Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Forecasting company financial distress: C4.5 and adaboost adoption

Okfalisa*[1)], Elvia Budianita[1)], Rezi Yuliani[1)], Ladda Suanmali[2)], Megawati[3)], Hidayati Rusnedy[4)] and Saktioto[5)]

[1)]Department of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
[2)]Department of Business Computer, Suan Dusit University, Bangkok, Thailand
[3)]Department of Information Systems, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
[4)]Information Technology, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
[5)]Department of Physics, Faculty of Science, Universitas Riau, Pekanbaru, Indonesia

Received 9 May 2021
Revised 25 June 2021
Accepted 22 July 2021

Abstract

Financial pressure is one of the factors that determine the survival of a business. In order to minimize the exhausted risk, economic-financial analytics and forecasting have been taken into account. Therefore, this study aims to cover the Altman model to monitor and assess the financial situation based on the financial report's balance sheet and income statement to predict the financial distress status into Health, Undefined, and Distress condition. Here, the integration of the C4.5 algorithm and Adaboost carried out five Altman's worth attributes for optimally undermining the financial distress index, which includes working capital to total assets (X1), retained earnings to total assets (X2), earnings before interest, and taxes to total assets (X3), market value of equity to book value of total liabilities (X4) and sales to total assets (X5). Furthermore, the Knowledge of Data Discovery (KDD) executed 755 data records of financial reports from the Indonesia Stock Exchange during the Year 2016-2019 to analyze its accuracy and error rate using this combining approach. The Confusion Matrix showed that algorithms C4.5 and AdaBoost forecast were 13.52% and 62.17% more precise than the original C4.5 and Altman's model, respectively, in ratio training tested data 90%:10%. This study, therefore, revealed the substantial contribution of C4.5 and Adaboost to company financial distress forecasting.

Keywords: Adaboost, C4.5 Algorithm, Financial distress, Datamining, Prediction

1. Introduction

Financial distress is defined as a company's worst financial condition according to the unstable values of profitability, lack of solvency capability, deferred payments of obligation or interest, deferred dividend payments and other defaults financial claims and insufficient cash flow, therefore, making the company suffer a decline in bankrupt [1, 2]. Indeed, the global financial crisis due to Covid-19 and digitalization disruption severely impacted company losses and even led to bankruptcy, including Indonesia's economic situation [3]. The lack of a financial risk advisory system to correctly deal with financial problems and disruptions has worsened the global and internal economic climate through financial deterioration and distress.

Therefore, a company should accustom itself to a thoroughly developed internal management structure and financial distress prediction (FDP) scheme. FDP administers the social problem handling and decision-making recommendations to solve unemployment, economic depression, and financial crisis issues [4]. Here, the most adopted statistical FDP analysis models include Altman's Z -score, Beaver's univariate, and Ohlson's Logit [5]. Charalambakis and Garrett deployed a multi-period logit model in examining the probability of profitability, leverage, the ratio of retained earnings-total assets, size, the liquidity ratio, an export dummy variable, the tendency to pay out dividends, and the growth rate in real Gross Domestic Product (GDP) as critical success factors in predicting the financial distress for Greek private firms [6]. Similarly, Kisman and Krisandi investigated using traditional accounting/financial variables and the logit models to create distress prediction models [7]. However, Chen et al. studied the significant contribution of Corporate Government (CG), including management and ownership, to complete the traditional financial bankruptcy prediction models in Taiwan. A logit model estimation with the dynamic distress threshold (DDT) values found the positive improvements of FDP through the inclusion of CG over only using the traditional financial variables [8]. MacCarthy tried to enhance Altman Z-score's advantages with Beneish M-model by considering the possible emergence of manipulating management in detecting financial fraud and company bankruptcy manifestation. This research revealed that the simultaneous use of these two models provides a better financial statements audit in predicting the bankruptcy instead of the Altman Z-score model alone [9]. Turk and Kurklu have successfully demonstrated the significant contribution of the easily computable and understandable Altman (Z-Score) and Springate (S-Score) models in forecasting the bankruptcy and financial failures of 166 companies in Istanbul Stock Exchanges. This research revealed that Altman (Z-Score) and Springate (S-Score) models equip similar results in determining the financial failure level. Indeed, both models fulfill the consideration of non-financial variables such as management policies and strategies instead of financial statements in predicting the financial risks of company failures [10]. To accomplish the beneficence of Altman scoring, Sugiyarti and

*Corresponding author. Tel.: +62823 8680 1721
Email address: okfalisa@gmail.com

doi: 10.14456/easr.2022.31

Murwaningsari convinced the higher accuracy of the Altman Z-score model than the Grover model in predicting the insolvency of retail companies on the Indonesia stock exchange [11]. However, Indriyanti and Gustyana analyzed the accuracy and error rate calculation of Altman Z-Score, Springate, Grover, Zmijewski, and Zavgren using the confusion matrix formula. Thus, they found that the Springate model performed the most accurate and lowest error in foreseeing the financial statements of each retail trading company in Indonesia and then following by Altman Z-Score, Grover, Zmijewski, and Zavgren, respectively [12]. The previous reviews indicated the positive challenges of Altman Z-Score compare to others in predicting financial distress. Nonetheless, the integration of the Altman Z-score with others models in terms of financial and non-financial variables consideration increases the values of this model in forecasting the company failure leveling index.

In addition, the breakthroughs artificial intelligence (AI) models evolve in this system through neural networks (NNs), decision tree, case-based reasoning, and support vector machine (SVM), which enhances the efficacy of such an approach [13]. However, these initial models adhere to single classifier models, which in some instances are unstable. Therefore, the current FDP focuses more on the various classifiers designed by random forest, boosting, bagging, and AI approaches. It indicates that hybrid FDP with AI models is always more outstanding than statistical ones. The considerable ensemble of classifiers further improves the FDP performance. Instead of conventional statistical, it integrates the complementary information output that considers the restricted assumption such as linearity, normality, independence variables and their correlations, and pre-existing functional forms of variables [14]. Unfortunately, the financial distress concept drift has failed to be considered and dynamically updated [15].

For instance, Salehi et al. compared the accuracy and error rate of data mining methods, including SVM, ANN, K-Nearest Neighbors (KNN), and Naïve Bayesian Classifier (NBC), in predicting the company financial distress based on the five variables of Altman Z-score estimation. The experiments revealed that ANN outperformed this comparison. Thus, it is pursued by SVM, KNN, and NBC, respectively [14]. Thus, Salehi and Pour have successfully applied the ANN approach with 5 inputs and 1 output using 840 data to classify the bankruptcy and non-bankruptcy of manufacturing companies in the Tehran Stock Exchange [16]. In similar cases, Fathi et al. tried to investigate the role of data mining models such as SVM, decision-making tree, and neural network to predict the production company failure. The study revealed that data mining models delivered a predictive ability of up to 92.4 percent whereby the Altman model yielded 82.41 percent prediction [17]. Therefore, this study tries to apply the potential classification algorithms in data mining that significantly drop the imbalanced class of the FDP dataset in predicting the accuracy level for financial distress instead of statistical FDP models.

Data mining is a knowledge discovery process to discover hidden patterns and exciting insights of large amounts of data and information repositories dynamically streamed to the system [18]. Furthermore, it provides the strength of algorithms in load prediction and pattern identification in many ways within supervised and unsupervised learning [19]. For instance, Agustina et al. carried out a study on the effective utilization of Support Vector Regression (SVR/SVM) in identifying the most contributor's foodstuff based on the demographic bonus [20]. Srinidhi et al determined the polarity of the sentence in sentiment analysis classification using the combination of a recurrent neural network (RNN) and SVM [21]. Okfalisa et al. compared the power of the linear regression and K-Nearest Neighbors in solving the scholarship recipient problem [22]. Furthermore, Okfalisa et al. applied the correlation of Modified K-Nearest Neighbors and K-Nearest Neighbors for classifying the data of Conditional Cash Transfer Implementation Unit [23] and Earthquake Building Structure Strength Prediction [24]. Meanwhile, Waseem et al. observed the application of decision trees, Naïve Bayes, and Random forest algorithms for analyzing diabetes and blood pressure datasets [25]. The comparison amongst the data mining classifier, including the C4.5 decision tree, Naïve Bayes, SVM, and Random Forest, identified that the C4.5 decision tree revealed the highest accuracy, specificity, sensitivity and proposed the most valuable data classification [26] and [27]. Husejinovic supported the positive role of C4.5 to detect credit card fraud instead of Naïve Bayesian [28]. It indicates that the C4.5 algorithm handles discrete or continuous type attributes, making it easier to group values based on predetermined criteria. In addition, it deals with training data with missing attribute values and prunes trees built by removing unaffected branches. [29]. The C.45 algorithm has advantages in tree diagram analysis that is easy to understand and make. This is because it requires less experimental data than other classification algorithms, may be validated using statistical techniques, faster time computation than other classification techniques, and its accuracy to match other classification techniques [30]. However, some disadvantages overlap, such as too many classes, designing optimal decision trees, high dependence on the tree design, and unreliable trees for small fluctuations in a small dataset [25].

To overcome the above weaknesses, the Adaboost ensemble is introduced to increase the accuracy of data distribution. Adaboost and its variants have been successfully applied in several fields due to their solid theoretical foundation, accurate predictions, and extraordinary simplicity. Lestari and Alamsyah increased the accuracy of C4.5 by 0.83% using information gain ratio and AdaBoost for Chronic Kidney Disease classification case [31]. Meanwhile, Damrongsakmethee and Neagoe enhanced C4.5 with Adaboost for Credit Scoring Modelling [32]. Therefore, this study tried to adopt the C4.5 with Adaboost for predicting the financial distress in Indonesia's company. Here, the basic calculation of the FDP system explored the power of Altman's model.

2. Materials and methods

This study follows the Knowledge Data Discovery (KDD) activities from data preparation and collection, pre-processing, transformation, and data mining with C4.5 and Adaboost [22]. It ends with the evaluations to examine the pattern of data or information provided in forecasting the company's financial distress. The analysis used a total of 755 secondary data records of financial reports from the Indonesia Stock Exchange between the years 2016-2019. A total of 10 parameters were determined from the finance ratio by the Altman model and adds on with identity code. Table 1 explains the finance report parameters.
Finance ratio formula by Altman model.

$$Z = 1{,}2\ X_1 + 1{,}4\ X_2 + \ 3{,}3X_3 + 0{,}6\ X_4 + 1{,}0\ X_5 \qquad (1)$$

Where Z score index is calculated by considering the values of parameter X1 as working capital to total assets, X2 as retained earnings to total assets, X3 as earnings before interest and taxes to total assets, X4 as market value of equity to book value of total liabilities and X5 as sales to total assets. The Altman model categorizes the Z score of the company into three classes [33]: the class > 2,99 as a company with healthy finance, < 1,81 as a company with distress finance and an undefined condition allocated to the company with values between 1,81 and 2,99.

Table 1 Finance report parameters

No	Parameters	Information	Type of Data
1	Code	Company code	Discrete
2	Total Assets	Total company assets	Continue
3	Networking Capital	Networking capital obtained from total current assets that are reduced by current liabilities	Continue
4	Current Assets	Total current assets	Continue
5	Current Liabilities	Current Liabilities	Continue
6	Retained Earnings	Retained retained earnings	Continue
7	Earnings Before Interest and Taxes (EBIT)	Earnings before interest and taxes	Continue
8	Market Value Equity	Capital market value	Continue
9	Value of Total Debt	The value of debt obtained from total current liabilities plus non-current liabilities	Continue
10	Non-current Liabilities	Noncurrent liabilities	Continue
11	Sales	Total sales	Continue

The procedure of the C4.5 algorithm follows the four steps as below [34].

1. Computing the entropy of the initial information from the sample data set S

$$\text{Entropy (S)} = \sum_{i=1}^{m} p_i log_2 p_i \tag{2}$$

where pi is the percentage of the class, *i* samples in all samples, and m is the number of classes. If all data has the same class label, *m = 1, pi = 1*, the entropy is zero. Meanwhile, if each data has its class label, *pi = 1/m*, the entropy value is the largest, *Entropy (S)= log2 m*.

2. Calculating the entropy of the separation of the sample dataset S

Assume that the attribute partitions S to be SL and SR. Find the entropy divided as the weighted entropy of each subset, which is given as:

$$\text{Entropy}_A\ (S) = \frac{|S_L|}{|S|}\text{Entropy}(S_L) + \frac{|S_R|}{|S|}\text{Entropy}(S_R) \tag{3}$$

A is an attribute of C, SL, and SR is a subset of the set S separated by A. |S| is the number of samples in S. |SL| and |SR| is the number of models in SL and SR, respectively.

3. Obtaining the gain values of attribute A

To determine whether the selected attribute A can reduce the overall entropy effectively, the information obtained from feature A can be defined as follows:

$$\text{Gain (A)} = \text{ Entropy (S)– EntropyA (S)} \tag{4}$$

4. Calculating the ratio to the acquisition of information

The C4.5 algorithm introduces distinct information values to avoid being overfitted and normalizing information retrieval. It can be written as:

$$SplitInfo\ (A) = \sum_{i=1}^{k} \frac{|S_i|}{|S|}\log_2 \frac{|S_i|}{|S|} \tag{5}$$

For attribute A, the information acquisition ratio is written as:

$$GainRatio\ (A) = \frac{Gain\ (A)}{SplitInfo\ (A)} \tag{6}$$

The information acquisition ratio needs to be calculated for each Decision Tree node. The attributes with the maximum information retrieval ratio are selected and stored in the appropriate nodes.

A hybrid C4.5 and Adaboost resumes C4.5 procedures as a learning-based algorithm to develop the decision tree and discover the maximum iteration. Meanwhile, Adaboost acts as an ensemble machine learning method that improves classification performance. In addition, it handles the performance improvements by associating the weights with training data within different classification conditions and replace the training error rate with confidence values.

The steps in the Adaboost algorithm are as follows [35].

1. Initializing the sample weighted. Where t=1,… T using the distribution of data sample D_t.

$$D_1(i) = \frac{1}{m} \text{for } i = 1, \ldots, m \tag{7}$$

2. Computing the error of base weak learner ht:

$$\epsilon_t = \sum_{i:h_t(x_i)\neq y_i} D_t(i) \tag{8}$$

If error rates of the base classifiers ϵt ≥ 1⁄2, then set the number of boosting rounds at *T=t-1*, cancel the loop and go straight to the output.

3. Determining the weight of ht, with the following equation:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{9}$$

4. Updating the training sample weight with the following equation:

$$D_{t+1}(i) = \frac{D_t(i)}{z_t} \times \begin{cases} e^{-\alpha_t}\ if\ h_t(x_i) = y_i \\ e^{\alpha_t}\ if\ h_t(x_i) \neq y_i \end{cases} \tag{10}$$

Where zt is a normalizing factor that activates Dt + 1 (i) to be the distribution. Output is calculated with the following equation:

$$H\ (x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{11}$$

Herein, the AdaBoost algorithm uses the train error rate of a base classifier in the process of boosting through the updating instance weights and in the process of decision making in determining the vote weights. Confusion Matrix and Rapid Miner comparison testing are carried out to evaluate the accuracy of class prediction. It measures the percentages of accuracy and error-rate by considering the values of TP (True Positive) as the quantity of correctly categorized data, actual (yes) and predicted class (yes), TN (True Negative) as the quantity of correctly classified data, actual (no) and predicted class (no). Furthermore, it considers FN (False Negative) as the quantity of incorrectly categorized data, actual (yes), predicted class (no), FP (False Positive) as the quantity of incorrectly categorized data, actual class (no), predicted class (yes), P as the total numbers of TP and FN and N as the whole numbers of FP and TN using the following formula [24].

$$\text{Accuracy} = \frac{TP+TN}{P+N} \times 100\% \tag{12}$$

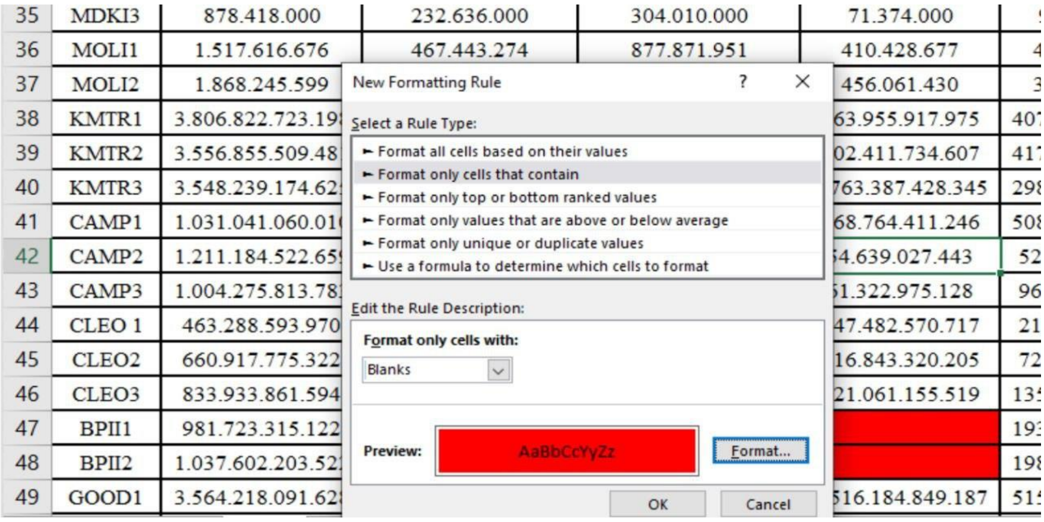
$$\text{Error}-\text{rate} = \frac{FP+FN}{P+N} \times 100\% \tag{13}$$

Besides, K-fold validation (k=10) is used to accomplish the prediction accuracy analysis instead of Confusion Matrix alone. K-fold validation is adopted to scrutinized the potential data model by considering the parameters of K [23]. The data training and tested data partition utilization simulates at 50:50, 60:40, 70:30, 80:20, 90:10. Therefore, 90:10 is defined as the highest partition performance.

3. Results and discussions

3.1 Data preparation

A total of six records of missing values, three records of duplicating data, no inconsistency, and outlier data were identified during data pre-processing using excel tools. It revealed a reduction from 755 data records to a total of 746 data. Figure 1 shows an example of the missing values investigation.



35	MDKI3	878.418.000	232.636.000	304.010.000	71.374.000	
36	MOLI1	1.517.616.676	467.443.274	877.871.951	410.428.677	4
37	MOLI2	1.868.245.599			456.061.430	3
38	KMTR1	3.806.822.723.19			63.955.917.975	407
39	KMTR2	3.556.855.509.48			02.411.734.607	417
40	KMTR3	3.548.239.174.62			763.387.428.345	29
41	CAMP1	1.031.041.060.01			68.764.411.246	50
42	CAMP2	1.211.184.522.65			4.639.027.443	52
43	CAMP3	1.004.275.813.78			1.322.975.128	96
44	CLEO 1	463.288.593.970			47.482.570.717	21
45	CLEO2	660.917.775.322			16.843.320.205	72
46	CLEO3	833.933.861.594			21.061.155.519	13
47	BPII1	981.723.315.122			🟥	19
48	BPII2	1.037.602.203.52			🟥	19
49	GOOD1	3.564.218.091.62			516.184.849.187	51

New Formatting Rule

Select a Rule Type:

- Format all cells based on their values
- Format only cells that contain
- Format only top or bottom ranked values
- Format only values that are above or below average
- Format only unique or duplicate values
- Use a formula to determine which cells to format

Edit the Rule Description:

Format only cells with: Blanks

Preview: AaBbCcYyZz

Figure 1 The missing values inspection

Table 2 shows the data integration with five Altman parameters (X1, X2, X3, X4, and X5) (Please refer to Equation 1) following Altman's model and the results calculated from the Finance ratio of the Z-score index.

Table 2 Finance ratio integration

No	X1	X2	X3	X4	X5	Index Status
1	-0,33	0,11	0,084	0,443	0,61	Distress
2	0,245	0,053	0,073	1,17	0,343	Distress
3	0,267	0,095	0,077	0,962	0,476	Distress
4	0,191	0,116	0,089	1,074	0,526	Undefined
5	0,148	0,103	0,045	1,032	0,595	Distress
6	0,124	0,146	0,073	1,147	0,747	Undefined
7	0,019	0,132	0,059	0,636	0,759	Distress
8	0,07	0,144	0,070	0,546	0,78	Distress
9	-0,154	-0,229	0,012	0,754	0,846	Distress
10	0,143	-0,112	0,054	2,045	0,835	Undefined
11	0,042	0,124	0,161	0,893	1,212	Undefined
…	…	…	…	…	…	…
701	-0,005	-0,245	-0,062	0,758	0,186	Distress
702	-0,036	-0,410	-0,144	0,582	0,194	Distress
703	0,150	0,131	-0,041	0,630	0,510	Distress
704	0,195	0,118	-0,063	0,685	0,425	Distress
705	0,191	0,151	0,815	0,442	0,302	Healthy
706	0,215	0,155	0,072	0,443	0,807	Distress
707	0,193	0,375	0,118	2,427	1,349	Healthy
708	0,248	0,440	0,159	3,001	1,130	Healthy
…	…	…	…	…	…	…
741	0,030	0,339	0,049	0,954	0,432	Distress
742	-0,016	0,309	0,016	0,807	0,356	Distress
743	-0,057	-0,050	-0,085	1,272	0,040	Distress
744	-0,094	-0,082	-0,029	1,492	0,088	Distress
745	-0,131	-0,029	0,055	1,369	0,079	Distress
746	-0,142	0,066	0,102	1,430	0,229	Distress

3.2 C.45 and Adaboost for data mining

The execution of the C4.5 and Adaboost procedure revealed the weighted entropy, and gained values for a maximum of 10 times iteration, and came to a stop at the fifth iteration when the condition error value was ≥ 0.5. The iteration analysis result is depicted in Table 3 and DT in Figure 2.

Table 3 C4.5 and Adaboost iteration analysis

No	X	Iteration #1			Iteration #2		
		Value Attribute	Entropy	Gain	Value Attribute	Entropy	Gain
1	X1	<=0,164 >0.164	1,2752 1,4363	0,1799	<=0,158 >0,158	1,359 1,511	0,127
2	X2	<=0,111 >0,111	1,2643 1,5062	0,1487	<=0,07 >0,07	1,327 1,573	0,086
3	X3	<=0,046 >0,046	1,3274 1,4916	0,1270	<=0,046 >0,046	1,396 1,520	0,107
4	X4	<=1,163 >1,163	1,1477 1,3800	0,2705	<=1,142 >1,142	1,231 1,425	0,233
5	X5	<=0,565 >0.565	1,3011 1,4634	0,1540	<=0,747 >0.747	1,313 1,554	0,108

No	X	Iteration #1			…	Iteration #2		
		Value Attribute	Entropy	Gain	…	Value Attribute	Entropy	Gain
1	X1	<=0,165 >0.165	1,5203 1,3722	0,1199	…	<=0,171 >0.171	1,5649 1,4249	0,0764
2	X2	<=0,135 >0,135	1,5660 1,4394	0,0651	…	<=0,133 >0,133	1,5740 1,4702	0,0508
3	X3	<=0,076 >0,076	1,5583 1,3830	0,0652	…	<=0,049 >0,049	1,5692 1,4453	0,0638
4	X4	<=1,285 >1,285	1,5251 1,3245	0,1371	…	<=1,270 >1,270	1,5100 1,3652	0,1352
5	X5	<=0,606 >0.606	1,5380 1,3996	0,0947	…	<=0,578 >0.578	1,5591 1,4442	0,0719

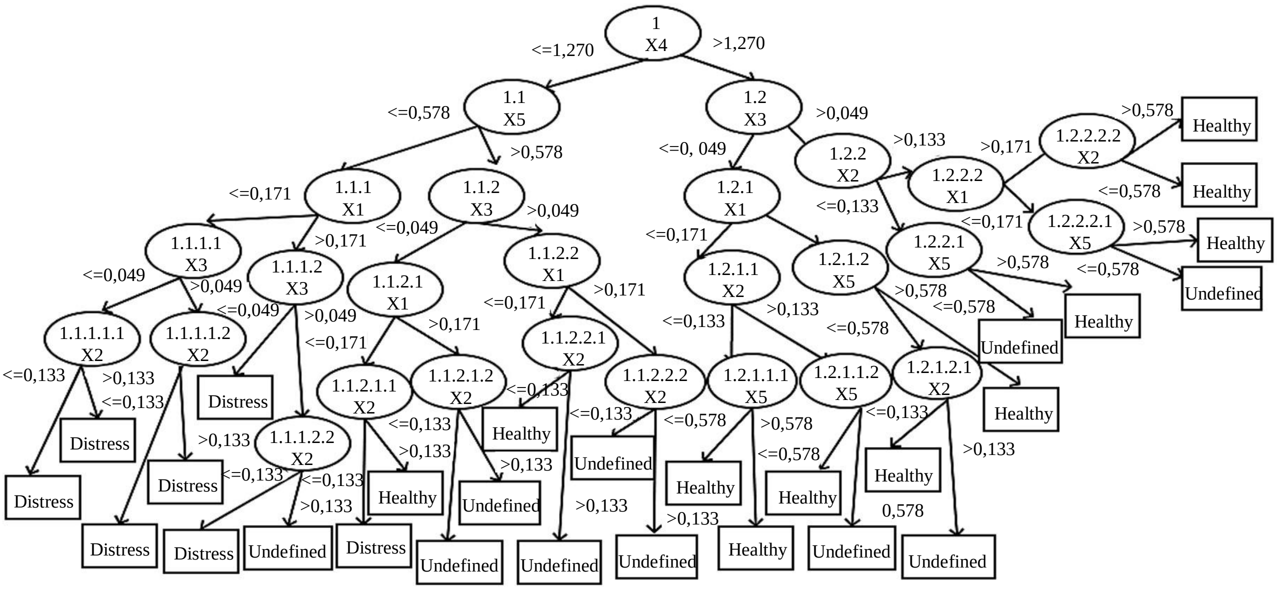


Figure 2 DT for iteration #5

3.3 Prediction evaluation

Table 4 explained the comparisons of accuracy and error rate values for company financial distress using Altman's model, C.45, and the combining of C.45 and AdaBoost, following the Confusion Matrix formula at Equations (12) and (13). Furthermore, it illustrated that Adaboost could enhance the accuracy of C4.5 and Altman's model up to 13.52% and 62.17%, respectively, for the highest performance of data split simulation at 90:10. The values of error-rate calculation for Adaboost ensembles also showed a significant linear contribution in boosting the C4.5 performance. Compared to previous studies, similar works were also supported by Lestari and Alamsyah [31] and Damrongsakmethee and Neagoe [32]. In addition, Beigi and Amin-Naseri discovered that Adaboost potentially reduced the misclassification cost by at least 14% compared to the C4.5 decision tree, naïve Bayes, bayesian network, neural network, and artificial immune system for detecting the credit card fraud real-time data mining [36]. In addition to applying the data mining analysis (Artificial Intelligence), Altman's Z-score prediction and classification are improved using the country-specific estimation that incorporates additional variables [34].

Table 4 Confusion matrix comparison analysis.

Data Partition	Actual Class	Altman's Model			C.45			C.45 and Adaboost		
		Prediction Class								
		Healthy	Undefined	Distress	Healthy	Undefined	Distress	Healthy	Undefined	Distress
90:10	Distress	2	23	0	23	0	2	19	5	1
	Healthy	0	16	0	4	0	12	2	13	1
	Undefined	0	33	0	2	0	31	0	1	32
	Accuracy		24,32%			72,97%			86,49%	
	Error-rate		75,68%			27.03%			13,51%	
80:20	Distress	0	51	0	4	46	1	47	3	1
	Healthy	0	33	0	3	21	8	16	8	8
	Undefined	0	66	0	1	15	50	5	1	60
	Accuracy		22%			50,34%			77,18%	
	Error-rate		78%			49,66%			22,82%	
70:30	Distress	4	76	0	44	13	19	57	13	6
	Healthy	1	48	0	17	11	21	19	14	15
	Undefined	0	99	0	3	20	76	1	5	93
	Accuracy		23,21%			58,48%			73,54%	
	Error-rate		76,79%			41,52%			26,46%	
60:40	Distress	0	102	0	84	18	0	79	16	7
	Healthy	0	65	0	43	22	0	22	19	23
	Undefined	0	132	0	31	101	0	2	9	121
	Accuracy		21,74%			35,45%			73,49%	
	Error-rate		78,26%			64,55%			26,51%	
50:50	Distress	0	127	0	0	127	0	83	0	45
	Healthy	0	81	0	0	81	0	40	0	0
	Undefined	0	164	0	0	164	0	17	14	133
	Accuracy		21,77%			21,77%			57,91%	
	Error-rate		78,23%			78,23%			42,09%	

Table 5 showed that C4.5 outperform at K parameters and accuracy level in 8 and 85.12%, respectively. Concurrently, C4.5 and Adaboost revealed the optimum accuracy (86.49%) at parameters K=10. This result supports the analysis of confusion matrix calculation.

Table 5 K-fold validation comparison analysis.

K values	C4.5	C4.5 dan *AdaBoost*
5	84,18 %	73,49 %
6	84,58 %	73, 49 %
7	83,65 %	73,54 %
8	85,12 %	77, 18 %
9	84, 72 %	77, 18 %
10	83,78%	86, 49%

4. Conclusions

Based on the results of this study, it is evident that C4.5 and Adaboost increased the accuracy and optimum error rate in forecasting the company's financial distress, compared to Altman's model and original C4.5. Furthermore, it revealed that Adaboost intensifies the performance of C4.5 and Altman's model to provide a more accurate analysis of financial distress up to 2.71% and 13.52% based on K-fold validation and Confusion Matrix tested, respectively. Therefore, enabling the company's managers to immediately detect the abnormal financial flows that trigger the emergence of a bankrupt situation. In addition to the Index Status, this prediction is equipped with the performance analysis of Altman's considered variables which includes working capital to total assets (X1), retained earnings to total assets (X2), earnings before interest, and taxes to total assets (X3), market value of equity to book value of total liabilities (X4) and sales to total assets (X5). The analysis revealed that the Altman model successfully predicted 2 Healthy and 16 Undefined companies. Meanwhile, C4.5 and Hybrid C4.5 and Adaboost precisely predicted 23 Healthy and 31 Distress company and 19 Healthy, 13 Undefined, and 32 Distress company, respectively. Therefore, this prediction enables the company to accurately identify the bankruptcy company level and take curative action to minimize the possibilities of risks and organizational distress.

5. Acknowledgements

The authors express gratitude to the Faculty of Science and Technology of the Universitas Islam Negeri Sultan Syarif Kasim Riau and Central Bank Indonesia for their generous assistance in contributing significantly to the feasibility and efficacy of this study. Furthermore, the entire contributors, including the Faculty of Management Science, Suan Dusit University, collaborate.

6. References

[1] Geng R, Bose I, Chen X. Prediction of financial distress: an empirical study of listed Chinese companies using data mining. Eur J Oper Res. 2015;241(1):236-47.

[2] Mselmi N, Lahiani A, Hamza T. Financial distress prediction: the case of French small and medium-sized firms. Int Rev Financ Anal. 2017;50:67-80.

[3] Ashraf S, Felix EGS, Serrasqueiro Z. Do traditional financial distress prediction models predict the early warning signs of financial distress?. J Risk Financ Manag. 2019;12(2):55.

[4] Al-Fatih S, Ahsany F, Alamsyah AF. Legal protection of labor rights during the coronavirus disease 2019 (Covid-19) pandemic. J Pembaharuan Hukum. 2020;7(2):100.

[5] Sun J, Li H, Fujita H, Fu B, Ai W. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. Inform Fusion. 2020;54:128-44.

[6] Charalambakis EC, Garrett I. On corporate financial distress prediction: what can we learn from private firms in a developing economy? Evidence from Greece. Rev Quant Finan Acc. 2019;52(2):467-91.

[7] Kisman Z, Krisandi D. How to predict financial distress in the wholesale sector: lesson from Indonesian stock exchange. J Econ Bus. 2019;2(3):18.

[8] Chen CC, Chen CD, Lien D. Financial distress prediction model: the effects of corporate governance indicators. J Forecast. 2020;39(8):1238-52.

[9] Maccarthy J. Using Altman Z-score and Beneish M-score models to detect financial fraud and corporate failure: a case study of Enron Corporation. Int J Finance Account. 2017;6(6):159-66.

[10] Turk Z, Kurklu E. Financial failure estimate in Bist companies with Altman (Z-Score) and Springate (S-Score) models. J Econ Admin Sci. 2017;1(1):1-14.

[11] Sugiyarti L, Murwaningsari E. Comparison of bankruptcy and sustainability prediction: Altman Z score versus Grover model. Selangor Bus Rev. 2020;5(2):56-72.

[12] Indriyanti ND, Gustyana TT. Analysis of bankruptcy prediction using Altman Z-Score, Springate Grover, Zmijewski and Zavgren in retail trade sub sectors registered in Indonesia stock exchange period 2015-2019. Int J Advan Res Econ Finance. 2021;3(1):21-31.

[13] Liang D, Tsai CF, Lu HY, Chang LS. Combining corporate governance indicators with stacking ensembles for financial distress prediction. J Bus Res. 2020;120(5):137-46.

[14] Salehi M, Shiri MM, Pasikhani MB. Predicting corporate financial distress using data mining techniques. Int J Law Manag. 2016;58(2):216-30.

[15] Shen F, Liu Y, Wang R, Zhou W. A dynamic financial distress forecast model with multiple forecast results under unbalanced data environment. Knowl Base Syst. 2020;192:105365.

[16] Salehi M, Pour MD. Bankruptcy prediction of listed companies in Tehran Stock Exchange. Int J Law Manag. 2016;58(5):545-61.

[17] Fathi S, Saif S, Heydari Z. Predicting bankruptcy of companies using data mining models and comparing the results with Z Altman model. Int J Finance Manag Account. 2018;3(10):33-46.

[18] Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3th ed. Burlington: Morgan Kaufmann; 2011.

[19] Noyunsan C, Katanyukul T, Saikaew K. Performance evaluation of supervised learning algorithms with various training data sizes and missing attributes. Eng Appl Sci Res. 2018;45(3):221-9.

[20] Agustina SD, Mustakim, Okfalisa, Bella C, Ramadhan MA. Support vector regression algorithm modeling to predict the availability of foodstuff in Indonesia to face the demographic bonus. J Phys Conf Ser. 2018;1028:012240.

[21] Srinidhi H, Siddesh GM, Srinivasa KG. A hybrid model using MaLSTM based on recurrent neural networks with support vector machines for sentiment analysis. Eng Appl Sci Res. 2020;47(3):232-40.

[22] Okfalisa, Fitriani R, Vitriani Y. The comparison of linear regression method and K-Nearest neighbors in scholarship recipient. 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD); 2018 Jun 27-29; Busan, South Korea. New York: IEEE; 2018. p. 194-9.

[23] Okfalisa, Gazalba I, Mustakim, Reza NGI. Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE); 2017 Nov 1-2; Yogyakarta, Indonesia. New York: IEEE; 2017. p. 294-8.

[24] Okfalisa O, Nugraha S, Saktioto S, Zulkifli Z, Fauzi S. The prediction of earthquake building structure strength: modified k-nearest neighbour employment. Indonesian J Electr Eng Informat. 2020;8(4):733-45.

[25] Waseem SN, Laith TR, Hazim SM, Aoras SE. Analysis and prediction blood pressure and disease by applying decision tree, naïve base and random forest algorithms. Indian J Public Health Res Dev. 2020;11(4):1736-40.

[26] Baswardono W, Kurniadi D, Mulyani A, Arifin DM. Comparative analysis of decision tree algorithms: random forest and C4.5 for airlines customer satisfaction classification. J Phys Conf Ser. 2019;1402(6):066055.

[27] Pah CEA, Utama DN. Decision support model for employee recruitment using data mining classification. Int J Emerg Trends Eng Res. 2020;8(5):1511-6.

[28] Husejinovic A. Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers. Period Eng Nat Sci. 2020;8:1-5.

[29] Abellan J, Mantas CJ, Castellano JG. AdaptativeCC4.5: Credal C4.5 with a rough class noise estimator. Expert Syst Appl. 2018;92:363-79.

[30] Rahim R, Zufria I, Kurniasih N, Simargolang MY, Hasibuan A, Sutiksno DU, et al. C4.5 Classification data mining for inventory control. Int J Eng Technol. 2018;7(2.3):68-72.

[31] Lestari A, Alamsyah. Increasing accuracy of C4.5 algorithm using information gain ratio and Adaboost for classification of chronic kidney disease. J Soft Comput Explor. 2020;1(1):32-8.

[32] Damrongsakmethee T, Neagoe VE. C4.5 Decision tree enhanced with AdaBoost versus multilayer perceptron for credit scoring modeling. Adv Intell Syst Comput. 2019;1047:216-26.

[33] Mselmi N, Lahiani A, Hamza T. Financial distress prediction: the case of French small and medium-sized firms. Int Rev Financ Anal. 2017;50:67-80.

[34] Meng X, Zhang P, Xu Y, Xie H. Construction of decision tree based on C4.5 algorithm for online voltage stability assessment. Int J Electr Power Energ Syst. 2020;118:105793.

[35] Wang F, Li Z, He F, Wang R, Yu W, Nie F. Feature learning viewpoint of Adaboost and a new Algorithm. IEEE Access. 2019;7:149890-9.

[36] Beigi S, Amin Naseri M. Credit card fraud detection using data mining and statistical methods. J AI Data Min. 2020;8(2):149-60.