EARLY WARNING SIGNS OF FINANCIAL DISTRESS USING RANDOM FOREST AND LOGIT MODEL

Valentino Budhidharma *, Roy Sembel **, Edison Hulu **, Gracia Ugut **

* *Corresponding author*, Universitas Pelita Harapan, Tangerang, Indonesia
Contact details: Universitas Pelita Harapan, Jl. M. H. Thamrin Boulevard 1100, Lippo Village, Tangerang 15811, Indonesia
** Universitas Pelita Harapan, Tangerang, Indonesia

OPEN ACCESS

How to cite this paper: Budhidharma, V., Sembel, R., Hulu, E., & Ugut, G. (2023). Early warning signs of financial distress using random forest and logit model. *Corporate & Business Strategy Review, 4*(4), 69–88. https://doi.org/10.22495/cbsrv4i4art8

ISSN Online: 2708-4965
ISSN Print: 2708-9924

Received: 23.01.2023
Accepted: 27.09.2023

JEL Classification: G32
DOI: 10.22495/cbsrv4i4art8

Abstract

The purpose of this study is to develop a new model to explain financial distress in Indonesia. There have been many theories, variables, and estimation methods used by previous studies about early warning signs of financial distress. Unfortunately, there are few studies on this subject using a combination of theories, random forests (RF) as the machine learning algorithm, and logit as the statistical method, especially in Indonesia. By using the RF, it is expected the study can get an improved combination of classification and regression tree (CART) and bagging (Breiman, 1996). The samples used are most sectors in Indonesia Stock Exchange (IDX) from 2005 to 2020, excluding the financial sector. The results show that cash to total assets (CTA), retained earnings to total assets (RETA), quick assets to total assets (QATA), earnings before tax to current liabilities (EBTCL), total liability to total assets (TLTA), total sales (TS), book value per share (BVPS), and market to book ratio of the firm (MB) have a negative significant association with the probability of firms in distress. While current assets to total assets (CATA), quick assets to current liabilities (QACL), total liabilities to market value of total assets (TLMTA), total assets (TA), and interest rate (INTEREST) have a positive significant association with the probability of firms in distress. In conclusion, to avoid financial distress firms must have good selling while maintaining enough cash flow to fulfill their short-term liabilities. Firms must also keep on growing to become bigger so they can withstand more crises. This condition must be supported by a conducive interest rate. Another result shows that combining theories, random forests, and logit can be used to build a new financial distress prediction model. The second result is a new enlightenment since this method can be used to develop many new financial study models, not only using logit estimates but also other estimation methods.

Keywords: Financial Distress, Random Forests, Logit

Authors' individual contribution: Conceptualization — V.B.; Methodology — V.B.; Software — V.B.; Validation — R.S., E.H., and G.U.; Formal Analysis — R.S. and E.H.; Investigation — V.B.; Resources — V.B.; Data Curation — V.B.; Writing — Original Draft — V.B.; Writing — Review & Editing — V.B.; Visualization — R.S.; Supervision — R.S., E.H., and G.U.; Project Administration — G.U.; Funding Acquisition — G.U.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

Acknowledgements: The Authors appreciate Khoirunurrofik, Maydison C. Ginting, and seminar participants at the Universitas Pelita Harapan for their helpful comments and suggestions.

1. INTRODUCTION

Since 1973, when Indonesia Stock Exchange (IDX) was first times open, Indonesia had experienced two times economic recessions in 1998 and 2020, and one capital market crash in 2008. During a financial crash, some firms experience financial distress. Early warning signals can be beneficial so that firms can do preventive action to prevent bankruptcy, while for investors the signals can protect them from bad investments by cashing out.

During a recession or crash in the capital market, the ability to predict firms in distress is good to have. Financial distress is a condition where a company cannot pay its obligations, either short-term debts or long-term debts. In the worst case, a company will go into bankruptcy. A bankrupt company will experience a big loss to the shareholders and the stakeholders. An ability to predict or give early warning signals of financial distress will be beneficial to market players and firms. That is why this topic has been studied by many economists for decades (Ashraf et al., 2019).

The ability to predict the probability of financial distress is beneficial to investors. Investors can predict firms that have good financial health and performance. In the capital market, having the ability to detect firms in distress can prevent investors from losing money while having the ability to ensure firms' financial health can give assurance about good firms to invest money.

Machine learning has recently become a new trend in data analytics. Traditional data analytics uses statistical modelling as its core methodology, while data mining uses machine learning algorithms. Both methods have their strengths and weaknesses. The research gaps of the study are there are few studies: 1) covering the economic recession of 2020 and the capital market crash of 2008 in Indonesia yet, 2) covering most sectors, excluding the financial sector, on the IDX in Indonesia, 3) that develop a new model using a combination of theories, machine learning algorithm, and statistical methods, and 4) that compare the results with the existing model in Indonesia. The research aim is to develop a new better model for financial distress prediction in Indonesia, and the research question is what is the better model to predict the probability of financial distress in Indonesia?

Firms do not enter financial distress in an instant. A study in the UK shows that firms take up to three years before they go bankrupt (Tinoco & Wilson, 2013). The significance of this study is the development of a new financial distress prediction model and the introduction of a new methodology to develop models. Most studies about financial distress use theories and statistical methods to develop a model. This study combines theories, machine learning algorithms, and statistical methods, which can give a better approach to developing new models. By collecting all financial, market, and macroeconomic variables that have been used in previous financial distress studies, the importance of the variables is analyzed using Random Forests as the machine learning algorithm. Then, multiple Logit models are developed by combining all variables sorted by the most important. The best model is picked by maximizing specificity and sensitivity and minimizing miss-calculation error and root-mean-square error (RMSE).

The main finding is a new financial distress model specifically for Indonesia. Another result shows that combining theories, random forests, and Logit can be used to build a new financial distress prediction model. The second result is a new enlightenment since this method can be used to develop many new financial study models, not only using Logit estimates but also other estimation methods.

The structure of this paper is as follows. Section 1 presents the introduction. Section 2 reviews the relevant literature: financial distress, market, and macroeconomic effects. Section 3 analyzes the methodology that has been used to conduct empirical research on financial distress. Section 4 presents the results and analysis: machine learning classification with random forests, choosing the best model with logit, and comparison with Altman's (1968) model. Section 5 presents the conclusions and suggestions.

2. LITERATURE REVIEW

This study aims to develop a model to explain the probability of financial distress in Indonesia. The list of all variables used in this study is presented in Table A.1, Appendix A.

2.1. Financial variables effect on financial distress

According to Beaver (1966), some financial ratios can be used to predict firms' financial health, even though not all ratios can predict distressed firms well. Beaver started the study to predict the probability of firms in distress. At that time, studies about financial distress were so limited that some early financial distress researchers did not use any theories to back up their pickup ratios. Instead, they based their decisions on phenomena.

Liquidity ratios measure the ability of firms to fulfil their short-term debts (Almansour, 2015). Beaver (1966, 1968) says that current assets to total assets (CATA), cash to current liabilities (CCL), current ratio (CR), cash to total assets (CTA), quick assets to current liabilities (QACL), working capital to total assets (WCTA), and quick assets to total assets (QATA) in a univariate model cannot predict financial distress well (Beaver, 1966, 1968). While Altman (1968) says retained earnings to total assets (RETA) and WCTA can predict financial distress well, without saying the effect is positive or negative (Altman, 1968). The difference may be caused by Beaver evaluating the ratios stand-alone, while Altman assessed them in a model. Springate (1978, as cited in Salsabila et al., 2022) says that earnings before tax to current liabilities (EBTCL) has a strong association with financial distress. Ohlson (1980) says that WCTA is negatively significant. Theodossiou (1991) says RETA and WCTA are negatively significant. Almansour (2015) and Ashraf et al. (2020) say that WCTA is negatively significant, CCL is not associated, and RETA and CR are positively significant with financial distress. On the contrary, Zmijewski (1984) claims that CR does not affect the probability of financial distress. The difference may be caused by the different estimation models used since Zmijewski used a Hazard model. Therefore, the hypotheses are:

H1a: CATA is negatively significant to the probability of firms in distress.

H1b: CCL is negatively significant to the probability of firms in distress.

H1c: CR is negatively significant to the probability of firms in distress.

H1d: CTA is negatively significant to the probability of firms in distress.

H1e: EBTCL is negatively significant to the probability of firms in distress.

H1f: QACL is negatively significant to the probability of firms in distress.

H1g: QATA is negatively significant to the probability of firms in distress.

H1h: RETA is negatively significant to the probability of firms in distress.

H1i: WCTA is negatively significant to the probability of firms in distress.

Profit is essential for firms to live and sustain. Firms with high big loss will lead to distress (Almansour, 2015). Beaver (1966) says that net income to total assets (NITA) is the second best in explaining financial distress, while net income to net worth (NINW) and net income to total debts (NITD) are not significant. Ohlson (1980), Zmijewski (1984), Theodossiou (1991), Shumway (2001), Campbell et al. (2008), and Almansour (2015) say NITA is negatively significant to financial distress. Campbell et al. (2008, 2011) says that net income over the market value of total assets (NIMTA) is negatively significant. Almansour (2015) says net income to total sales (NITS) is positively significant and NITA is not. Ohlson (1980) says that measurement of change in net income (CHIN) is negatively significant to financial distress. Beaver (1966) says that NITA is significant to financial distress. Ohlson (1980), Zmijewski (1984), Theodossiou (1991), Shumway (2001), and Campbell et al. (2008) say that NITA is negatively significant to financial distress. Last but not least, according to Ross et al. (2016), ROE is an important factor for. Therefore, the hypotheses are:

H2a: CHIN is negatively significant to the probability of firms in distress.

H2b: NIMTA is negatively significant to the probability of firms in distress.

H2c: NINW is negatively significant to the probability of firms in distress.

H2d: NITA is negatively significant to the probability of firms in distress.

H2e: NITS is negatively significant to the probability of firms in distress.

H2f: NITD is negatively significant to the probability of firms in distress.

H2g: ROE is negatively significant to the probability of firms in distress.

Leverage ratios compare equities to debts in firms and measure firms' ability to fulfil their obligations. Beaver (1966) says that total debt to total assets (TDTA) is significant to financial distress. Altman (1968) says that earnings before interest and tax to total assets (EBITTA) is significant to financial distress. Ohlson (1980) and Ashraf et al. (2020) say total liability to total assets (TLTA) is significantly positive on financial distress, while current liabilities to current assets (CLCA) is not. Ohlson (1980) also says that funds provided by operational funds divided by total liabilities (FUTL) is significantly negative on financial distress. Zmijewski (1984) and Theodossiou (1991) say TDTA is positively. While Shumway (2001) says TLTA is positively significant and EBITTA is negatively significant to financial distress. Blums (2003) says that total debt to market equity (TDME) is significantly positive to financial distress. Campbell et al. (2008, 2011) says that total liabilities to market value of total assets (TLMTA) is positively significant to financial distress. Beaver (1966) also says that TDTA is significant to financial distress. However what Beaver (1966) uses is TLTA (current plus long-term liabilities to total assets, CLLTLTA). Shumway (2001) and Campbell et al. (2008) also support that TLTA is positively significant to financial distress. And lastly, Almansour (2015) says both EBITTA and total debt to total equity (TDTE) variables are not significant. Therefore, the hypotheses are:

H3a: CLCA is positively significant to the probability of firms in distress.

H3b: EBITTA is positively significant to the probability of firms in distress.

H3c: FUTL is positively significant to the probability of firms in distress.

H3d: TDME is positively significant to the probability of firms in distress.

H3e: TDTA is positively significant to the probability of firms in distress.

H3f: TDTE is positively significant to the probability of firms in distress.

H3g: TLMTA is positively significant to the probability of firms in distress.

H3h: TLTA is positively significant to the probability of firms in distress.

Inventory turnover ratios measure how effectively firms can manage their inventories. Firms with high turnover ratios mean they can sell many of their inventory stocks during the whole year (Niko, 2022). Beaver (1966, 1968) says account receivables to total sales (ARTS), current assets to total sales (CATS), cash to total sales (CTS), inventory to total sales (ITS), net worth to total sales (NWTS), quick assets to total sales (QATS), total assets to total sales (TATS), and working capital to total sales (WCTS) do not affect the probability of financial distress. However, Beaver (1966) assessed the ratios individually. There is a possibility if the ratios are used in a model, the result can be different. Therefore, the hypotheses are:

H4a: ARTS is negatively significant to the probability of firms in distress.

H4b: CATS is negatively significant to the probability of firms in distress.

H4c: CTS is negatively significant to the probability of firms in distress.

H4d: ITS is negatively significant to the probability of firms in distress.

H4e: NWTS is negatively significant to the probability of firms in distress.

H4f: QATS is negatively significant to the probability of firms in distress.

H4g: TATS is negatively significant to the probability of firms in distress.

H4h: WCTS is negatively significant to the probability of firms in distress.

Liability ratios show the comparison of debts to total assets in a firm. Firms with high liability ratios have a higher probability of financial distress (Stotz, 2020). Beaver (1966) says current liabilities to total assets (CLTA) and long-term liabilities (debt) to total assets (LTLTA) do not affect financial distress. Theodossiou (1991) says the opposite that LTLTA is positively significant. The difference might be caused by the different methods used. Beaver (1966) evaluated the ratios individually, while Theodossiou (1991) assessed them in a model. Therefore, the hypotheses are:

H5a: CLTA is positively significant to the probability of firms in distress.

H5b: LTLTA is positively significant to the probability of firms in distress.

Beaver (1966) considers firms as a big tank of liquid assets, supplied by inflows and drained by outflows. Larger tanks mean a lower probability of financial distress. Larger net inflows also mean a smaller chance of failure, while larger net outflows mean the opposite. Beaver (1966, 1968) says that in a univariate model cash flow to net worth (CFNW), cash flow to assets (CFTA), and cash flow to total sales (CFTS) are not significant, while cash flow to total debts (CFTD) is significant to financial distress. Therefore, the hypotheses are:

H6a: CFNW is negatively significant to the probability of firms in distress.

H6b: CFTA is negatively significant to the probability of firms in distress.

H6c: CFTD is negatively significant to the probability of firms in distress.

H6d: CFTS is negatively significant to the probability of firms in distress.

Activity ratios measure how efficiently firms utilize their assets to generate sales. According to Almansour (2015), activity ratios can only have a significant effect when they are used in a model. Altman (1968) and Shumway (2001) say total sales to total assets (TSTA) is not significant to financial distress. While Almansour (2015) says that TSTA has a positive significant association with financial distress. Therefore, the seventh hypothesis is:

H7: TSTA is negatively significant to the probability of firms in distress.

Solvency ratios measure the ability of firms to continue operating as stable companies in the long term. These ratios indicate the ability of firms to pay all of their obligations even if all assets got sold (Almansour, 2015). Altman (1968) and Almansour (2015) say market value equity to book value of total debt (MVEBVOD) is positively significant to financial distress, while Shumway (2001) says the opposite. The difference in the result could be due to the different estimation models used by both authors: Altman (1968) used a multiple discriminant analysis (MDA), while Shumway (2001) used the hazard model. Almansour (2015) says that earnings before interest and tax to interest (EBITI) does not affect financial distress either. Therefore, the hypotheses are:

H8a: EBITI is negatively significant to the probability of firms in distress.

H8b: MVEBVOD is negatively significant to the probability of firms in distress.

There are some methods to measure the logarithm of market cap (SIZE), which are financial ratios or market proxies. Hereby, SIZE is measured as one of the financial ratios. Firms' size is a very important proxy to predict the probability of financial distress (Shumway, 2001). Dogan (2013) and Abeyrathna and Priyadarshana (2019) say that both total assets (TA) and total sales (TS) can be used to measure firms' size. Trujillo-Ponce et al. (2014) say TA is negatively significant to financial distress. While Dogan (2013) and Abeyrathna and Priyadarshana (2019) say that TS is positively significant. Akpan et al. (2021) supports the idea that TS represents firms' value. Firms with big values tend to have less chance of distress (Azhar et al., 2019). Nursal et al. (2023) also supports this idea by saying sales growth is negatively significant to distress. Therefore, the ninth hypothesis is:

H9a: TA is negatively significant to the probability of firms in distress.

H9b: TS is negatively significant to the probability of firms in distress.

2.2. Market variables effect on financial distress

Market variables have a strong association with the probability of financial distress. Shumway (2001) proposes that combining both accounting ratios and market variables can improve the accuracy of a model. Shumway (2001) says that relative SIZE and the difference between the logarithm of the closing price (CP) and market price are negatively significant to financial. Farah Freihat (2019) indirectly supports this argument by saying that SIZE has a positive effect on CP. Indupurnahayu et al. (2023) also supports indirectly supports this argument through his study in palm oil plantation firms that SIZE has a significant positive effect on firm value. While Andreou et al. (2021) says that there is a positive significant association between distress risk and CP crash. Campbell et al. (2008, 2011) says market to book ratio of the firm (MB) is positively significant to financial distress, while the stock of cash and short-term investments to market value of total assets (CASHMTA) is negatively significant. Graham and Buffett (1973) say book value per share (BVPS) is an important factor to determine whether firms are worth investing in his popular Graham's number formula. Graham is often called the "father of value investing". He was a famous investor and mentor of Warren Buffet (WallStreetMojo, 2021). Lastly, according to Doshi et al. (2018), during high uncertainty, small firms tend to reduce their capital expenditure (CAPEX). Since uncertainty (UNCERTAINTY) is expected to affect financial distress, CAPEX is also analyzed. Therefore, the hypotheses are:

H10a: BVPS is negatively significant to the probability of firms in distress.

H10b: CASHMTA is negatively significant to the probability of firms in distress.

H10c: CP is negatively significant to the probability of firms in distress.

H10d: MB is negatively significant to the probability of firms in distress.

H10e: SIZE is negatively significant to the probability of firms in distress.

H10f: CAPEX is negatively significant to the probability of firms in distress.

2.3. Macroeconomic variables effect on financial distress

Macroeconomic factors can affect firms' performance (Issah & Antwi, 2017). Chen et al. (1968) says that INTEREST is one of the macroeconomic variables. According to Oktavia and Handayani (2018), exchange rate (USDIDR) does not affect the Composite

Stock Price Index (IHSG), while gross domestic product (GDP) and Dow Jones industrial average (DJIA) have a strong positive association with the IHSG. Prawoto and Putra (2020) say that inflation rate (INFLATION) has a negative significant effect on IHSG in the short and long term, while USDIDR and oil price (OIL) have a positive significant effect on IHSG in the short and long. IHSG is the composite price of all stock prices in Indonesia. During global financial distress, the IHSG price goes down deeply. According to Sniazhko (2019), UNCERTAINTY can greatly affect decision-making in firms, while decision-making is important for the future of firms. Therefore, the hypotheses are:

H11a: DJIA is negatively significant to the probability of firms in distress.

H11b: GDP is negatively significant to the probability of firms in distress.

H11c: IHSG is negatively significant to the probability of firms in distress.

H11d: INFLATION is positively significant to the probability of firms in distress.

H11e: INTEREST is positively significant to the probability of firms in distress.

H11f: OIL is positively significant to the probability of firms in distress.

H11g: UNCERTAINTY is negatively significant to the probability of firms in distress.

H11f: USDIDR is negatively significant to the probability of firms in distress.

3. RESEARCH METHODOLOGY

3.1. Data

The population data for this study are all firms listed on Indonesia Stock Exchange (IDX) from 1973 to 2022. The data range of time is quarterly data from 2005 to 2020. The sampling method for distressed firms is purposive random sampling. All non-distressed firms' data are taken during the period of the observations. Firms in the financial sector are excluded.

Most sample data for this study are downloaded from the S&P CapitalIQ and S&P CapitalIQ Pro. This range of time experiences one capital market crash in 2008 and the latest economic recession in 2020. The total sample is about 727 firms.

The inflation rate is downloaded from the official Bank Indonesia website (https://www.bi.go.id/id/default.aspx). The oil price and interest rate, some important macroeconomic variables, are downloaded from http://investing.com. The uncertainty data are downloaded from https://worlduncertaintyindex.com/. Distressed firms are defined as firms that are delisted (Campbell et al., 2011; Shumway, 2001).

The sample firms in distress are taken from the S&P website. The sample size is determined by the total number of firms in distress during the observation time. Firms that volunteered to go private, merged, and do not have enough data are excluded. Only distressed firms that have pair sectors and time in the winning firms are selected.

Table 1. Delisted firms from Indonesia Stock Exchange

Notation	*Description*	*Year*	*Sector*
BORN	Borneo Lumbung Energi & Metal Tbk.	2020	Basic industry
CKRA	Cakra Mineral Tbk.	2020	Basic industry
SAIP	Surabaya Agung Industri Pulp dan Kertas Tbk.	2013	Basic industry
SOBI	Sorini Agro Asia Corporindo Tbk.	2017	Basic industry
CPGT	PT Citra Maharlika Nusantara Corpora Tbk.	2017	Infrastructure
INVS	Inovisi Infracom Tbk.	2017	Infrastructure
SIMM	Surya Intrindo Makmur Tbk.	2012	Infrastructure
ATPK	Bara Jaya Internasional Tbk.	2019	Mining
BRAU	Berau Coal Energy Tbk.	2017	Mining
CPDW	Indo Setu Bara Resources Tbk.	2013	Mining
KARK	Dayaindo Resources Internasional Tbk.	2013	Mining
SIAP	Sekawan Intipratama Tbk.	2019	Mining
TKGA	PT Permata Prima Sakti Tbk.	2017	Mining
ASIA	PT Asia Natural Resources Tbk.	2014	Trade
DAJK	PT Dwi Aneka Jaya Kemasindo Tbk.	2018	Trade
GREN	Evergreen Invesco Tbk.	2020	Trade
IATG	Infoasia Teknologi Global Tbk.	2009	Trade
TMPI	PT Sigmagold Inti Perkasa Tbk.	2019	Trade
TRUB	Truba Alam Manunggal Engineering Tbk.	2018	Trade

3.2. Methods

This study combines a machine learning algorithm and a statistical method. Random forests (RF) are used to classify which variables are the most important to explain financial distress, while logit is chosen as the binary regression estimation. RF is chosen because the algorithm is an improved combination of classification and regression tree (CART) and bagging (Breiman, 1996). With RF, when building the trees, every time a split is needed, only a random sample of variables is selected. Therefore, trees in RF are de-correlated compared to trees in the bagging algorithm (Breiman, 2001). Other alternatives that might improve the classification ability are to use XGBoost, Tree Net, or Artificial Neural Network (ANN) as the machine learning classification algorithm. XGBOOST and Tree Net are gradient-boosting algorithms (Chen & Guestrin, 2016; Shrivastava et al., 2020), while ANN is a mimic of the human neural system (Nisbet et al., 2018).

First, the dataset is split into 70% for the training set and 30% for the testing set. The training set is used for RF classification, training the Logit model, and choosing the best model with Logit. The testing set is used for assessing the model's fit and stability. Synthetic minority over-sampling technique (SMOTE) is used on the training dataset as synthetic data since the dataset is heavily unbalanced, and the ratio of distressed firms and good firms is about 3%. Using SMOTE can help to improve the model performance (Chawla et al., 2002).

After the variables are ranked, then multiple logit regression models are developed using the most important variables, the first and second most important variables, the first, second, and third most important variables, and so on until all variables are used.

The first logit model is:

$$Z_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \varepsilon \tag{1}$$

The second logit model is:

$$Z_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \varepsilon \tag{2}$$

And so on until all variables are used:

$$Z_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \cdots + \beta_n X_{ni,t} + \varepsilon \tag{3}$$

where, X_1, X_2 …, and X_n are variables from Table A.1. after they are ranked by RF. So, X_1 is the highest rank variable, X_2 is the second highest, and so on until all variables are included. In total, there will be n logit models with n as the number of variables. The best model is chosen by considering the highest in sensitivity and specificity, and the lowest in miss-calculation error and RMSE.

4. RESULTS AND DISCUSSION

4.1. Machine learning classification

Before SMOTE is applied, the total number of observations is 491 for distressed firms and 19,261 for good firms. The ratio of distressed firms and good firms is about 3%. Figure 1 shows the comparison result of RF without and with synthetic data. The green chart is the error rate for predicting distressed firms, the black chart is the error rate for the Out of Bound (OOB) dataset, and the red chart is the error rate for predicting good firms. The left chart shows a high error rate for predicting distressed firms. While the right chart is RF with synthetic data.

Figure 1. Comparison of RF error rate without and with synthetic data

After synthetic data is applied, the number of distress firm observations is 7,365, and the total number of observations is 26,135. The ratio between distressed and good firms becomes 38%. The right green chart shows the error rate decreases significantly. Adding more synthetic data does not improve accuracy anymore.

The RF number of tree iterations shows that after around four hundred trees, the error rates are stable. This study uses 500 RF-tree iterations to make sure the error rates are already stable and good. The RF's mean-decrease-accuracy shows how much accuracy loss will happen if a variable is excluded. The higher the accuracy number, the more important the variable is for better classification results (Taboada & Redondo, 2020). Table 2 shows the classification results by RF from the training dataset.

Table 2. List of all RF variable ranks

Variables	*Mean decrease accuracy*	*Variables*	*Mean decrease accuracy*
EBTCL	50.54414	*CTS*	19.24962
INFLATION	39.46973	*NITS*	18.99505
MB	32.54651	*UNCERTAINTY*	18.87879
INTEREST	32.14873	*EBITI*	18.592
QACL	31.46134	*QATS*	18.29528
CATA	28.8712	*EBITTA*	18.05097
TS	26.89155	*ITS*	18.00768
QATA	26.66743	*CAPEX*	17.91215
CHIN	26.27414	*NINW*	17.35308
TA	25.4627	*GPTA*	17.17282
TLMTA	25.30869	*NWTS*	16.95392
BVPS	25.29105	*CLCA*	16.9207
TLTA	25.08681	*CATS*	16.88183
CTA	24.87195	*OIL*	16.77278
RETA	24.51097	*TSTA*	16.11806
TDME	24.26558	*TATS*	15.93986
MVEBVOD	24.23434	*USDIDR*	14.81129
TDTA	23.99288	*GDP*	14.73785
TDTE	23.19509	*IHSG*	14.58462
CR	23.16231	*ROE*	14.42523
CFTD	22.95291	*NITA*	14.29648
SIZE	22.7621	*DJIA*	13.6449
CASHMTA	22.74921	*NITD*	13.47134
CLTA	22.21944	*LTLTA*	11.5901
CFTS	21.76162	*WCTA*	11.12383
CLLTLTA	21.52076	*NIMTAAVG*	10.83268
CCL	21.42584	*WCTS*	8.567204
CFNW	21.13743	*ARTS*	8.359568
CFTA	21.00478	*CAPTL*	0
FUTL	20.05034	*FATA*	0
CP	19.29218		

The next step is to validate the RF results with the testing dataset, zero is for healthy firms and one is for distressed firms. Using synthetic data there is an increase in the prediction accuracy from 0.9867 to 0.9978. There is a slight decrease in sensitivity from 0.9999 to 0.9996 and a significant improvement in specificity from 0.5172 to 0.9934. The overall results show there is a prediction accuracy improvement after using synthetic data.

Table 3. Confusion matrix RF results using test dataset

No synthetic data				*With synthetic data*			
Actual distress	*Predicted default*			*Actual distress*	*Predicted default*		
	0	1	*Row total*		0	1	*Row total*
0	9304	126	9430	0	9337	24	9361
1	1	135	136	1	4	3610	3614
Column total	9305	261	9566	Column total	9341	3634	12975
Accuracy	0.9867			Accuracy	0.9978		
Sensitivity	0.9999			Sensitivity	0.9996		
Specificity	0.5172			Specificity	0.9934		

4.2. Choosing the best model

The RF results give guidance on how to build the model. The Logit regression models are looped starting from the most important variable alone, then the second loop model contains the most important variable and the second most important one, and so on until all the variables are included in the model. Table 4 shows the results of the model fit assessment for all logit models. Sensitivity is the ability of a model to identify firms in distress, while specificity is the ability to identify firms that are not in distress (Swift et al., 2020). Considering the highest in sensitivity and specificity, and lowest in miss-calculation error and RMSE, a logit model with fifteen variables is chosen as the best model. Mc-Fadden Pseudo R^2 and AIC criterion cannot be used since the model with the highest R^2 and lowest AIC includes almost all variables available, meaning that there are no variables selections.

Table 4. Model fit assessment for all logit models

Number of vars	*R^2*	*AIC*	*Sensitivity*	*Specificity*	*MisClassError*	*Optimal*	*RMSE*
15	0.1983	17396.06	0.6488	0.7959	0.2465	0.3592	1.076348
56	0.3138	15060.85	0.6349	0.8686	0.1968	0.4	1.100834
7	0.1748	17921.97	0.6113	0.8011	0.2532	0.3686	1.091409
60	0.3298	14754.21	0.6102	0.8784	0.1956	0.42	1.111269
55	0.3171	14995.97	0.5955	0.8806	0.199	0.43	1.11916
46	0.3089	15123.87	0.5938	0.8911	0.1928	0.43	1.12507
33	0.2767	15811.83	0.5935	0.8687	0.2086	0.4091	1.117335
16	0.2095	17203.28	0.5924	0.8412	0.2296	0.39	1.110925
57	0.3149	15083.3	0.5864	0.8932	0.1913	0.44	1.12439
38	0.2866	15571.42	0.5821	0.8724	0.2102	0.429	1.124843
27	0.278	15815.75	0.5806	0.8762	0.2057	0.4298	1.122062
39	0.2901	15545.82	0.5795	0.8672	0.2134	0.4186	1.12138
17	0.2083	17189.58	0.5793	0.8435	0.2326	0.38	1.118932
40	0.285	15617.24	0.5775	0.8802	0.2057	0.4284	1.128635
18	0.2142	17175.66	0.5718	0.8502	0.2273	0.4091	1.117677
23	0.2411	16483.94	0.5714	0.8591	0.224	0.4144	1.127335
30	0.2782	15802.52	0.5683	0.8719	0.2126	0.428	1.126033
54	0.3145	14962.93	0.5678	0.8883	0.2039	0.4299	1.137303
47	0.3023	15291.37	0.5662	0.904	0.1907	0.4399	1.138199
35	0.2856	15576.3	0.5652	0.872	0.2155	0.4198	1.131738
28	0.2812	15613.1	0.5641	0.8801	0.2115	0.4194	1.137359
41	0.2943	15469.61	0.5599	0.8821	0.2078	0.4398	1.132696
49	0.3119	15098.66	0.5581	0.8985	0.1964	0.45	1.138592
45	0.3022	15305.61	0.5565	0.9035	0.1932	0.4499	1.140662
42	0.2897	15537.16	0.5511	0.8891	0.2062	0.439	1.140326
59	0.3268	14708.88	0.5508	0.9053	0.1965	0.4599	1.148795
36	0.2883	15554.64	0.5479	0.8739	0.2181	0.4298	1.136686
20	0.2133	17176.06	0.5423	0.8655	0.2251	0.4095	1.134945
58	0.3189	14944.11	0.5404	0.8971	0.203	0.4499	1.14596
52	0.3148	15069.89	0.5372	0.902	0.1987	0.45	1.145627
37	0.2885	15597.79	0.5335	0.8891	0.2097	0.4598	1.143955
26	0.2476	16464.81	0.5328	0.8823	0.2149	0.4499	1.142338
43	0.2928	15462.34	0.53	0.9022	0.2032	0.4698	1.152785
19	0.2153	17048.09	0.5264	0.8622	0.2343	0.4089	1.144568
21	0.2184	17056.21	0.5224	0.8662	0.2305	0.4174	1.143119
22	0.2407	16477.71	0.5222	0.8743	0.2276	0.4398	1.151456
10	0.1817	17834.5	0.5213	0.8513	0.2414	0.3956	1.13876
61	0.3226	14897.92	0.5207	0.919	0.1916	0.4798	1.157807
44	0.3026	15335.44	0.52	0.9168	0.1922	0.4799	1.155547
34	0.2788	15824.32	0.5196	0.8961	0.2078	0.4496	1.149904
50	0.2993	15332.51	0.518	0.926	0.1895	0.4898	1.164617
29	0.284	15717.59	0.5163	0.8865	0.215	0.4398	1.146962
51	0.3104	15095.14	0.5153	0.915	0.1981	0.4799	1.16204
48	0.3156	14983.04	0.5113	0.9131	0.2002	0.48	1.162424
25	0.2434	16582.67	0.5067	0.8794	0.2233	0.4501	1.149128
11	0.1849	17766.31	0.5064	0.8483	0.2478	0.3958	1.143342
13	0.1863	17847.61	0.4986	0.8623	0.2366	0.4138	1.144011
24	0.2417	16606.94	0.4968	0.8868	0.221	0.4585	1.155657
9	0.1857	17685.01	0.4949	0.8528	0.2496	0.3988	1.152563
31	0.2819	15752.08	0.4903	0.9061	0.2085	0.4687	1.163466
8	0.1866	17825.34	0.4899	0.8554	0.2442	0.4068	1.145348
32	0.2846	15589.92	0.4777	0.8993	0.2211	0.4698	1.17291
14	0.1987	17474.24	0.4612	0.8652	0.2483	0.4208	1.165
53	0.3113	15199.69	0.4101	0.9475	0.1981	0.5699	1.200476
12	0.1916	17603.76	0.2936	0.9283	0.2511	0.4897	1.244091
5	0.0901	19754.39	0.0379	0.985	0.2857	0.7399	1.346916
6	0.0949	19658.82	0.0371	0.9868	0.2843	0.7572	1.347247
1	0.0356	20947.61	0	0.99	0.2917	0.767	1.358841
2	0.0574	20511.02	0	0.9995	0.2825	0.8924	1.358653
3	0.0622	20379.17	0	0.9998	0.2843	0.9244	1.360998
4	0.0687	20314.85	0	0.9998	0.2788	0.9346	1.35494

4.3. Results and empirical interpretation

A collinearity test is done to check if there is a multicollinearity problem in the model. According to Akinwande et al. (2015), a variable is free from multicollinearity problems if the VIF is below five. The multicollinearity problem should be avoided to prevent bias in the results. The results show that all independent variables are free from the multicollinearity problem. Table 5 shows the variance inflation factor (VIF) result.

Table 5. Logit VIF result

CHIN	*CTA*	*QATA*	*CATA*	*QACL*	*RETA*	*EBTCL*	*BVPS*	*TS*	*TA*	*MB*	*TLMTA*	*TLTA*	*INFLAT*	*INTER*
1.004192	1.260479	2.232282	1.951306	1.418492	1.474363	1.102143	9.007351	1.173342	7.157831	1.084181	3.290031	2.137651	4.795393	4.898029

Table 6 shows the logit regression coefficients estimate. The final decision is to pick a logit model with fifteen variables. This model is chosen because it has the best overall performance with a sensitivity of 0.6488, specificity of 0.7959, and a miss-calculation error of 0.2465. The "***" sign shows that a variable is significant at 1%. McFadden's Pseudo R^2 for this model is 19.83%, meaning there are still lots of chances for improvement.

Table 6. Logit regression coefficients estimate

Variables	Estimate	Prob
(Intercept)	-1.66008280	0.0000***
CATA	1.07663299	0.0000***
CTA	-2.27373434	0.0000***
EBTCL	-0.00263719	0.0000***
QACL	0.13933203	0.0000***
RETA	-0.25675622	0.0000***
QATA	-1.68795491	0.0000***
CHIN	0.00602292	0.4120
TLMTA	0.00002072	0.0000***
TLTA	-0.50438502	0.0000***
TS	-0.00000783	0.0000***
TA	0.00000003	0.0000***
BVPS	-205.11169542	0.0000***
MB	-38.92267190	0.0000***
INFLATION	-0.67184567	0.6070
INTEREST	20.09866693	0.0000***

CATA, *CTA*, *EBTCL QACL*, *RETA*, and *QATA* are liquidity variables. Liquidity measures the ability of firms to pay their short-term debts. Firms with higher liquidity ratios tend to have a lower chance of experiencing distress. *CTA*, *QATA*, *RETA*, and *EBTCL* are negative and significant, while *CATA* and *QACL* are positive and significant. Therefore:

- *H1a* that *CATA* is negatively significant to the probability of firms in distress is rejected.
- *H1d* that *CTA* is negatively significant to the probability of firms in distress is not rejected.
- *H1e* that *EBTCL* is negatively significant to the probability of firms in distress is not rejected.
- *H1f* that *QACL* is negatively significant to the probability of firms in distress is rejected.
- *H1h* that *RETA* is negatively significant to the probability of firms in distress is not rejected.
- *H1g* that *QATA* is negatively significant to the probability of firms in distress is not rejected.

The results are different from Beaver (1966) that says *CATA*, *QATA*, and *QACL* do not affect financial distress. The difference in result might be caused by the univariate model used by Beaver. The result of *RETA* strengthens Altman's (1968) study that says *RETA* has a negative significant effect on financial distress. For *EBTCL*, the result also strengthens Springate (1978, as cited in Salsabila et al., 2022) that says *EBTCL* has a negative significant effect on financial distress. The results imply that firms should maintain their *CTA*, *RETA*, *QATA*, and *EBTCL* high enough while maintaining *QATA* and *CATA* low to prevent financial distress.

CATA measures the portion of funds invested in current assets compared to the total assets (Corporate Finance Institute [CFI], 2022). Current assets are quick assets with the addition of inventories. Current assets have some functionality for firms' financials, such as paying for the capex and daily needs of firms. Having a too-high *CATA* is not a good sign since it includes some inventories that do not sell well. Firms that have bad selling tend to be in financial distress.

CTA measures a portion of assets a company holds in the form of cash and the efficiency of cash flows. *CTA* is not related to income or profitability. There are some good reasons for firms to have extra cash on the balance sheet. Having lots of cash means firms have a better ability to pay their short-term debts and have good financial health. Moreover, firms in some sectors have lower capital expenditures than firms in other sectors, so their cash increases much. However, having too much cash has its bad reasons too. An abundant cash reserve can indicate that firms do not utilize their funds well. If there is too much idle cash, it implies that the investments and businesses might not run well. Firms need to maintain their CTA high enough to meet their short-term debts and needs, but not too much.

EBTCL measures the ratio of earnings before tax and the short-term debts of firms. Firms that can make high earnings mean the firm runs well. Earnings are like the life support of firms. Earnings come from sales, either goods or services. Nevertheless, investors need to be aware of earnings management. Mahrani and Soewarno (2018) describe earnings management as a practice of manipulating the reported earnings for some motivations and purposes. Therefore the earnings information given by the management cannot be trusted that it is the real financial condition of the firms. In normal conditions, firms having high *EBTCL* should have a lower chance to experience financial distress.

QACL measures the portion of funds invested in quick assets compared to the current liabilities (CFI, 2022). Quick assets are parts of current assets excluding the inventory. Having a too-high *QACL* is not a good sign since it includes some inventories that do not sell well. Firms that have bad selling tend to be in financial distress. By having a positively significant sign, firms must maintain the *QACL* ratio to be low enough just to fulfil their short-term obligations. Putting too many funds in QA will lead firms to financial distress.

RETA measures a portion of the profit that firms hold for use in business growth and investment. Retained earnings are used for paying debts, firm operations, and expansion (OCBC NISP, 2022b). High *RETA* means firms are less dependent on debts and equity financing, indicating a financial health firm. When firms focus on expansion, they usually do not pay dividends or they pay only a small amount of dividends to increase retained earnings. Dividends are usually one important factor that investors consider when investing. Firms that pay dividends can be considered healthy firms, even though dividends are not the only factors. Firms need to maintain *RETA* high enough, but not too high. A too-high *RETA* means firms focus most of their profit on expansions and ignore their shareholders. This strategy works during the initial stages of a company or the expansion stage, but bigger firms need to consider sharing dividends too.

QATA measures a portion of assets a company holds that can be immediately converted into cash. Firms can use quick assets to fulfil their short-term debts and for business operations. The results of this study show that quick assets can be used to measure firms' health. Having high quick assets means firms can generate money. Similar to *CTA*, even though firms should aim for high *QATA*, having the ratio value too high can indicate a bad funds allocation. Firms need to balance the quick assets so that they are not too high and not too low.

CHIN is a profitability variable. Profit is important for firms to live and sustain. Firms with high negative profits will lead to distress. The result shows that *CHIN* is not significant. Therefore:

• *H2a* that *CHIN* is negatively significant to the probability of firms in distress is rejected.

The result says that *CHIN* has no effect on financial distress, which is different from a study done by Ohlson (1980) saying that *CHIN* is negatively significant to financial distress.

Ohlson (1980) defines *CHIN* as the net income growth ratio, while net income is defined as the remaining profit after deducted by all expenses and losses at that term (OCBC NISP, 2022a). Net income plays an important role in the balance sheet. Firms can share net income as dividends or use it as retained earnings. Even though *CHIN* is not significant to financial distress, firms need to maintain net income growth for long-term sustainability.

TLTA and *TLMTA* are leverage ratios. They measure how much of firms' assets are liabilities (yCHARTS, 2023). The results show that *TLTA* is negatively significant, while *TLMTA* is positively significant. Therefore:

• *H3g* that *TLMTA* is positively significant to the probability of firms in distress is not rejected.

• *H3h* that *TLTA* is positively significant to the probability of firms in distress is rejected.

The results disagree with Campbell's et al. (2008, 2011) study about *TLTA* but agree on *TLMTA*. The results imply that firms that maintain low *TLMTA* and high *TLTA* tend to have a lower chance of financial distress.

Campbell et al. (2008) propose *TLMTA* as the better modification of *TLTA*. *TLMTA* measures total liability to the market value of total assets, while *TLTA* measures total liability to the total assets. A high *TLMTA* indicates that firms have too much debt compared to the market value of their total assets and the shareholder equities are low. Firms that are rapidly growing often have a high *TLMTA*, and they can be just fine in growth markets (yCHARTS, 2023). However, when a recession happens, firms will have difficulties fulfiling their financial obligations, which may lead to financial distress (BDC, 2023).

TS and *TA* are size variables. *TS* have a negative significant association with financial distress, while TA has a positive significant association. Therefore:

• *H9a* that *TA* is negatively significant to the probability of firms in distress is rejected.

• *H9b* that *TS* is negatively significant to the probability of firms in distress is not rejected.

The result agrees with Shumway's (2001) that size is important in financial distress and disagrees with Trujillo-Ponce et al.'s (2014) study that *TA* is negatively significant to financial distress. The results also disagree with Abeyrathna's (2019) study that *TS* is positively significant.

TS are the total amount of sales generated from the business operations. As financial variables, Dogan (2013) and Abeyrathna (2019) support the idea that *TA* and *TS* can be used to measure the size of firms. Big firms tend to have a lower chance of financial distress. Earnings and net income come from sales. Sales are also important for selling targets and growth. If sales are high, earnings and net income will be high too, which will lead to selling achievement and good growth. It is easy very easy to tell when firms have low sales, either goods or services, sooner or later they will be in financial distress.

As simple as the name says, *TA* is the total assets of firms. Bigger firms should have a lower chance of financial distress. There is a saying "too big to fail," meaning noticeably big firms can withstand stronger lots of financial problems. Of course, there is no such thing as immortality in this world. Even companies with huge assets such as Lehman Brothers could still go bankrupt if the managements are wrong. Lehman Brothers went bankrupt during the financial crisis in 2008 that was caused by securitizing lots of mortgage packages for onward sales (Backhouse, 2023).

BVPS and *MB* are market variables. Market variables can affect the probability of financial distress. The results show that *BVPS* and *MB*. Therefore:

• *H10a* that *BVPS* is negatively significant to the probability of firms in distress is not rejected.

• *H10d* that *MB* is negatively significant to the probability of firms in distress is not rejected.

The results support Graham and Buffett (1973), that *BVPS* has a negative significant effect on financial distress. The results disagree with Campbell's et al. study (2008, 2011) about *MB* and *TLTA* but agree on *TLMTA*. The results imply that firms that maintain high *BVPS* and *TLTA* and low *TLMTA* tend to have a lower chance of financial distress. As a comparison, *MB* for blue-chip companies tends to be higher than smaller ones in the US. Firms with high *BVPS* and *MB* have a lower chance of financial distress.

BVPS measures total equity compared to the number of outstanding shares (Ross et al., 2016). The higher the *BVPS* is, the higher the equity firms have, meaning firms have a strong equity base. Investors can use BPVS to measure if a stock price is undervalued by comparing *BVPS* to the market value per share (MVPS). If *BVPS* is higher than the MVPS, then the stock is deemed undervalued and vice versa (HSB Investasi, 2023). Hence, having a high *BVPS* means investors believe the stock has a high value. Since investors' trust is important in the capital market, owning a high investors' trust can lead lower chance of financial distress.

MB is usually used by investors to measure the market's perception of a stock. A high *MB* means the market's perception is good; the market wants to pay a high price for a stock. It can also mean a stock is overvalued, and many investors want to invest in that stock. It is a sign of good firms. Firms with a high *MB* tend to have a lower chance of financial distress.

INFLATION and *INTEREST* are macroeconomic variables. Macroeconomic variables can affect firms' performance, which is related to financial distress. The results show that only *INTEREST* is positively significant. Therefore:

- *H11d* that *INFLATION* is positively significant to the probability of firms in distress is rejected.
- *H11e*: *INTEREST* is positively significant to the probability of firms in distress is not rejected.

INFLATION is defined as the overall continuous increase in the price of goods and services during a certain interval. The results say that *INFLATION* does not affect financial distress, while *INTEREST* does. Low and stable inflation is the prerequisite for sustained economic growth that can give benefit society. The government needs to control the inflation rates by considering high and unstable inflation rates are bad for the economy (Bank Indonesia, 2023a).

Nowadays, the *INTEREST* rate is defined as BI-7 Day Reverse Repo Rate (BI7DRR), replacing BI Rates since 19 August 2016 in Indonesia (Bank Indonesia, 2023b). Interest rates are premium rates used when banks give loans or when customers save money in banks. High-interest rates will lead people to save more money in the bank, making few people want to invest or do business. High-interest rates also make the premium for borrowing money higher, resulting in higher costs for running businesses. In a long time, higher costs mean lower profits and growth rates (Curry & Adams, 2023). This condition is harmful to running businesses and may lead firms to financial distress.

There are not enough statistical data to support the remaining hypotheses. Therefore, all of them are rejected.

Studies on financial distress using a combination of theories, RF as a machine learning algorithm, and logit as a statistical method are rare. The second result of this study shows that the combined method can give a new perspective result on financial distress methodology. Most studies only use theoretical frameworks and statistical methods to develop and analyze the empirical model. This study also supports Shrivastava's et al. (2020) study that RF is a good classification and prediction performance. By combining methods, this study gets the best of each method, while minimizing the drawbacks, proven by overall better accuracy results than Altman's 1968 model. This result is important since it gives a better methodology for future studies. This methodology gives a chance to improve the accuracy and explanatory power of current studies in many sectors, not only in finance but also in many other subjects.

4.4. Comparison with Altman's (1968) model

The last step of this study is to compare the results with the Altman 1968 model. To be fair and robust, the comparison is done using the testing dataset. SMOTE is not applied to this dataset, and the model has never seen this dataset. Altman's (1968) study has a weakness in that most firms fall in the grey zone area. Therefore, this study will only divide into Safe Zone and Distress Zone. The optimal cutoff for the probability is 0.3585. Thus, any stocks with a probability of default of 0.3585 or higher will be predicted to default. Table 7 shows the comparison table between the new model and Altman 1968's model.

Table 7. Comparison matrix benchmark between the new model and the Altman's (1968) model

This model				***Altman Z 1968***			
Actual distress	*Predicted default*			*Actual distress*	*Predicted default*		
	0	1	Row total		0	1	Row total
0	14,943	47	14,990	0	10,524	35	10,559
1	3,820	135	3,955	1	8,239	147	8,386
Column total	18,763	182	18,945	Column total	18,763	182	18,945

Table 8 shows the model fit assessment between the new model and the Altman's (1968) model. Overall, the new model has higher a sensitivity of 0.7417582 and a specificity of 0.7964078, while the miss-classification error and the RMSE are also smaller at 0.4517933. The comparison shows that combining RF and logit can give an overall more accurate model.

Table 8. Model fit assessment

Model	***SENSITIVITY***	***SPECIFICITY***	***MISCLASSERROR***	***RMSE***
This model	0.7417582	0.7964078	0.2041	0.4517933
Altman's (1968)	0.8076923	0.5608911	0.4367	0.6608615

5. CONCLUSION

During a recession or crash in the capital market, the ability to predict firms in distress is important. Financial distress is a condition where a company cannot pay its obligations, either short-term debts or long-term debts. In the worst case, a company will go into bankruptcy. The capability to predict the probability of financial distress is beneficial to investors.

This study tries to develop a new financial distress prediction model using a combination of theories, random forests, and logit. The research data is quarterly data from 2005 to 2020 in IDX, excluding the financial sector. The results show that *CTA, RETA, QATA, EBTCL, TLTA, TS, BVPS*, and MB have a negative significant association with the probability of firms in distress. While *CATA*, *QACL*, *TLMTA, TA*, and *INTEREST* have a positive significant association with the probability of firms in distress.

Firms with good selling and enough cash flow tend to have a lower possibility of financial distress. Moreover, growth is also an important factor since bigger firms can withstand more crises. A favorable interest rate can even further lower the probability of financial distress.

Another result shows that combining theories, random forests, and logit can be used to build a new financial distress prediction model. The second result is a new enlightenment since this method can be used to develop many new financial study models, not only using Logit estimates but also other estimation methods. This result is important since it gives a better methodology for future studies. This methodology gives a chance to improve the accuracy and explanatory power of current studies in many sectors, not only in finance but also in many other subjects.

The Mc-Fadden Pseudo R^2 in this study is still low at about 19%, meaning there are still many limitations and unexplained factors for financial distress. There are some limitations to the studies. First, this study does not use corporate governance and political risk variables. The data for these two variables are hard to get. Hopefully, in the future, the availability of data will be better. Second, this study does not use research and development variables (R&D). At the time of this study, many firms do not report their R&D funds, so it is hard to include these variables. Hopefully, in the future, many firms will be more responsible by reporting their R&D funding. Third, this study only uses one machine learning algorithm, random forests. There are still many machine learning algorithms that can be compared to random forests, such as XGBoost, Support Vector Machine (SVM), etc. Comparing and using these different algorithms are expected to give better comprehensive results.

REFERENCES

1. Abeyrathna, S. P. G. M., & Priyadarshana, A. J. M. (2019). Impact of firm size on profitability. *International Journal of Scientific and Research Publications, 9*(6), Article p9081. https://doi.org/10.29322/IJSRP.9.06.2019.p9081
2. Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics, 5*(7), 754–767. https://doi.org/10.4236/ojs.2015.57075
3. Akpan, M., Dhillon, G., & Trottier, K. (2021). Cluster analysis of share price: How firm characteristics relate to accounting metrics. *Risk Governance and Control: Financial Markets & Institutions, 11*(4), 8–25. https://doi.org/10.22495/rgcv11i4p1
4. Almansour, B. Y. (2015). Empirical model for predicting financial failure. *American Journal of Economics, Finance, and Management, 1*(3), 113–124. https://www.academia.edu/12140129/Empirical_Model_for_Predicting_Financial_Failure
5. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x
6. Andreou, C. K., Andreou, P. C., & Lambertides, N. (2021). Financial distress risk and stock price crashes. *Journal of Corporate Finance, 67*, Atricle 101870. https://doi.org/10.1016/j.jcorpfin.2020.101870
7. Ashraf, S., Félix, E. G. S., & Serrasqueiro, Z. (2020). Development and testing of an augmented distress prediction model: A comparative study on a developed and an emerging market. *Journal of Multinational Financial Management, 57–58*, Article 100659. https://doi.org/10.1016/j.mulfin.2020.100659
8. Ashraf, S., Félix, E. G. S., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress? *Journal of Risk and Financial Management, 12*(2), Article 55. https://doi.org/10.3390/jrfm12020055
9. Azhar, A., Srijayanti, S., Zenita, R., & Nanda, S. T. (2019). The effect of financial distress on firm value: The mediating role of investment opportunity set. In *Proceeding of the International Conference on Business, Economy, Social and Technology* (200–211). Putrajaya, Malaysia. https://rrs.mmu.edu.my/upload_files/conference/3872/ICBEST%202019%20Conference%20Proceeding.pdf#page=206
10. Backhouse F. (2023, March 24). *Bankruptcy of Lehman brothers.* Encyclopedia Britannica. https://www.britannica.com/event/bankruptcy-of-Lehman-Brothers
11. Bank Indonesia. (2023a, June 30). *Inflasi.* https://www.bi.go.id/id/fungsi-utama/moneter/inflasi/default.aspx
12. Bank Indonesia. (2023b). *BI-7 day reverse repo rate (BI7DRR).* https://www.bi.go.id/id/fungsi-utama/moneter/bi-7day-rr/default.aspx
13. BDC. (2023). *Debt-to-asset ratio.* https://www.bdc.ca/en/articles-tools/entrepreneur-toolkit/financial-tools/debt-asset-ratio
14. Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111. https://doi.org/10.2307/2490171
15. Beaver, W. H. (1968). Alternative Accounting Measures as Predictors of Failure. *The Accounting Review, 43*(11), 113–22. https://www.jstor.org/stable/244122
16. Blums, M. (2003). *D-score: Bankruptcy prediction model for middle market public firms.* Macalester College.
17. Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. https://doi.org/10.1007/BF00058655
18. Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
19. Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance, 63*(6), 2899–2939. https://doi.org/10.1111/j.1540-6261.2008.01416.x
20. Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2011). Predicting financial distress and the performance of distressed stocks. *Journal of Investment Management, 9*(2), 14–34. http://nrs.harvard.edu/urn-3:HUL.InstRepos:9887619
21. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953
22. Chen, N.-F., Roll, R., & Ross, S. (1986). Economic forces and the stock market. *The Journal of Business, 59*(3), 383–403. https://doi.org/10.1086/296344

23. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). New York, United States. https://doi.org/10.1145/2939672.2939785
24. Corporate Finance Institute (CFI). (2022). *Quick assets.* https://corporatefinanceinstitute.com /resources/accounting/quick-assets/
25. Curry, B, & Adams, M. (2023, June 21). *What happens when the fed raises rates?* https://www.forbes.com /advisor/investing/fed-raises-interest-rates/
26. Dogan, M. (2013). Does firm size affect the firm profitability? Evidence from Turkey. *Research Journal of Finance and Accounting, 4*(4), 53–59. https://core.ac.uk/download/pdf/234629457.pdf
27. Doshi, H., Kumar, P., & Yerramilli, V. (2018). Uncertainty, capital investment, and risk management. *Management Science, 64*(12), 5769–5786. https://doi.org/10.1287/mnsc.2017.2815
28. Farah Freihat, A. R. (2019). Factors affecting price to earnings ratio (P/E): Evidence from the emerging market. *Risk Governance and Control: Financial Markets & Institutions, 9*(2), 47–56. https://doi.org/10.22495/rgcv9i2p4
29. Graham, B., & Buffett, W. E. (1973). *The intelligent investor* (4th rev. ed.). HarperBusiness Essentials.
30. HSB Investasi. (2023). *Book value per share.* https://www.hsb.co.id/glosarium/b/book-value-per-share
31. Indupurnahayu, Walujadi, D., Lysandra, S., Hurriyaturrohman, & Endri, E. (2023). Determinant of firm value: Evidence of oil palm plantation companies. *Corporate & Business Strategy Review, 4*(2), 124–131. https://doi.org/10.22495/cbsrv4i2art11
32. Issah, M., & Antwi, S. (2017). Role of macroeconomic variables on firms' performance: Evidence from the UK. *Cogent Economics & Finance, 5*(1), Article 1405581. https://doi.org/10.1080/23322039.2017.1405581
33. Mahrani, M., & Soewarno, N. (2018). The effect of good corporate governance mechanism and corporate social responsibility on financial performance with earnings management as mediating variable. *Asian Journal of Accounting Research, 3*(1), 41–60. https://doi.org/10.1108/AJAR-06-2018-0008
34. Niko. (2022, May 31). *Cara menghitung rasio perputaran persediaan (Inventory turnover)* [How to calculate the inventory turnover ratio (Inventory turnover)]. Mekari Jurnal. https://www.jurnal.id/id/blog/rasio-perputaran-persediaan/
35. Nisbet, R., Miner, G., Yale, K., Elder, J. F., & Peterson, A. F. (2018). *Handbook of statistical analysis and data mining applications* (2nd ed.). Academic Press.
36. Nursal, E. Z., Kristanti, F. T., & Murti, G. T. (2023). Effect of financial ratios, macroeconomics, and corporate governance on financial distress. *Asean International Journal of Business, 2*(1), 85–96. https://journal.adpebi.com/index.php/AIJB/article/view/462
37. OCBC NISP. (2022a). *Net income (Laba Bersih) dan bedanya dengan gross* profit [Net income and how it differs from gross profit]. http://www.ocbcnisp.com/id/article/2022/03/21/net-income-adalah
38. OCBC NISP. (2022b, December 8). *Mengenal apa itu retained earning, cara hitung & contohnya* [Know what is retained earning, how to calculate & examples]. http://www.ocbcnisp.com/id/article/2022/12/08/retained-earning-adalah
39. Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131. https://doi.org/10.2307/2490395
40. Oktavia, S., & Handayani, W. (2018). Effect of rupiah exchange rate, GDP growth, and Dow Jones index on composite stock price in Indonesia Stock Exchange. *Journal of Accounting and Strategic Finance, 1*(1), 23–32. https://tinyurl.com/4xbpvmuz
41. Prawoto, N., & Putra, B. A. (2020). The factors affecting the composite stock price index (CSPI) of Indonesia Stock Exchange. *Jurnal Ekonomi & Studi Pembangunan, 21*(1). https://doi.org/10.18196/jesp.21.1.5032
42. Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2016). *Fundamental of corporate finance* (11th ed.). McGraw-Hill Education.
43. Salsabila, R., Wicaksono, A. T. S., & Shamakov, N. (2022). The role of early warning system components in financial distress: Evidence from Indonesian Islamic banks. *Maliki Islamic Economics Journal, 2*(2), 123–133. https://doi.org/10.18860/miec.v2i2.18977
44. Shrivastava, A., Kumar, N., Kumar, K., & Gupta, S. (2020). Corporate distress prediction using random forest and tree net for India. *Journal of Management and Science, 1*(1), 1–10. https://doi.org/10.26524/jms.2020.1
45. Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business, 74*(1), 101–124. https://doi.org/10.1086/209665
46. Sniazhko, S. (2019). Uncertainty in decision-making: A review of the international business literature. *Cogent Business & Management, 6*(1), Article 1650692. https://doi.org/10.1080/23311975.2019.1650692
47. Stotz, A. (2020, August 10). *What is liabilities to assets ratio?* Valuation Master Class. https://valuationmasterclass.com/what-is-liabilities-to-assets-ratio/
48. Swift, A., Heale, R., & Twycross, A. (2020). What are sensitivity and specificity? *Evidence-Based Nursing, 23*(1), 2–4. https://doi.org/10.1136/ebnurs-2019-103225
49. Taboada, F. M., & Redondo, J. I. (2020). *Variable importance plot (mean decrease accuracy and mean decrease Gini)* [Figure]. PlosONE. https://doi.org/10.1371/journal.pone.0230799.g002
50. Theodossiou, P. (1991). Alternative models for assessing the financial condition of business in Greece. *Journal of Business Finance & Accounting, 18*(5), 697–720. https://doi.org/10.1111/j.1468-5957.1991.tb00233.x
51. Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394–419. https://doi.org/10.1016/j.irfa.2013.02.013
52. Trujillo-Ponce, A., Samaniego-Medina, R., & Cardone-Riportella, C. (2014). Examining what best explains corporate credit risk: Accounting-based versus market-based models. *Journal of Business Economics and Management, 15*(2), 253–176. https://doi.org/10.3846/16111699.2012.720598
53. WallStreetMojo. (2021). *Graham number.* https://www.wallstreetmojo.com/grahams-number/
54. yCHARTS. (2023). *Liabilities to assets ratio.* https://ycharts.com/glossary/terms/liabilities_to_assets
55. Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22*, 59–82. https://doi.org/10.2307/2490859

APPENDIX A. VARIABLES

Table A.1. List of variables used in this study

Group	Variables	Description
Activity	*TSTA*	Total sales to total assets
Cash Flow	*CFNW*	Cash flow to net worth
Cash Flow	*CFTA*	Cash flow to assets
Cash Flow	*CFTD*	Cash flow to total debts
Cash Flow	*CFTS*	Cash flow to total sales
Leverage	*CLCA*	Current liabilities to current assets
Leverage	*EBITTA*	Earnings before interest and tax to total assets
Leverage	*FUTL*	Funds provided by operational funds divided by total liabilities
Leverage	*TDTA*	Total debt to total assets
Leverage	*TDTE*	Total debt to total equity
Leverage	*TDME*	Total debt to market equity
Liability	*CLLTLTA*	Current plus long-term liabilities to total assets
Liability	*CLTA*	Current liabilities to total assets
Liability	*LTLTA*	Long-term liabilities (debt) to total assets
Liquidity	*CATA*	Current assets to total assets
Liquidity	*CCL*	Cash to current liabilities
Liquidity	*CR*	Current ratio = CACL (Current assets to current liabilities) (Ross et al., 2016)
Liquidity	*CTA*	Cash to total assets
Liquidity	*EBTCL*	Earnings before tax to current liabilities
Liquidity	*QACL*	Quick assets to current liabilities
Liquidity	*QATA*	Quick assets to total assets
Liquidity	*RETA*	Retained earnings to total assets
Liquidity	*WCTA*	Working capital to total assets
Macroeconomics	*DJIA*	Dow Jones industrial average
Macroeconomics	*GDP*	Gross domestic product
Macroeconomics	*IHSG*	Composite stock price index
Macroeconomics	*INFLATION*	Inflation rate
Macroeconomics	*INTEREST*	Interest rate
Macroeconomics	*OIL*	Oil price
Macroeconomics	*UNCERTAINTY*	Uncertainty
Macroeconomics	*USDIDR*	Exchange rate
Market	*BVPS*	Book value per share
Market	*CASHMTA*	The stock of cash and short-term investments to market value of total assets
Market	*CP*	The logarithm of the closing price
Market	*MB*	Market to book ratio of the firm
Market	*SIZE*	The logarithm of market cap
Market	*CAPEX*	Capital expenditure
Profitability	*CHIN*	(Nit – NIt-1)/(|NIt| + |NIt-1|) where NIt is the net income for the most recent period
Profitability	*NIMTA*	Net income over the market value of total assets
Profitability	*NINW*	Net income to net worth
Profitability	*NITA*	Net income to total assets = ROA (return on assets) (Ross et al., 2016)
Profitability	*NITD*	Net income to total debts
Profitability	*NITS*	Net income to total sales
Profitability	*ROE*	ROE (return on equity)
Size	*TA*	Total assets
Size	*TS*	Total sales
Solvency	*EBITI*	Earnings before interest and tax to interest
Solvency	*MVEBVOD*	Market value equity to book value of total debt
Solvency	*TLMTA*	Total liabilities to market value of total assets
Solvency	*TLTA*	Total liability to total assets
Turnover	*ARTS*	Account receivables to total sales
Turnover	*CATS*	Current assets to total sales
Turnover	*CTS*	Cash to total sales
Turnover	*ITS*	Inventory to total sales
Turnover	*NWTS*	Net worth to total sales
Turnover	*QATS*	Quick assets to total sales
Turnover	*TATS*	Total assets to total sales
Turnover	*WCTS*	Working capital to total sales

APPENDIX B. CONFUSION MATRIX FROM RANDOM FORESTS (RF)

Table B.1. Without synthetic data

Confusion	***Matrix***	***Statistics reference***
Prediction	0	1
0	9304	126
1	1	135
Accuracy	0.9867	
95% CI	(0.9842, 0.9889)	
No information rate	0.9727	
P-value [Acc > NIR]	< 0.00000000000000022	
Kappa	0.674	
Mcnemar's test p-value	< 0.00000000000000022	
Sensitivity	0.9999	
Specificity	0.5172	
Pos pred value	0.9866	
Neg pred value	0.9926	
Prevalence	0.9727	
Detection rate	0.9726	
Detection prevalence	0.9858	
Balanced accuracy	0.7586	
'Positive' class	0	

Table B.2. With synthetic data

Confusion	***Matrix***	***Statistics reference***
Prediction	0	1
0	9304	126
1	1	135
Accuracy	0.9978	
95% CI	(0.9969, 0.9986)	
No information rate	0.7199	
P-Value [Acc > NIR]	< 0.00000000000000022	
Kappa	0.9946	
Mcnemar's test p-value	0.0003298	
Sensitivity	0.9996	
Specificity	0.9934	
Pos pred value	0.9974	
Neg pred value	0.9989	
Prevalence	0.7199	
Detection rate	0.7196	
Detection prevalence	0.7215	
Balanced accuracy	0.9965	
'Positive' class	0	

Figure B.1. Variable importance RF graph

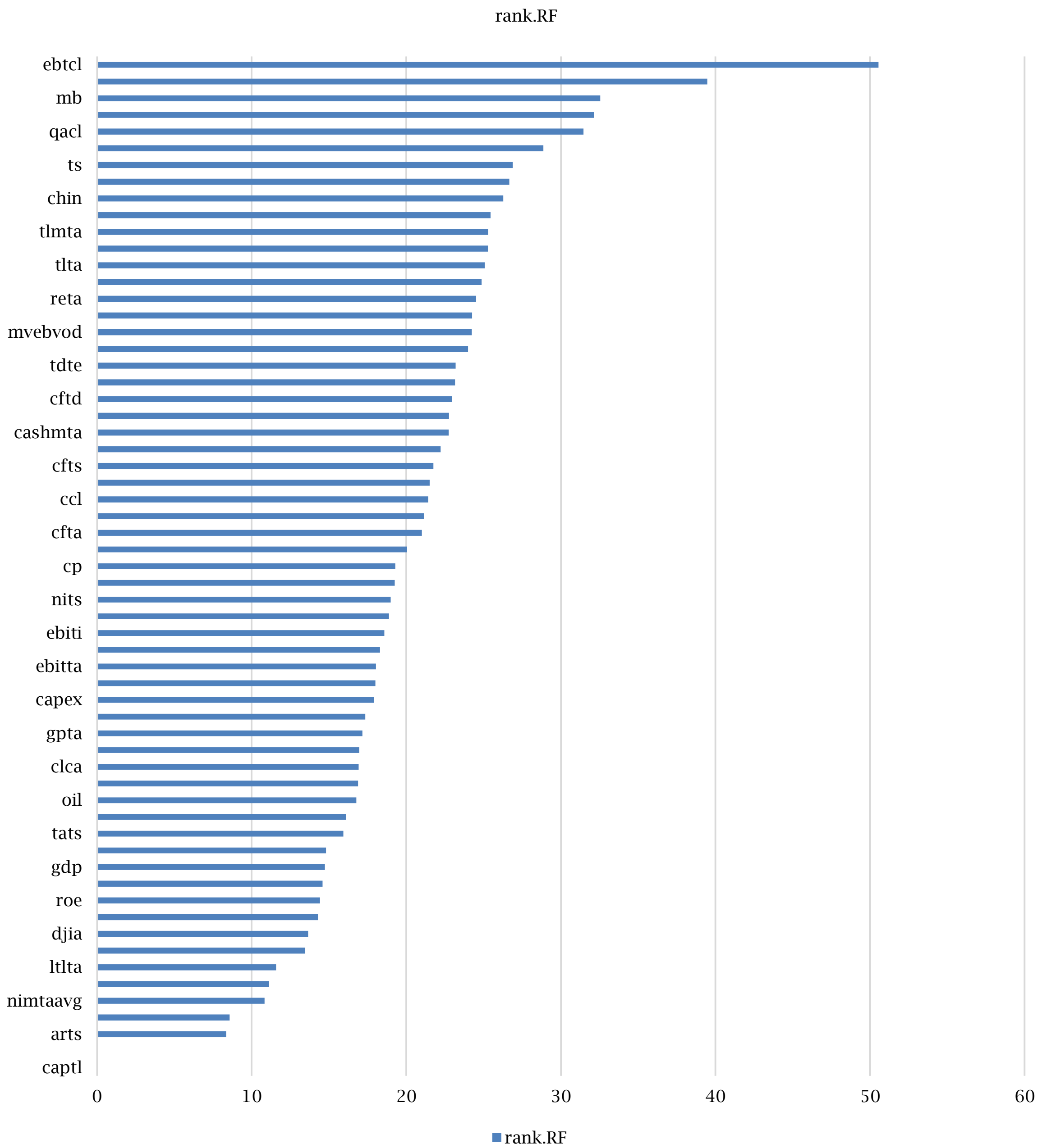

Table B.3. Variable importance RF table

Variables	*Rank.RF*	*Importance number*
ebtcl	50.54413949	1
inflation	39.46973157	2
mb	32.54651071	3
interest	32.14873304	4
qacl	31.46134016	5
cata	28.87120179	6
ts	26.89154753	7
qata	26.66742618	8
chin	26.27413958	9
ta	25.4626954	10
tlmta	25.30869094	11
bvps	25.29105354	12
tlta	25.08680763	13
cta	24.87195428	14
reta	24.51096992	15
tdme	24.26558438	16
mvebvod	24.23433787	17
tdta	23.99287942	18
tdte	23.19508618	19
cr	23.16231164	20
cftd	22.95291363	21
size	22.76209883	22
cashmta	22.74920857	23
clta	22.2194391	24
cfts	21.76162249	25
clltlta	21.52075585	26
ccl	21.42584361	27
cfnw	21.1374344	28
cfta	21.00478345	29
futl	20.05033825	30
cp	19.29217542	31
cts	19.24962324	32
nits	18.99505422	33
uncertainty	18.87879003	34
ebiti	18.59199584	35
qats	18.29528176	36
ebitta	18.05096654	37
its	18.00768042	38
capex	17.91214894	39
ninw	17.35307699	40
gpta	17.17281682	41
nwts	16.95391717	42
clca	16.92069928	43
cats	16.88182758	44
oil	16.77278018	45
tsta	16.11805766	46
tats	15.93985812	47
usdidr	14.81129162	48
gdp	14.73785024	49
ihsg	14.58461786	50
roe	14.42522463	51
nita	14.29648146	52
djia	13.64490232	53
nitd	13.47134003	54
ltlta	11.59009665	55
wcta	11.12383317	56
nimtaavg	10.83268106	57
wcts	8.56720352	58
arts	8.359567777	59
captl	0	60
fata	0	61

Table B.4. Multiple logit model fit assessment (Part 1)

No.	R^2	*Rank.RF*	*Vars*	*Aic*	*Sensitivity*	*Specificity*	*misClassError*	*Optimal*	*MAE*	*MSE*	*RMSE*
1	0.0356	1	ebtcl	20947.6075	0	0.99	0.2917	0.767	1.27738808825405	1.84644815712282	1.35884074016156
2	0.0574	2	inflation	20511.0215	0	0.9995	0.2825	0.8924	1.28172426986354	1.84593801810993	1.35865301608245
3	0.0622	3	mb	20379.1679	0	0.9998	0.2843	0.9244	1.2840198954215	1.85231475577095	1.36099770601238
4	0.0687	4	interest	20314.8471	0	0.9998	0.2788	0.9346	1.27853590103303	1.83586277260553	1.35494013617043
5	0.0901	5	qacl	19754.3913	0.0379	0.985	0.2857	0.7399	1.26425200867236	1.81418186455809	1.34691568576437
6	0.0949	6	cata	19658.8182	0.0371	0.9868	0.2843	0.7572	1.26539982145135	1.81507460783063	1.34724704780921
7	0.1748	7	ts	17921.974	0.6113	0.8011	0.2532	0.3686	0.969009054967479	1.19117459507716	1.09140945344869
8	0.1866	8	qata	17825.3365	0.4899	0.8554	0.2442	0.4068	1.03379670960337	1.31182247162352	1.14534818794265
9	0.1857	9	chin	17685.0068	0.4949	0.8528	0.2496	0.3988	1.03940823874506	1.32840198954215	1.15256322583282
10	0.1817	10	ta	17834.4982	0.5213	0.8513	0.2414	0.3956	1.02767504144879	1.29677337074353	1.1387595754783
11	0.1849	11	tlmta	17766.3067	0.5064	0.8483	0.2478	0.3958	1.02971559750032	1.30723122050759	1.14334212749622
12	0.1916	12	bvps	17603.7556	0.2936	0.9283	0.2511	0.4897	1.14832291799515	1.54776176508098	1.24409073828278
13	0.1863	13	tlta	17847.6086	0.4986	0.8623	0.2366	0.4138	1.03609233516133	1.30876163754623	1.14401120516638
14	0.1987	14	cta	17474.2417	0.4612	0.8652	0.2483	0.4208	1.05445733962505	1.35722484376993	1.16499993294846
15	0.1983	15	reta	17396.0558	0.6488	0.7959	0.2465	0.3592	0.956000510139013	1.15852569825277	1.07634831641657
16	0.2095	16	tdme	17203.2802	0.5924	0.8412	0.2296	0.39	1.00229562555796	1.23415380691238	1.11092475303793
17	0.2083	17	mvebvod	17189.5839	0.5793	0.8435	0.2326	0.38	1.00969264124474	1.25200867236322	1.11893193374897
18	0.2142	18	tdta	17175.6553	0.5718	0.8502	0.2273	0.4091	1.01096798877694	1.24920290779237	1.11767746143168
19	0.2153	19	tdte	17048.0874	0.5264	0.8622	0.2343	0.4089	1.03787782170642	1.31003698507843	1.14456847111845
20	0.2133	20	cr	17176.0551	0.5423	0.8655	0.2251	0.4095	1.0315010840454	1.28810100752455	1.1349453764497
21	0.2184	21	cftd	17056.205	0.5224	0.8662	0.2305	0.4174	1.03813289121286	1.30672108149471	1.14311901458016
22	0.2407	22	size	16477.7052	0.5222	0.8743	0.2276	0.4398	1.0491008799898	1.32585129447775	1.15145616263831
23	0.2411	23	cashmta	16483.9361	0.5714	0.8591	0.224	0.4144	1.02346639459253	1.27088381583982	1.12733482862893
24	0.2417	24	clta	16606.9419	0.4968	0.8868	0.221	0.4585	1.05726310419589	1.33554393572248	1.15565736086544
25	0.2434	25	cfts	16582.6724	0.5067	0.8794	0.2233	0.4501	1.04859074097692	1.32049483484249	1.1491278583528
26	0.2476	26	clltlta	16464.8136	0.5328	0.8823	0.2149	0.4499	1.04501976788675	1.30493559494962	1.14233777620703
27	0.278	27	ccl	15815.7534	0.5806	0.8762	0.2057	0.4298	1.02665476342303	1.25902308379033	1.12206197858689
28	0.2812	28	cfnw	15613.1006	0.5641	0.8801	0.2115	0.4194	1.04106619053692	1.29358500191302	1.13735878328389
29	0.284	29	cfta	15717.5857	0.5163	0.8865	0.215	0.4398	1.05024869276878	1.3155209794669	1.14696162946583
30	0.2782	30	futl	15802.5169	0.5683	0.8719	0.2126	0.428	1.02767504144879	1.26795051651575	1.1260330885528
31	0.2819	31	cp	15752.0813	0.4903	0.9061	0.2085	0.4687	1.07256727458232	1.35365387067976	1.1634663169511
32	0.2846	32	cts	15589.9232	0.4777	0.8993	0.2211	0.4698	1.07728606045147	1.37571738298686	1.17290979320102
33	0.2767	33	nits	15811.8274	0.5935	0.8687	0.2086	0.4091	1.01989542150236	1.24843769927305	1.11733508817769
34	0.2788	34	uncertainty	15824.3175	0.5196	0.8961	0.2078	0.4496	1.05726310419589	1.32228032138758	1.14990448359313
35	0.2856	35	ebiti	15576.3046	0.5652	0.872	0.2155	0.4198	1.03264889682438	1.280831526591	1.13173827654233
36	0.2883	36	qats	15554.6412	0.5479	0.8739	0.2181	0.4298	1.03698507843387	1.29205458487438	1.13668578986208
37	0.2885	37	ebitta	15597.7863	0.5335	0.8891	0.2097	0.4598	1.04948348424946	1.30863410279301	1.14395546364053
38	0.2866	38	its	15571.4155	0.5821	0.8724	0.2102	0.429	1.02754750669557	1.26527228669813	1.12484322760913
39	0.2901	39	capex	15545.817	0.5795	0.8672	0.2134	0.4186	1.0220635123071	1.25749266675169	1.12137980486171

Table B.4. Multiple logit model fit assessment (Part 2)

No.	R^2	*Rank.RF*	*Vars*	*Aic*	*Sensitivity*	*Specificity*	*misClassError*	*Optimal*	*MAE*	*MSE*	*RMSE*
40	0.285	40	ninw	15617.2352	0.5775	0.8802	0.2057	0.4284	1.03405177910981	1.27381711516388	1.1286350673109
41	0.2943	41	gpta	15469.6052	0.5599	0.8821	0.2078	0.4398	1.03762275219997	1.28299961739574	1.13269573028053
42	0.2897	42	nwts	15537.1551	0.5511	0.8891	0.2062	0.439	1.04706032393827	1.30034434383369	1.1403264198613
43	0.2928	43	clca	15462.3425	0.53	0.9022	0.2032	0.4698	1.06287463333758	1.32891212855503	1.15278451089309
44	0.3026	44	cats	15335.4414	0.52	0.9168	0.1922	0.4799	1.07154699655656	1.33528886621604	1.15554699870496
45	0.3022	45	oil	15305.6072	0.5565	0.9035	0.1932	0.4499	1.05394720061217	1.30110955235302	1.14066189221566
46	0.3089	46	tsta	15123.8745	0.5938	0.8911	0.1928	0.43	1.03647493942099	1.26578242571101	1.12506996480708
47	0.3023	47	tats	15291.3678	0.5662	0.904	0.1907	0.4399	1.05241678357352	1.29549802321133	1.13819946547665
48	0.3156	48	usdidr	14983.0439	0.5113	0.9131	0.2002	0.48	1.07550057390639	1.35123071036858	1.16242449663132
49	0.3119	49	gdp	15098.6603	0.5581	0.8985	0.1964	0.45	1.04999362326234	1.29639076648387	1.13859157140911
50	0.2993	50	ihsg	15332.5113	0.518	0.926	0.1895	0.4898	1.08340772860605	1.35633210049739	1.16461671828005
51	0.3104	51	roe	15095.142	0.5153	0.915	0.1981	0.4799	1.07613824767249	1.35033796709603	1.16204043264253
52	0.3148	52	nita	15069.8872	0.5372	0.902	0.1987	0.45	1.05688049993623	1.31246014538962	1.14562652962893
53	0.3113	53	djia	15199.685	0.4101	0.9475	0.1981	0.5699	1.1215406198189	1.44114271138885	1.20047603532468
54	0.3145	54	nitd	14962.9323	0.5678	0.8883	0.2039	0.4299	1.04476469838031	1.2934574671598	1.13730271570932
55	0.3171	55	ltlta	14995.9748	0.5955	0.8806	0.199	0.43	1.02678229817625	1.2525188113761	1.11915986855145
56	0.3138	56	wcta	15060.8518	0.6349	0.8686	0.1968	0.4	1.00752455044	1.21183522509884	1.10083387715806
57	0.3149	57	nimtaavg	15083.3035	0.5864	0.8932	0.1913	0.44	1.03647493942099	1.26425200867236	1.12438961604613
58	0.3189	58	wcts	14944.1135	0.5404	0.8971	0.203	0.4499	1.05509501339115	1.31322535390894	1.1459604504122
59	0.3268	59	arts	14708.884	0.5508	0.9053	0.1965	0.4599	1.06159928580538	1.31972962632317	1.14879485824196
60	0.3298	60	captl	14754.2077	0.6102	0.8784	0.1956	0.42	1.01964035199592	1.23491901543171	1.11126910126742
61	0.3226	61	fata	14897.9189	0.5207	0.919	0.1916	0.4798	1.07448029588063	1.34051779109807	1.15780732036815

Table B.5. Final model logistic regression result

	Estimate	*Std. Error*	*Z value*	*Prob*	*Significant at 1%*
(Intercept)	-1.6600828049	0.1085414957	-15.2940000000	0.00000000000000020000	***
chin	0.0060229209	0.0073418777	0.8200000000	0.41200000000000000000	
cta	-2.2737343428	0.3160277611	-7.1950000000	0.00000000000062580000	***
qata	-1.6879549130	0.1560264142	-10.8180000000	0.00000000000000020000	***
cata	1.0766329862	0.1092520226	9.8550000000	0.00000000000000020000	***
qacl	0.1393320255	0.0086571032	16.0950000000	0.00000000000000020000	***
reta	-0.2567562222	0.0251937732	-10.1910000000	0.00000000000000020000	***
ebtcl	-0.0026371950	0.0001864036	-14.1480000000	0.00000000000000020000	***
bvps	-205.1116954214	22.3645373890	-9.1710000000	0.00000000000000020000	***
ts	-0.0000078279	0.0000003723	-21.0270000000	0.00000000000000020000	***
ta	0.0000000319	0.0000000038	8.5120000000	0.00000000000000020000	***
mb	-38.9226718993	5.9189741088	-6.5760000000	0.00000000004835480000	***
tlmta	0.0000207180	0.0000037915	5.4640000000	0.00000004647941340000	***
tlta	-0.5043850247	0.0666254805	-7.5700000000	0.00000000000003720000	***
inflation	-0.6718456684	1.3055821299	-0.5150000000	0.60700000000000000000	

Figure B.2. Logit data plot