INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2021.15.166 Volume 15, 2021

Establishing Causality to Detect Fraud in Financial Statements

Kiran Maka[1]
Department of CSE
Faculty of Engineering and Technology
Annamalai University
Chidambaram, India

S. Pazhanirajan[2]
Department of CSE
Faculty of Engineering and Technology
Annamalai University
Chidambaram, India

Sujata Mallapur[3]
Department of CSE
Faculty of Engineering and Technology
Sharanbasva University
Kalaburagi, India

Received: March 17, 2021. Received: August 27, 2021. Accepted: September 23, 2021. Published: September 27, 2021.

Abstract— In this work, two approaches have been presented to derive the important variables that an auditor should watch out for during the audit trials of a financial statement. To achieve this goal, machine learning modeling is leveraged. In the first approach, important features or variables are derived based on ensemble method and in the second approach, an explainable model is used to corroborate and expand the conclusions derived from the ensemble method. A dataset of financial statements that was labeled manually is utilized for this purpose. Four important measures, namely, random forest recommendations of first approach, random Forest Explaner -pvalue, random Forest Explainer-first multi-way importance plot and random Forest Explainer-second multi-way importance plot, are employed to derive the important features. A final list of six variables is derived from these two approaches and four measures.

Key-Words: - Financial statements, fraud, machine learning, data mining.

I. INTRODUCTION

Regulators in banking and financial industry like SEBI and RBI requires all the financial institutes especially involved in banking and trading under their jurisdiction, to publish financial statements every year. These financial statements are prepared by the finance and accounting departments of the organization. The statements provide information about financial condition of the organization about investments, assets, liabilities, interest earned etc. Growth or fall in the value of assets, income or liabilities is important factors that are used in determining the correct value of the organization. The financial condition [1] of the organization is referred by investors, creditors or rating agencies for different purposes like grating loans, investments, recognitions etc. Hence some of the organizations may manipulate the financial statements to their advantage. The manipulations in figures reported or the fraud present in the financial statements, if not detected, will result in

unexpected loss of revenue, funds or reputation to the organizations which are planning to invest, grant loan or partner with such companies. There have been multiple events of such frauds in developed markets as well as in emerging markets [2-4]. Hence it is essential to classify any financial statement for fraud. Usually, the classification models developed for classification of financial statements as part of digital transformation of the organization for this purpose will not add much value as frequency of usage of such models are less in a year. These models may used once or twice by an organization depending upon the need like granting loans, investing in equity or while considering for recognition with awards. However, the regulators can use these models more frequently in a day to day business as the volumes of these reports are high. But since financial industry is a highly regulated industry, using models to determine the fraud in financial statement may not be completely acceptable. Hence the purpose of modeling is not just to predict the fraud but also determine the important factors driving the fraud in a typical financial statement.

The models can be used for classifying the statements for fraud as well as for deriving important variables or features from patterns of data and are helpful in discriminating the data between fraudulent and genuine statements. Since there are many models that can be used in developing a model, each model will follow its own algorithm to suggest a list of important variables for a given dataset. Many past works in determining the fraud in financial statements can be referred in ref [5-6].

Most popular machine learning algorithms adopted by the modelers are logistic regression, random forests, decision trees, boosting algorithms like XGboost or Adaboost, Neural network models etc. These algorithms can be categorized as transparent models and black box models. The transparent models like logistic regression and decision trees not only provides the classification score for each of the financial statement, but also lists down quantifiable important features from the overall training dataset. However, these models have limitation in getting good accuracy for the prediction. On the other hand, the black box models like random forests, boosting algorithms like XGboost or Adaboost, Neural network models or discriminant analysis [7] provide very good accuracy in predictions, but do not explain the reason behind the scores. Other methods like support vectors [8] and Zipf's law [9] were also used to classify financial statements for fraud. It is not possible to explain with good confidence what is driving the predicted score. Hence there is a need to explain the black box models in a better way so that auditor can use these variables or features while manually auditing the financial reports.

Some of the important works in the classification of financial statements and applications of various algorithms are neural networks, decision trees and belief networks [4, 10], Back propagation method with neural networks [11, 12, 15, 16], neural network models with risk assessment [13], preprocessing methods for detection of fraud in financial statements [14], importance of distribution of digits and its place values using Benford's law [17] and other miscellaneous methods [18-20].

Most of the previous works focused on increasing the accuracy of models in the detection of fraud than aiding the auditor with important variables to watch out for during the audit trials. Since the machine learning models cannot be used as decision makers, the models with good accuracy won't help completely to prevent the fraud. Rather, machine learning models must be used to derive important factors that are helpful to verify if there is indeed any fraud in the published financial statements. In this work, an attempt is made, with the help of machine learning models, to establish a procedure in deriving important features that will aid the auditor to detect fraud in financial statements.

In Sec. II, two approaches have been highlighted. The first one with ensemble method using 38 models, and second approach is validating the ensemble approach with explainable models. In Sec. III, simulation results are presented along with

the procedure to derive important features. In Sec. IV, important conclusions are presented.

II. Dataset and ML Models

In this section, two important aspects of modeling are discussed in detail. Two most important things in building a model are dataset and algorithm. In the first part of this section, details about the dataset are provided and in the second part, details about algorithms are outlined.

The data has been procured for five years. Some of the companies have financial statements released for all the five years and some have not. In this dataset, there are around 14,000 financial statements of 3,500 firms for the period of five years. In the dataset each financial statement has been converted into a row or a vector and each row represents a financial statement of one firm for one year. The data procured from the market data about financial statements do not have the label about if certain financial statement is genuine or fraudulent. Hence each of the records has been labeled manually with the help of auditors [21]. In some cases, auditor reports and comments were also considered during the labeling of financial statements. Of all the 14,000 records, only 358 records have been labeled as fraudulent based on the comments made by auditors. The comments made by auditors like non conformance, no adherence to principals, adverse impact or report etc, were treated as fraudulent for the respective financial reports. The remaining financial statements other than 358 were marked as genuine records. Dataset with 17 variables has been used for training and testing the models. Some of the important variables used in training the model are:

Interest earned (interest_earned)

Altman Z-Score (az_score)

Ratio of Total debts to Total assets (tdebts_tasset or td_ta)

Ratio of Debt to Equity (debt_equity)

Total Assets (tasset or ta)

Total Liability (tliability or tl)

Return on Equity (roequity or roe)

Ratio of Total Accruals to Total assets (total_accruals_ta)

Ratio of Investment to Sales (inv_to_sales)

Ratio of Total sales to Total assets (sales_tassets or sales_ta)

Ratio of accounts receivables to sales (ac_recv_to_Sales)

Beniesh M-Score (m_score)

Total Sales (sales)

Total accounts receivables (ac_recvbl)

Ratio of PPE and Total assets (ppe_tasset)

Ratio of Fixed asset and Total asset (fixedAsset_tasset)

Gross Margin (gross_margin).

Since the data has 14,000 records in total and only 358 for fraudulent class, it results in a heavily skewed or imbalanced dataset. Since the model is going to classify each record either as fraudulent or genuine, the dataset must have around 50% of representation for each class. Therefore the dataset has been sampled into two sub datasets separately from the master dataset. The datasets are:

1. Under sampled dataset (US)
2. Over sampled dataset (OS)

In under sampled dataset, the number of records of majority class is reduced to a number close to that of number of records of minority class. For example in this case, only around 358 records can be sampled out from the majority class since minority class has only 358 records. The other records of majority class are omitted from training of a model. Similarly, when creating a over sampled dataset, number of records of minority class are sampled with duplication several times to a number close to that of number of records of majority class. For example in this case, around 13,284 records can be sampled out from the minority class with each record being sampled multiple times randomly to make the number of records equal to majority class (13,642). However, in this work, the dataset has some quality issues

like incomplete or missing data and hence some records were not considered in the dataset after cleaning. Therefore the 14,000 records were reduced to 4,960 records for training and 1,240 records for testing after cleaning. After performing under sampling and over sampling operations, oversampled dataset has 9,656 records and under sampled dataset has 264 records with nearly 50% class distribution.

CARET library was chosen as the library for the machine learning algorithms along with oversampling (OS) and under sampling (US) strategies. The models considered in this work are:

- M1 C5.0 - US
- M2 Boosted Classification Trees - US
- M3 Bagged CART - US
- M4 Boosted Generalized Linear Model - US
- M5 Boosted Logistic Regression - US
- M6 Parallel Random Forest - US
- M7 Boosted Generalized Additive Model - US
- M8 eXtreme Gradient Boosting TREE - US
- M9 eXtreme Gradient Boosting DART-US
- M10 Stochastic Gradient Boosting -US
- M11 Model Averaged Neural Network -US
- M12 AdaBoost.M1- US
- M13 Bagged MARS - US
- M14 Bagged MARS using gCV Pruning - US
- M15 Bagged Flexible Discriminant Analysis - US
- M16 Bayesian Generalized Linear Model US
- M17 Boosted Tree - US
- M18 CART rpart - - US
- M19 CART rpart1SE - US
- M20 CART rpart2 - US
- M21 Conditional Inference Tree - US
- M22 Boosted Classification Trees - OS
- M23 Boosted Logistic Regression - OS
- M24 Parallel Random Forest - OS
- M25 Boosted Generalized Additive Model - OS
- M26 Boosted Generalized Linear Model - OS
- M27 Stochastic Gradient Boosting - OS
- M28 Model Averaged Neural Network - OS
- M29 AdaBoost.M1 - OS
- M30 Bagged MARS - OS
- M31 Bagged Flexible Discriminant Analysis - OS
- M32 Bagged MARS using gCV Pruning - OS
- M33 Bayesian Generalized Linear Model - OS
- M34 Boosted Tree - OS
- M35 J48 - OS
- M36 CART rpart - OS
- M37 CART rpart1SE - OS
- M38 CART rpart2 - OS

Of the 38 models, top two models are chosen based on accuracy, sensitivity and precision. The importance of features extracted by each of the two selected algorithms is analyzed and finally a list of important features that aids the detection of fraud in the financial statements is prepared. In order to validate the list of important features, another model of random forest that explains the model with different important measures is developed and list of important features are extracted. Two approaches are finally combined and analyzed to derive the final list of features. This set of final features must be watched out by the auditor during the audit trial process to detect fraudulent financial statements.

Algorithm proposed:

1. Prepare the dataset by manually adding labels based on comments made by auditors and generate M-scores.
2. Split the dataset for train and test purposes
3. Generate two sub-sets of train set as under sampled and over sampled sub-sets.
4. Train all the 38 models with appropriate training set (under sampled or over sampled sets).
5. Select top two models based on accuracy, sensitivity and precision. Select one model from under sampled and one from over sampled datasets.
6. Extract most significant features with scale between 0 to 100 for these two top models.
7. Select the most significant features from all the important features based on importance

score.

8. Run a randomForestExplain model on the dataset if one of the models is based on random forest. Otherwise, chose appropriate explainable models depending on the selected model type.
9. Use other methods like p-value, first multi-way importance plot and second multi-way importance plot to derive important features from the appropriate dataset. If the random forest model in the first approach with 38 models is selected from under sampled dataset used models, then use the under sampled dataset for randomForestExplain as well.
10. Now combine all the four methods to finalize the list most significant common features that will aid the auditor to detect the fraud during audit trials.

III. The Simulation Results

In this section, simulation results are presented for two approaches, namely, an ensemble method and an explainable random forest method. Simulations are performed to determine the important factors that are useful to classify if a given financial statement indeed has any manipulations. The factors that are derived from the analysis are useful for the auditors to focus on while checking the statements for any fraudulency. The two approaches followed are:

1. Ensemble of several machine learning methods to determine the causal factors
2. Explainability model of random forest method.

In the first approach, nearly 38 models have been trained and tested. Of the 38 models, two models are found to be accurate. The two models are:

1. M6 - Parallel Random Forest – (Undersampled)
2. M27 - Stochastic Gradient Boosting – (Oversampled)

Based on the analysis, the models M6 and M27 are found to have the better performance in terms of accuracy, sensitivity and precision both on training as well as test sets. The important features were derived from the Caret library and scale for the importance features has been set between minimum as zero and maximum as 100. The features with a score of 100 are the most important features and that are with zero are least significant features.

Table 1. Important features derived by Random forest model

Feature importance rank - M6	Feature name	Importance score
1	interest_earned	100
2	az_score	91.21
3	tdebts_tasset	66.636
4	debt_equity	42.854
5	tasset	42.209
6	tliability	33.853
7	roequity	31.749
8	total_accruals_ta	29.16
9	inv_to_sales	28.833
10	sales_taseets	27.955
11	ac_recv_to_Sales	23.08
12	m_score	20.529
13	sales	17.95
14	ac_recvbl	12.081
15	ppe_tasset	10.347
16	fixedAsset_tasset	8.519
17	gross_margin	0

Table 1 shows the list of important features when a random forest model was used on the dataset on training and testing dataset. Similarly, Table 2 shows the list of important features when a stochastic gradient boosting method was used.

Table 2. Important features derived by Stochastic Gradient Boosting method

Feature importance rank – M27	Feature name - M27	Importance score - M27
1	interest_earned	100
2	az_score	45.49
3	fixedAsset_tasset	36.42
4	debt_equity	35.73
5	gross_margin	30.85
6	ac_recv_to_Sales	22.44
7	inv_to_sales	22.28
8	tliability	20.29

9	tasset	19.5
10	tdebts_tasset	19.45
11	total_accruals_ta	19.26
12	sales	17.3
13	sales_tasset	17.09
14	roequity	16.73
15	m_score	16.23
16	ac_recvbl	11.81
17	ppe_tasset	0

Table 3. Important features derived by Stochastic Gradient Boosting method

Feature importance rank	Feature name - M6	Feature name - M27
1	interest_earned	interest_earned
2	az_score	az_score
3	tdebts_tasset	fixedAsset_tasset
4	debt_equity	debt_equity
5	tasset	gross_margin
6	tliability	ac_recv_to_Sales
7	roequity	inv_to_sales
8	total_accruals_ta	tliability
9	inv_to_sales	tasset
10	sales_taseets	tdebts_tasset
11	ac_recv_to_Sales	total_accruals_ta
12	m_score	sales
13	sales	sales_tasset
14	ac_recvbl	roequity
15	ppe_tasset	m_score
16	fixedAsset_tasset	ac_recvbl
17	gross_margin	ppe_tasset

Table 3 shows the comparison of importance features of both random forest method and the stochastic gradient boosting method. It can be observed that top 5 features derived by random forest method are interest_earned (100), az_score (91.21), tdebts_tasset (66.636), debt_equity (42.854) and tasset (42.209). Similarly, the top 5 features derived by stochastic gradient boosting method are interest_earned (100), az_score (45.49), fixedAsset_tasset (36.42), debt_equity (35.73) and gross_margin (30.85). It can be observed that the features interest_earned, az_score and debt_equity are ranked similarly with ranks 1, 2 and 4 respectively by both random forest method and stochastic gradient boosting methods. Hence these three features are considered as the most important features at ranks 1, 2 and 4 for this dataset. However, the important scores for these variables are not same and not nearer to each other. For example, random forest method estimated higher contribution by az_score with a score of 91.21 and stochastic gradient boosting method estimated a score of 45.49. There is a conflict of two features between random forest and stochastic gradient boosting method at ranks 3 and 5. Random forest derived tdebts_tasset (66.636) at rank 3 and stochastic gradient boosting method derived fixedAsset_tasset (36.42) at rank 3. But by carefully observing the importance scores, random forest recommendation has a higher score than stochastic gradient boosting method. Hence tdebts_tasset can be considered as a third most significant feature. Similarly, random forest derived tasset (42.209) at rank 5 and stochastic gradient boosting method derived gross_margin (30.85) at rank 5. Again, random forest recommendation has a higher score than stochastic gradient boosting method. Hence tasset can be considered as a fifth most significant feature. Final and most significant features are listed in Table 4.

Table 4. Top 5 important features derived by Random forest method

Feature importance rank	Feature name
1	interest_earned
2	az_score
3	tdebts_tasset
4	debt_equity
5	Tasset

Since in the first approach, all the top variables were selected from random forest method, another explainablity model of random forest has been

developed as a second approach to validate the first approach. In the second approach, only random forest model has been trained and tested for extracting the important features using random Forest Explainer library.

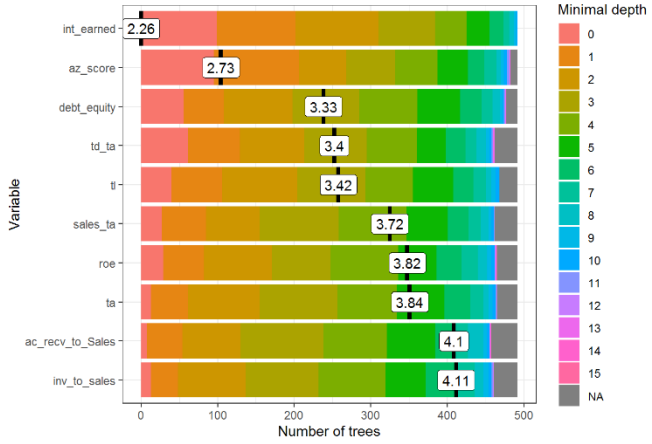

Fig. 1: List of important features and Distribution of minimal depth

From Fig. 1 and Table 5, it can be noticed that the list of top 5 important features are interest_earned, az_score, debt_equity, tdebts_tasset or td_ta and sales_tasset or sales_ta. The important features were selected based on the p-values in the ascending order. The plot shows the distribution of minimal depth with respect to number of trees in the forest. Mean of distribution is indicated with a vertical line mark along with a label of value on it. X-axis indicates the number of trees from zero to maximum in which a variable was used for splitting.

Table 5. List of important features explained by random Forest Explainer

	Feature	mean_min_depth	no_of_nodes	accuracy_decrease	gini_decrease	no_of_trees	times_a_root	p_valu
1	ac_recv	4.4051	1178	0.0015	5.5838	450	0	0.998
2	ac_recv_to_Sa	4.0957	1301	0.0016	6.7812	457	8	0.27
3	az_score	2.7315	1447	0.0226	12.541	482	95	
4	debt_equity	3.332	1400	0.0105	8.9311	477	56	0.000
5	fixedAsset_ta	4.3839	1144	0.001	5.7426	464	2	
6	gross_margin	4.459	1030	0.0016	5.5872	446	4	
7	int_earned	2.2642	1669	0.0258	14.945	492	99	
8	inv_to_sales	4.1088	1235	0.0012	6.9401	461	13	0.905
9	m_score	4.2815	1216	0.0001	6.4921	452	24	0.969
10	ppe_ta	4.4701	1130	0.0017	5.3195	450	0	
11	roe	3.8231	1327	0.0053	7.8567	465	29	0.09
12	sales	4.3283	1222	0.0025	6.1231	461	0	0.95
13	sales_ta	3.7208	1328	0.0011	7.5074	462	27	0.086
14	ta	3.8373	1294	0.0053	7.2955	465	13	0.34
15	td_ta	3.3956	1336	0.0108	9.1432	462	61	0.055
16	tl	3.4193	1311	0.0063	8.1712	468	40	0.1
17	total_accruals	4.191	1193	0.0041	6.5745	456	29	0.994

Table 6. Top 5 important features derived based on p-value

Feature importance rank	Feature name
1	interest_earned
2	az_score
3	debt_equity
4	tdebts_tasset
5	sales_tasset

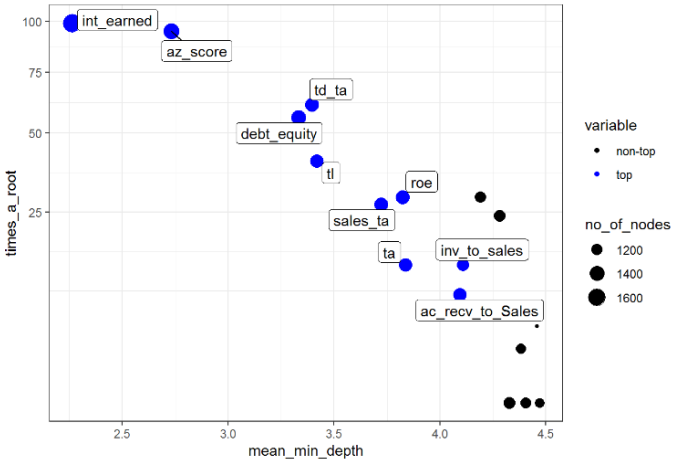

Fig. 2: Important features from first multi-way importance plot

Fig. 2 shows the list of important features derived from first multi-way importance plot. The criteria used for selection are: mean depth of first split, number of trees in which the root is split and total number of nodes in the forest that was used for split. Most significant features recommended from first multi-way importance plot are listed in Table 7.

Table 7. Top 5 important features derived by first multi-way importance plot

Feature importance rank	Feature name
1	interest_earned
2	az_score
3	debt_equity
4	tdebts_tasset
5	Tliability

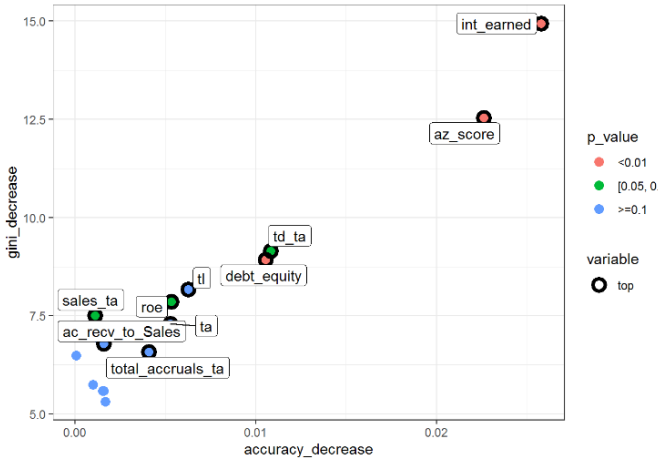

Fig. 3: Important features from second multi-way importance plot

Fig. 3 shows the list of important features derived from second multi-way importance plot. The criteria used for selection are: accuracy_decrease, gini_decrease and p_value assuming a binomial distribution for number of splits at nodes for each feature. Most significant features recommended from second multi-way importance plot are listed in Table 8. Figs. 4-6 show the correlation plots for important measures, rankings and frequent interactions. These plots provide important information about the relation between various measures considered in selecting the significant features.

Table 8. Top 5 important features derived by second multi-way importance plot

Feature importance rank	Feature name
1	interest_earned
2	az_score
3	tdebts_tasset
4	debt_equity
5	tliability

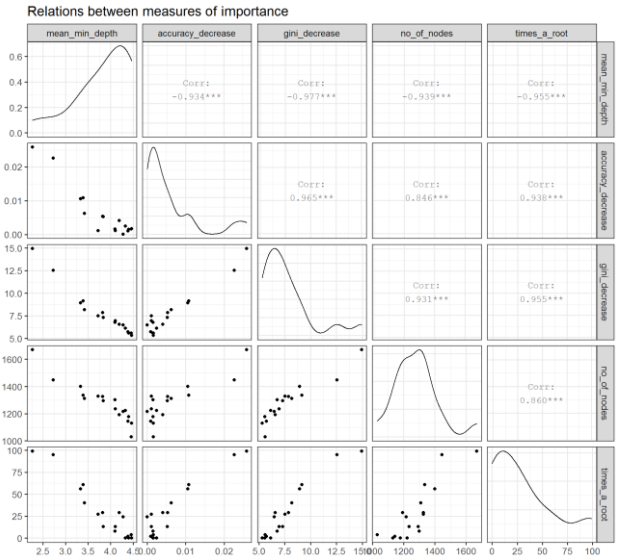

Fig. 4: Correlation between important measures

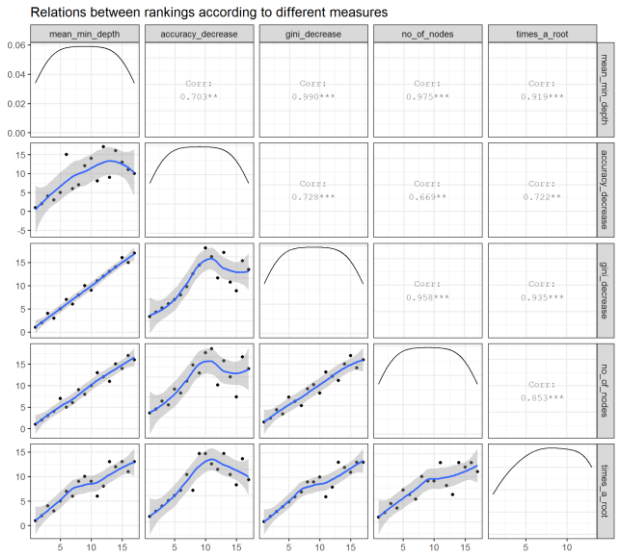

Fig. 5: Correlation between rankings due to important measures

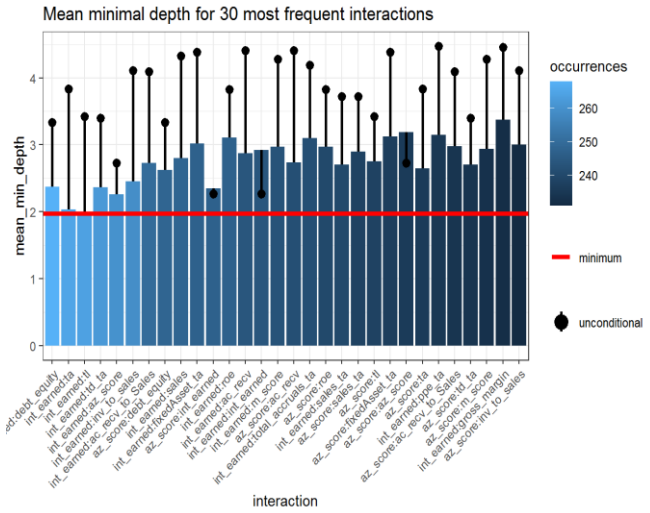

Fig. 6: Plot of 30 most important and frequent interactions

Table 9. Summary of recommendations from various methods

Feature importance rank	Random Foest - Built in algorithm	randomForestExplainer - p-value	randomForestExplainer-first multi-way importance plot	randomForestExplainer - Second multi-way importance plot
1	interest_earned	interest_earned	interest_earned	interest_earned
2	az_score	az_score	az_score	az_score
3	tdebts_tasset	debt_equity	debt_equity	tdebts_tasset
4	debt_equity	tdebts_tasset	tdebts_tasset	debt_equity
5	tasset	sales_tasset	tliability	tliability

Table 9 shows the recommendation about the features to be focused on during the audits to determine the manipulations or fraud in the financial statements by various methods like random forest algorithm, -pvalue, random Forest Explainer-first multi-way importance plot and random Forest Explainer-second multi-way importance plot. Of the recommendations provided by all the four methods, interest_earned, az_score, tdebts_tasset and debt_equality is common. From the remaining features in the list of top-5, the tasset and sales_tasset has vote of one each and tliability has two votes. Therefore, the tliability can also be added to the list of top 5 recommended by the first approach. The final list of features to watch for by an auditor during the audit of financial statements is listed in Table 10.

Table 10: Top 6 important features derived by second multi-way importance plot

Feature importance rank	**Feature name**
1	interest_earned
2	az_score
3	tdebts_tasset
4	debt_equity
5	Tasset
6	Tliability

IV. Conclusion

In this work, important features to be observed by the auditor during the audit of financial statements are derived using two approaches. In first approach, many algorithms were trained and tested and two out of 38 models were chosen to be the best models based on accuracy, sensitivity and precision on training and test sets both on over sampled data and under sampled data. The two methods chosen are: Random forest model and stochastic gradient boosting methods. Both these models were chosen as the final models and important features recommended by these algorithms were analyzed. Based on the importance score estimated by random forest model and stochastic gradient boosting methods, further analysis was carried out to finalize the list of top 5 important features. In second approach another model, namely, random Forest Explainer was used to assess the random forest model on the under sampled data as well based on three measurement approaches, namely, p-value, first multi way importance plot and second multi way importance plot since the random forest model of first approach also had built on under sampled dataset. Based on the recommendations derived from all the four methods namely, random forest of first approach, random Forest Explaner-pvalue, random Forest Explainer-first multi-way

importance plot and random Forest Explainer-second multi-way importance plot, a final list of 6 valriables are derived. It is concluded from this research work that variables like interest_earned, az_Score, tdebts_tasset, debt_equity, tasset and tliability are to be watched by the auditor

during audit of financial statements to identify the fraudulent financial statements.

REFERENCES:

[1] W.H. Beaver, Financial ratios as predictors of failure, Journal of Accounting Research 4 (1966) 71–111.
[2] http://en.wikipedia.org/wiki/Enron.
[3] http://en.wikipedia.org/wiki/MCI_Inc.
[4] Kirkos E, Spathis C, Manolopoulos Y. Data Mining Techniques for the Detection of Fraudulent Financial Statements. Expert Systems with Applications 2007; 32: 995–1003.
[5] Glancy FH, Yadav SB. A Computational Model for Financial Reporting Fraud Detection. Decision Support Systems 2011; 50: 595–601.
[6] Jans M, Werf JM, Lybaert N, Vanhoof K. A Business Process Mining Application for Internal Transaction Fraud Mitigation. Expert Systems with Applications 2011; 38: 13351–13359.
[7] Neuroshell 2.0, Ward Systems Inc. http://www.wardsystems.com.
[8] Cecchini M, Aytug H, Koehler G, Pathak P. Detecting Management Fraud in Public Companies. Management Science 2010; 56: 1146-1160.
[9] S.-M. Huang, D.C. Yen, L.-W. Yang, J.-S. Hua, An investigation of Zipf's Law for fraud detection, Decision Support Systems 46 (1) (2008) 70–83.
[10] Hoogs B, Kiehl T, Lacomb C, Senturk, D. A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative of Financial Statement Fraud. Intelligent Systems in Accounting, Finance and Management 2007; 15: 41-56.
[11] J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model selection, Information & Management 29 (6) (1995) 297–303.
[12] M.J. Cerullo, V. Cerullo, Using neural networks to predict financial reporting fraud: Part 1, Computer Fraud & Security 5 (1999) 14–17.
[13] T.G. Calderon, J.J. Cheh, A roadmap for future neural networks research in auditing and risk assessment, International Journal of Accounting Information Systems 3
[14] (4) (2002) 203–236.
[15] E. Koskivaara, Different pre-processing models for financial accounts when using neural networks for auditing, Proceedings of the 8th European Conference on
[16] Information Systems, vol. 1, 2000, pp. 326–3328, Vienna, Austria.
[17] E. Koskivaara, Artificial neural networks in auditing: state of the art, The ICFAI Journal of Audit Practice 1 (4) (2004) 12–33.
[18] B. Busta, R. Weinberg, Using Benford's law and neural networks as a review procedure, Managerial Auditing Journal 13 (6) (1998) 356–366.
[19] E.H. Feroz, T.M. Kwon, V. Pastena, K.J. Park, The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach, International Journal of Intelligent Systems in Accounting, Finance, and Management 9 (3) (2000) 145–157.
[20] Lokanan, M.E. (2017), “Theorizing Financial Crimes as Moral Actions”, European Accounting Review, Vol. 0 No. 0, pp. 1–38.
[21] Lokanan, M.E. and Sharma, S. (2018), “A Fraud Triangle Analysis of the Libor Fraud”, Journal of Forensic and Investigative Accounting, Vol. 10 No. 2, pp. 187–212.
[22] Li, Z. (2016). "Anomaly detection and predictive analytics for financial risk management." Available at: https://rucore.libraries.rutgers.edu/rutgers-lib/49363/ (accessed 21 September 2018).
[23] https://github.com/AdroMine/Predicting-Fraud-in-Financial-Statements.
[24] Pocock M, Chandler M, Bonney R, Thornhill I, Albin A, August T, et al. A Vision for Global Biodiversity Monitoring With Citizen Science. Advances in Ecological Research. 2018.
[25] Gouraguine A, Moranta J, Ruiz-Frau A, Hinz H, Reñones O, Ferse SCA, et al. Citizen science in data and resource-limited areas: A tool to detect long-term ecosystem changes. PLOS ONE. 2019;14: e0210007. pmid:30625207
[26] Beatriz M., Alejandro O., Julio G., and Juan A. Citizen science for predicting spatio-temporal patterns in seabird abundance during migration.
[27] https://doi.org/10.1371/journal.pone.0236631, (Aug 2020)

AUTHORS PROFILE

Kiran Maka completed his Bachelor of engineering in computer science and engineering affiliated to VTU Belagavi from SDMCET Dharwad and M.Tech in computer science and engineering from Dept of PG studies VTU Belagavi and currently pursuing Ph.D in computer science and engineering from Annamalai University, Chidambaram

Dr. S. Pazhanirajan completed his M.E computer science in Annamalai University. He has a Ph.D degree in the area of medical image and signal processing . he has published more than five papers in reputed journals and has attended more than ten conferences. He has taken role of reviewer for many conferences.

Dr. Sujata V Mallapur is currently working as Head and professor in department of Information and Technology, Faculty of Engineering and Technology(Exclusively for Women), Sharnbasva University, Kalaburagi. She completed her Ph.D from PDACE, Kalaburagi under VTU , Belagavi,. She is member of IEEE, ISTE and ISTE. She has published 15 research papers in National and International Journals and conferences. She is guiding 5 Ph.D students.