Advance Journal of Econometrics and Finance
Vol-3, Issue-1, 2025

Advance Journal of Econometrics and Finance

Online ISSN
2959-8990
Print ISSN
2959-8982
https://ajeaf.com/index.php/Journal/About

[redacted]

Name of Publisher: SCHOLAR CRAFT EDUCATION & RESEARCH HUB

Review Type: Double Blind Peer Review

Journal Frequency: Quarterly Research Journal

Machine Learning Meets Financial Forensics: Predicting Financial Statement Fraud with Decision Tree Using Beneish M-Score Ratios in Pakistan

Dr. Umar Farooq[1], Dr. Adeel Nasir[2], Dr. Kanwal Iqbal Khan[3*]

Dr. Umar Farooq[1]
Associate Professor, Department of Management Sciences, NAMAL University Mianwali

Dr. Adeel Nasir[2]
Associate Professor, Department of Management Sciences, Lahore College for Women University, Lahore

Dr. Kanwal Iqbal Khan[3*]
Assistant Professor, Department of Management Sciences, University of Engineering and Technology, New Campus, Kala Shah Kaku, Pakistan

Corresponding Author*

Abstract

This research developed a financial statement fraud prediction model in case Pakistani non-financial firms listed on the Pakistan Stock Exchange (PSX). The auditor opinion is taken as a class variable, while eight ratios used in Beneish M-score were taken as predictors to define fraudulent reporting. Six different variants of Decision tree models are also deployed to proposed final models. However, the imbalanced data problem is addressed using oversampling through the SMOTE algorithm before applying the decision tree models. Results showed that random forest provided the highest predictability, i.e., 83%, among other selected models. Random forest outperformed other evaluation matrices, including individual class true positive rate, f-score, ROC, and PRC. Detailed analysis also explored how inflating receivables and internal pressure contribute to the predictability of adverse and/or qualified opinions. The findings suggested that stakeholders use the proposed random forest model to identify the potential Financial statement fraud (FSF) in the case of Pakistani non-financial firms.

Keywords: Financial statement fraud; Fraudulent reporting; Decision tree models; Auditor opinion; Random forest model Pakistani non-financial firms

Introduction

Financial statement fraud (FSF) is often planned, committed, and concealed by expert preparators in almost all organizations, including SMEs, national, international, and even multinational firms. FSF is defined as deliberate misrepresentation, concealment, or omission of financial information from financial statements (Soltani et al., 2023). The consequences of such frauds are severe and affect various stakeholders relying on financial statements to make decisions (Riskiyadi, 2024). The 2024 annual report of ACFE also explores that FSF is among the most expensive compared to other types of fraud (ACFE, 2024). Stakeholders are always keen to predict the quality of financial statements before making their decisions. Therefore, most of the literature on FSF develops some prediction model of fraudulent reporting (Albizri et al., 2019; Amiram et al., 2018; Mangala & Soni, 2023).

The literature on developing FSF prediction models in developed economies is well established. For instance, Beneish (1999a) developed a widely used M-score model using AAER enforcement data. Similarly, Dechow et al. (2011) proposed an F-score model to predict material accounting misstatements. Both models often measure the probability of fraudulent reporting (Achmad et al., 2022; Hołda, 2020; Irwandi et al., 2019; Nurcahyono et al., 2021). In particular, most of the studies used Beneish M-score in this respect. However, applying these models to predict financial misstatements in developing countries may provide varied results due to different regulatory frameworks. Therefore, a financial misstatement detection model must be proposed in developing countries like Pakistan.

Although weak governance and regulatory framework make it more pertinent to develop FSF prediction models in developing countries like Pakistan, limited research has been done in this respect (Khan et al., 2023). This research fills this gap and creates a financial statement fraud prediction model for Pakistani non-financial firms listed at PSX while focusing on three perspectives. First, fraud in financial statements is measured using auditors' opinions. Second, the response variables are the same as the eight financial ratios of Beneish M-score. Third, six decision tree-based machine learning algorithms are applied to develop the final prediction model. Thus, the outcome of this research is significant and provides a classification model to predict fraudulent reporting so that the stakeholders can make their own decisions.

Literature Review

ACFE reported 1921 occupation fraud cases that led to a total loss of $3.1 billion in 2024 (ACFE, 2024). Their report also showed that only 5% of cases were related to fraudulent financial statement reporting, but the median loss per case was $766,000 compared to $145,000 overall median losses per case. Karpoff et al. (2008) indicated that victims lost approximately 29% of equity from financial reporting misconduct. Financial misconduct deprives investors of equity and harms employees and workers through cuts and lost retirement plans (Oppel, 2001). These statistics show that FSF is a costly event. Financial Statement Fraud (FSF) occurs when a company deliberately overstates/understates assets and liabilities, losses, and expenses. Financial statement frauds are detrimental to the organizational reputation and carry the substantial cost of the audit, loss of credibility, and trust among stakeholders such as customers, investors, regulators, suppliers, and the general public (Yousefi Nejad et al., 2024). It leads to a lack of business opportunities, economic unrest, low investor attraction, and damaging market share. Misreporting of financial statements creates grave concerns for organizational leadership and weakens internal control systems (Mandal & Amilan, 2023). It distorts the market integrity and leads to the insufficient allocation of economic resources (Bao, 2023).

The severe outcomes of financial statement fraud led to the importance of developing an early detection model to reduce its adverse consequences. Regulators and auditors also need such detection models to assess the quality of financial statements. Therefore, most of the literature on fraudulent reporting is intended to develop some fraud prediction models. For instance, Beneish's M-score developed a manipulation detection model using eight financial ratios (Beneish, 1999b). In literature, the M-score model is one of the famous models often used to detect earnings manipulation. For instance, Handoko & Natasya (2019) and Knežević et al. (2021) used M-score models to detect earnings manipulation. Similarly, the Deshaw F-score and Altman Z score are other models mostly adopted to predict the FSF (Saleh et al., 2021).

However, it is notable that these fraud prediction models are proposed for developed economies while limited work has been done on developing models for developing economies like Pakistan. Furthermore, the market dynamics differ in developing and developed countries, leading to other conceptions and motivations of earnings manipulation. Developing a modified version of the prediction model that caters explicitly to developing countries' market structure needs is a significant requirement. Though some of the FSF prediction models have been developed for Malaysia (Hasnan et al., 2013, 2014; Hussain et al., 2016) and Indonesia (Fitri et al., 2019; Mustafida, 2020; Sukmadilaga et al., 2022), however, the literature on Pakistan is very limited. This research will use data from non-financial firms listed at PSX to develop an FSF prediction model.

Selecting relevant response variables, predictors, and prediction models is essential in developing a prediction model. In literature, FSF has been measured from different proxies such as enforcements (M. Firth et al., 2005; Romano & Guerrini, 2012), fraudulent restatements (BenYoussef & Khan, 2017; Othman et al., 2021), auditor opinion (Azad et al., 2023; M. A. Firth et al., 2014) and earnings management (Hasnan et al., 2013; Perols & Lougee, 2011) Young (2020) discussed various potential measurements of financial statement frauds such as management bias, earnings management, abnormal changes in accounts receivables, speech patterns of senior managers, Bradford law and the authors' opinion. However, this research uses auditor opinion to measure fraudulent activities in financial statements. This is because the auditor's opinion on financial statements provides a more factual picture of the quality of earnings.

Similarly, it is also essential to decide about the predictors. Different types of response variables have been used in the literature to develop FSF. These variables are often taken from financial statements (Gepp et al., 2021), corporate governance (Du, 2021; Nasir et al., 2019), stock market information (Cox & Weirich, 2002; Peasnell et al., 2011), textual analysis (Othman et al., 2012; Singh Yadav & Sora, 2021) and qualitative analysis (Widuri & Gautama, 2020) of the company. However, most studies used Financial Statement ratios to develop their prediction model, perhaps due to its strong predictability and data availability. This research also used financial statement data to define the predictors. We adopted the financial ratios used in Beneish M-score and applied machine learning algorithms to detect earnings manipulation in the case of non-financial firms listed at PSX. M-Score used eight ratios calculated using data from the financial statements.

The third important decision is about the selection of machine learning algorithms. In the literature, various machine learning models such as logistics, neural networks, decision trees, support vector machines, and many others have been applied to develop fraud detection models. For instance, (Hidayattullah et al., 2020) used metaheuristic optimization approaches of machine learning to predict financial statement fraud patterns. They used two classification methods: support vector machines (SVM) and backpropagation neural networks. They found 96.15% accuracy from the Genetic Algorithm (GA) optimizer with the SVM classifier. Chi et al. (2019) used machine learning techniques such as SVM and Chi-square Automatic Interaction Detector (CHAID) to predict financial statement fraud. They deployed C5.0-SVM, SVM-SVM, C5.0-CHAID, and SVM-CHAID to check the accuracy of the prediction of financial statement fraud. They indicate a maximum accuracy of 83.15% with the C5.0-SVM model. (Alden et al., 2012) used ten-fold cross-validation in GA and MARLEDA to detect financial statement fraud patterns. They found the accuracy of training classification to be 75.5 (GA) and 74.2 (MARLEDA).

Though various machine learning models have been used in literature, this research selected decision tree models due to their high predictability and ability to manage outliers and abnormalities in the data. Specifically, we used the WEKA tool for machine learning to use six decision tree models, including the C4.5 model, Logistic Model Tree (LMT), Naïve Based Tree (NBTree), random forest, and REP Tree methods. It is intended to apply these six models and select the best-fitted model to detect earnings manipulation in the case of non-financial firms listed at PSX. In short, the study aims to develop an FSF prediction model using eight ratios of the Beneish model while applying six decision tree models for a developing country, i.e., Pakistan.

The outcome of this research will be significant. Financial statement fraud (FSF) is the primary national concern, and our model will help to identify it before its occurrence. Many firms are audited, and reports are submitted to the nominated agencies, but fraudsters deploy some gaps to gain out of wrongdoing. WorldCom and Enron are prime examples; despite being audited by reputable firms, they managed to betray company stakeholders by cooking their books. Our study will apply to all the firms of Pakistan and generate a model that an ordinary investor can use to identify or estimate how much sense a particular firm makes towards making fraudulent activity. Financial statement fraud and misreporting have impeded the corporate sector's growth and curbed the trust of stakeholders. This study uses the well-established variables of Beneish M-score and identifies the most appropriate machine learning model to predict the fraud and misreporting of financial statements. The study results will apply to our industry and be generalizable in all emerging economies. The study will help the regulatory authority with the early signs of misappropriation in the financial statements. Furthermore, it will also help investors analyze the temporary artificial bubble created by the overstated assets and revenues. All the capital market stakeholders will benefit from the prediction model. Furthermore, we aim to develop a web application to predict the early signs of misappropriation of financial statements.

Methodology

Sample

The data is collected for non-financial firms listed at the Pakistan Stock Exchange (PSX) for 14 years, from 2010 to 2023. This research focuses on non-financial firms and excludes the financial sector because of their distinguished operations and financial reporting practices. Financial statement data is obtained from the annual publications of the State Bank of Pakistan to compute the eight variables used in the Beneish M-score model. Auditor opinion data were obtained from Datastream, covering auditor assessments of the balance sheet, income statement, and cash flow statement. The dataset included five types of audit opinions: UNQ (unqualified opinion), QUA (qualified opinion), AOP (Adverse Opinion), NOP (No opinion), and Null values.

Table 1: Sample Selection

	Excluded	Firms	Observations
Total Number of Observations		404	5188
Loss of Two years (2010 and 2011) data to calculate lag values	*724*		
Delisted firm data	*174*		
Zero Sales Observations	*444*		
Missing Values	*532*		
Firms having Unqualified opinions for all the sample years	*1775*		
Auditor opinion with no opinion or disclaimer	*369*		
Total observations excluded			(4018)
Final Sample		**151**	**1170**

Initially, the dataset comprised 5,188 firm-year observations from 404 non-financial firms listed on the PSX. However, the final sample is finalized after applying some exclusion criteria, as shown in Table 1. For instance, the first two years of data were lost due to calculating lagged values for specific variables. Similarly, we excluded delisted firms, firms with zero sales, and observations with missing values. Firms that had received an Unqualified Opinion (UNQ) throughout the sample period were removed, as our analysis focuses on firms that received at least one non-unqualified opinion during the study period. After applying these exclusions, the final sample comprised 1,170 firm-year observations for 151 non-financial firms listed on the PSX.

Variables of the Study

Dependent Variable

Various proxies have been used to calculate fraud in financial statements. However, this study used auditor opinion as a proxy for fraudulent financial reporting. Auditor opinion can be divided into four categories: unqualified, qualified, adverse, and no opinion/disclaimer. An unqualified opinion refers to the auditor's opinion that financial statements are prepared according to the applicable accounting standards. Conversely, qualified opinion shows that the financial statements contain material misstatements or deviations from standards, though not pervasive. However, an adverse opinion shows that some significant red flags indicate potential fraud or misstatement. A disclaimer reflects the auditor's inability to express an opinion due to insufficient evidence or uncertainty. This study treats an unqualified opinion as non-fraudulent (coded as 0), whereas qualified and adverse opinions are considered indicators of potential fraud (coded as 1). Thus, the dependent variable is a binary measure where 1 represents suspected fraudulent financial reporting, and 0 denotes non-fraudulent reporting.

Independent Variables

This study incorporates eight financial ratios from the Beneish M-score model. These ratios are the Days Sales Receivable Index (DSRI), Gross Margin Index (GMI), Asset Quality Index (AQI), Sales Growth Index (SGI), Depreciation Index (DEPI), Sales, General, and Administrative Expenses Index (SGAI), Total Accruals to Total Assets (TATA) and Leverage Index (LVGI). Here, the high values of DSRI indicate that receivables increase faster than sales (Khan et al., 2021). This is a red flag regarding potential signs of inflated revenues. GMI shows a change in gross profit margins, and its decreasing value may pressure managers to manipulate the earnings (Khan et al., 2017). An increase in AQI also indicates more investments in less verifiable assets, which can be another red flag regarding asset misappropriation. SGAI is related to operating costs, and its higher value will pressure the managers and motivate them to inflate their earnings (Rashid et al., 2022). DEPI shows the changes in depreciation expenses where its decreasing value may be deliberate to increase the decreasing profits (Rasheed et al., 2021). So, its lower value can be an indicator of the fraudulent activity. SGI is about sales growth, and its increasing value will pressure the managers to maintain its standards over time (Khan et al., 2023). TATA is related to accruals and earnings quality. A high level of accruals indicates that the earnings are not backed by cash and are considered a common fraud proxy. Lastly, LVGI is related to leverage, and its high value puts external pressure on it, which can also motivate fraud. Notably, a high leverage can pressure managers to commit fraud to avoid covenant violations by the debt providers. Table 2 provides the details of these variables and the calculation formula. This research developed a fraud prediction model using these variables in the case of Pakistani non-financial listed firms.

Model Selection

We employed six different decision tree algorithms using the WEKA machine learning tool to develop a predictive model for financial statement fraud, as presented in Table 3. The decision tree is a famous machine learning algorithm recognized for its robustness and high predictability in classification problems. This research employed six variations of the decision tree model utilizing different algorithms on the leaves and nodes. For instance, J48 is the most famous decision tree model in WEKA, and the C4.5 decision tree model was developed with pruning/uprunning options to avoid overfitting problems. Similarly, the LMT model develops a decision tree using a logistic model at the leaves level, and NBTree utilizes a naïve based algorithm. Each algorithm has pros and cons and may provide different accuracy rates within various contexts. We aim to select the best-fitting decision tree model to predict the fraudulent reporting indicated by the qualified or adverse auditor's opinion. Therefore, we applied these six different versions of the decision tree to compare their performances and to select the best-fitted model.

Table 2: List of Variables and their Formula

Meaning	Formula	Interpretation
Days Sales in Receivables Index (DSRI)	(Receivables_t/Sales_t) / (Receivables_t-1/Sales_t-1)	High DSRI suggests revenue inflation via receivables.
Gross Margin Index (GMI)	[(Sales_t-1 – COGS_t-1) / Sales_t-1] / [(Sales_t – COGS_t) / Sales_t]	Declining margins may pressure management to manipulate earnings.
Asset Quality Index (AQI)	[1 – (Current Assets_t + PPE_t) / Total Assets_t] / [1 – (CurrentAssets_t-1 + PPE_t-1) / TotalAssets_t-1]	A high AQI indicates increasing intangible/less verifiable assets.
Sales Growth Index (SGI)	Sales_t / Sales_t-1	High SGI shows growth pressure — a common motive for fraud.
Depreciation Index (DEPI)	(Depreciation_t-1 / (Depreciation_t-1 + PPE_t-1)) / (Depreciation_t / (Depreciation_t + PPE_t))	A lower rate (high DEPI) may suggest earnings manipulation via slower depreciation.
SG&A Expense Index (SGAI)	(SGA_t / Sales_t) / (SGA_t-1 / Sales_t-1)	High SGAI may reflect inefficiency masked by inflated revenue.
Total Accruals to Total Assets (TATA)	(Total Debt_t / Total Assets_t) / (Total Debt_t-1 / Total Assets_t-1)	Rising leverage increases pressure to manipulate earnings.
Leverage Index (LEVI)	(Income From Operations_t – Cash Flow From Operations_t) / Total Assets_t	High TATA signals earnings are not supported by cash, which is a major red flag.

Table 3: Selected Decision Tree Algorithms

Model	Description
J48	A pruned/unpruned C4.5 decision tree algorithm
LMT	A 'logistic model tree' that uses logistic functions at leaves
NB Tree	A decision tree that uses naive Bayes at leaves.
Random Forest	Constructing a forest of random trees.
Random Tree	A decision tree that uses K attributes at the node.
REP Tree	Fast decision tree learner. Builds a tree using information gain and prunes.

However, before applying these models, we addressed data cleaning issues in our data. Table 1 shows that our final sample comprised 1170 observations, including 153 observations of adverse or qualified opinion, while all other observations were unqualified. This resulted in an imbalanced data problem, and developing a prediction model for such a dataset can lead to misleading conclusions. Often, prediction algorithms do not produce accurate predictions of low-number class, i.e., fraud. To solve this problem, we applied an oversampling method through the SMOT algorithm that increased the number of minority class values, i.e., fraud observations based on 5 nearest neighbors. We used this SMOTE algorithm in WEKA. Therefore, the final sample after oversampling consisted of 1017 fraud and non-fraud observations. We also tested the performance of decision tree models through a 10-fold validation method. Further, confusion and different evaluation matrices are used to evaluate the model performances.

Results

Table 4 presents the descriptive statistics of the eight ratios used to develop the prediction model. On average, the DSRI value is high, indicating that receivables are increasing faster than sales. This shows a red flag. The GMI value is also greater than 1, indicating that, generally, firms are good at generating gross profit. However, its mean value is less than 1, and the standard deviation is also high. This shows that some firms may face pressure to generate more gross profit that may lead them towards income manipulation. AQI's mean and standard deviation values are also high, indicating that some firms have been investing more in less tangible assets, possibly due to poor asset quality or some asset misappropriation method. The SGI value of the selected sample is also high, showing that managers may be pressured to maintain fast-paced sales growth, which can motivate them to follow accounting irregularities for personal gains.

However, in the case of DEPI, the average value is greater than 1, indicating that managers are not slowing down the depreciation to inflate earnings. However, a possible reason can be the manipulation of depreciation methods to mask profits, especially in extreme values. Similar arguments can be made in the case of SGAI as its value is greater than 1, indicating the expenses are increasing faster than sales increase. Conversely, the accruals are not generally rising as the mean value of TATA is negative, indicating conservative earnings quality. Finally, the LVGI value is greater than 1, showing that the selected sample has external pressure due to high leverage. This may motivate the managers to account for irregularities to avoid covenant violations. Overall, it is notable that in most of the variables, standard deviations are very high, which shows there would be some observations where a high chance of accounting manipulations can be expected. Similarly, higher mean values than median indicate that these variables are positively skewed, and most observations lie on the right side. Therefore, using these variables in developing a financial statement fraud prediction model is useful.

Table 4: Descriptive Statistics

	DSRI	GMI	AQI	SGI	DEPI	SGAI	TATA	LVGI
Mean	2.588	1.484	1.528	1.674	1.266	1.368	-0.028	1.026
Median	1.034	0.962	0.983	1.091	1.000	1.021	-0.023	1.002
Std.	25.805	29.923	24.019	12.173	4.301	3.619	0.156	0.442
Min	0.000	-181.611	-113.610	0.017	-1.762	0.001	-2.604	0.114
Max	806.420	973.170	772.082	324.172	133.218	101.945	1.384	13.716

Initially, we had imbalanced data, with only 153 fraud observations. To solve this imbalanced data problem, we applied the SMOTE algorithm using 5 nearest neighbor parameters in WEKA. This resulted in the balanced data where both the categories were 1017. Table 5 provides the confusion matrix found after applying six decision tree models in WEKA. This matrix will help identify how well each model predicts fraud and non-fraud observations. Results showed that Random Forest (RF) best predicted most fraud cases correctly among six algorithms. Random Forest (RF) predicted 880 observations correctly, while in 137 cases, this model misclassified fraud cases into the non-fraud category. Similarly, RF predicted the non-fraud cases correctly in most cases compared to other decision tree models. It was found that 818 cases of non-fraud observations were correctly predicted as non-fraud, while 199 cases were misclassified as fraud.

Table 6 further evaluates the six models using different evaluation matrices. We presented the results of six model sorted by their performances. In other words, high-performing models are presented first, and so on. It can be viewed that true positive (TP) rates of both fraud and non-fraud are highest in the case of RF. Table 6 shows that in the case of RF, the TR of fraud and non-fraud is 80.4% and 86.5%, respectively, while the overall weighted average TR is 83.5%. This shows that RF predicted 83.5% of the observations correctly overall. After the RF, the LMT model performed well and predicted 77.2% of the observations correctly, with the second-highest predictability of fraud cases. These conclusions are also consistent when interpreting other evaluation matrices. For instance, ROC and PRC are two other famous measures of model performance. Results show that RF outperformed ROC and PRC compared to other decision tree models. Therefore, overall, it is concluded that the best-fitting model among the six selected decision tree models is the RF, especially in predicting fraud cases correctly comparatively.

Table 5: Confusion Matrix

J48	**0**	**1**	
	700	317	**0**
	195	822	**1**
LMT	**0**	**1**	
	734	283	**0**
	180	837	**1**
NB Tree	**0**	**1**	
	695	322	**0**
	242	775	**1**
Random Forest	**0**	**1**	
	818	199	**0**
	137	880	**1**
Random Tree	**0**	**1**	
	753	264	**0**
	217	800	**1**
REP Tree	**0**	**1**	
	739	278	**0**
	247	770	**1**

Table 6: Evaluation Matrices

		TP	**FP**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC**	**PRC**
	0	0.804	0.135	0.857	0.804	0.83	0.671	0.885	0.885
Random Forest	1	0.865	0.196	0.816	0.865	0.84	0.671	0.903	0.881
	Ave.	0.835	0.165	0.836	0.835	0.835	0.671	0.894	0.883
	0	0.722	0.177	0.803	0.722	0.76	0.548	0.766	0.695
LMT	1	0.823	0.278	0.747	0.823	0.783	0.548	0.746	0.624
	Ave.	0.772	0.228	0.775	0.772	0.772	0.548	0.756	0.66
	0	0.74	0.213	0.776	0.74	0.758	0.528	0.761	0.694
Random Tree	1	0.787	0.26	0.752	0.787	0.769	0.528	0.775	0.689
	Ave.	0.764	0.236	0.764	0.764	0.763	0.528	0.768	0.69
	0	0.688	0.192	0.782	0.688	0.732	0.5	0.751	0.716
J48	1	0.808	0.312	0.722	0.808	0.763	0.5	0.783	0.684
	Ave.	0.748	0.252	0.752	0.748	0.747	0.5	0.767	0.7
	0	0.727	0.243	0.749	0.727	0.738	0.484	0.761	0.721
REP Tree	1	0.757	0.273	0.735	0.757	0.746	0.484	0.769	0.689
	Ave.	0.742	0.258	0.742	0.742	0.742	0.484	0.765	0.705
	0	0.683	0.238	0.742	0.683	0.711	0.447	0.752	0.729
NB Tree	1	0.762	0.317	0.706	0.762	0.733	0.447	0.773	0.714
	Ave.	0.723	0.277	0.724	0.723	0.722	0.447	0.763	0.722

One of the distinguishing features of RF is that it can also provide the ranking of the selected features based on their importance in predicting the class variable. Table 7 presents the results of ranking the selected ratios based on the decrease in average impurity. Results showed that DSRI has the highest decreased impurity and ranked first in the list. This indicates that receivables are the most critical factor in signaling fraudulent reporting. Similarly, GMI is the second most important factor, indicating that the internal pressure of gaining targeted gross profit margins often leads to fraudulent reporting. The third important variable is asset quality, which shows that less investment in tangible assets is also essential in predicting fraudulent activities. Interestingly, the variable of external pressure is ranked at the lowest in the list of selected variables. This concludes that receivables, internal pressure, and asset quality are more important than external pressure debt holders exert when developing financial statement fraud prediction models.

Table 7: Importance of Features based on Ave. impurity Decreased

Feature Name	Ave. Impurity Decrease	No. of nodes used
DSRI	0.42	2881
GMI	0.37	3773
AQI	0.33	3345
SGI	0.32	2838
DEPI	0.29	2615
SGAI	0.28	2335
TATA	0.27	2877
LVGI	0.26	2301

It is also important to discuss why RF performed well compared to other models. This can be due to its ensemble learning approach, which segregates results based on several outputs. Such an ensemble approach controls the risk of overfitting problems often faced in other decision models focusing on a single tree, such as J48. Similarly, RF is good at dealing with outliers and extreme values. Since our data contained high volatility in the case of most of the variables, RF might be coping with outliers efficiently and providing a more accurate prediction comparatively. Another helpful feature of RF is that it selects the random subset to avoid the dominancy of a single feature, leading to more generalized pattern identification. In short, we suggest using RF to predict fraudulent financial reporting for non-financial firms listed at PSX. This will help the various stakeholders, such as auditors, regulators, shareholders, and investors, to identify the potential chances of fraudulent reporting so they can make their decisions promptly.

Conclusion

This research provided a financial statement fraud prediction model for non-financial firms listed at PSX from 2010 to 2023 using auditor opinion as a proxy of fraudulent reporting. To develop a prediction model, we borrowed the eight financial ratios used by the Beneish Mscore model. Benish MScore is one of the most used proxies of fraudulent reporting. However, the M-score model was developed for a developed country, and applying the same model to other countries of different nature, such as in the case of developing countries like Pakistan, may provide distorted results. Therefore, this research intended to create a model for developing countries like Pakistan using the same variables to establish the Benish M-score. We mainly utilized six different decision tree models after removing the imbalanced data problem in the final sample. After applying various exclusion criteria, the final sample consists of 1017 observations for 151 companies. Results showed that the random forest (RF) best predicted fraud and non-fraud cases. Similarly, the logistic model tree (LMT) showed the second-highest predictability in this respect. Results further concluded that receivables quality concerning sales, internal pressure, and asset quality are essential factors in predicting fraudulent reporting. It is also concluded that external pressure measured with external debt has the least capability of predicting fraudulent reporting. It is suggested that stakeholders can predict potential fraudulent reporting using our proposed RF model to make timely and informed decisions.

However, this research has some limitations. This research primarily focuses on eight features used in the Beinish MScore model. Using information from other sources can increase the predictability of the model. For instance, the selected ratios focus on one of the factors of the fraud triangle theory and do not represent any proxy of rationalization. Enhancing the information sources using fraud models will provide a more efficient and robust prediction model. Similarly, we only relied on decision tree models and applied a single classifier approach. Using other models and different classifying approaches, such as hybrid modeling or assembling modeling, can also increase the predictability of the class variable. Therefore, future studies should include more quantitative and qualitative variables using different machine learning methods and applying approaches in this respect.

References

ACFE. (2024). Association of Certified Fraud Examiners The Nations Occupational Fraud 2024: A Report To The Nations. In *Association of Certified Fraud Examiners*. https://www.acfe.com/-/media/files/acfe/pdfs/rttn/2024/2024-report-to-the-nations.pdf

Achmad, T., Hapsari, D. I., & Pamungkas, I. D. (2022). Analysis of Fraud Pentagon Theory to Detecting Fraudulent Financial Reporting using F-Score Model in State-Owned Companies Indonesia. *WSEAS Transactions on Business and Economics*, *19*, 124–133. https://doi.org/10.37394/23207.2022.19.13

Albizri, A., Appelbaum, D., & Rizzotto, N. (2019). Evaluation of financial statements fraud detection research: a multi-disciplinary analysis. *International Journal of Disclosure and Governance*, *16*(4), 206–241. https://doi.org/10.1057/s41310-019-00067-9

Alden, M. E., Bryan, D. M., Lessley, B. J., & Tripathy, A. (2012). Detection of financial statement fraud using evolutionary algorithms. *Journal of Emerging Technologies in Accounting*, *9*(1), 71–94. https://doi.org/10.2308/jeta-50390

Amiram, D., Bozanic, Z., Cox, J. D., Dupont, Q., Karpoff, J. M., & Sloan, R. (2018). Financial reporting fraud and other forms of misconduct: a multidisciplinary review of the literature. *Review of Accounting Studies*, *23*(2), 732–783. https://doi.org/10.1007/s11142-017-9435-x

Azad, A., Salehi, M., & Lari Dashtbayaz, M. (2023). The materiality of identified misstatements by auditors and earnings management. *Management Research Review*, *46*(10), 1405–1426. https://doi.org/10.1108/MRR-01-2022-0073

Bao, Y. (2023). The Impact of Financial Fraud on Economic Stability: An Extensive Economic Analysis. *Journal of Internet Banking and Commerce*, *28*(4), 1–3.

Beneish, M. D. (1999a). The Detection of Earnings Manipulation. *Financial Analysts Journal*, *55*(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296

Beneish, M. D. (1999b). The Detection of Earnings Manipulation. *Financial Analysts Journal*, *55*(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296

BenYoussef, N., & Khan, S. (2017). Identifying fraud using restatement information. *Journal of Financial Crime*, *24*(4), 620–627. https://doi.org/10.1108/JFC-07-2016-0046

Chi, D.-J., Chu, C.-C., & Chen, D. (2019). Applying Support Vector Machine, C5.0, and CHAID to the Detection of Financial Statements Frauds. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11645 LNAI*, 327–336. https://doi.org/10.1007/978-3-030-26766-7_30

Cox, R. A. k., & Weirich, T. R. (2002). The stock market reaction to fraudulent financial reporting. *Managerial Auditing Journal*, *17*(7), 374–382. https://doi.org/10.1108/02686900210437471

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

Du, M. (2021). Corporate governance: five-factor theory-based financial fraud identification. *Journal of Chinese Governance*, *6*(1), 1–19. https://doi.org/10.1080/23812346.2020.1803036

Firth, M. A., Lan Mo, P. L., & Wong, R. M. K. (2014). Auditors' reporting conservatism after regulatory sanctions: Evidence from china. *Journal of International Accounting Research*, *13*(2), 1–24. https://doi.org/10.2308/jiar-50711

Firth, M., Mo, P. L. L., & Wong, R. M. K. (2005). Financial statement frauds and auditor sanctions: An analysis of enforcement actions in China. *Journal of Business Ethics*, *62*(4), 367–381. https://doi.org/10.1007/s10551-005-0542-4

Fitri, F. A., Syukur, M., & Justisa, G. (2019). Do the fraud triangle components motivate fraud in Indonesia? *Australasian Accounting, Business and Finance Journal*, *13*(4), 63–72. https://doi.org/10.14453/aabfj.v13i4.5

Gepp, A., Kumar, K., & Bhattacharya, S. (2021). Lifting the numbers game: identifying key input variables and a best-performing model to detect financial statement fraud. *Accounting and Finance*, *61*(3), 4601–4638. https://doi.org/10.1111/acfi.12742

Handoko, B. L., & Natasya. (2019). Fraud diamond model for fraudulent financial statement detection. *International Journal of Recent Technology and Engineering*, *8*(3), 6865–6872. https://doi.org/10.35940/ijrte.C5838.098319

Hasnan, S., Abdul Rahman, R., & Mahenthiran, S. (2013). Management motive, weak governance, earnings management, and fraudulent financial reporting: Malaysian evidence. *Journal of International Accounting Research*, *12*(1), 1–27. https://doi.org/10.2308/jiar-50353

Hasnan, S., Rahman, R. A., & Mahenthiran, S. (2014). Determinants of fraudulent financial reporting: Evidence from Malaysia. *Jurnal Pengurusan*, *42*(June 2021), 103–117. https://doi.org/10.17576/pengurusan-2014-42-09

Hidayattullah, S., Surjandari, I., & Laoh, E. (2020). Financial statement fraud detection in indonesia listed companies using machine learning based on meta-heuristic optimization. *2020 International Workshop on Big Data and Information Security, IWBIS 2020*, 79–84. https://doi.org/10.1109/IWBIS50925.2020.9255563

Hołda, A. (2020). Using the Beneish M-score model: Evidence from non-financial companies listed on the warsaw stock exchange. *Investment Management and Financial Innovations*, *17*(4), 389–401. https://doi.org/10.21511/imfi.17(4).2020.33

Hussain, A. R. M., Sanusi, Z. M., Mahenthiran, S., & Hasnan, S. (2016). Management motives and firm financial misstatements in Malaysia. *International Journal of Economics and Financial Issues*, *6*(4), 18–25.

Irwandi, S. A., Ghozali, I., Faisal, & Pamungkas, I. D. (2019). Detection fraudulent financial statement: Beneish m-score model. *WSEAS Transactions on Business and Economics*, *16*, 271–281.

Karpoff, J. M., Lee, D. S., & Martin, G. S. (2008). The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis*, *43*(3), 581–611. https://doi.org/10.1017/S0022109000004221

Khan, K. I., Qadeer, F., & Ghafoor, M. M. (2017). Debt specialization within profitability sub-groups: A new perspective of debt structure choices. *FWU Journal of Social Sciences*, *11(2),* 92-108.

Khan, K. I., Qadeer, F., Mata, M. N., Neto, J. C., Sabir, Q., Filipe, J. A., & Martins, J. N. (2021). Core predictors of debt specialization: A new insight to optimal capital structure. *Mathematics, 9(9),* 975.

Khan, K. I., Sherazi, S. K. A., & Hamid, H. (2023). Striving towards better governance and a knowledge-based economy: Impact of intangible resources on firm performance. *International Journal of Business Governance and Ethics, 17(3),* 246-278.

Knežević, S., Špiler, M., Milašinović, M., Mitrović, A., Milojević, S., & Travica, J. (2021). Using Beneish M-Score and Altman Z-Score models to detect financial fraud and company failure. *Tekstilna Industrija*, *69*(4), 20–29. https://doi.org/10.5937/tekstind2104020k

Mandal, A., & Amilan, S. (2023). Preventing financial statement fraud in the corporate sector: insights from auditors. *Journal of Financial Reporting and Accounting*, *23*(1), 56–80. https://doi.org/10.1108/JFRA-02-2023-0101/FULL/XML

Mangala, D., & Soni, L. (2023). A systematic literature review on frauds in banking sector. *Journal of Financial Crime*, *30*(1), 285–301. https://doi.org/10.1108/JFC-12-2021-0263

Mustafida, N. (2020). Determinants of Employee Whistleblowing Intentions in Indonesia: Applying Theory of Planned Behavior. *The Indonesian Journal of Accounting Research*, *23*(02). https://doi.org/10.33312/ijar.476

Nasir, N. A. B. M., Ali, M. J., & Ahmed, K. (2019). Corporate governance, board ethnicity and financial statement fraud: evidence from Malaysia. *Accounting Research Journal*, *32*(3), 514–531. https://doi.org/10.1108/ARJ-02-2018-0024

Nurcahyono, N., Hanum, A. N., Kristiana, I., & Pamungkas, I. D. (2021). Predicting fraudulent financial statement risk: The testing dechow f-score financial sector company inindonesia. *Universal Journal of Accounting and Finance*, *9*(6), 1487–1494. https://doi.org/10.13189/ujaf.2021.090625

Oppel, R. A. (2001). *Employees' Retirement Plan Is a Victim as Enron Tumbles - The New York Times*. https://www.nytimes.com/2001/11/22/business/employees-retirement-plan-is-a-victim-as-enron-tumbles.html

Othman, I. W., Hasan, H., Tapsir, R., Rahman, N. A., Tarmuji, I., Majdi, S., Masuri, S. A., & Omar, N. (2012). Text readability and fraud detection. *ISBEIA 2012 - IEEE Symposium on Business, Engineering and Industrial Applications*, *99*, 296–301. https://doi.org/10.1109/ISBEIA.2012.6422890

Othman, I. W., Slack, R., & Stratling, R. (2021). Predicting income-decreasing forced financial restatement. *Universal Journal of Accounting and Finance*, *9*(6), 1571–1581. https://doi.org/10.13189/ujaf.2021.090634

Peasnell, K., Talib, S., & Young, S. (2011). The fragile returns to investor relations: Evidence from a period of declining market confidence. *Accounting and Business Research*, *41*(1), 69–90. https://doi.org/10.1080/00014788.2011.549638

Perols, J. L., & Lougee, B. A. (2011). The relation between earnings management and financial statement fraud. *Advances in Accounting*, *27*(1), 39–53. https://doi.org/10.1016/j.adiac.2010.10.004

Rasheed, W., Khan, K. I., Naqvi, S. M. W. A., & Mahmood, S. (2021). Consistency of firm's debt policy: An intrinsic role of business group affiliation. *Lahore Journal of Business,* 10*(1),* 1-27.

Rashid, S., Khan, K. I., Nasir, A., & Rashid, T. (2022). Unveiling living dead: Characteristics and consequences of zombie firms. *Cogent Business & Management, 9(1),* 2121240.

Riskiyadi, M. (2024). Detecting future financial statement fraud using a machine learning model in Indonesia: a comparative study. *Asian Review of Accounting*, *32*(3), 394–422. https://doi.org/10.1108/ARA-02-2023-0062/FULL/XML

Romano, G., & Guerrini, A. (2012). Corporate governance and accounting enforcement actions in Italy. *Managerial Auditing Journal*, *27*(7), 622–638. https://doi.org/10.1108/02686901211246778

Saleh, M. M. A., Aladwan, M., Alsinglawi, O., & Almari, M. O. S. (2021). Predicting Fraudulent Financial Statements Using Fraud Detection Models. *Academy of Strategic Management Journal*, *20*(3), 1–17.

Singh Yadav, A. K., & Sora, M. (2021). Fraud detection in financial statements using text mining methods: A review. *IOP Conference Series: Materials Science and Engineering*, *1020*(1). https://doi.org/10.1088/1757-899X/1020/1/012012

Soltani, M., Kythreotis, A., & Roshanpoor, A. (2023). Two decades of financial statement fraud detection literature review; combination of bibliometric analysis and topic modeling approach. *Journal of Financial Crime*, *30*(5), 1367–1388. https://doi.org/10.1108/JFC-09-2022-0227

Sukmadilaga, C., Winarningsih, S., Handayani, T., Herianti, E., & Ghani, E. K. (2022). Fraudulent Financial Reporting in Ministerial and Governmental Institutions in Indonesia: An Analysis Using Hexagon Theory. *Economies*, *10*(4). https://doi.org/10.3390/economies10040086

Widuri, R., & Gautama, Y. (2020). Computer-assisted audit techniques (CAATs) for financial fraud detection: A qualitative approach. *Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020*, *August*, 771–776. https://doi.org/10.1109/ICIMTech50083.2020.9211280

Young, S. D. (2020). Financial Statement Fraud: Motivation, Methods, and Detection. *Corporate Fraud Exposed*, 321–339. https://doi.org/10.1108/978-1-78973-417-120201021

Yousefi Nejad, M., Sarwar Khan, A., & Othman, J. (2024). A panel data analysis of the effect of audit quality on financial statement fraud. *Asian Journal of Accounting Research*, *9*(4), 422–445. https://doi.org/10.1108/AJAR-04-2023-0112/FULL/PDF