JIATIS

Journal of International Accounting, Taxation and Information Systems

https://jiatis.com/index.php/journal

Online ISSN 3048-085X

Application of Classification Algorithm on Financial Data to Improve Financial Distress Prediction

Dea Amellia Kartika Dewi[1], Sriwidharmanely Sriwidharmanely[2]*

[1,2]*University of Bengkulu, Bengkulu, Indonesia*

E-mail: [1)] deaa98991@gmail.com, [2)] nely.bachsin72@gmail.com

ARTICLE INFO

Article History
Received : 16.04.2025
Revised : 29.04.2025
Accepted : 06.05.2025
Article Type: Research Article

*Corresponding author:
Dea Amellia Kartika Dewi
deaa98991@gmail.com

ABSTRACT

The COVID-19 pandemic has intensified financial distress across various industries in Indonesia, especially in sectors like Accommodation and food & beverage, Other services, and Transportation & Warehousing. This situation highlights the urgent need for an accurate early warning system that can predict financial distress using reliable classification algorithms for business continuity. This research compares the Performance of the Support Vector Machine (SVM) and Decision Tree classification algorithms in predicting financial distress. The study utilizes secondary data from annual financial reports of companies listed on the Indonesia Stock Exchange (IDX) from 2019 to 2023. The research focuses on the Accommodation and food & beverage, Other services, and Transportation & Warehousing sectors. Data is collected using a purposive sampling method, ensuring balance across observations. A quantitative data analysis approach with an experimental design is applied to evaluate the classification performance. The results indicate that the Decision Tree algorithm outperforms SVM in all key Performance metrics: accuracy, precision, recall, and F1-score. The Decision Tree achieves perfect classification results, while SVM exhibits lower predictive ability, particularly in recall and F1-score. These findings suggest that the Decision Tree is more effective for financial distress prediction in this dataset. The study contributes to financial risk assessment by demonstrating the practical application of machine learning in financial classification tasks. Future research can enhance this model by incorporating larger datasets and developing application-based implementations to improve decision-making processes in corporate financial management.

Keywords: Financial Distress, Support Vector Machine, Decision Tree

1. Introduction

The global crisis marked by the COVID-19 pandemic has severely affected the Indonesian economy. Sharp economic fluctuations, triggered by tight monetary policy and supply chain disruptions, have pushed many large and small companies into financial distress. Wang et al. (2020). This condition is characterized by the company's difficulty meeting its financial obligations due to a significant decline in revenue. As a result, the risk of bankruptcy is increasingly threatening, potentially triggering investment losses, unemployment, and worsening overall economic conditions. As Abedin et al. (2021) researched, the pandemic has increased financial distress cases in many companies. This condition is further exacerbated by prolonged economic instability, which can hamper economic growth and weaken people's purchasing power.

The increase in financial distress cases has urged companies to have an effective early warning system. Moreover, the findings of the BPS indicate that a notable decrease in revenue in most companies is an alarm for all business actors. 3 out of 10 company sectors, namely the Accommodation and food & beverage, Other services, and Transportation & Warehousing sectors, experienced a drastic decline in demand due to economic uncertainty that began with the emergence of the COVID-19 case (Badan Pusat Statistik, 2020). In a digital era characterized by rapid technological development, using classification algorithms based on machine learning to predict financial distress emerges as a solution that can be considered. With its ability to analyze data in a complex and accurate manner, this model can help companies identify early signs of financial distress and take preventive action before the problem worsens. By predicting the likelihood of financial distress early on, companies can take appropriate mitigation measures, such as economic restructuring, improving operational efficiency, finding sources of capital to improve operational efficiency, or seeking additional funding.

Machine learning applications in economics and finance have grown exponentially in the past decade. Recent research by Tran et al. (2022) indicates that Machine Learning models yield significantly better outcomes than traditional statistical models, thus becoming a major driver of innovation in the financial sector. Classification algorithms are one effective way to secure and manage financial data. Classification algorithms are one of the important components in data science that are used to predict the category or label of data. Along with technological advancements and the growing complexity of the data at hand, the ability to classify data accurately is becoming increasingly crucial. Data scientists use classification to detect patterns and trends in data, which can subsequently be leveraged to make more informed and improved decisions. Classification algorithms minimize errors and increase the accuracy of predictions based on the data obtained. Data that is then processed into information will go through a classification and analysis process to determine whether a company has a level of financial distress.

In artificial intelligence, Machine Learning combines statistics and computer science to create algorithms that improve efficiency through exposure to relevant data rather than relying on explicit instructions. Machine Learning algorithms develop a population of models based on samples, defined as training data (Bahzad Taha Jijo, 2021), to produce forecasts or decisions without explicit programming. So that the data to be tested later will be more ready to be processed. Various data analysis methods, especially those leveraging machine learning, have become widely used in processing financial data to predict a company's financial health. Among the multiple algorithms, decision tree and support vector machine classification are two popular methods because of their capacity to manage complex data and user-friendly interpretation of results. These algorithms map the relationships between financial variables and generate decisions incrementally through decision trees and vectorization.

The logic behind Decision Tree algorithms plays a crucial role in predicting the likelihood of a company experiencing financial distress. This method can build complex predictive models by classifying financial data into distinct risk categories. Each branch in the Decision Tree represents a specific decision point or economic condition. This structure enables the model to identify patterns that signal potential financial problems. According to S. Y. Kim & Upneja (2014), Decision Tree and Support Vector Machine (SVM) are among the ten most popular data classification algorithms. Non-parametric prediction methods like Decision Trees are especially useful because they overcome the assumption limitations of MDA and logit models (Frydman et al, 1985). Bahzad Taha Jijo (2021), also found that Decision Tree was the most accurate method in their study. It achieved an impressive accuracy rate of 99.93%, outperforming other methods such as K-Nearest Neighbors (KNN), Linear Regression, Support Vector Machine (SVM), and Naïve Bayes.

Support Vector Machine (SVM) is an algorithm in machine learning that is highly effective in classifying data, including predicting the likelihood of financial distress in a company. SVM works by finding the optimal hyperplane to separate data into two or more classes, such as healthy companies and those experiencing financial difficulties. SVM's logical qualifications allow the model to learn complex patterns in corporate financial data, such as financial ratios, sales growth, and other factors. As such, SVM can identify the unique characteristics of companies that are likely to experience financial distress, even when the data is high-dimensional and non-linear. The advantage of SVM in predicting financial distress lies in its ability to handle imbalanced data, good generalization, and resistance to overfitting. As explained through research by Chen et al. (2011) on the topic of classification research, the results show that the SVM model excels in predicting the bankruptcy of German companies compared to the frequently used logit model. D. S. Kim & Shin (2021) SVM was also used to predict bankruptcy in 2,320 medium-sized companies within the Korea Credit Guarantee Fund from 1996 to 1999. The findings indicate that SVM offers more accurate prediction results than the Artificial Neural Network (ANN) and Random Forest models. These studies, as described, show disparate results, thus requiring further research to assess the effectiveness of the two leading classification algorithms in each study, namely Decision Tree and Support Vector Machine (SVM), to determine which classification algorithm performance is superior.

This study aims to determine the most effective classification algorithm for predicting financial distress in the Accommodation and food & beverage, Other services, and Transportation & Warehousing sectors listed on the Indonesia Stock Exchange (IDX) from 2019-2023, by comparing the performance of a Support Vector Machine (SVM) and Decision Tree using an experimental approach. Through this approach, the developed model can significantly contribute to business decision-making, especially in the early detection of financial decision-making and financial distress. Furthermore, this research will also discuss the performance evaluation of the resulting prediction model by measuring the accuracy, precision, recall, and F1 Score of the classification results. Thus, this research applies classification algorithms and emphasizes the importance of model testing to ensure its use in the business world. This research can help identify a more accurate and reliable financial distress prediction model that various parties can apply to mitigate the risk of financial failure and assist overall economic stability.

2. Literature Review

2.1. Probability Theory

In modern finance, probability theory models and measures uncertainty in volatile financial market conditions. This theory provides a robust framework for modelling uncertainty in financial data and estimating the likelihood of financial distress (Sheldon M. Ross, 2010). High entropy signifies high uncertainty, while low entropy signifies low uncertainty based on the information obtained. Information provides a robust framework for understanding various problems in data compression, learning, classification, and inference (Carrà, 2018). Information processing concerns how data is collected, processed, and transformed into useful information. Further, Bishop (2006) applies information in machine learning algorithms, defining information in probability and uncertainty. The lower the likelihood of an event, the higher the information it contains.

2.2. Financial Distress

Financial distress is when an entity (individual, company, or country) has difficulty meeting its financial obligations. Decreased liquidity, difficulty paying debts, and potential bankruptcy characterize this condition. As explained by Piatt and Piatt (2002), economic distress is a company's unhealthy financial condition. The company has difficulty paying off obligations.

Rismaniar and Aulia (2022) stated that financial distress refers to a situation in which a company experiences a decline in profits and business activities that do not align with its obligations, causing the company to incur losses if it continuously faces this condition and cannot address it. The company will cause the bankruptcy of the company. One can see that a company is experiencing financial distress if it experiences the following: negative net operating revenue for two straight years, stops dividend payments, restructures the company, and reduces employee labour.

2.3. Classification Algorithm

Classification algorithms are methods to classify data into specific classes based on the data's attributes. Classification algorithms are one of the methods in data mining, which is the process of extracting information or knowledge from large and complex data. Classification algorithms usually predict the class or label of new data and forecast the category or label of new data based on historical or training data with known classes. Classification algorithms fall under the category of supervised learning, which is learning that uses already labelled data as guidance. Classification algorithms can recognize financial data patterns, trends, and anomalies and predict future financial behaviour or outcomes. In their research, Charbuty and Abdulazeez (2021) explain that classification algorithms can handle large amounts of information and find relationships between input and output attributes in data mining.

Classification algorithms have many applications in the industrial world, such as classifying spam emails, recognizing faces or voices, detecting disease or fraud, recommending products or services, etc.

Classification algorithms evolve with technological developments and needs, such as deep learning, ensemble learning, or transfer learning. Classification algorithms can also help audit, verify, and validate financial data and detect and prevent fraud. There are two main types of classification algorithms: parametric and non-parametric. Parametric classification algorithms assume that the data follows a particular distribution, such as normal, binomial, or Poisson, and use parameters estimated from the data to create a classification model. A non-parametric classification algorithm does not assume that the data follows a particular distribution and uses the data as the classification model.

2.4. Hypothesis

2.4.1. Decision Tree and Support Vector Machine in Predicting Financial Distress

Data mining, particularly with the Decision Tree method, is becoming a highly relevant approach in the modern financial context because of its capability to process data uncertainty and provide easy-to-interpret results. The Decision Tree, first introduced in the 1960s, serves as a powerful tool for classification and prediction while simplifying the interpretation of complex data into easy-to-understand models. With an accuracy rate of 99.93%, this method is proven superior to other algorithms such as KNN, Linear Regression, SVM, and Naïve Bayes (Bahzad Taha Jijo, 2021). The study explains this superiority by integrating probability theory, which models uncertainty in financial data. It measures this uncertainty through entropy, where a low entropy value indicates low uncertainty and provides a strong basis for informed decision-making (Sheldon M. Ross, 2010). In this context, the study uses information to process and transform raw data into valuable insights, which underlies the effectiveness of machine learning algorithms, including Decision Trees, in accurately predicting financial distress (Bishop, 2006). Combining the Decision Tree's ability to generate interpretative models and the probability theory approach to quantifying uncertainty makes this algorithm ideal for complex financial research.

H$_1$: The Decision Tree classification algorithm performs better than the Support Vector Machine (SVM) classification algorithm.

2.4.2. Accuracy rate

Accuracy measures the proportion of correct predictions out of the total test data. The higher the accuracy, the better the model correctly classifies the data. Pattern Recognition and Machine Learning (Bishop, 2006) mentions that SVM relies heavily on the linear separation of data in high-dimensional space. SVM can struggle to build an optimal model if the data cannot be separated linearly with a clear margin. Decision Tree, on the other hand, can capture non-linear patterns and form probability-based rules from the dataset. According to Sheldon M. Ross (2010), the conditional probabilities used in node splitting allow the Decision Tree to divide the dataset into more structured categories compared to vector-based approaches such as SVM. If the dataset has many irrelevant features, SVM can suffer from overfitting because the model tends to fit the data into higher dimensions without considering the relevance of the features. In contrast, the Decision Tree implements a pruning technique, which allows the removal of insignificant branches based on the probability of occurrence of the pattern, thereby improving the overall accuracy of the model.

H$_{1a}$: The accuracy of the Decision Tree classification algorithm is higher than that of the Support Vector Machine (SVM) classification algorithm.

2.4.3. Precision Rate

Precision measures the proportion of correct positive predictions to the total one prediction. It is essential in scenarios where errors in classifying the positive class have a significant impact. In datasets with an imbalanced class distribution, SVMs are often biased towards the majority class due to their margin optimization method (Bishop, 2006). It can lead to more errors in detecting the minority class. Decision Tree is more flexible in dealing with imbalanced distributions by adjusting the probability weights at each node. Sheldon M. Ross (2010) explains that the conditional probability applied to each branch allows the model to be more adaptive in dealing with differences in proportion between classes. Based on probability theory, when the number of positive class examples is smaller, rule-based approaches such as Decision Tree can be more accurate in avoiding excessive false positives. This is because the attribute selection mechanism in Decision Tree is based on entropy minimization or information gain maximization, which ensures that the most relevant features for predicting the positive class are given greater weight.

H1b: The precision of the Decision Tree classification algorithm is higher than that of the Support Vector Machine (SVM) classification algorithm.

2.4.4. Recall Rate

Recalling measures how many positive data are correctly classified compared to the total number of positive data. A high recall value is important for applications where detecting all instances of the positive class is more important than avoiding misclassification. SVM works by constructing a hyperplane that separates two classes based on the maximum margin. However, when the data distribution is imbalanced or there are many outliers, SVM may fail to recognize some positive class instances because it tries to optimize the global margin (Bishop, 2006). Decision Trees, on the other hand, can adjust parameters such as tree depth and number of branches to be more flexible in capturing all possible categories in the data. Sheldon M. Ross (2010) mentioned that the conditional probability of the decision tree branching allows this model to handle complex data distributions without sacrificing recall. In cases where the positive class is more challenging to predict, a Decision Tree can provide better results in detecting all instances of that class. The gain ratio or Gini index-based splitting mechanism allows the model to focus on classes with low probabilities without sacrificing interpretability.

H1c: The recall of the Decision Tree classification algorithm is higher than that of the Support Vector Machine (SVM) classification algorithm.

2.4.5. F1 Score Rate

F1 Score is the harmonic means of precision and recall, a more balanced measure of model performance, especially when there is an imbalance between positive and negative classes in the dataset. Since Decision Trees are often better at capturing complex patterns in real datasets, they are more likely to produce higher F1 Scores than SVMs. Bishop (2006) notes that if the dataset has high noise, SVM can experience performance degradation because it relies heavily on margin-based separation. SVM produces less consistent predictions if there is an overlap between classes or unstable data distribution. Decision Tree is more noise-resistant than SVM because it can adjust the model to the dataset structure through pruning mechanisms and more significant feature selection. Sheldon M. Ross (2010) explains that the probabilistic strategy in tree-based decision-making allows this model to balance precision and recall, resulting in a better F1 Score. In general, since F1 Score is a combination of precision and recall, the superiority of the Decision Tree in both metrics directly contributes to the high F1 Score value compared to SVM.

H1d: The F1 Score of the Decision Tree classification algorithm is higher than that of the Support Vector Machine (SVM) classification algorithm.

3. Methodology

3.1. Research Design

This research uses a quantitative approach with experimental methods. The independent variables used are classification algorithms, namely Support Vector Machine (SVM) and Decision Tree. The performance of these algorithms is assessed based on accuracy, precision, recall, and F1-Score. Meanwhile, the dependent variable is financial distress, measured using five key financial ratios. These ratios include profitability, liquidity, leverage, and operational efficiency. The purpose of this study is to compare the ability of the two algorithms to predict financial distress.

3.2. Research Sample

The population of this study includes all companies listed on the Indonesia Stock Exchange (IDX) in the Accommodation and Food & Beverage, Other Services, and Transportation & Warehousing sectors, with complete annual financial reports for the period 2019–2023. Based on the availability of key financial ratios—including profitability, liquidity, leverage, and operational efficiency—a total of 131 companies were selected as the final sample from an initial population of 197 companies. The sample distribution by sector is as follows: 12 companies from the Industrial Services sector (out of 18), 12 companies from the Food & Staples Retailing sector (out of 14), 55 companies from the Food & Beverage sector (out of 95), 32 companies from the Consumer

Services sector (out of 49), all 9 companies from the Transportation & Infrastructure sector, and 10 companies from the Transportation & Logistics sector (out of 12).

3.3. Data Collection Tools and Procedure

The data obtained will be cleaned and converted into a numeric format suitable for analysis. This study will divide the data into training data and test data with a proportion of 70:30 each. The Support Vector Machine (SVM) and Decision Tree classification models will be built using training data of 70% of the total sample. The performance of the classification algorithm model will be assessed based on 4 performance metrics, namely accuracy, precision, recall, and F1-Score. Data analysis in this study uses Google Colab with the Python programming language.

3.4. Financial Distress Measurement

The Z-Score model uses five key financial ratios that cover important aspects of a company's financial health, such as profitability, liquidity, leverage, and operational efficiency. These ratios provide a broader and more in-depth view. Compared to qualitative methods that rely on interviews or observations (e.g., management or leadership assessments), the Z-Score provides more objective and measurable results. Research shows that the Z-Score model has an accuracy of **94%** in predicting bankruptcy in the first year before financial distress occurs. The Z-Score calculation is in the formula below.

$$\textbf{Z-Score} = 1.2*X_1 + 1.4*X_2 + 3.3*X_3 + 0.6*X_4 + 1.0*X_5$$

Outcome Indicators:

a. >2,99 = 0 (Healthy)
b. 1,81-2,99 = grey zone (Neutral to Distress)
c. <1,81 = 1 (Financial Distress)

3.5. Classification Algorithm Measurement

Classification algorithm performance refers to the indicators used to assess the algorithm's ability to classify data accurately and efficiently. This study evaluates this performance using various metrics, such as Accuracy, Precision, Recall, and F1-score. For binary classification, it presents the confusion matrix table in Table 1.

Table 1. Confusion Matrix

Data Class	Classified As Positive	Classified As Negative
Positive	True Positive (tp)	True Negative (fn)
Negative	False positive (fp)	False Negative (tn)

Table 1 identifies four evaluation terms that describe classification outcomes: true positive (tp), true negative (tn), false positive (fp), and false negative (fn). The definition of each term is as follows:

- True positive (tp): positive class data that is classified correctly.
- True Negative (tn): negative class data classified correctly.
- False positive (fp): negative class data classified as positive.
- False negative (fn): positive class data that is classified as a negative class.

Accuracy is a ratio describing the financial data level successfully detected during testing. This value reflects the extent to which system predictions are close to human predictions (Forest & Bayes, 2021). The formula for calculating accuracy is:

$$Accuracy = \frac{\mathrm{tp} + \mathrm{t}}{\mathrm{tp} + \mathrm{tn} + \mathrm{fp} + \mathrm{f}}$$

Precision is a measure of the accuracy of a system in accurately identifying positive and negative data. The study calculates the precision value by comparing the correct and total positive predictions. It explains the precision calculation process using the following formula:

$$Precision = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}}$$

Recall measures the extent to which a model can detect all positive data in a dataset. This metric is crucial when missing positive data, which is riskier than making false positive predictions (Powers, 2011). A high recall value indicates that the model can capture most of the positive data available, making it useful in cases where misclassification of the positive class needs to be minimized, such as in fraud detection or medical diagnosis. (Powers, 2011). This study calculates recall using the following formula:

$$Recall = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}}$$

F1 Score is the harmonic means of precision and recall, balancing the two to assess the model's overall performance. The F1 Score is beneficial when there is a class imbalance in the dataset, as it ensures that precision and recall are considered equally (Powers, 2020). The F1 Score formula is:

$$F1\ Score = \frac{\mathrm{Precission\ x\ Recall}}{\mathrm{Precission} + \mathrm{Recall}}$$

3.6. Research Procedure

The research procedure is shown in the figure below:

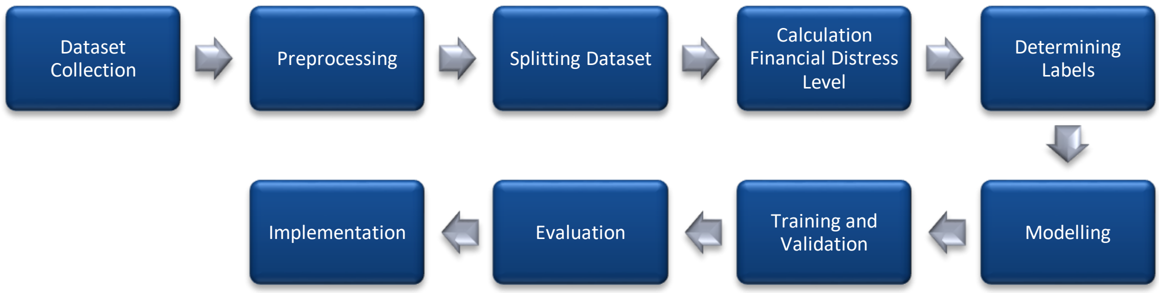


Figure 1. Research Procedure Flowchart

3.6.1 Dataset Collection

Data was collected from annual reports of companies in the Accommodation and Food & Beverage, Other Services, and Transportation & Warehousing sectors listed on the Indonesia Stock Exchange from 2019-2023 to find X_1, X_2, X_3, X_4, and $_5$.

3.6.2 Preprocessing

In this study, several preprocessing stages were carried out, namely: Cleaning to separate data that has special characters and columns or lists that need to be cleaned, NaN Value Handling to separate numeric data that cannot be represented, Financial Ratio Counting to calculate the level of financial ratios, Updating Data Frame to modify data in the DataFrame, such as changing values in certain cells, adding or deleting columns, or changing column names, and Removal Special Characters to remove special characters.

3.6.3 Splitting Dataset

After labeling and preprocessing, the data will be divided into training data and test data. Training data and test data get a proportion of 70:30 each from the samples used in the study.

3.6.4 Calculation Financial Distress Level

At this stage the algorithm will start to apply the calculation of the company's financial distress level. The formula used is Z-score with the calculation:

$$\textbf{Z-Score} = \mathbf{1.2{*}X_1 + 1.4{*}X_2 + 3.3{*}X_3 + 0.6{*}X_4 + 1.0{*}X_5}$$

3.6.5 Determining Labels

In this study, the level of financial distress is classified based on a certain score. Companies with a score greater than or equal to 2.99 are categorized as non-financial distress and given a code of 0. Meanwhile, companies with a score of less than 2.99 are classified as experiencing financial distress and given a code of 1. This classification aims to facilitate the labeling process in the analysis of financial distress prediction using a classification algorithm.

3.6.6 Modelling

At this stage, researchers apply the Decision Tree and Support Vector Machine algorithms to predict financial distress based on company financial data. Decision Tree is a decision tree-based classification method that divides data into branches based on the most influential features. The structure of the Decision Tree model consists of several main components: root, internal, and leaf nodes. Meanwhile, the SVM algorithm searches for the optimal hyperplane that separates the data into two classes: companies experiencing financial distress and those that are not. This model uses the concept of support vectors, namely data points that are near the decision boundary, to determine the most optimal separation. This study conducted a hyperparameter tuning process to optimize model performance using the Grid Search method to find the best combination of parameters.

3.6.7 Training and Validation

The training set is used to train the models, namely Decision Tree and Support Vector Machine. The validation set is useful during the training process to test the model's performance on data not used in training, and helps in tuning hyperparameters. The purpose of this data division is to avoid overfitting, which is a condition where the model only works well on training data but cannot generalize to new data.

3.6.8 Evaluation

After the model is tested, the model performance is measured using several evaluation metrics, namely Accuracy, Precision, Recall, and F1 Score. Then the Confusion Matrix is also used to analyze the model performance in more detail, which shows the number of correct and incorrect predictions for each class.

3.6.9 Implementation

At the implementation stage, the final model is applied to the test data, so that the model can predict financial distress using the inputted dataset without going through the previous process.

4. Results and Discussion

4.1. Results

Table 2 shows the data gathered from the population following the sample criteria.

Table 2. Research Sample

No	Business Sector	Population	Samples Accepted	Sample Rejected	Sample Rejection Information
1	Accommodation and Food & Beverage	109	67	42	41 IPO Companies Less Than 5 Years, 1 Company Suspended from July 2022 – July 2024
2	Other Services	67	44	23	20 IPO Companies Less than 5 years old, 1 Company Suspended since 2019, 2 Companies not available on IDX/Official Website
3	Transportation & Warehouse	21	19	2	1 IPO Company Less than 5 years old, 1 stopped trading on June 18, 2021
Total		197	130	67	

Table 2 presents the distribution of samples across three business sectors: Accommodation and Food & Beverage, Other Services, and Transportation & Warehousing. Out of a total population of 197 companies, 130 companies were accepted as research samples, while 67 companies were rejected. Rejections occurred due to reasons such as the company being listed on the stock exchange for less than five years, the company being suspended from trading, or the unavailability of data on the IDX or official website. The 130 accepted samples were further divided into two groups for model development and evaluation. Following a 70:30 ratio, 91 samples were allocated as training data, while 39 samples were used as testing data. This division aims to ensure the model is trained on an adequate amount of data while also being evaluated on unseen data to test its generalization and predictive performance.

4.1.1. Data Preparation

a. Dataset Collection

Data was collected from annual reports of companies in the Accommodation and food & beverage, Other services, and Transportation & Warehousing sectors to find X_1, X_2, X_3, X_4, and $_5$. X_1 is obtained by dividing total Working Capital by Total Assets, where Working Capital is obtained by total Current Assets minus Total Current Liabilities. X_2 is obtained by dividing the amount of Retained Earnings by Total Assets. X_3 is obtained by dividing Total Earnings Before Interest And Tax by Total Assets. X_4 is obtained by dividing the Market Value of Equity by the Book Value of Total Liabilities. X_5 is obtained by dividing total Sales by Total Assets.

Perusahaan	Tahun	Working Capital	Total Assets	Retained Earning	Earnings Before Inte	Market Value of Equi	Book Value of Total	Sales	Sector
PT ASTRA GRAPHI/	2019	Rp1.153.430.000.00	Rp2.896.840.000.00	Rp1.286.778.000.00	Rp852.405.000.000,	Rp250.873.173.000,	Rp1.270.830.000.00	Rp4.771.800.000.00	INDUSTRIAL SERV
PT ASTRA GRAPHI/	2020	Rp1.128.765.000.00	Rp2.288.831.000.00	Rp1.365.325.000.00	Rp539.241.000.000,	Rp1.085.768.302.50	Rp726.053.000.000,	Rp3.348.871.000.00	INDUSTRIAL SERV
PT ASTRA GRAPHI/	2021	Rp1.285.080.000.00	Rp2.655.278.000.00	Rp1.430.200.000.00	Rp514.197.000.000,	Rp971.121.960.000,	Rp1.027.625.000.00	Rp3.299.105.000.00	INDUSTRIAL SERV
PT ASTRA GRAPHI/	2022	Rp1.308.332.000.00	Rp2.677.651.000.00	Rp1.495.769.000.00	Rp593.340.000.000,	Rp1.240.878.060.00	Rp984.429.000.000,	Rp2.909.972.000.00	INDUSTRIAL SERV
PT ASTRA GRAPHI/	2023	Rp1.357.806.000.00	Rp26.828.130.000.0	Rp1.594.449.000.00	Rp641.746.000.000,	Rp1.200.414.645.00	Rp890.911.000.000,	Rp2.968.952.000.00	INDUSTRIAL SERV
PT Berkah Prima Pe	2019	Rp42.685.188.863,0	Rp81.987.669.814,0	Rp13.775.827.818,0	Rp48.068.598.673,0	Rp219.450.000.000,	Rp7.111.855.116,00	Rp100.093.362.672,	INDUSTRIAL SERV
PT Berkah Prima Pe	2020	Rp51.747.880.189,0	Rp92.305.660.771,0	Rp23.275.050.316,0	Rp35.887.551.827.0	Rp292.600.000.000,	Rp7.974.519.192,00	Rp74.179.874.751,0	INDUSTRIAL SERV
PT Berkah Prima Pe	2021	Rp46.583.397.790,0	Rp89.325.700.727,0	Rp19.417.184.749,0	Rp131.320.519.406,	Rp125.400.000.000,	Rp8.746.850.883,00	Rp109.018.092.634,	INDUSTRIAL SERV
PT Berkah Prima Pe	2022	Rp50.740.186.383,0	Rp95.882.590.139,0	Rp26.690.250.490,0	Rp44.477.725.122,0	Rp122.056.000.000,	Rp8.018.755.151,00	Rp131.320.519.406,	INDUSTRIAL SERV
PT Berkah Prima Pe	2023	Rp50.385.366.819,0	Rp103.992.480.787,	Rp29.629.656.242,0	Rp48.997.687.719.0	Rp145.464.000.000,	Rp13.292.487.394,0	Rp143.083.261.635,	INDUSTRIAL SERV
T Dyandra Media I	2019	Rp141.997.973.414,	Rp1.215.041.391.93	Rp735.252.031.388,	Rp297.890.690.584,	Rp504.209.784.922,	Rp390.067.438.922,	Rp980.145.831.991,	INDUSTRIAL SERV
PT Dyandra Media I	2020	-Rp21.601.248.097,(	Rp1.043.313.799.00	Rp518.410.541.352,	Rp46.260.180.559,0	Rp235.013.035.345,	Rp466.243.114.050,	Rp284.181.538.459,	INDUSTRIAL SERV
PT Dyandra Media I	2021	-Rp58.295.291.856,(	Rp959.596.413.102,	Rp216.038.954.543,	Rp130.030.601.511,	Rp320.472.320.925,	Rp478.290.525.186,	Rp284.181.538.459,	INDUSTRIAL SERV
PT Dyandra Media I	2022	Rp53.156.904.789,0	Rp1.088.333.283.16	Rp248.229.280.691,	Rp368.133.130.961,	Rp418.750.499.342,	Rp577.622.763.324,	Rp1.210.481.160.29	INDUSTRIAL SERV
PT Dyandra Media I	2023	Rp78.382.461.125,0	Rp1.203.642.535.21	Rp168.945.200.422,	Rp393.151.807.195,	Rp363.201.963.715,	Rp611.350.297.949,	Rp1.329.121.702.30	INDUSTRIAL SERV
PT Island Concepts I	2019	Rp203.360.257.040,	Rp369.071.617.773,	Rp52.353.743.065,0	Rp67.063.956.212,0	Rp74.103.000.000,0	Rp129.953.534.558,	Rp203.245.833.500,	INDUSTRIAL SERV
PT Island Concepts I	2020	Rp117.896.702.527,	Rp370.846.674.317,	Rp54.381.343.445,0	Rp29.602.680.098,0	Rp80.641.500.000,0	Rp125.712.975.552,	Rp143.693.572.364,	INDUSTRIAL SERV
PT Island Concepts I	2021	Rp127.631.152.488,	Rp371.158.059.902,	Rp51.885.806.973,0	Rp26.870.277.498,0	Rp111.154.500.000,	Rp126.763.216.728,	Rp140.656.758.698,	INDUSTRIAL SERV
PT Island Concepts I	2022	-Rp9.234.896.257,0(	Rp371.248.007.154,	Rp33.418.381.692,0	Rp22.617.013.586,0	Rp65.385.000.000,0	Rp150.145.263.915,	Rp171.934.673.596,	INDUSTRIAL SERV
PT Island Concepts I	2023	-Rp17.803.783.487.(	Rp359.741.973.622,	Rp28.488.588.174,0	Rp19.300.394.291.0	Rp54.487.500.000,0	Rp144.506.670.308,	Rp181.373.119.160,	INDUSTRIAL SERV
Tanah Laut Tbk	2019	Rp18.171.596.605,0	Rp38.827.752.319,0	-Rp48.341.812.641,(	Rp2.210.615.398,00	Rp778.239.500,00	Rp6.482.378.525,00	Rp4.800.000.000,00	INDUSTRIAL SERV
Tanah Laut Tbk	2020	Rp12.340.318.548,0	Rp48.633.324.937,0	-Rp48.292.111.788,(	Rp2.235.668.992,00	Rp415.438.400,00	Rp8.075.961.521,00	Rp4.800.000.000,00	INDUSTRIAL SERV
Tanah Laut Tbk	2021	Rp12.948.182.314,0	Rp65.165.122.716,0	-Rp49.341.460.514,(	Rp1.970.589.773,00	Rp113.288.296.000,	Rp11.598.116.951,0	Rp4.800.000.000,00	INDUSTRIAL SERV

Figure 2. Raw Dataset

b. Preprocessing

In this research, several preprocessing stages were carried out: Cleaning, NaN Value Handling, Financial Ratio Counting, Updating the Data Frame, and Removing Special Characters.

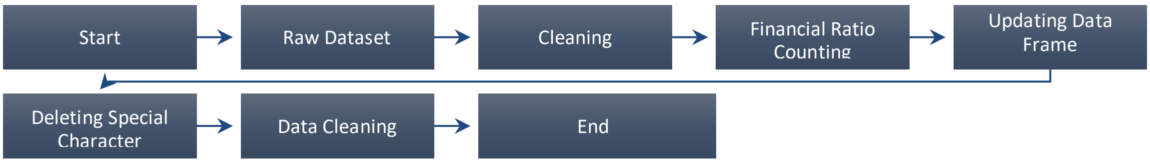


Figure 3. Data Preprocessing Flowchart

c. Cleaning

At the cleaning stage, to reduce noise or random errors in the data, a process of removing elements or special characters such as dots, commas, rupiah symbols, etc. will be carried out. In this process, the study uses scraping data. After successfully entering the data, the system checks whether it contains special characters. If it does, the system deletes these characters and converts the data into cleaned data. If not, the system immediately converts it into cleaned data.

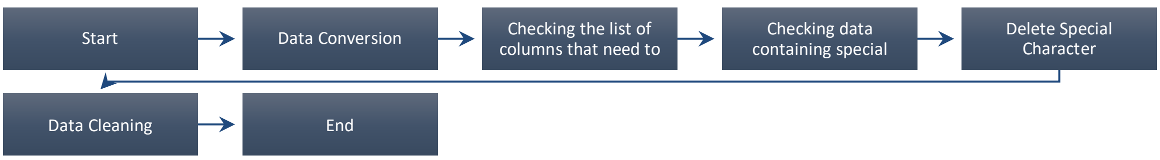


Figure 4. Stages of the Cleaning Process

d. Financial Ratio Counting

Create new features based on financial ratios. The data collected and gone through the cleaning process and NaN Value Handling will be calculated into five financial ratios, namely X1, X2, X3, X4, and X5, which are then stored in five new columns.

e. Splitting Dataset

After the labelling and preprocessing, the data will be divided into training and test data. Training data is used to test the model's accuracy; the model will learn to recognize patterns from this data by adjusting weights or parameters to make good predictions. It covers most of the dataset, 70%. Test data is used to measure model performance after the training process; the model has never seen this data before, so evaluation of test data reflects the model's ability to generalize to new data. Covers 30% of the total dataset. The main purpose of separating training data and test data is to ensure the model memorizes the data (overfitting) and works well on new data. The total data used in this study consists of 130 observations per year over 5 years, resulting in 650 observations. Out of this total, 455 observations were used as training data, while the remaining 95 observations were used as test data.

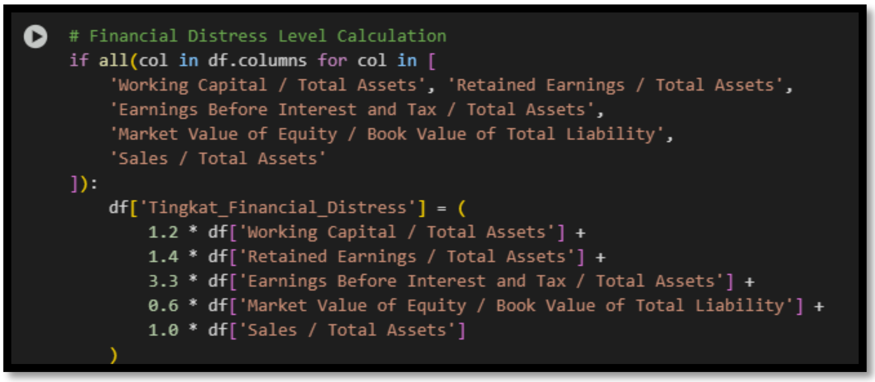
4.1.2. Implementation of Financial Distress Calculations

a. Calculation of Financial Distress Level

The level of Financial Distress is calculated using the Z-Score formula, namely:

$$\textbf{Z-Score} = 1.2*X_1 + 1.4*X_2 + 3.3*X_3 + 0.6*X_4 + 1.0*X_5$$

The program code used can be seen in picture 4 below:



```python
# Financial Distress Level Calculation
if all(col in df.columns for col in [
    'Working Capital / Total Assets', 'Retained Earnings / Total Assets',
    'Earnings Before Interest and Tax / Total Assets',
    'Market Value of Equity / Book Value of Total Liability',
    'Sales / Total Assets'
]):
    df['Tingkat_Financial_Distress'] = (
        1.2 * df['Working Capital / Total Assets'] +
        1.4 * df['Retained Earnings / Total Assets'] +
        3.3 * df['Earnings Before Interest and Tax / Total Assets'] +
        0.6 * df['Market Value of Equity / Book Value of Total Liability'] +
        1.0 * df['Sales / Total Assets']
    )
```

Figure 5. Source Code Z Score Counting

b. Determining Labels

The results of calculating the level of Financial Distress are then labelled with the following conditions:

Table 3. Financial Distress Labels

Level of Financial Distress	Code	Information
≥2,99	0	Non-Financial Distress
<2,99	1	Financial Distress

4.1.3. Modelling

a. Decision Tree

In this stage, researchers apply the Decision Tree algorithm to predict financial distress based on company financial data. Decision Tree is a decision tree-based classification method that divides data into branches based on the most influential features. The Decision Tree model structure consists of several main components: root, internal, and leaf nodes. The root node is the starting point for decisions that divide the dataset based on the most significant financial attributes. Internal nodes represent conditions that further separate data based on specific attribute values. Leaf nodes are the final classification result, determining whether a company is experiencing financial distress. In building the model, pruning techniques are used to avoid overfitting by reducing the decision tree's complexity. In addition, adjustments are made to parameters such as maximum tree depth (max_depth) and minimum samples per split (min_samples_split) so that the model can work optimally. Next, the model will be evaluated using accuracy, precision, recall, and F1-score metrics to measure its prediction performance. The program code used can be seen in Figures 6 and 7.

```
# Splitting Data into Training Data and Test Data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Defining Models and Grid Search
model = DecisionTreeClassifier(random_state=42)

# Hyperparameter Decision Tree
param_grid = {
    'max_depth': [3, 5, 10, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 5],
    'criterion': ['gini', 'entropy']}

# Calculating Training Start Time
start_time = time.time()

# Performing a Grid Search
if X_train.shape[1] > 0:
    grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=2, scoring='accuracy', n_jobs=-1)
    grid_search.fit(X_train, y_train)
    best_model = grid_search.best_estimator_
else:
    print("Tidak bisa melakukan GridSearch karena tidak ada fitur!")
```

Figure 6. Source Code for Pruning Technique

```
# Creating and Training a Decision Tree Model
model = DecisionTreeClassifier(max_depth=5, random_state=42)
model.fit(X_train, y_train)

# Calculating Model Training Completion Time
end_time = time.time()
training_time = end_time - start_time
print(f"Training time: {training_time:.4f} seconds")
```

Figure 7. Source Code Model Decision Tree

b. Support Vector Machine

The SVM algorithm finds an optimal hyperplane that separates data into two classes: companies experiencing financial distress and those not. This model uses the concept of support vectors, namely data points near the decision boundary, to determine the most optimal separation. The study carried out a hyperparameter tuning process to optimize model performance using the Grid Search method to find the best parameter combination. In addition, data standardization with Min-Max Scaling or Standard Scaling is applied so that all numerical variables have a uniform scale, thereby increasing the effectiveness of class separation by SVM. The program code used can be seen in Figures 8 and 9.

```
# Splitting training and test data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Data standardization
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

#Calculating Training Start Time
start_time = time.time()

#Determining the Model Before the Grid Search Process is Performed
model = SVC(class_weight='balanced')

# Hyperparameter SVM
param_grid = {
    'C': [0.1, 1, 10],
    'kernel': ['linear', 'rbf'],
    'gamma': ['scale', 'auto']
}
```

Figure 8. Source Code for Standardization of SVM Model Data

```
# Using Stratified K-Fold Cross Validation
cv = StratifiedKFold(n_splits=3)
grid_search = GridSearchCV(SVC(class_weight='balanced'), param_grid, cv=cv, scoring='accuracy')
grid_search.fit(X_train_scaled, y_train)

# Building and Training Support Vector Machines
svm_model = GridSearchCV(SVC(), param_grid, cv=3, scoring='accuracy', n_jobs=-1)
svm_model.fit(X_train.values, y_train)

# Calculating Model Training Completion Time
end_time = time.time()
training_time = end_time - start_time
print(f"Training time: {training_time:.4f} seconds")
```

Figure 9. Source Code Model SVM

4.1.4. Diagram Training and Validation

a. Decision Tree

The following are the results of accuracy visualization training in the form of diagrams from the Decision Tree model, which was formed according to Figure 10.

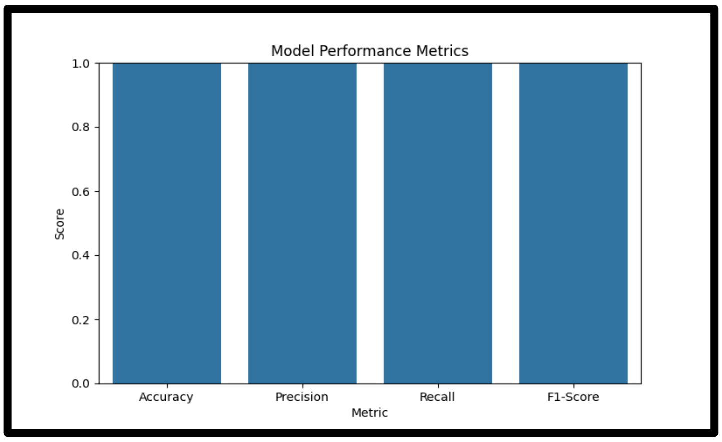


Figure 10. Model Performance: Metrics Decision Tree

Based on the bar diagram displaying the evaluation results of the Decision Tree model in predicting financial distress, the model achieved maximum performance with Accuracy, Precision, Recall, and F1-Score scores of 1.0 or 100%, respectively. Accuracy measures the extent to which the model can predict correctly from all available data. With a value of 1.0 (100%), the model makes no errors in predicting financial distress or non-distress. Precision indicates how many of the "positive" predictions (financial distress) are cases of financial distress. A score of 1.0 means no false positives, meaning that all instances predicted to experience financial distress experience financial distress. Recalling (Sensitivity) measures the extent to which the model succeeds in capturing all true positive cases. A score of 1.0 means no false negatives, which means the model successfully detects all companies truly experiencing financial distress. F1-Score is the harmonic average between Precision and Recall, which provides an idea of the balance between the two metrics. With a value of 1.0, this shows that the model has a perfect balance between accuracy in detecting financial distress and avoiding errors in predictions.

This decision tree model performs perfectly well in predicting financial distress. However, this perfect score could indicate overfitting, especially if the model is tested only on a limited data set and is not validated with other data. It is important to conduct cross-validation or testing with new data (out-of-sample testing) to ensure the model is reliable.

b. Support Vector Machine

The following results show the accuracy visualization training in the form of diagrams from the Decision Tree model, which is formed according to Figure 11.

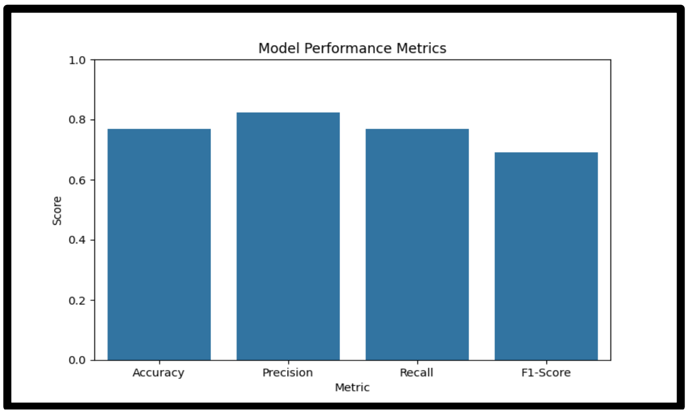


Figure 11. Model Performance: Metrics Support Vector Machine

The bar chart above shows an evaluation of the performance of the Support Vector Machine (SVM) model in predicting financial distress based on four main metrics: Accuracy, Precision, Recall, and F1-Score. Accuracy with a value of 0.769 (76.92%) means that the model successfully predicted financial distress and non-distress in around 77% of all test data. This value is quite good, but there is still a possibility that the model will experience errors in some predictions. Precision, with a score of 0.824 (82.39%), means that when the model predicts a company experiencing financial distress, 82.39% of the predictions are correct. False positives (errors in predicting a company is distress even though it is not) still exist, but not so much. Recall (Sensitivity) has a value of 0.769 (76.92%), which means that of all companies that experienced financial distress, the model only detected around 76.92% of them. Still, 23.08% of financial distress cases are missing (false negative). F1-Score with a value of 0.690 (69.03%) shows that even though precision is quite high, there is still a trade-off because recall is not optimal. The model needs to be improved to recognize more cases of financial distress without too many false positives or negatives.

Conclusion & Recommendations The SVM model performs well, with an accuracy of almost 77%. Precision is higher than recall, so the model is more careful in providing positive predictions (financial distress), but some distress cases are still undetected. The low F1-score (69%) indicates that the model could still be improved to be more balanced in detecting financial distress without losing accuracy. If an imbalanced dataset occurs, the model can be tuned to improve performance by optimizing parameters, using additional data, or testing balancing methods.

4.1.5. Evaluation

This study evaluates the model's performance for each training experiment by calculating its accuracy, precision, recall, and f1 score. Additionally, we calculate a confusion matrix to analyze the model's performance in more detail, showing the number of correct and incorrect predictions for each class. The evaluation results help determine the model's suitability for use on new data. Figures 12 and 13 display the confusion matrix results of this study.

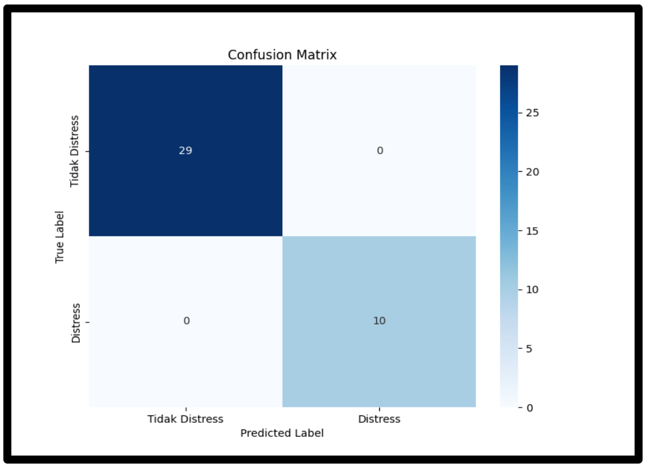


Figure 12. Confusion Matrix Dataset on Decision Tree

In this confusion matrix, there are four main quadrants, each showing the number of model predictions:

a. True Negative (TN) = 29. The model correctly predicted 29 companies that did not experience financial distress as 'Not Distressed.' It means that the model is not wrong in recognizing truly healthy companies.
b. False Positive (FP) = 0. No False Positive cases mean no companies are healthy but incorrectly classified as distressed. It means the model does not make mistakes in predicting distress for healthy companies.
c. False Negative (FN) = 0. There are no False Negative cases, which means no companies are experiencing financial distress but are predicted not to be in distress. In other words, the model does not miss a single case of distress.
d. True Positive (TP) = 10. The model successfully classified ten companies experiencing financial distress as 'Distressed, ' meaning it accurately recognizes companies experiencing financial difficulties.

The Decision Tree model performs excellently in this case because there are no errors in predictions (no False Positives or False Negatives). The model is 100% accurate because all predictions are correct. It shows that the Decision Tree effectively detects financial distress in this dataset.

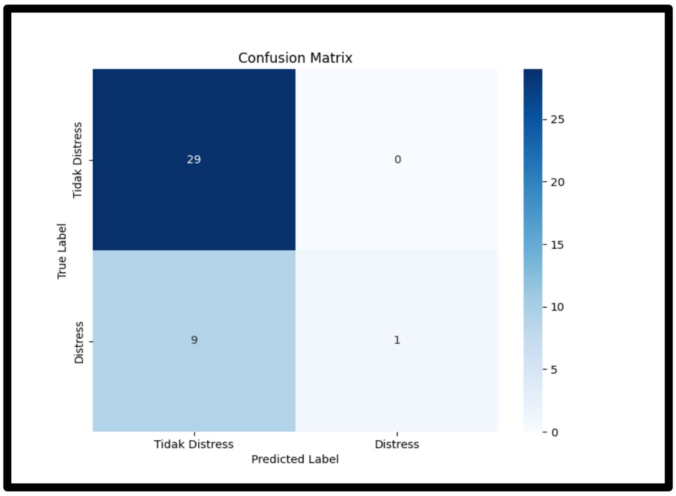


Figure 13. Confusion Matrix Dataset pada Support Vector Machine

In this confusion matrix, there are four main quadrants, each showing the number of model predictions:

a. True Negative (TN) = 29. Twenty-nine companies not experiencing financial distress were correctly classified as 'Not Distressed.' It shows that the model is particularly good at recognizing healthy companies.
b. False Positive (FP) = 0. There are no False Positives, meaning that no healthy companies are incorrectly classified as distressed. It shows that the model does not overpredict distress categories.
c. False Negative (FN) = 9. Nine companies experiencing financial distress were incorrectly classified as 'Not Distressed'. ' It means the model can detect companies experiencing financial distress less accurately, so many distress cases are missed.
d. True Positive (TP) = 1. Only one company truly experiencing distress is correctly classified as 'Distressed.' It shows that the model has difficulty correctly identifying distressed companies.

Confusion Matrix Model SVM is particularly good at recognizing healthy companies because there are no False Positives. However, the model has difficulty recognizing companies experiencing financial distress, as seen from the number of False Negatives (9 out of 10 distressed companies are undetected). The model is biased towards the 'Not Distressed' class, which can occur if the dataset is unbalanced (for example, the number of healthy companies is much greater than the number of distressed companies). The model accuracy may seem high, but the recall for the distress class is exceptionally low, which is dangerous if used to detect companies in financial distress.

4.1.6. Implementation

Based on the entire series of research carried out, this research has reached the implementation stage of the model created so that the model can predict financial distress using the dataset that has been input without going through the previous process. Figures 14 and 15 show the program code of Decision Tree model, and the Support Vector Machine model in predicting financial distress, which is the input.

```
# Predicting with the Best Model
best_model = grid_search.best_estimator_
y_pred = best_model.predict(X_test)

# Computerization of Evaluation Scores
y_pred = best_model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
```

Figure 14. Financial Distress Decision Tree Level Prediction Program Code

```
# Predicting with the Best Model
best_model = grid_search.best_estimator_
X_test_scaled = scaler.transform(X_test)
y_pred = best_model.predict(X_test.values)
df_grouped.loc[X_test.index, 'Predicted_Label'] = y_pred

# Computerization of Evaluation Scores
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average="weighted", zero_division=1)
recall = recall_score(y_test, y_pred, average="weighted", zero_division=1)
f1 = f1_score(y_test, y_pred, average="weighted", zero_division=1)
```

Figure 15. Financial Distress Level Prediction Program Code Support Vector Machine

Based on financial distress prediction analysis, the Decision Tree (DT) shows very optimal performance, with accuracy, precision, recall, and F1-score reaching 100%, which means this model can classify all companies correctly without errors. Accuracy shows the overall level of prediction accuracy, precision measures how many "distress" predictions are correct, recall shows the model's ability to detect all companies experiencing distress, and the F1-score is a balance between precision and recall. Meanwhile, the Support Vector Machine (SVM) shows lower performance, with an accuracy of 76.9%, precision of 82.4%, recall of 76.9%, and F1-score of 69%, which shows that this model still experiences classification errors, especially in detecting companies that are genuinely experiencing distress. DT is superior because it can handle patterns in data well and works more effectively in rule-based classification. These results show that DT can be a more reliable choice for financial distress analysis. However, in the case of more complex and diverse data, SVM can still be an alternative that needs to be further optimized.

4.2. Discussion

The results of this study indicate that Decision Tree has superior performance compared to Support Vector Machine (SVM) in predicting financial distress. This superiority is investigated through evaluation based on four main metrics, namely accuracy, precision, recall, and F1-score, where Decision Tree consistently produces higher values compared to SVM. Based on the test results, the Decision Tree algorithm shows an accuracy of 1.0 (100%), while the SVM algorithm shows an accuracy of 0.7692 (76.92%). This means that Decision Tree is able to classify all data correctly, while SVM still experiences 23.08% of misclassifications. In the precision metric, Decision Tree also obtained a perfect score of 1.0, indicating that all positive predictions generated are truly positive cases. In contrast, SVM only achieved a precision of 0.7692, meaning that around 23.08% of positive predictions generated were incorrect. In addition, in the recall metric, Decision Tree again recorded a score of 1.0, indicating success in detecting all positive cases in the data. In contrast, SVM only obtained a recall of 0.6410 (64.10%), indicating that 35.90% of positive cases were not detected. Finally, in terms of F1-score — which is a harmonization between precision and recall — Decision Tree obtained a perfect value of 1.0, while SVM only achieved 0.6903 (69.03%). Thus, it can be concluded that the Decision Tree algorithm is more consistent and effective in classifying financial distress compared to SVM based on the four evaluation metrics.

This advantage can be explained through the probability theory that is the basis for creating the classification model. Decision Tree divides the dataset based on the probability of each feature appearing in a particular class, using the entropy and information gain approaches to determine the most informative attributes in decision making (Sheldon M. Ross, 2010). With this mechanism, Decision Tree is able to capture non-linear patterns in financial data, making it more effective in classifying companies experiencing financial distress. On the other hand, although SVM is known as a robust algorithm in data separation, the results of this study indicate that the model is less effective in handling datasets with unbalanced distributions. SVM works by finding the optimal hyperplane that separates the classes, but when the data has complex patterns or an ill-defined minority class, this model can suffer from higher classification errors (Bishop, 2006). This is clearly seen in this study, where the recall and F1-score of SVM are lower than Decision Tree, indicating its inability to capture all instances of financial distress in the dataset. This limitation of SVM is also confirmed in previous studies, which found that SVM often has difficulty detecting complex financial patterns, especially when the data has many interacting variables.

From a probability perspective, the Decision Tree has the advantage of building a probabilistic rule-based classification model, which allows the model to be more flexible in grouping data based on its unique characteristics (Sheldon M. Ross, 2010). In the context of financial distress prediction, Decision Tree can analyse the relationship between various economic variables with distress or non-distress status deterministically. It explains why this model can achieve extremely high performance in this study. In contrast, SVM, which relies on margin-based separation, often fails to detect minority classes with unclear distributions, resulting in lower recall values than the Decision Tree. It aligns with the findings (D. S. Kim & Shin, 2021), which state that Decision Tree is more effective in financial data classification than SVM because of its ability to handle datasets with unbalanced distributions and more complex patterns.

With these results, the study has confirmed all hypotheses. Decision Tree is proven to have higher accuracy than SVM (H1a), better precision in handling imbalanced class distributions (H1b), superior recall in capturing all instances of financial distress (H1c), and higher F1-score in maintaining a balance between recall and precision (H1d). These results are also consistent with the research (Bahzad Taha Jijo, 2021), which found that Decision Tree has an accuracy of up to 99.93% in financial classification, surpassing other algorithms such as KNN, SVM, and Naïve Bayes. In addition, the study (Halina Frydman, Edward I. Altman, 1985) also supports this finding, stating that the Decision Tree is more effective than traditional statistical models in predicting bankruptcy. Thus, this study concludes that the Decision Tree is a more accurate and reliable model for financial distress analysis than SVM, especially in dataset conditions with complex financial patterns and unbalanced class distributions.

Moreover, the superior performance of the Decision Tree model may also be attributed to industry-specific characteristics inherent in the sectors analyzed. For example, the Food & Beverage and Consumer Services sectors are often marked by fluctuating revenue streams, variable cost structures, and seasonal trends, which tend to create non-linear and high-dimensional financial patterns. These conditions align well with the Decision Tree's strength in capturing complex relationships without assuming a specific data distribution. In contrast, sectors like Transportation & Infrastructure typically display more consistent financial indicators with fewer outliers, limiting the separation capacity of SVM, especially when minority class instances—such as financially distressed companies—are sparsely represented. These structural differences between sectors affirm that the effectiveness of a classification algorithm is not only dependent on its theoretical framework but also on its compatibility with the underlying economic realities of each industry. Therefore, selecting an appropriate model should be grounded in a comprehensive understanding of both data characteristics and sector-specific financial behavior.

5. Conclusion

This study was conducted before, during, and after the COVID-19 pandemic in the period 2019-2023, which was marked by increasing financial uncertainty and increasing vulnerability of companies in various sectors in Indonesia. The disruptive impact of the pandemic on revenue streams, operational stability, and market demand has increased the need for accurate and early prediction of financial distress. This study confirms that Decision Tree outperforms Support Vector Machine (SVM) in predicting financial distress among companies listed on the Indonesia Stock Exchange. The superior accuracy, precision, recall, and F1 score of Decision Tree demonstrate its strong capacity to reliably classify financial conditions, especially when dealing with complex financial patterns and imbalanced class distributions. These advantages reflect the theoretical foundation of Decision Tree in probability, where entropy and information gain are used to group data based on the most informative features. This allows for better pattern recognition and better differentiation of financial distress status. In contrast, SVM's reliance on margin-based separation is less effective in contexts where the data does not have clear boundaries.

This study has limitations that offer directions for future research. Expanding the scope to include other sectors—such as banking, manufacturing, and technology—would increase the generalizability of the findings. Furthermore, applying ensemble learning methods such as Random Forest or Gradient Boosting could further improve accuracy and reduce the risk of overfitting. Developing practical applications, such as predictive systems or interactive dashboards, would allow the findings to be directly used by stakeholders, including companies, investors, and regulators.

These findings imply that Decision Tree-based models can be adopted by companies and financial regulators as reliable early warning tools to detect financial distress, enabling proactive strategies in risk mitigation. Moreover, this supports the increasing relevance of machine learning in enhancing financial decision-making processes, especially in sectors vulnerable to economic shocks. Overall, the results underscore the practical utility of the Decision Tree model as an early warning tool in financial risk analysis. Its effectiveness in identifying at-risk companies makes it well-suited to enhancing proactive financial strategies, especially in uncertain economic environments. The study also reinforces the growing value of machine learning techniques in financial decision-making.

6. References

Abedin, M. Z., Moon, M. H., Hassan, M. K., & Hajek, P. (2021). Deep learning-based exchange rate prediction during the COVID-19 pandemic. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04420-6

Badan Pusat Statistik. (2020). Katalog: 3101028. *Analisis Hasil Survei Dampak COVID-19 Terhadap Pelaku Usaha*, vi+ 22 halaman.

Bahzad Taha Jijo, A. M. A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. In *Journal of Machine and Computing* (Vol. 4, Issue 1). https://doi.org/10.53759/7669/jmc202404020

Carrà, S. (2018). Entropy and Information. In *Frontiers Collection: Vol. Part F929*. https://doi.org/10.1007/978-3-319-95459-2_2

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Chen, S., Härdle, W. K., & Moro, R. A. (2011). Modeling default risk with support vector machines. *Quantitative Finance*, *11*(1), 135–154. https://doi.org/10.1080/14697680903410015

Halina Frydman, Edward I. Altman, D.-L. K. (1985). Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance*, *40*(1), 269–291. https://doi.org/10.1111/j.1540-6261.1985.tb04949.x

Kim, D. S., & Shin, S. (2021). The economic explainability of machine learning and standard econometric models-an application to the u.S. mortgage default risk. *International Journal of Strategic Property Management*, *25*(5), 396–412. https://doi.org/10.3846/ijspm.2021.15129

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling*, *36*, 354–362. https://doi.org/10.1016/j.econmod.2013.10.005

Piatt, H. D., & Piatt, M. B. (2002). Predicting corporate financial distress: Reflections on choice-based sample bias. *Journal of Economics and Finance*, *26*(2), 184–199. https://doi.org/10.1007/bf02755985

Powers, D. M. W. (2020). *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. *May*.

Sheldon M. Ross. (2010). Introduction to Probability Models. In *Journal of the American Statistical Association* (Vol. 93, Issue 441). https://doi.org/10.2307/2669658

Tran, K. L., Le, H. A., Nguyen, T. H., & Nguyen, D. T. (2022). Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. *Data*, *7*(11). https://doi.org/10.3390/data7110160

Wang, G., Ma, J., Chen, G., & Yang, Y. (2020). Financial distress prediction: Regularized sparse-based Random Subspace with ER aggregation rule incorporating textual disclosures. *Applied Soft Computing Journal*, *90*, 106152. https://doi.org/10.1016/j.asoc.2020.106152