IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.16, No.3, July 2022, pp. 281~290
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.76575

Financial Distress Prediction with Stacking Ensemble Learning

Muhammad Fadhlil Hadi*[1], De-Ron Liang[2], Tri Kuntoro Priyambodo[3], Azhari SN[4]
[1]Master Program in Computer Science, FMIPA UGM, Yogyakarta, Indonesia
[2]National Central University, Zhongli, Taiwan R.O.C.
[3, 3]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
e-mail:[*1]**fadhlil.1990@gmail.com,** [2]drrliang@csie.ncu.edu.tw, [3]mastri@ugm.ac.id,
[3]arisn@ugm.ac.id

Abstrak

Rasio Keuangan telah digunakan secara luas pada penelitian sebelumnya untuk membangun model prediksi kesulitan keuangan mereka. Rasio Altman menjadi yang paling sering digunakan untuk memprediksi terutama dalam studi akademis, namun Rasio Altman sangat tergantung pada validitas data dalam laporan keuangan, maka diperlukan variabel lain untuk menilai kemungkinan manipulasi laporan keuangan. Tak satu pun dari studi sebelumnya mencoba untuk menggabungkan lima Rasio Altman dengan Beneish M-Score. Kami menggunakan Stacking Ensemble Learning untuk melakukan klasifikasi perusahaan krisis dan melakukan analisis yang komprehensif. Wawasan ini membantu investasi publik membuat keputusan pinjaman dengan tidak hanya mencampur semua informasi indikator keuangan, tetapi juga menilai secara cermat berdasarkan kondisi jangka panjang, jangka pendek, dan juga kemungkinan manipulasi laporan keuangan.

Kata kunci*—Rasio Altman, Beneish M-Score, Prediksi Kesulitan Keuangan, Stacking Ensemble Learning*

Abstract

Previous studies have used financial ratios extensively to build their predictive model of financial distress. The Altman ratio is the most often used to predict, especially in academic studies. However, the Altman ratio is highly dependent on the validity of the data in financial statements, so other variables are needed to assess the possibility of manipulation of financial statements. None of the previous studies combined the five Altman Ratios with the Beneish M-Score. We use Stacking Ensemble Learning to classify crisis companies and perform a comprehensive analysis. This insight helps the investment public make lending decisions by mixing all the financial indicator information and assessing it carefully based on long-term and short-term conditions and possible manipulation of financial statements.

Keywords—*Altman Ratio, Beneish M-Score, Prediction of Financial Distress, Stacking Ensemble Learning*

1. INTRODUCTION

Financial Distress Prediction (FDP) is a dynamic research topic that has been going on for a long time and is changing as the corporate business world changes. It is needed to predict

Received July 26th,2022; Revised July 28th, 2022; Accepted July 31th, 2022

whether a company will face financial trouble or vice versa. It plays an important role in investment decision-making for investors, managerial decision-making for firms, credit decision-making for creditors, customer credit rating for banks, and so on. [1].

Many researchers have developed models for conducting FDP, and the most famous FDP method is using Altman Z-score [3]. It is able to predict bankruptcies 1-year period to the bankruptcy. The formula consists of five financial ratios that can help to determine the company's financial healthiness. In particular. The latest research explores the effect of spitting five Altman Variables (based on their behavior) into two sets of features as Long-Term and Short-Term. Ratio retained earnings to total assets is becoming a Long-term feature because it represents the accumulated "health" condition of a company from the very beginning until now. The rest four Altman Variables will belong to the Short-term feature set since they represent periodic performance [22].

However, it is possible that this prediction could be wrong, especially if there is manipulation in the financial statements [22]. Beneish M-Score [24] is one of the well-known studies that measure manipulation in financial statements and is used on various data from around the world. This score can be calculated and analyzed several years ago before the company was declared a fraud [14, 21, 24, 25, 32]. This research will explore the effect of adding Beneish M-Score to measure predictions due to Earning Management of financial statements in Financial Distress Prediction. We split five Altman into two sets of features as Long-Term and Short-Term and Beneish M-Score to Represent the possibility of earnings management.

Many studies have been carried out using the prediction results of a group of multiple classifiers and then proposing various ensemble mechanisms to make the final prediction [15, 16, 17, 18, 22]. Ensemble methods were used in this study to combine decisions from separated features to improve the overall performance. The approach expected the model to learn from each Long-Term, Short-Term behavior, and Beneish M-Score. This approach also wants to prove that Beneish M-Score can help Altman Variables to predict bankruptcy.

2. METHODS

2.1 Experiment Architecture

Figure 1 Experiment Architecture

In the main experiment, the author makes a prediction model using Stacking Ensemble Learning using 324 company datasets with a proportion of 50% crisis companies: and 50% normal companies. Then the results will be compared with the baseline Altman model. The model created will be evaluated using the DET Curve and Wilcoxon test [22].

2. 1.1 Experiment Dataset

The authors choose crisis and normal sample companies based on events that cause major losses, while a normal company is a company that has never experienced a single crisis event during its existence. Crisis companies collected from TEJ are all companies that had (at least) 1-crisis event that caused a loss in the first year of crisis after a certain normal period. We calculate the Altman ratios and the beneish M-score using Taiwan Economic Journal (TEJ) data. The list of companies we used in this experiment must have M-Score data for at least 3-year before the crisis year and Altman Ratios data for at least 1-year before the crisis year. The long-term (Z2_T1) and short-term (Z1_T1, Z3_T1, Z4_T1, Z5_T1) variables are defined based on the behavior of the accounting terms[22].

Table 1 Altman Feature

Features Name	Description
Z2_T1 (Long Term)	Retained Earnings/Total Assets at 1-year prior to bankruptcy
Z1_T1 (Short Term)	Working Capital/Total Assets at 1-year prior to bankruptcy
Z3_T1 (Short Term)	Earnings before interest and tax/Total Assets at 1-year prior to bankruptcy
Z4_T1 (Short Term)	Market value equity/ Book value of total debt at 1-year prior to bankruptcy
Z5_T1 (Short Term)	Sales/Total Assets at 1-year prior to bankruptcy

The M-score T-1, M-score T-2, and M-score T-3 are calculated using the original Beneish Formula [24]:

$$M-Score = -4.84 + 0.92 * DSRI + 0.528 * GMI + 0.404 * AQI + 0892 * SGI \\ + 0.115 * DEPI - 0.172 * SGAI + 4.679 * TATA - 0.327 * LEVI \quad (1)$$

Table 2 Beneish M-Score Feature

Features Name	Description
M-score T-1 (Beneish)	M-score calculation at 1-year prior to distress
M-score T-2 (Beneish)	M-score calculation at 2-year prior to distress
M-score T-3 (Beneish)	M-score calculation at 3-year prior to distress

The experiment dataset consists of 162 crisis companies and 162 normal companies (324 data). Each dataset is split using 10-fold sampling to produce the training set and testing set, which will be used in prediction to get the result.

2. 2 Experiment Design

In this research, we use Altman Ratios (T-1) and Beneish M-Score (T-1, T-2, T-3) as a feature and implement multiple classifiers to build the model. The possibility of earning management itself is only obtained when the entire Beneish ratio is calculated to get the M-score, and every crisis company has a different pattern of manipulation, so we can't train the M-score between years in one base classifier. This separation is expected to give the M-score its own predictive ability.

Table 3 Model Description

Model	Feature
Baseline Model	Split 5 Altman into :
	LongTerm : Z2
	ShortTerm : Z1, Z3, Z4, Z5
Proposed Model	LongTerm : Z2
	ShortTerm : Z1, Z3, Z4, Z5
	Beneish: M-Score T-1
	Beneish: M-Score T-2
	Beneish: M-Score T-3

2. 2.1 Base Classifier Selection

The authors will find the best algorithm for each base classifier (N). Several algorithms (X) are compared and then used in financial distress prediction: K-Nearest Neighbor, Support Vector Machine, Logistic Regression, Linear Discriminant Analysis, Bagging Tree, and Naïve Bayes; those six algorithms implemented in each feature as illustrated in Figure 2.

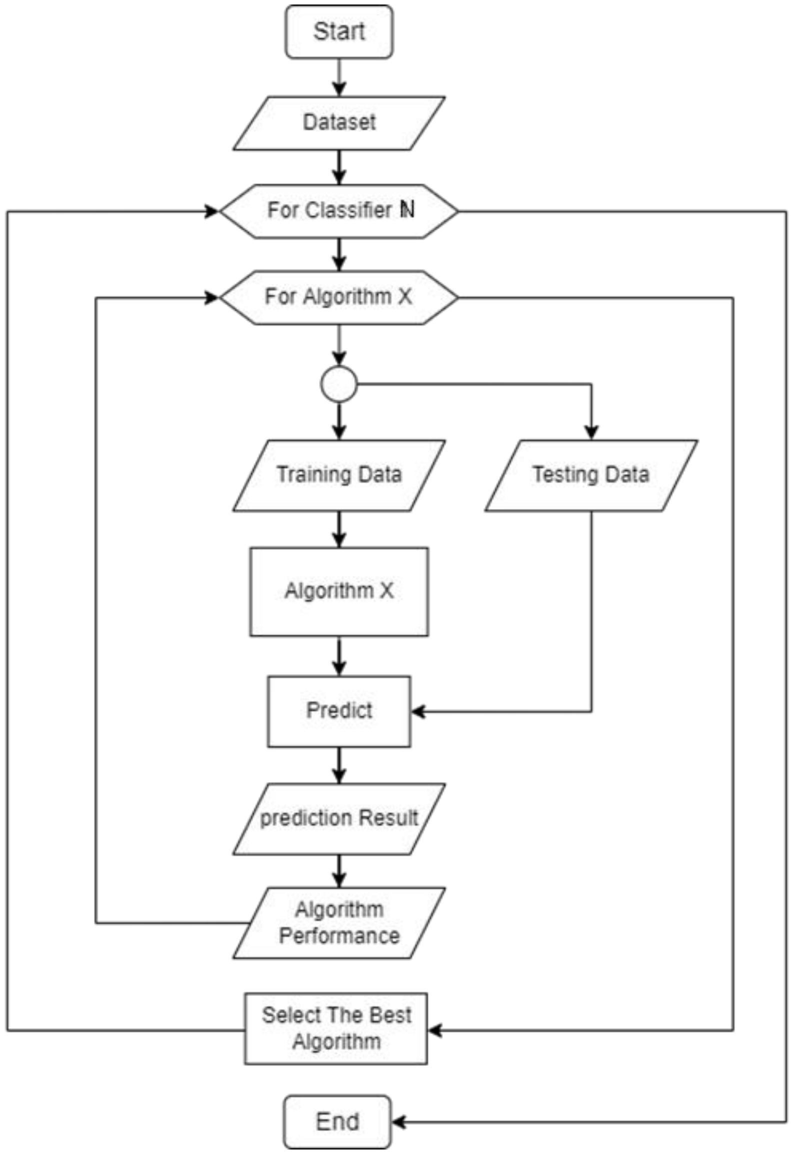

Figure 2 Base Classifier Selection

For each feature (Long-term, Short-Term, M-Score T-1, M-Score T-2, M-Score T-3), the best algorithm performance will be selected as base classifiers in the Stacking Ensemble Learning for the baseline model, as well as the proposed model. Then the baseline model will be compared against the proposed model.

2. 2.2 Baseline Model

Ensemble learning using the stacking generalization approach used to build the baseline model. The view of the different sample sets is built based on long-term and short-term behavior. The ensemble learning from different feature sets aims to maximize the performance of the classifier. This approach introduces a model to learn from each feature to generate the corresponding views. Each feature is considered to contain some information to learn the target concept. The approach is supposed to learn from each feature and optimize the outcome. The experiment using the baseline model shown in Figure 3.

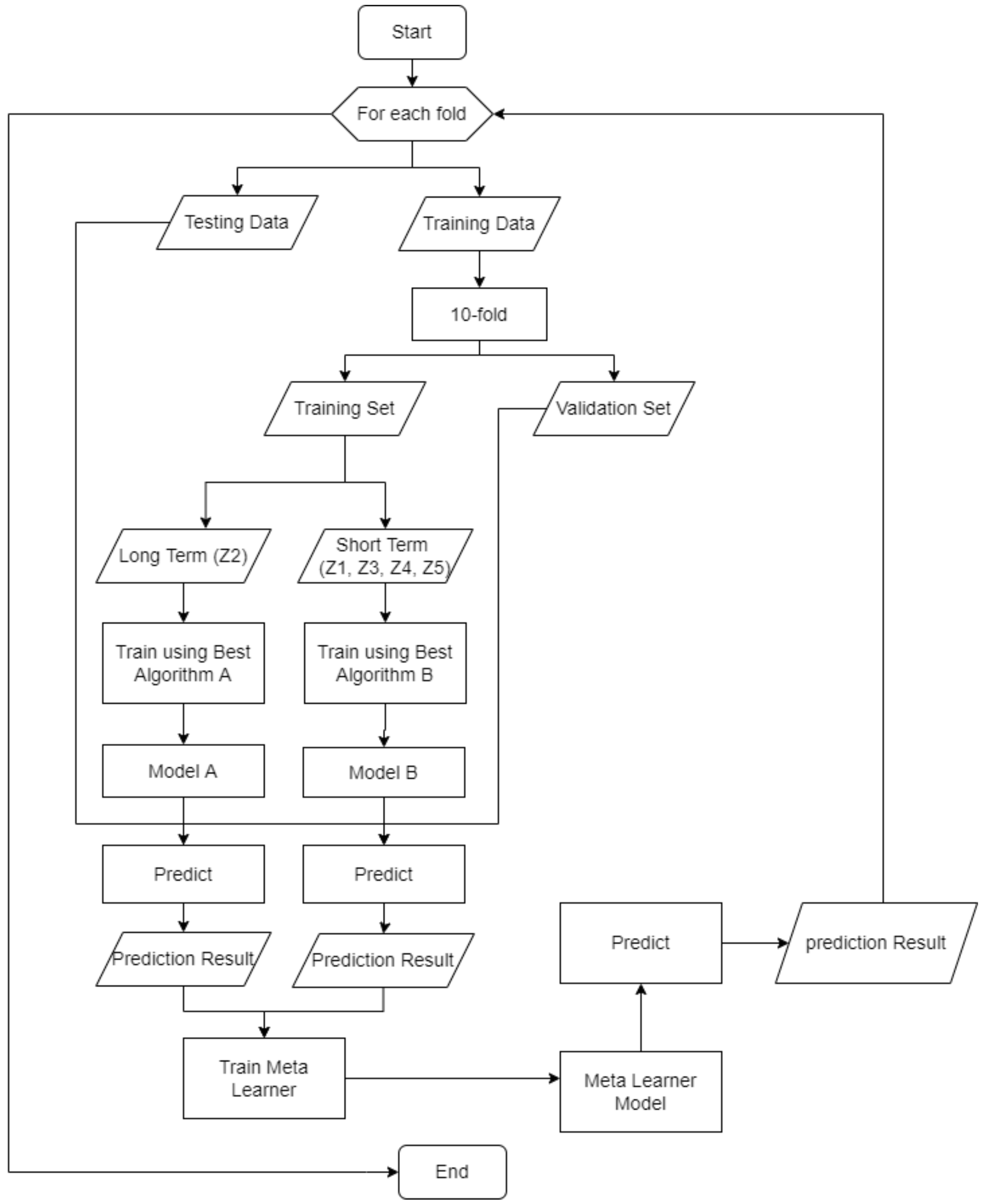

Figure 3 Baseline Model

2. 2.3 Proposed Model

Ensemble learning using the stacking generalization approach used to build the proposed model. The difference between the baseline model and the proposed model is: that there are five basic classifiers, and the algorithm is determined based on a selection in the model-building process. The experiment using the proposed model shown in Figure 4.

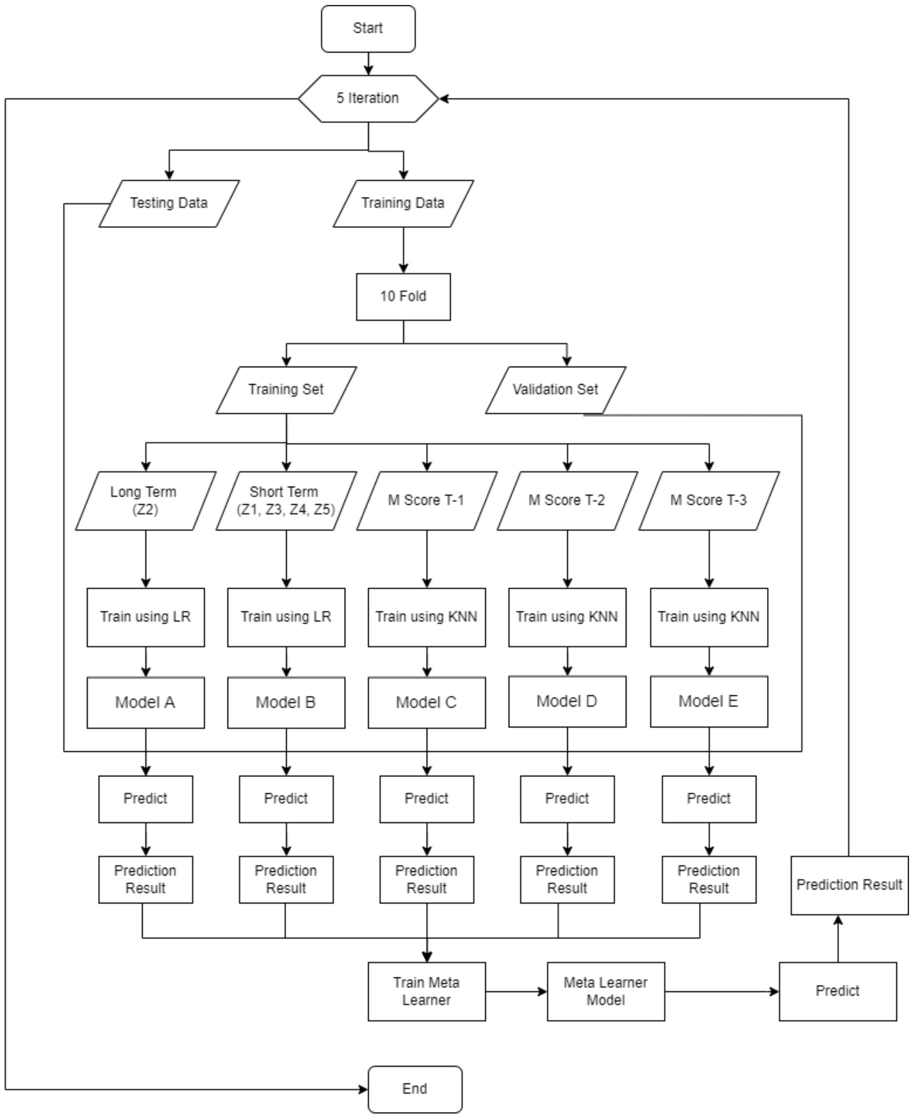

Figure 4 Proposed Model

2. 3 *Result Overview*

This section will discuss the result overview. The result of model testing will generate a DET Curve for each model and the result of the Wilcoxon test, as shown in Figure 5.

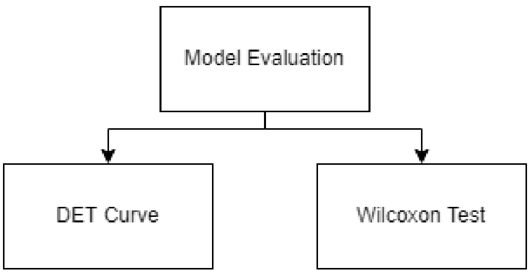

Figure 5 Result Overview

3. RESULTS AND DISCUSSION

The result of the proposed model performed using Stacking Ensemble Learning (implement LR for Long-Term and Short-Term classifier, and KNN for M-Score T-1, T-2, and T-3 classifier) is shown in figure 6.

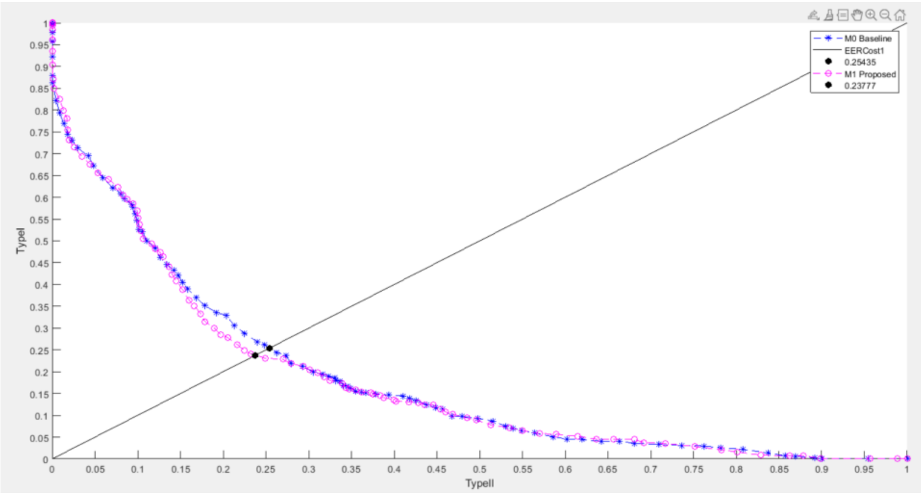

Figure 6 DET Curve

The DET curve shows that the Proposed model has a lower misclassification cost on the EER point (0.23) compared to the baseline model (0.25). It means that when type I error = type II error, the proposed model outperforms compare the baseline model. EER is not the evaluation of the only metric to be considered then. The Wilcoxon test was performed to evaluate how significantly different the proposed model is performing better. The Wilcoxon test results are shown in Table 4.

Table 4 Wilcoxon Test

Cost Ratio	M0	M1
1	0.004	1
1.5	0.000	1
2	0.165	1
2.5	0.491	1
3	0.394	1
3.5	1	0.397
4	0.888	1
4.5	0.718	1
6	1	0.329
7	0.855	1

Wilcoxon test result shows that the proposed model has the best performance as well compared to the baseline, especially in cost ratios 1 to 3, 4 to 4.5, and 7. The significant difference between the proposed and baseline happens when the cost ratio is less than 1.5. It is shown by the p-value is < 0.05; it indicates a high level of significant difference.

4. CONCLUSIONS

In this research, a hypothesis is tested to examine the impact of splitting the Altman Variables into long-term and short-term behavior and adding the Beneish M-Score (three years before the crisis) in financial distress prediction. This hypothesis is used to confirm that the M-Score does help to improve the model performance.

The implication of the research shows that the m-score can be categorized as long-term because the possibility of manipulation that occurs in a company does not take place in one period only, the company will try to manipulate it subtly and continuously so that it is not easily detected, and this is the reason why a long-term manipulation investigation is needed.

REFERENCES

[1] J. Sun, H. Li, Q. H. Huang, and K. Y. He, “Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches,” *Knowledge-Based Systems*, vol. 57, pp. 41–56, 2014.

[2] Y. S. Huang and C. Y. Suen, “Behavior-knowledge space method for combination of multiple classifiers,” in *IEEE Computer Vision and Pattern Recognition*, 1993.

[3] E. I. Altman, “Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy,” *The Journal of Finance*, 1968.

[4] D. Liang, C.-C. Lu, C. F. Tsai, and G. A. Shih, “Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study,” *European Journal of Operational Research*, 2016.

[5] P. du Jardin, D. Veganzones, and E. Séverin, “Forecasting Corporate Bankruptcy Using Accrual-Based Models,” *Computational Economics*, 2019.

[6] K. Valaskova, R. Fedoroko, "Beneish M-Score: A Measurement of fraudulent financial transactions in global environment," SHS Web of Conferences 92, 02064, 2021

[7] E. I. Altman, M. Iwanicz-Drozdowska, E. K. Laitinen, and A. Suvas, “Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman’s Z-Score Model,” *SSRN Electronic Journal*, 2014.

[8] C. Brentani, “Financial statement analysis and financial ratios,” in *Portfolio Management in Practice*, Elsevier, pp. 149–163, 2004.

[9] P. A. Griffin, “Financial Statement Analysis,” in *Finding Alphas: A Quantitative Approach to Building Trading Strategies*, 2015.

[10] E. J. Allen, C. R. Larson, and R. G. Sloan, “Accrual reversals, earnings and stock returns,” *Journal of Accounting and Economics*, 2013.

[11] S. Tian and Y. Yu, “Financial ratios and bankruptcy predictions: An international evidence,” *International Review of Economics and Finance*, 2017.

[12] J. Barugahare, A. Amirkhanian, F. Xiao, S. Amirkhanian, “Predicting the dynamic modulus of hot mix asphalt mixtures using bagged trees ensemble,” *Construction and Building Materials*, 260, p.120468, 2020.

[13] H. Frydman, E. I. Altman, and D.-L. Kao, “Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress,” *The Journal of Finance*, vol. 40, no. 1, pp. 269–291, Mar. 1985.

[14] Tarjo, N. Herawati, “Application of Beneish M-Score Models and Data Mining to Detect Financial Fraud,” *Global Conferences on Business and Social Science*, Vol.211, pp. 924-930, 2015.

[15] M. P. Sesmero, A. I. Ledezma, and A. Sanchis, “Generating ensembles of heterogeneous classifiers using Stacked Generalization,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2015.

[16] W. Jiang, Z. Chen, Y. Xiang, D. Shao, L. Ma, and J. Zhang, “SSEM: A Novel Self-Adaptive Stacking Ensemble Model for Classification,” *IEEE Access*, vol. 7, pp. 120337–120349, 2019.

[17] S. Džeroski and B. Ženko, “Is Combining Classifiers with Stacking Better than Selecting the Best One?,” *Machine Learning*, vol. 54, no. 3, pp. 255–273, Mar. 2004.

[18] E. Menahem, L. Rokach, and Y. Elovici, “Troika – An improved stacking schema for classification tasks,” *Information Sciences*, vol. 179, no. 24, pp. 4097–4122, Dec. 2009.
[19] M.-Y. Chen, “Predicting corporate financial distress based on integration of decision tree classification and logistic regression,” *Expert Systems with Applications*, vol. 38, no. 9, pp. 11261–11272, Sep. 2011.
[20] A. J. Scott, D. W. Hosmer, and S. Lemeshow, “Applied Logistic Regression,” *Biometrics*, 1991.
[21] R. Septiani, S. Musyarofah, R. Yuliana, "Beneish M-Score Reliability as a Tool For Detecting Financial Statements Fraud," *International Colloquium on Forensics Accounting and Governance*, Vol.1, no.1 , 2020
[22] A.N. Fadillah, D. Liang, C.-C. Lu, “Financial Distress Prediction Based on Long-Term and Short-Term Behaviour of Taiwan List Companies,” *National Central University*, 2020.
[23] D. Novitasari, D. Liang, "The Role of Comprehensive Income and Accrual in Predicting Bankruptcy," *National Central University*, 2018
[24] M.D. Beneish, M.C. Lee, C. Nichols, "Earnings Manipulation and Expected Returns.," *Financial Analysts Journal*, vol. 69 no. 2, pp. 57-82, 2013.
[25] L. Svabova, K. Kramarova, J. Chutka, L. Strakova, "Detecting earnings manipulation and fraudulent financial reporting in Slovakia," *Oeconomia Copernicana*, vol.11 no.3, pp 485-508, 2020.
[26] M. Ayu, R.R. Gamayuni, M. Urbański, "The impact of environmental and social costs disclosure on financial performance mediating by earning management," *Polish Journal of Management Studies*, vol.21, no, 1, pp. 74-86, 2020.
[27] A. Siekelova, A. Androniceanu, P. Durana, K. Frajtova Michalikova, "Earnings management (EM), initiatives and company size: An empirical study," *Acta Polytechnica Hungarica*, vol. 17, no. 9, pp. 41-56, 2020.
[28] P. M. Healy, J.M. Wahlen, , "A Review of the Earnings Management Literature and its Implications for Standard Setting.," *Accounting Horizons*, vol. 13, pp. 365-383, 1999
[29] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
[30] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998
[31] C. Cortes and V. Vapnik, “Support Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
[32] J. MacCarthy, “Using Altman Z-score and Beneish M-score Models to Detect Financial Fraud and Corporate Failure: A Case Study of Enron Corporation.”, *International Journal of Finance and Accounting,* International Journal of Finance and, 2017.
[33] M. Warshavsky, “Analyzing Earnings Quality as a Financial Forensic Tool,” *Financial Valuation and Litigation Expert Journal*, vol. 39, pp. 16-20, 2012.
[34] I. Pustylnick, “Combined Algorithm of Detection of Manipulation in Financial Statements,” *Far Eastern Federal University*, 2009.
[35] K. Qin, S. Shi, P. Suganthan, M. Loog, “Enhanced Direct Linear Discriminant Analysis for Feature Extraction on High Dimensional Data,” *Twentieth National Conference on Artificial Intelligence*, pp. 851-855, 2005.
[36] F. Barboza, H. Kimura, and E. Altman, “Machine Learning Models and Bankruptcy Prediction”. *Expert Systems with Applications: An International Journal,* vol. 83, pp 405–417, 2017.
[37] A. Vieira, J. Duarte, B. Ribeiro, and J. Carvalho das Neves, “Accurate Prediction of Financial Distress of Companies with Machine Learning Algorithms,” *International Conference on Adaptive and Natural Computing Algorithms*, vol.5495, pp. 569-576, 2009.

[38] D. Liang, C. Tsai, H. Lu, L.Chang "Combining Corporate Governance Indicators with Stacking Ensembles for Financial Distress Prediction," Journal of Business Research, vol.120, pp. 137-146, 2020.