

Application of Decision Tree in Analyzing Indonesia Stock Exchange (IDX) Companies' Valuation

Dionisius Darryl Hermansyah 13519058¹

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

¹13519058@std.stei.itb.ac.id

Abstract—Stock is the most common choice when it comes to invest our money in. Investing in the stock market, requires a skill to distinguish between undervalued and overvalued stocks in order to achieve consistent returns. Fundamental analysis is a tool that stockholders commonly use to evaluate stock value. Unfortunately, this time-consuming method needs to analyze dozens of economic parameters that exist in a company. Thus, a more robust and effective solution in analyzing companies' stock valuation, specifically in Indonesia Stock Exchange (IDX), is required. Decision tree is a tree-based algorithm that can be applied to solve numerous classification problems. It is a potential tool to be used as a model to predict the valuation of a stock. The experiment result proves that a stock valuation process can be modelled in a decision tree. The Gini importance in the decision tree shows that the most essential parameter when determining a valuation of a stock is PER or price/earnings ratio, EPS or earnings per share, PBR or price book ratio, and ROA or return on assets. By evaluating each attribute in the decision tree, it could help stockholders to select various potential stock in IDX and obtain a better return.

Keywords—decision tree, IDX, stock, valuation

I. INTRODUCTION

Equity stocks are one of the main choice for people, coming from a various background, to invest their money in. With dispensable wages emerging and individuals looking past ordinary sparing plans, people especially millennials are jumping in the stock market even if they only have a little knowledge.

Analysis is a very crucial process to follow if someone is going to invest in a stock. It is surely not an easy task to do. Stock selection process in the equity market, Indonesian Stock Exchange (IDX), for example, is a dreary and tedious task because of the wide variety of stock that are available for purchase. Apart from that, there are a lot of external factors that also affect the decision making process when buying a stock. The problem is, not all stocks or companies have a promising prospect for the future.

To evaluate a prospect of a company's stock, one may use a technical or fundamental analysis. Technical analysis is based only on stock price or transaction volume data. On the other hand, fundamental analysis is a method of evaluation a company and determining the intrinsic value of correlating stock. By

analyzing the fundamental of the company, we can determine its valuation, whether it is undervalued, overvalued, or just right. Take a sample stock from IDX, for example BBKA (Bank Central Asia). By analyzing its profit, net profit, dividend, return on capital, and other indicators, we can be sure if BBKA is undervalued or overvalued. Thus, fundamental analysis is more of a logical way to analyze stock, but it is more complex. Based on those problems, a more robust and effective solution in analyzing companies' stock valuation, specifically in IDX, is required.

Decision tree is a algorithmic tool, based on discrete mathematics theory, that are often used in data mining and data analysis process. This specific algorithm, uses a tree-like graph as a model for decision making. In consequence, decision tree is a potential and powerful tool that can help us in evaluating IDX stock valuation.

This paper will discuss about determining IDX companies' stock valuation utilizing some main components involved in fundamental analysis using decision tree.

II. BASIC THEORY

A. Stock

A share or stock is an object that represents the ownership of a company. Any company has to start its activities having a specific proprietors' capital when it is enrolled with the authorities. This indicates who owns the company. The ownership is reflected in its shareholding pattern.



Figure 1. Indonesia Stock Exchange (IDX) as the stock market authority in Indonesia (Source: Tirta.ID)

Stock can be bought and sold in the stock market. Stock market is also referred as capital market. In Indonesia, Indonesia Stock Exchange (IDX) acts as a main stock market authority.

Buying a stock means that we are buying the company or the business. There are two main income from stock, which are capital gain and dividend. Capital gain is obtained by the movement of stock prices. It is affected by the forces of demand and supply. If the prices are rising, it must be that there are more buyers than seller, *vice versa* [9].

When an individual or a company bought a stock, they are called as stockholder or shareholder. They will be granted privileges depending on the value and class of stock that they hold [4].

B. Stock Valuation

Stock valuation is a process of determining the current and/or the projected worth of a stock in a give period of time. Commonly, there are 2 ways to get a value of a stock, which are absolute and relative valuation. Absolute valuation calculates the present worth of business and forecasting the future income of the company. On the other hand, relative valuation compares the stock value to its competitors in the same industry [1].

Each stock in the market are unique, and all of them have their own intrinsic value or valuation.

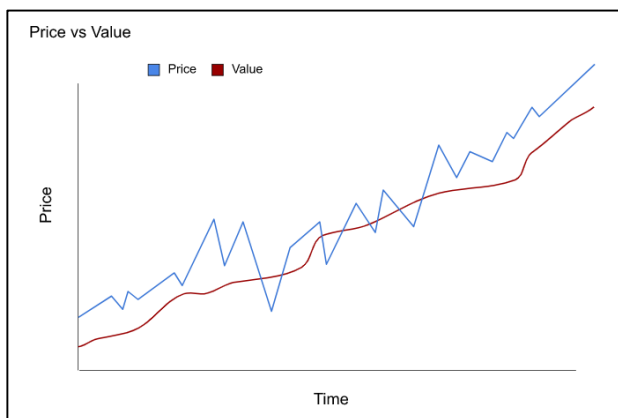


Figure 2. Overvalued stock (Source: Cliffcore)

Investing in the stock market, requires all of the investors to distinguish between undervalued and overvalued stocks in order to achieve consistent returns. Undervalued stock means that the price of the stock is way below the true value of it. On the contrary, overvalued stock means that the price of the stock is far more expensive than the true value.

The three scenarios of stock valuation are as follows:

- The current stock price is higher than its fair value, meaning that the stock is overvalued.
- The current stock price is equal to its fair value, meaning that the stock is fair-valued
- The current stock price is lower than its fair value, meaning that the stock is undervalued.

There are a lot of ways to assess the value of a stock, one of the methods is applying valuation ratios and metrics by doing fundamental analysis [2].

C. Technical Analysis

Technical analysis uses a stock's past price and buying or selling volumes to determine and forecast the future price of the

stock. The principle of technical analysis lies in the collective actions on the market, which are buying and selling. This method makes an assumption that all known fundamentals are factored into stock price. Thus, it is not effective to measure a company's stock intrinsic value. According to technical analysis, it is better to use stock charts to identify underlying patterns and trends that will be used to predict the stock future price point. Technical analysis aims to capitalize on pricing opportunities, patterns, and trends that are revolving around the market activity around each existing stock. This method is considered as a backward looking methodology since it is based on historical price data on the stock market [6].

D. Fundamental Analysis

Fundamental analysis is a technique to evaluate a stock by measuring the valuation of a stock. Valuation can be inferred from the intrinsic value of a stock. Fundamental analysis works by measuring and analyzing overall economy and industry conditions to the financial and management of companies. Some factors that could be used as an indicator are, but not limited to, earnings, assets, expenses, and liabilities. Fundamental analysis aims to establish a value for the stock that would be suitable for all of the underlying factors mentioned above.

This specific technique requires analyzing different financial, micro-economic, and macro-economic factors stock determine the valuation of a stock. By evaluating the overall economic and company conditions, the stock value could be determined almost precisely. When assessing the company, there are some important ratios that could be used as the main tools in determining stock valuation, such as [10]:

a. Earnings per Share (EPS)

EPS is calculated as follows:

$$EPS = \frac{\text{Net income} - \text{Dividend}}{\text{Number of outstanding shares}}$$

EPS shows how much each share is assigned with the company's profit.

b. Dividend Payout Ratio

Compares dividends paid out to the company's net income. It indicates for retained earnings, which is income that is not paid out, rather retained for company's growth.

c. Price / Earnings Ratio (P/E)

The formula for P/E is as follows:

$$PE = \frac{\text{Market price per share}}{EPS}$$

P/E shows how closely the price of the stock followed the earnings per share. High P/E ratio means that the stockholders were expecting the stock price to increase, *vice versa*.

d. Return on Assets (ROA)

ROA is calculated by:

$$ROA = \frac{\text{Net income} + \text{Interest expense}}{\text{Total assets}}$$

It shows how well the company was using its assets. Higher ROA means that the assets were effectively and efficiently used by the company, *vice versa*.

e. *Return on Equity (ROE)*

The formula for ROE is:

$$ROE = \frac{\text{Post-tax earnings}}{\text{Shareholder equity}}$$

It shows how well the shareholders' money was used and how profitable the money that was invested. Higher ROE indicates a more robust company regarding the shareholder's assets.

f. *Debt/Equity Ratio*

This is given by:

$$DE = \frac{\text{Total liabilities}}{\text{Shareholders' equity}}$$

The higher the DE ratio, means that the credit that is available is utilized well and fully.

g. *Market Capitalization*

To calculate a company's market capitalization, the formula is:

$$MC = \frac{\text{Number of shares}}{\text{Price per share}}$$

This indicators measure the amount of shares traded in the market. Companies in stock market can be categorized based on its capitalization as small-cap, medium-cap, and large-cap.

h. *Price/Sales Ratio*

The formula for price/sales ratio is as follows:

$$PS = \frac{\text{Share price}}{\text{Revenue over 1 year}}$$

This shows if the share price of the stock represents the value of the stock or not.

i. *Price/Book Ratio*

This is given by:

$$PB = \frac{\text{Stock price}}{\text{Total assets} - \text{Intangible assets} - \text{Liabilities}}$$

Price/book ratio compares the stock's intrinsic value with the share price. This is one of the most important indicator for determining valuation of a stock.

E. *Comparison between multiple analysis*

As explained, there are two main analysis in the stock market, which are technical and fundamental analysis. Both are used for researching and also forecasting the future prince of a unique stock. Apart from that, they are considered as an opposite from

each other. While fundamental analysis is based on news, articles, companies' economic data, and other unstructured data, technical analysis uses a more structured past stock movement data [6].

F. *Tree*

Tree is an undirected graph that is fully connected and have no circuits at all [3]. If G is a tree with N number of nodes, then G Is a tree if and only if:

- a. G is connected and have n-1 edges
- b. G has no circuit and an addition of an edge will make a circuit in G
- c. All of the edges in G are bridges

There are some important terms in a tree, such as:

1. Root : the top node of a tree.
2. Child : a node connected directly to another node moving further from the root.
3. Parent : The reverse of a child.
4. Siblings : a group of nodes with a single parent.
5. Ancestor: a node is a descendant to its ancestors.
6. Leaf : a node without child.
7. Degree : number of a node's children.
8. Edge : connection between nodes.
9. Path : continuous connection of a node.
10. Level : number of edges between node.
11. Depth : level minus one.
12. Height : Max distance from root to a leaf.

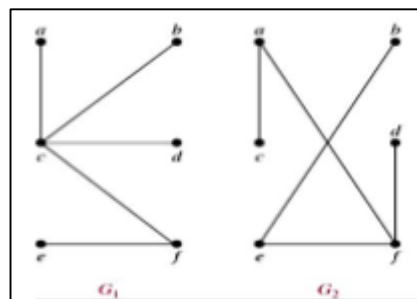


Figure 3. Graph G1 and G2 are trees (Souce: K.H. Rosen)

There are some types of tree such as n-ary trees. N-ary trees is a tree where all of its nodes have n or less children. Such tree are useful to build a tree model such as binary tree [7].

G. *Decision Tree*

Tree have a lot of application and usage, whether to help use solve complex problems or even a small tasks in our daily life. With different approaches, trees can be used as a basic solution in almost everything. One of the tree application is decision tree.

Decision tree is a tree-based algorithm that can be used to solve regression and classification problems. It is mainly used to make decision based on some deciding factors. It uses a tree as a representation in which each leaf node corresponds to a label or a condition, meanwhile the attributes are represented in the internal node of the tree. Boolean function or any discrete attributes can be represented in the decision tree.

Decision tree consists of a structure that includes a root node, branches, and leaf nodes, similar with ordinary trees. Each internal node denotes a test case for an attribute. Each branch

depicts the outcome of the test case. Finally, the leaf node holds a label for corresponding class. The root of the tree compares the values of the root attribute to the record's attribute. We will follow the branch corresponding to the value on the basis of the comparison and move to the next node. By continuing to compare the record's attribute values with other node until we reaches the end or the leaf, the result will be the class shown in the leaf.

Decision tree is utilizing two popular attribute selection measurement to help identify the attribute for the root node in each level. This process is also known as attribute selection. The two measures are information gain and Gini index.

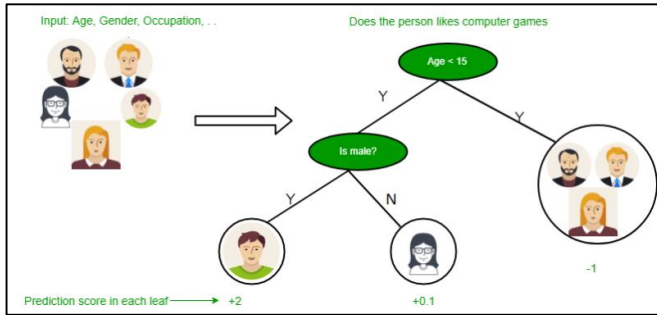


Figure 4. Example of a decision tree (Source: GeeksforGeeks)

Information gain is used when we are using a node in the tree to partition the training instances into subsets. Information gain is a measure of the change in entropy.

$$Gain(S, A) = Entropy(S) - \sum Val(A) \frac{|S_v|}{|S|} Entropy(S_v)$$

where S is a set of instances, A is an attribute, S_v is the subset of S with A = v, and Val(A) is the set of all possible values of A. Gini index is a metric to measure how often a randomly chosen element would be incorrectly identified. An attribute with lower Gini value should be preferred. The formula is as follow:

$$GI = 1 - \sum_j P_j^2$$

where P_j is the proportion of the samples that belongs to a class for a particular node.

In a decision tree, feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. Mathematically, we can calculate the importance or Gini importance by:

$$fi_i = \frac{\sum_j \text{node } j \text{ splits on feature } i \cdot ni_j}{\sum_k \text{all nodes } ni_k}$$

where fi_i is the importance of feature i and ni_j is the importance of node j. ni_j could be calculated as:

$$ni_j = w_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)}$$

where:

- ni_j : the importance of node j
- W_j : weighted number of samples reaching node j
- C_j : the impurity value of node j

- left(j) : child node from left split on node j
- right(j) : child node from right split on node j

A decision tree can be converted into a linear form as a collection of decision rules. The outcome of the leaf node and the condition along the path can be generalized into “if condition_b and condition_b then outcome.”

Once a decision tree is built, it can be applied to new data to classify it. Decision tree is often used to train a model which can predict a class or value of a target variables by learning decision rules. This rules is made out of the training data [5].

Decision tree, like other algorithms, have its own advantage and disadvantages. The advantages of decision tree are as follows:

- a. Simple to understand and interpret
- b. Can generate important insights even with a little amount of data.
- c. Useful to determine worst, best, and average values for various scenarios.
- d. Can be combined into another algorithm such as random forest.

The disadvantages of decision tree are:

- a. Volatile and sensitive in the changing of data. A small change of data can result in a large change in tree structure.
- b. A more complex calculation is required compared to other algorithm.

In machine learning, tree based algorithms are considered one of the best and mostly used supervised learning methods. Trees are adaptable at solving any kind of problem at hand.

III. DECISION TREE MODELLING

Decision tree can be used as a model that can predict whether a stock in IDX is overvalued or undervalued. To build such decision tree, we can model it using the help of a simple machine learning.

In this case, we will use a financial data in .csv format obtained from Stockbit, which is a stock market platform in Indonesia. The financial data contains various companies stocks that are considered to be overvalued and undervalued. The data will be used to train a decision tree model and determine the comparing-rules of each node in the tree. Here are the example of the data:

Table 1. Example of the stock data used (Source: Stockbit)

Stock	EPS	DPR	PER	ROA	ROE	PBR	V
ADHI	92.1	10	221.3	0.9	5.9	0.82	O
BBCA	29.49	47.9	29.49	2.8	15.5	4.4	U
BBNI	414.56	25	20.56	0.8	7	1.07	O
BRIS	21.42	10	56.02	0.4	3.9	2.69	U
DUCK	32.07	0	16.83	3.1	4.5	0.51	O
FIRE	9.44	0	67.71	2.6	3.9	4.24	U
GIAA	-682.6	0	-0.5	-12	259.4	-1.6	O
ICBP	438.69	49.8	21.9	5	19	4.3	U
IRRA	39.07	28.9	39.0	25.5	35.8	10.5	O

The data itself consists of 7 main features which are stock, EPS (earnings per share), DPR (dividend payout ratio), PER (price/earnings ratio), ROA (return on assets), ROE (return on equity), and PBR (price book ratio) value, that had been

explained before. Meanwhile the target feature is valuation (V) that indicates if the stock is overvalued (O) or undervalued (U). The features that we will use as the training data are all of the main features except stock or stock name. All of this training data will be use to train the decision tree model such that it can be used to predict the corresponding stock valuation.

In the modelling process, we use Python, Pandas, and Scikit-learn package to train our decision tree model. Hereby are the codes:

```
import pandas as pd
from sklearn import tree
import matplotlib.pyplot as plt

df = pd.read_csv('data.csv')

X = df.drop(['stock','valuation'], axis=1)
y = df['valuation']

clf = tree.DecisionTreeClassifier()
clf.fit(X,y)

fn=['EPS', 'DPR', 'PER', 'ROA', 'ROE', 'PBR']
cn=['undervalued', 'overvalued']
fig, axes = plt.subplots(nrows = 1,
                          ncols = 1,
                          figsize = (4,4),
                          dpi=300)

tree.plot_tree(clf,
               feature_names = fn,
               class_names= cn,
               filled = True);
```

IV. RESULT AND ANALYSIS

After we train the model, we can look at the decision tree that is generated by the model. The picture of the decision tree can be produce by Sklearn automatically. The resulting decision tree is shown in the fig. 5.

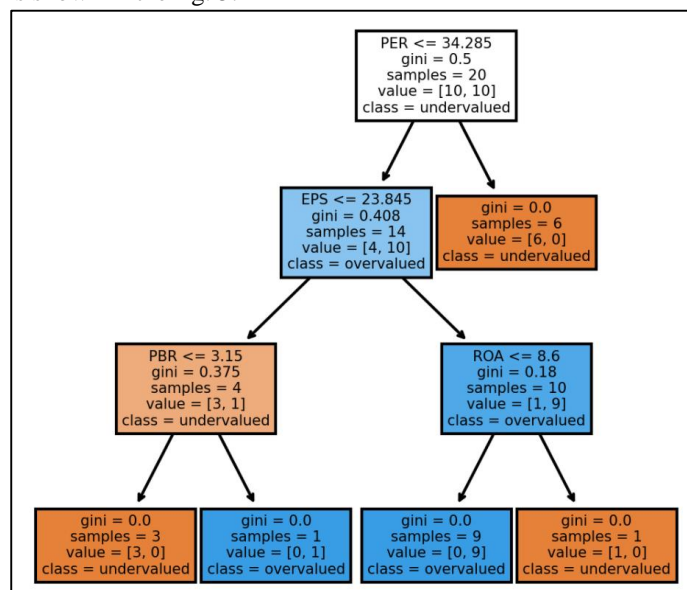


Figure 5. Decision tree result

As we can see, the decision tree is a binary tree. The 3-leveled decision tree has 9 nodes in total. There are 4 leaves that show the result of each corresponding attribute after doing the data-

checking. We could also obtain the Gini value in each node as shown in the figure 5.

Each node in the tree, from the root to the leaf, will analyze each given value in the data. By doing this, the decision making process will be based on the parameter that exist in each corresponding node.

The result indicates that although there are 6 different training features, the decision tree is minimizing the uses of the available feature to generate a more effective and efficient tree algorithm. Based on the resulting tree, we can observe that PER or price/earnings ratio, EPS or earnings per share, PBR or price book ratio, and ROA or return on assets are more important feature compared to the others. As explained in the basic theory, we can calculate feature importance in decision tree by:

$$f_i = \frac{\sum_j \text{node } j \text{ splits on feature } i . n_{ij}}{\sum_k \text{all nodes } n_{ik}}$$

where f_i is the importance of feature i and n_{ij} is the importance of node j .

With the formula, we get that the importance of each feature are as follows:

Table 2. Decision tree features' importance

Feature	Gini importance
EPS	0.24142857
DPR	0
PER	0.42857143
ROA	0.18
ROE	0
PBR	0.15

As we can see, the most important feature to predict a stock valuation using decision tree is PER or price/earnings ratio. On the other hand, DPR and ROE does not have any importance at all for the tree. Note that we can get a high variability of result by using different dataset, and this decision tree might only perform best in Indonesia Stock Exchange (IDX).

Furthermore, the decision tree can be linearized as following rules:

- If PER is less than or equal 34.285 then the stock is undervalued.
- If PER is more than 34.285 and EPS is less than or equal 23.845 and PBR is less than or equal 3.15 then the stock is overvalued.
- If PER is more than 34.285 and EPS is less than or equal 23.845 and PBR is more than 3.15 then the stock is undervalued.
- If PER is more than 34.285 and EPS is more than 23.845 and ROA is less than or equal 8.6 then the stock is overvalued.
- If PER is more than 34.285 and EPS is more than 23.845 and ROA is more than 8.6 then the stock is undervalued.

As we can observed, the corresponding tree produces 5 unique rules when determining valuation of a stock.

For a faster analysis, we can simplify the resulting tree into:

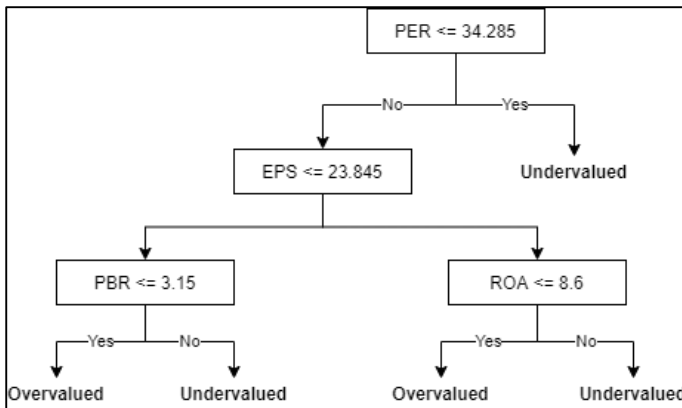


Figure 6. Simplified decision tree result

We can use the simplified decision tree to predict a valuation of another stock that is not a part of the training data. For example, we will use Media Nusantara Citra (MNCN) stock data. The data of MNCN is as follows:

Stock : MNCN	ROA : 10.5
EPS : 130.54	ROE : 1.43
DPR : 14.0	PBR : 1.17
PER : 8.69	

If we take a look at the PER of MNCN, which is 8.69, because the PER is less than 34.285 then MNCN is undervalued. Undervalued stock means that it is still a great deal to buy that stock since the current price is still below the fair-value.

Another example is Kimia Farma (KAEF) stock, the economic data is:

Stock : KAEF	ROA : -0.1
EPS : -3.13	ROE : -0.3
DPR : 20	PBR : 2.9
PER : 397.54	

The PER of KAEF is high, around 397.54. According to the root of the tree, we have to move to the left child in order to check the EPS. The EPS is -3.13 which is lower than 23.845. Now, all we have to do is to check the PBR. The PBR of KAEF is 2.9, which is lower than 3.15. According to our decision tree model, KAEF stock is overvalued. An overvalued stock indicates if the current price is already beyond the fair-value, so we should reconsider if we want to buy it.

From the experiment, it is proven that we can train a decision tree model based on stock data to obtain a robust model in predicting stock valuation. This approach could be very useful for stockholders or stock enthusiast to help them in analyzing potential stock by determining the valuation of each stock. By filtering the stock and looking for an undervalued stock, we can surely invest our money and hope to get a great gain from it. In short, this will be helpful for investors in Indonesia to choose the stock for investments in order to earn better returns.

V. CONCLUSION

Decision tree, as an tree-based algorithm in discrete mathematics, can be applied to determine a valuation of a stock in Indonesia Stock Exchange (IDX), whether it is undervalued or overvalued. Result shows that the most important ratios when

determining a valuation of a stock using decision tree is PER or price/earnings ratio, EPS or earnings per share, PBR or price book ratio, and ROA or return on assets. By evaluating each value in the decision tree, it could help Indonesian stockholders when it comes to making a decision in buying various available stock in IDX.

For further research, the author suggests that the decision tree model could be developed and trained with a larger amount of data from various stock market around the world, so that it is not only limited to IDX market.

VI. ACKNOWLEDGMENT

The completion of this paper would not be possible without the blessings of Almighty God. The author would like to thanks Dra. Harlili S., M.Sc. as the lecturer of IF2120 Discrete Mathematics class, Dr. Ir. Rinaldi Munir, MT. for creating the useful paper reference website, family, also fellow friends for supporting the creation of this paper.

REFERENCES

- [1] A. Stotz, "What is Stock Valuation?", ValuationMasterClass, 2020, retrieved from www.valuationmasterclass.com on 4th December 2020.
- [2] Anonymous, "How to Determine Whether a Stock Is Undervalued or Overvalued", CliffCore, 2020, retrieved from www.clifcore.com on 4th December 2020.
- [3] H.K. Rosen, *Discrete Mathematics and Its Applications*, New York: McGraw-Hill International, 2012, 7th Edition.
- [4] T.M. Obamuyi, *Factors Influencing Investment Decisions in Capital Market: A Study of Individual Investors in Nigeria*, Organizations and Markets in Emerging Economics, 2013.
- [5] P. Gupta, "Decision Trees in Machine Learning", TowardsDataScience, 2017, retrieved from www.towardsdatascience.com on 4th December 2020.
- [6] R. Bowman, "Fundamental vs. Technical Analysis", CatanaCapital, 2019, retrieved from www.catanacapital.com on 4th December 2020.
- [7] R. Munir, *Matematika Diskrit*, Bandung: Departemen Teknik Informatika Institut Teknologi Bandung, 2003.
- [8] S. Ronaghan, "The Mathematics of Decision Trees, Random Forest, and Feature Importance, in Scikit-learn, and Spark", TowardsDataScience, 2018, retrieved from www.towardsdatascience.com on 4th December 2020.
- [9] T. D. Chaudhuri, *Stock Market Investing*, Calcutta: Calcutta Business School, 2018.
- [10] C.K. Vekantesh, *Fundamental Analysis as a Method of Share Valuation in Comparison with Technical Analysis*, International Journal of Applied Financial Management Perspectives, 2012.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 6 Desember 2020

Dionisius Darryl Hermansyah (13519058)