

Auditor Choice Prediction Model Using Corporate Governance and Ownership Attributes: Machine Learning Approach

Rahayu Abdul Rahman¹, Suraya Masrom², Nor Balkish Zakaria³, Sunarti Halid⁴

^{1,4}*Faculty of Accountancy, Universiti Teknologi Mara, Perak Branch, Tapah Campus, Malaysia.*

²*Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, Perak Branch, Tapah Campus, Malaysia.*

³*Accounting Research Institute, Universiti Teknologi Mara, Selangor, Malaysia.*

Abstract-External auditor is one of the governance mechanisms in mitigating corporate managerial misconduct and thereby enhance the credibility of accounting information. Thus, the main objective of this study is to develop machine learning prediction model on auditor choice of the firm which signal the quality of auditing and financial reporting processes. This paper presents the fundamental knowledge on the design and implementation of machine learning model based on four selected algorithms tested on the real dataset of 2,262 firm-year observations of companies listed on Malaysian stock exchange from 2000 to 2007. The performance of each machine learning algorithm on the auditor choice dataset has been observed based on three groups of features selection namely firm characteristics, governance and ownership. The findings indicated that the machine learning models present better accuracy performance with ownership features selection mainly with the Naïve Bayes algorithm.

Keywords-Auditor Choice, Machine Learning, Prediction, Malaysia

I. INTRODUCTION

Corporate scandals resulted from manipulation of financial statement has undermined the confidence of shareholders on the reliability of the accounting information. Prior studies highlight that high-quality auditor enhance quality of accounting numbers and information as such auditors are more likely to limit corporate managerial misconducts such as financial reporting fraud [1], tax fraud and aggressiveness [2][3], and earnings manipulation [4]-[7]. The empirical findings seem consistent with the agency theory that stresses on the important of the independence auditor in mitigating principal (shareholders)-agents (managers and board of director) conflicts results from information asymmetric and self interest. Although the board of directors and management of the firm are the agents that appointed by the shareholders to run the business on behalf of them, yet both of the parties have different motives and interests. Further, the separation of ownership and control of the firms also lead to the information asymmetric problem.

Due to the limited access on the business operation information, the financial statements act as a primary mechanism for shareholders to monitor the performance of directors and managers. [8] stress that stakeholders especially the shareholders and investors require financial information to be truly reliable. Given that external auditor acts as an important monitoring mechanism in the financial reporting process to ensure the reliability of accounting information in the financial statements.

Firm that hire high-quality auditors will promote confidence and reinforcing trust in financial information as the auditor provide independent assurance to shareholders on the compliance of financial statements with generally accepted accounting principles (GAAP)[9]. [10] argue that high-quality auditor enforces and protects shareholders and investors' rights by detecting and reporting GAAP violations, material misstatement and insiders expropriation. Given that, companies tend to choose a high-quality auditor to enhance their governance control system and reduce potential risks [11].

In audit market, external audit firms are categorised into Big-4 firms and non Big-4 firms. Big-4 firms' auditors have been recognised as high-quality auditors [12][13] as such auditors enjoy a brand-name reputation [14]. According to [12], Big 4 auditors have more expertise, resources and incentives to constraint unethical corporate activities such as opportunistic earnings management and tax aggressiveness, as well as reduce financial restatement, and GAAP violations of auditee firms in order to avoid loss of reputation and litigation risk. The ability of the Big-4 firms' auditors in determining and reporting of a breach in the accounting system of the client signal the accuracy and reliability of financial statement audited by them. In line with the argument, prior studies find that firms that audited by Big 4 auditors have high rating bonds issued [15], low cost of capital [16] and provide high quality of auditors report [13].

The prevalence and significant impact of high-quality auditors on financial reporting process has sparked the interest to study on auditor selection models [17]-[24].

Despite the growing body of literature on auditor choice prediction using traditional statistical method; logistic regression, very little attention has been devoted to construct auditor choice prediction model using machine learning algorithms [25]-[27]. Prior research that applied machine learning on prediction, classification and detection studies in financial sector highlight the effectiveness and accuracy of such methods to that of traditional statistical methods in problems such as in detection of corporate bankruptcy [28][29], tax fraud and aggressiveness [30][31] and financial statement fraud [32]. Despite the wider use of machine learning in various finance and accounting areas, yet study on machine learning prediction on auditor choice is still limited. Motivated by the limitation, this paper aims to develop auditor choice model using machine learning algorithms to predict auditor choice of the firm which signal the quality of auditing and financial reporting processes.

There are two main contributions of this paper. First, this study attempts to extend prior works [25]-[27] in constructing auditor choice model using machine learning algorithms in order to deepen current understanding on the effectiveness of machine learning approach in predicting firm auditor selection decision. Second, it provides another design and implementation approaches that extends the method used in [25]-[27] by including the elements of corporate governance and ownership.

The following section provides a review of prior studies on auditor selection prediction model. Section 3 provides a brief description on the data set of this study and the feature selection process. Section 4 describes the method and experimental results for the representative compared algorithms. Finally, section 5 discusses the conclusions and future research directions.

II. LITERATURE REVIEW

Prediction of auditor choice using traditional statistic; logistic regression is a well-documented topic [17]-[24]. For example, [23] study auditor choice prediction, using eight inputs from firms-specifics characteristic and corporate governance categories. Using 511 non financial firm listed on Vietnamese stock market from 2015 to 2017, the results show that ownership concentration, foreign ownership and size of the firm board significantly affect the auditor selection decision. Further, the findings indicate that the accuracy rate of logit regression model is 77.43%.

In Malaysia setting, [17] develop a logit regression model to predict auditor selection decision among Malaysian listed firms. The model uses five corporate governance inputs and six firms-specifics characteristic inputs. The results indicate that firms that have large composition of board and audit committee members, less concentrated ownership, non duality role, and lower composition of independent board member are more likely to choose big 4 auditors.

A number of studies [28]-[32] have demonstrated that data mining as well as artificial intelligence approaches as alternative methodologies for classification problems. The first attempt to use data mining classification techniques for auditor choice prediction is done by [25]. The study construct auditor choice prediction model using three data mining methods; decision trees, neural networks and k-nearest neighbor. Based on 10-fold cross validation test, the results show that the accuracy rate of decision tree model is 81.97% outperformed the other two methods, the neural network (78.45%) and the k-NN (73.21%). The finding also highlight that input factor namely firms' debt significantly affect auditor selection decision among UK and Irish firms. The results suggest that firms that have high debt level are more likely to choose high-quality auditors. [26] compared a neural networks auditor choice prediction model with decision trees and support vector machines models. The sample of the study are 338 UK and Irish firms from 2003 to 2005 which consists of 181 Big 4 auditors and 157 non Big 4 auditors. Among the evaluated models, decision trees model showed more accurate result (83.73%) than others models, the support vector machine (79.29%) and the neural network (75.44%). Consistent with [25] findings, the results also reveal that firm debt is a significant input for auditor choice prediction. Further, study by [27] aims to predict auditor selection decision of 95 Iranian firms listed on the Tehran Stock Exchange from 2005 to 2010 using a combination of particle swarm optimization (PSO) and classification and regression trees (CART). In line with [25] and [26], the study use sixteen input factors as predictors of firm auditor selection. The input factors consists of firm specific characteristics, firm financial ratios and auditor's fees. The results show that six firms specific characteristics; total assets, current assets, turnover, working capital, current debt and capital, two financial ratios; current ratio, and solvency ratio and audit fees are significantly affect the prediction of auditor choice with the prediction rate of 94.43%.

International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 07, July 2021)

III. RESEARCH METHOD

A. Sample of data

This study uses 2,262 of firms listed on Malaysian stock exchange from 2000 to 2007 as a sample dataset. Similar to prior research [17]-[32], this study uses size of audit firms to measure high-quality auditors. We codes as 1 for big 4 audit firms represent as high-quality auditors and 0 for non big 4 audit firms. Table 1 indicates the set of features (independent variables) used to develop auditor choice prediction model. This study uses thirteen features that derive from three main categories. Following [17][23] this study uses four features; firm size (SIZE), firm leverage (LEVERAGE), firm growth (GROWTH), firm profit (PROFIT) from the first category, namely firm specific characteristics.

Second category is corporate governance proxies by four features; size of board of director (BODSIZE), composition of independent directors on the board (BODIND), the proportion of independent directors on the audit committee (ACIND), different person hold CEO and chairman position (DIFFCEOCHAIR). Finally, the last category is ownership; director shareholdings (DIROWN), family ownership (FAMOWN), government and state ownership (GOVOWN), foreign ownership (FOROWN), and ownership concentration (BCKOWN).

The correlation coefficient of each feature to the dependent variable was identified as listed in the Table 1. Although all of the features are very weak correlation to the DV, there should be some information that could be used by the machine learning models to generate the prediction results.

TABLE 1
FEATURES IN THE DATASET

Features	Abbreviations	Measurement	Correlation
Firm size	SIZE	Natural log of total assets of firm	0.009
Firm debt	LEVERAGE	The ratio of total liabilities to total assets	0.073
Firm growth	GROWTH	Market to book value	0.08
Firm profit	PROFIT	Earnings (EBIT) to total assets	0.1
Independent director	BODIND	The proportion of independent directors on the board	0.029
Board size	BODSIZE	The number of directors on the board	0.083
Independent audit committee	ACIND	The proportion of independent directors on the audit committee	0.002
CEO/Chairman roles	DIFFCEOCHAIR	1 if CEO is not board chair and 0 otherwise	0.004
Director ownership	DIROWN	The proportion of managerial ownership	0.011
Family ownership	FAMOWN	The proportion of family ownership	0.005
Government ownership	GOVOWN	The proportion of government or state ownership	0.184
Foreign ownership	FOROWN	The proportion of foreign ownership	0.052
Blockholding ownership	BCKOWN	The proportion of largest owner's shareholding	0.032

B. The Algorithms of Machine Learning

Algorithms of Naïve Bayes, Decision Tree, Random Forest and Deep Learning were used in this research for the models. The hyper-parameters' settings for each algorithm are given in Table 2.

C. Software and Hardware

The experiments to run the machine learning models for performances comparisons were implemented with *RapidMiner Studio*. The processor of computer is Intel i7 processor with 16GB RAM.

TABLE 2
HYPER-PARAMETERS

Algorithm	Hyper-parameters
Naïve Bayes	No hyper-parameters
Decision Tree	Maximal depth=7
Random Forest	Number of Trees=140 Maximal depth=4
Deep Learning	No of hidden layers=2 Inner optimizer=ReLU Outer classifier=Softmax

IV. RESULTS AND DISCUSSION

The mean of results from five times of experiments have been recorded.

Table 3 lists the mean of prediction error and the total time to complete for all algorithms.

TABLE 3
THE PREDICTION ERROR OF EACH ALGORITHM WITH DIFFERENT TYPES OF FEATURE SETS

Algorithm	Firm Characteristics (+- std.dev) (Total time in second)	Governance (+- std.dev) (Total time in second)	Ownership (+- std.dev) (Total time in second)
Naïve Bayes	0.317 0.27 3	0.317 0.27 3	0.326 0.21 3
Deep Learning	0.317 0.27 7	0.315 0.20 7	0.310 0.33 8
Decision Tree	0.317 0.27 3	0.317 0.27 3	0.421 0.36 3
Random Forest	0.320 0.20 9	0.315 0.24 42	0.420 0.40 25

Most of the algorithms produced low prediction errors (less than 35%) except the Decision Tree and Random Forest for ownership features selections (42%). We consider the models with lower than 35% prediction error in less than 5 seconds time to complete as the efficient, which presented by the Naïve Bayes in all group of features selection. Although the ownership group has some features with very high correlation to the dependent variable, the prediction errors from all tested algorithms were slightly higher than the other two groups. To test the statistical differences among the mean validation scores produced by each machine learning model with the three groups of features selection, Kruskal Wallis test has been conducted. The p-values obtained from all models were less than 0.01 respectively at the significance level of 95%, which shows that all the prediction errors from the four machine learning algorithms were not generated in the same distribution.

Furthermore, it is interesting to present the graph of Receiver Operating Characteristic (ROC) curve at all the classification thresholds in a range[0,1]. The graph of ROC presents the True Positive Rate (TPR) on y axis and False Positive Rate (FPR) on x axis. TPR shows how frequent is the 1 value (auditor choice) to be correctly predicted as 1 while FPR is calculated as the ratio between the 0 value (not as auditor choice) that truly classified as 0.

ROC can be used to present the trade-off between TPR and FPR, hence is a reliable metric to represent the model accuracy. Figure 1, Figure 2 and Figure 3 present the ROC of all algorithms with the three group of features selection.

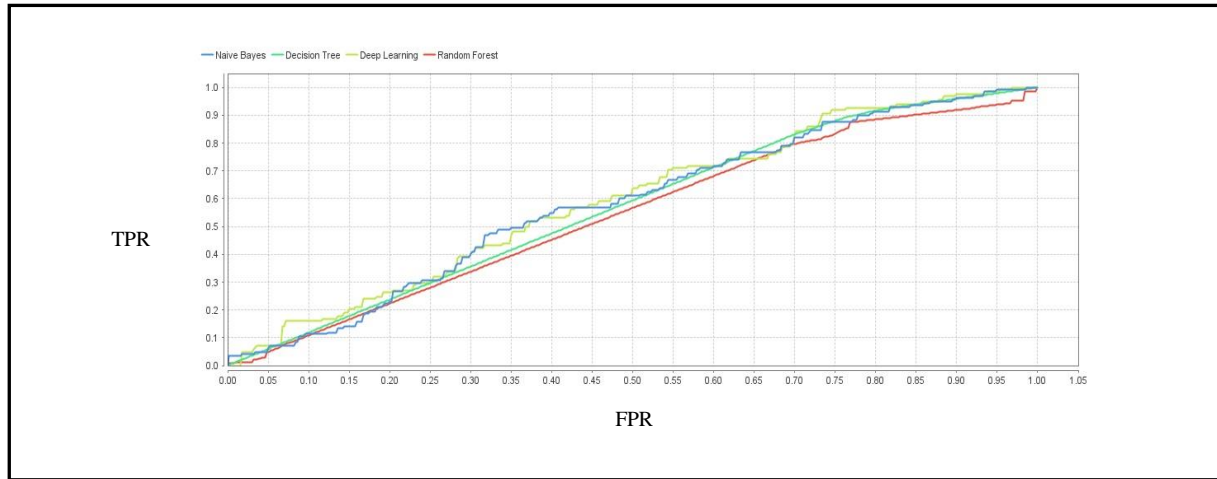


Figure 1 ROC Of All Algorithms With Firm Characteristics Features

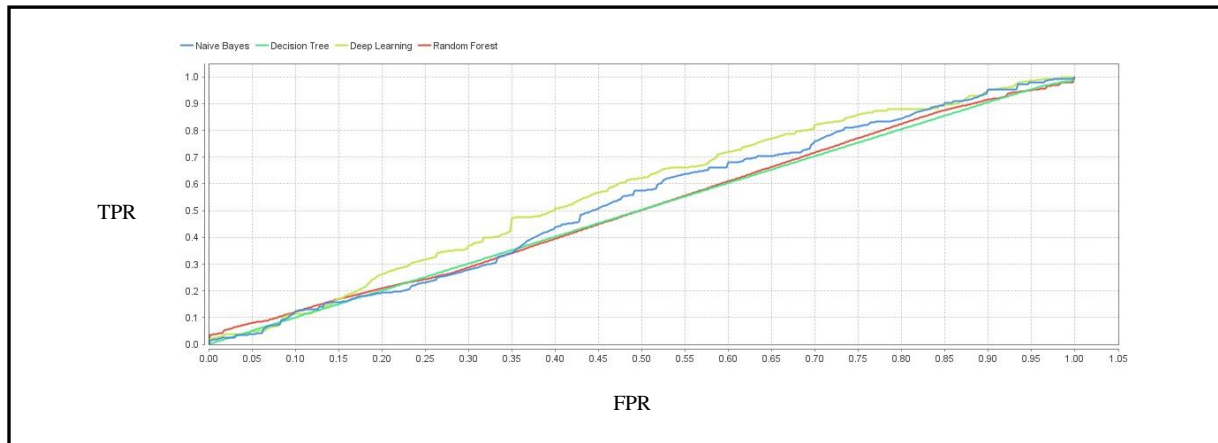


Figure 2 ROC Of All Algorithms With Governance Features

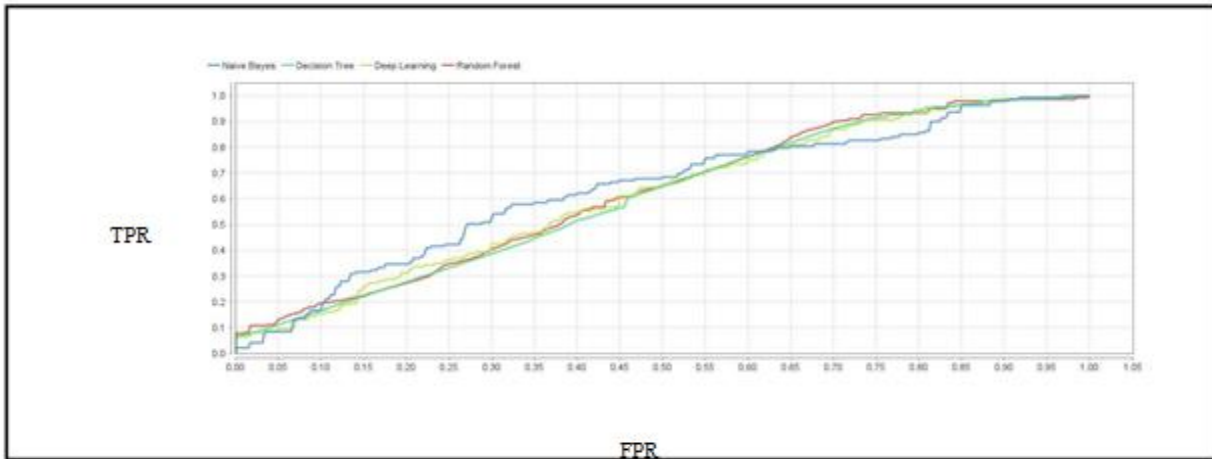


Figure 3 ROC Of All Algorithms With Ownership Features

It is difficult to measure the accuracy of algorithms from the ROC figures. Therefore the Area Under Curve (AUC) of the underneath ROC plots have to be calculated. AUC presents the probability estimation that the model will evaluate a randomly selected positive instance higher than a randomly negative instance. In contrast to other kinds of the accuracy metrics from the sensitivity or specificity formulations, AUC used to calculate the performance of a binary classifier at all thresholds for all the possible decision in each of the model, hence widely thought to be a better measure for the model. AUC value is between 0 to 1, where an excellent model range is between 0.7-0.88 while an outstanding model should reach more than 0.9. Table 4 presents the AUC values.

AUCs between 0.5-0.6 as in the Table 4 are considered as moderate ability of the machine learning models to predict the auditor choice of the company. All algorithms generated bigger AUC with the ownership features and the best algorithm with this feature is Naïve Bayes followed with Deep Learning. With Naïve Bayes, only three seconds were taken to generate the results, which is faster than the Deep Learning. Based on the test, we believed that the higher correlation by the two variables from the ownership features have influenced the model prediction ability.

TABLE 4
THE AUC OF EACH ALGORITHM WITH DIFFERENT TYPES OF FEATURE SETS

Algorithm	Firm Characteristics (+- std.dev) (Total time in second)	Governance (+- std.dev) (Total time in second)	Ownership (+- std.dev) (Total time in second)
Naïve Bayes	0.580 0.054 3	0.531 0.058 3	0.636 0.054 3
Deep Learning	0.592 0.063 7s	0.571 0.059 7s	0.613 0.082 8
Decision Tree	0.570 0.033 3	0.502 0.020 3	0.607 0.067 3
Random Forest	0.545 0.028 9	0.509 0.034 42	0.617 0.038 26

International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 07, July 2021)

V. CONCLUSION

This paper presents the review and empirical works for the design and implementation of machine learning classification model to detect the auditor choice decision by the Malaysian listed companies. Based on the real dataset of auditor choice decision, the performances of different machine learning models are extensively presented by means of three groups of features selection namely firm characteristics, governance and ownership. We present in this paper, the affect of correlation between dependent and independent variables to the machine learning ability. Although the correlation has no direct impact on the prediction error, it is shown in this study that it has the influences on the overall accuracy prediction that balance the trade-off between truly negative and truly positive, which is measured by the model AUC. Therefore, identifying some more independent variables should be conducted in the future research work as to improve the auditor choice machine learning prediction model.

Acknowledgement

This project was funded by the Ministry of Education, Malaysia for project under FRGS(600-IRMI/FRGS5/3 (140/2019). Thank you to the Ministry of Education, Malaysia and the Universiti Teknologi MARA for the support.

REFERENCES

- [1] C. S. Lennox, and J. Pittman, "Big Five audits and accounting fraud," Available at SSRN 1137829, 2008.
- [2] K. J. Klassen, P. Lisowsky, and D. Mescall, "The role of auditors, non-auditors, and internal tax departments in corporate tax aggressiveness," *The Accounting Review*, 91(1), 179-205, 2016.
- [3] A.L Martinez, and C. Hartmann, "Tax Aggressiveness and Big4 Audit Firms," Available at SSRN 3599215, 2020.
- [4] R.A. Rahman, N. Omar, A. H. Osman, and M. Zakaria, "Do Big Four Auditors Limit Real Earnings Management," 2020.
- [5] C. L. Becker, M. L. DeFond, J. Jiambalvo, and K. R. Subramanyam, "The effect of audit quality on earnings management," *Contemporary accounting research*, 15(1), 1-24, 1998.
- [6] A. T. Van, and A. Vanstraelen, "Earnings management and audit quality in Europe: Evidence from the private client segment market", *European accounting review*, 17(3), 447-469, 2008.
- [7] G.V. Krishnan, "Does Big 6 auditor industry expertise constrain earnings management?," *Accounting horizons*, 17, 1-1, 2003.
- [8] J. Cohen, G. Krishnamoorthy, and A. M. Wright, "Corporate governance and the audit process," *Contemporary accounting research*, 19(4), 573-594, 2002.
- [9] S.L. Watts, and J. L. Zimmerman, J. L. "Agency problems, auditing, and the theory of the firm: Some evidence," *The Journal of Law and Economics*, 26(3), 613-633, 1983.
- [10] D.P. Newman, E. R. Patterson, and J. R. Smith, "The role of auditing in investor protection," *The Accounting Review*, 80(1), 289-313, 2005.
- [11] B. P. Fan, and T. J. Wong, "Do external auditors perform a corporate governance role in emerging markets?," Evidence from East Asia. *Journal of Accounting Research*, 43(1), 35-72, 2005.
- [12] C. E. DeAngelo, "Auditor size and audit quality," *Journal of Accounting and Economics*, 3(3), 183-199, 1981.
- [13] D. C. Lennox, "Are large auditors more accurate than small auditors?," *Accounting and business research*, 29(3), 217-227, 1999.
- [14] E. V. Krishnan, "Audit quality and the pricing of discretionary accruals," *Auditing: A journal of practice & theory*, 22(1), 109-126, 2003.
- [15] F. C. Allen, "The effect of large-firm audits on municipal bond rating decisions," *Auditing*, 13(1), 115, 1994.
- [16] G. S. A. Ahmed, S. J. Rasmussen, and S. Y. Tse, "Audit quality, alternative monitoring mechanisms, and cost of capital: An empirical analysis," 2008.
- [17] H. A. W. Nasrudin, N. Mohamed, and N. A. Shafie, N. A. "Corporate Governance & Auditor Choice in Malaysia," In *SHS Web of Conferences* (Vol. 36, p. 00007). 2017.
- [18] I. Haniffa, A. C. Ahmad, K. A. Houghton, and N. Z. M. Yusof, "The Malaysian market for audit services: Ethnicity, multinational companies and auditor choice," *Managerial Auditing Journal*, 2006.
- [19] J. N.F.S.M. Nazri, M. Smith, and Z. Ismail, "The impact of ethnicity on auditor choice: Malaysian evidence," *Asian Review of Accounting*, 2012.
- [20] K. A.Habib, A. H.Muhammadi, and H. Jiang, "Political connections, related party transactions, and auditor choice: Evidence from Indonesia," *Journal of Contemporary Accounting & Economics*, 13(1), 1-19., 2017.
- [21] L. Wang, T.J.Wong, and L. Xia, "State ownership, the institutional environment, and auditor choice: Evidence from China," *Journal of accounting and economics*, 46(1), 112-134, 2008.
- [22] M. M. Ngo, T. Tran, and K.Ho, "Determinants of the choice of a big four auditor in the Vietnamese Stock Market," *Accounting*, 6(5), 671-680, 2020.
- [23] N. A.Soyemi, "Internal Corporate Governance Practices and Choice of External Auditor in Nigeria: A Logistic Regression Analysis," *Binus Business Review*, 11(1), 9-16, 2020.
- [24] O. Kirkos, C. Spathis, and Y. Manolopoulos, "Applying data mining methodologies for auditor P. selection," In *Proceedings 11th Pan-Hellenic Conference in Informatics (PCI)*, pp. 165-178, 2007.
- [25] Q. Kirkos, C. Spathis, C., and Y. Manolopoulos, "Support vector machines, Decision Trees and Neural Networks for auditor selection," *Journal of Computational Methods in Sciences and Engineering*, 8(3), 213-224, 2008.
- [26] R. M. Salehi, S. Kamalahmadi, and M. Bahrani, "Prediction of Auditor Selection Using a Combination of PSO Algorithm and CART in Iran," *The Journal of Distribution Science*, 12(3), 33-41, 2014.
- [27] S. Barboza, H.Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, 83, 405-417, 2017.
- [28] T. E. Erdogan, "Prediction of bankruptcy using support vector machines: an application to bank bankruptcy," *Journal of Statistical Computation and Simulation*, 83(8), 1543-1555, 2013.



International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 07, July 2021)

- [29] U. A. Rahman, S. Masrom, and N. Omar, "Tax avoidance detection based on machine learning of malaysian government-linked companies," *International Journal of Recent Technology and Engineering*, 8(2 Spec), 535-541, 2019.
- [30] V. A. Rahman, S. Masrom, N. Omar, and M. Zakaria, "An application of machine learning on corporate tax avoidance detection model," *IAES International Journal of Artificial Intelligence*, 9, 4, 721-725, 2020.
- [31] W. Perols, "Financial statement fraud detection: An analysis of statistical and machine learning algorithms," *Auditing: A Journal of Practice & Theory*, 30(2), 19-50, 2011.