin-Watson:                   2.054
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5372.982
Skew:                           3.056   Prob(JB):                         0.00
Kurtosis:                      14.312   Cond. No.                     1.01e+08
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```