

# Trend Analysis of The Effect of LQ45 Stocks on Stock Price Index Fluctuations using The C4.5 Algorithm with Correlation-Based Feature Selection and Information Gain

Muh Fitra Nur Asri<sup>a</sup>, Catur Edi Widodo<sup>b</sup>, & Eko Sedyono<sup>c</sup>

<sup>a</sup>Master Student of Information System Postgraduate School, Diponegoro University, Semarang, Indonesia

<sup>b</sup>Department of Physic, Diponegoro University, Semarang, Indonesia

<sup>c</sup>Faculty of Information Technology, Kristen Satya Wacana University, Salatiga, Indonesia

---

## Abstract

The research was conducted to reveal the effect of LQ45 stock on the accuracy of stock price index fluctuations using the C4.5 algorithm with Correlation-Based Feature Selection (CFS) and Information Gain (IG) techniques. This study used the superior C4.5 algorithm using a combination feature selection technique between Correlation-based Feature Selection (CFS) and Information Gain in the hope of getting accurate results. Analysis conducted on the LQ45 index through various stages that include data collection, manual pre-processing, validation methods, process features, decision tree model result, and classification accuracy performance. The result of test revealed that the implementation of the C4.5 algorithm using correlation-based feature selection (CFS) and information gain techniques can be applied well to LQ45 stocks. The accuracy generated from the original data (without the selection feature) was 77.857%, while the addition of features to the combination of Correlation-Based Feature Selection (CFS) and Information Gain had a large influence on the results of increasing data accuracy from the accuracy of the original data by 77.857% to 78.333%. Thus, the C4.5 calculation process with the Correlation-based Feature Selection (CFS) feature selection technique alone cannot improve the accuracy level, while when combined with the Information Gain technique, the accuracy processing results will be better (higher).

*Keywords:* C4.5; LQ45; index fluctuations; stock price

---

Received: 17 June 2023

Revised: 11 September 2023

Accepted: 3 October 2023

## 1. Introduction

The capital market is defined as a market of various types of financial instruments that are long-term in nature and can be traded (Junaedi and Salistia 2020), both debt securities (bonds) (Karle 2021), equities (shares) (Sucimanah, 2018), mutual funds (Sari 2019), derivative instruments, or others. For various companies or many other institutions (for example, the government), the capital market is a means of funding as well as a means of investing (Chirumalla 2021). Among the forms of funding found in the capital market are shares. Then, stocks are a financial instrument that is one indicator of the nation's progress. Stocks are also an investment regarding the ownership of a company. In the study of the capital market, an index of stocks or bonds is an imaginary portfolio that measures price changes in a market or part of that market.

When the stock index tends to rise, this shows that the prices of most of the stocks that the index measures move up (Madyan et al. 2020). And conversely, if the stock index tends to fall, this means that most of the index's constituent stocks also tend to move down. In Indonesia, there are several indices; the IDX notes that there are 36 indices that

---

\* Corresponding author.

E-mail address: [murascout@gmail.com](mailto:murascout@gmail.com)



currently exist, especially in Indonesia. Among the indexes that are widely used is the LQ45 index (Suwandi 2020). The LQ45 index is an index of a group of 45 selected stocks that meet the criteria of having liquidity, high market capitalization, high trading frequency, and relatively good growth prospects and financial conditions.

The LQ45 Index accounts for roughly 70% of trading capitalization on the Indonesia Stock Exchange (Syukur and Istiawan 2020). Therefore, the movement of the stocks included in this index has a major impact on the price of the Indonesia Composite Index (ICI) (Widodo 2018). The list of shares included in the LQ45 share list for the period August 2021 to January 2022 are Ace Hardware Indonesia Tbk (1), Adaro Energy Tbk. (2), AKR Corporindo Tbk. (3), Aneka Tambang Tbk. (4), Astra International Tbk. (5), Bank Central Asia Tbk. (6), Bank Negara Indonesia (Persero) Tbk. (7), Bank Rakyat Indonesia (Persero) Tbk. (8), Bank Tabungan Negara (Persero) Tbk. (9), Bank Mandiri (Persero) Tbk. (10), Bumi Serpong Damai Tbk (11), Bank BTPN Syariah Tbk. (12), Charoen Pokphand Indonesia Tbk. (13), Ciputra Development Tbk (14), Erajaya Swasembada Tbk. (15), XL Axiata Tbk. (16), Gudang Garam Tbk. (17), H.M. Sampoerna Tbk. (18), (19) Indofood CBP Sukses Makmur Tbk., Vale Indonesia Tbk. (20), Indofood Sukses Makmur Tbk. (21), Indah Kiat Pulp & Paper Tbk. (22), Indocement Tunggul Prakarsa Tbk. (23), Indo Tambangraya Megah Tbk. (24), Japfa Comfeed Indonesia Tbk. (25), Jasa Marga (Persero) Tbk. (26), Kalbe Farma Tbk. (27), Merdeka Copper Gold Tbk. (28), Medco Energi Internasional Tbk. (29), Mitra Keluarga Karyasehat Tbk.(30), Media Nusantara Citra Tbk. (31), Perusahaan Gas Negara Tbk.(32), Bukit Asam Tbk. (33), PP (Persero) Tbk. (34), Pakuwon Jati Tbk. (35), Semen Indonesia (Persero) Tbk. (36), Summarecon Agung Tbk. (37), Tower Bersama Infrastructure Tbk. (38), Pabrik Kertas Tjiwi Kimia Tbk. (39), Telkom Indonesia (Persero) Tbk. (40), Sarana Menara Nusantara Tbk. (41), Chandra Asri Petrochemical Tbk. (42), United Tractors Tbk. (43), Unilever Indonesia Tbk. (44), Wijaya Karya (Persero) Tbk. (45).

The list above shows 45 stocks with liquid conditions that affect the ICI index. However, in reality, the increase or decrease in the ICI was not accompanied by an increase in all 45 stocks; only a few of them showed similar action. It is important to know how far each stock is valued against the ICI index. Each increase or decrease forms its own pattern; this pattern can be a reference for investors to make entry or exit decisions from certain stocks. Pattern formation can be studied using data mining studies.

(Tukino 2019) conducted research to find out the pattern of company operations using the C4.5 method and produce rules that are useful for evaluation. This method is carried out using an analysis in the form of a decision tree using the WEKA program. The results of this study note that the performance of the method reaches 92.6%. In addition, (Cahyani and Muslim 2020) conducted research using Correlation-based Feature selection (CFS) in the C4.5 method to increase accuracy in a diagnosis or prediction. This research uses the WEKA program. The results of this study show that the performance of the method can increase accuracy by 0.5%.

## 2. Research Method and Materials

The study begins by identifying issues or cases that frequently arise while analyzing the tendency of LQ45 stocks to fluctuations in the Indonesia Composite Index, and then focuses on how the combination of CFS and Information Gain feature selection techniques are used in the application of the C4.5 algorithm in viewing the issuer's weight on the ICI movement so that it can be used as a reference for the purchase and sale of an issuer.

The second stage is literature study: we conducted a study on the relationship of the C4.5 algorithm and the application of combined techniques as a tool to be used in developing a system that is able to give weight to each variable (issuer) and formulate it into the tendency of price fluctuations (increasing, falling, constant) so that it makes it easier for investors to feel more confident in making investment decisions.

The third stage is data collection: collect information in the form of data taken from financial instruments, namely stocks, and data collection portals from yahoo.finance.com and idx.co.id. In research, stock data uses LQ45 stock data for the period August 2021 to January 2022.

The fourth stage is needs analysis: the process of analyzing and collecting data for the need to analyze stock price data that has ever existed and formulate it into a trend of rising, falling, and constant price fluctuations so that it makes it easier for investors to be more confident in making investment decisions. In the process of data analysis, users will use a combination of the C4.5 algorithm and selection techniques (CFS and Information Gain features) until the data is

generated, including what data is used. The data needed for the system to be developed is in the form of data from financial instruments (stocks).

The fifth stage is design: system design is carried out by synchronizing data or requirements that have been obtained from the needs analysis process into a programming language consisting of data structures, software architectures, procedural algorithms (C4.5, CFS, Information Gain) used in this study, and representations of the system interface.

The sixth stage is evaluation: evaluation of this system is carried out in two ways, namely testing with the black box testing method and validating data on the Rapid Miner application (Andrews, Alhaddad, and Boukhris 2019). Black box testing is carried out to determine the possibility of an error that will appear on the system and ensure that the system is working as it should (Cabana, Charrier, and Louis 2019). This is accomplished by the user analyzing the data management function and checking the possible combinations of questions and answers in the consulting function.

The seventh stage is implementation: in the previous section, a system will be implemented that is able to analyze the tendency of stock price movements and formulate them into price fluctuation tendencies of up, down, or steady using Rapid Miner. Extracting information from large-scale data sets can be done with data mining techniques. Then, data mining leads to information search activities that were previously unknown in large data groups (Lalo, Batarius, and Siki 2021). The final stage is documenting all stages of the research in a report and concluding.

### 3. Results and Discussion

#### 3.1. Results

##### a. Data Processing

Raw data is processed into data ready for processing using preprocessing and data normalization methods. Preprocessing is done to make the data more optimal when used in the classification process (Benhar, Idri, and L Fernández-Alemán 2020). The preprocessing process is carried out by first defining the data with the help of Microsoft Excel to form new data that will be ready to be processed.

The data obtained is not directly processed; it needs to be analyzed and pre-processed first so that it can represent the value of this system. The price fluctuation prediction system will be run with data containing fluctuation values (increase, decrease, stay). The determination of the ups and downs is based on the value of the previous row. In this case, ten rows of training data are taken for five variable stock issuers, namely ACES (Ace Hardware Indonesia Tbk.), ADRO (Adaro Energy Tbk.), AKRA (AKR Corporindo Tbk.), ANTM (Aneka Tambang Tbk.), ASII (Astra International Tbk.). For the variables ACES, ADRO, AKRA, ANTM, ASII as an attribute and ICI acts as a label.

**Table 1.** Row Table of Pre-processing Training Data

TIME	ACES	ADRO	AKRA	ANTM	ASII	IHSG
2015-12-30	825	515	7175	314	6000	4593
2016-01-04	800	494	7200	296	5900	4526
2016-01-05	785	499	7125	302	6125	4558
2016-01-06	755	500	7175	300	6075	4609
2016-01-07	755	485	7600	294	5975	4530
2016-01-08	745	493	7625	299	6050	4546
2016-01-11	735	486	7225	303	5950	4465
2016-01-12	750	481	7225	302	5925	4513
2016-01-13	805	473	7200	301	5900	4537
2016-01-14	790	464	7000	310	5875	4513
2016-01-15	790	449	6800	315	6000	4524

In table 1, the time information as of December 30, 2015 is used as a reference value to define the rising or falling prices contained on January 4, 2016, the starting date of the year.

Data lines before data pre-processing, as in Table 1, are then processed and analyzed. Then it is translated by eliminating the time variable (Time) in order to focus on the stocks contained in LQ45. All variables experience the addition of the word "prepro" on the front as a sign that the data has undergone data preprocessing. So it becomes prepro\_ACES, prepro\_ADRO, prepro\_AKRA, prepro\_ANTM, prepro\_ASII, and prepro\_IHSG.

The pre-processing process is carried out by forming it based on the closing price every day. If the closing price of H+1 is higher than H, it is categorized as "increase", if the closing price of H+1 is lower than H, then it is categorized as "decrease", whereas if the closing price of H+1 is the same as H, then it is categorized as "fixed". Data that was originally in the form of numbers, is processed to become a description of the condition of the stock price, namely; values "increase", "decrease," and "fixed". So as to generate fluctuation values or data ready for processing, as shown in the Table 2.

**Table 2.** Table of Rows of Data Training After Data Pre-processing

No	Prepro ACES	Prepro ADRO	prepro AKRA	prepro ANTM	prepro ASII	prepro IHSG
1	Decrease	Decrease	Increase	Decrease	Decrease	Decrease
2	Decrease	Increase	Decrease	Increase	Increase	Increase
3	Decrease	Increase	Increase	Decrease	Decrease	Increase
4	Fixed	Decrease	Increase	Decrease	Decrease	Decrease
5	Decrease	Increase	Increase	Increase	Increase	Increase
6	Decrease	Decrease	Decrease	Increase	Decrease	Decrease
7	Increase	Decrease	Fixed	Decrease	Decrease	Increase
8	Increase	Decrease	Decrease	Decrease	Decrease	Increase
9	Decrease	Decrease	Decrease	Increase	Decrease	Decrease
10	Fixed	Decrease	Decrease	Increase	Increase	Increase

The values in Table 2 represent the closing prices on each date listed. Where the ACES, ADRO, AKRA, ANT, ASII variables act as attributes and the ICI variable acts as a label.

*b. Formation of Decision Tree*

The next step in the data processing process is the formation of a decision tree, where the "Class" parameter is used by the decision tree model that will be assumed. The parameters used have a fairly high influence on the changes and movements of the ICI, and the variables selected as a reference for forming nodes are used to identify the class of each record, whether the issuer's price of a variable is "increase", "decreasing" or "fixed".

The selected variables are: (1) the "prepro\_ACES" variable to identify ACES issuer's price; (2) the "prepro\_ADRO" variable to identify ADRO's issuer's price; (3) the "prepro\_AKRA" variable to identify AKRA's issuer's price; (4) the "prepro\_ANTM" variable to identify issuer's price ANT; (5) the "prepro\_ASII" variable to identify the price of ASII issuers; and (6) the "prepro\_IHSG" variable to identify the ICI value.

Furthermore, for root selection, the gain calculation is carried out as in the following formula:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \tag{7}$$

Where S is the case set, where A is an attribute, where n is the number of partitions of A, where |Si| is the number of cases I, and with (S) is the number of cases in S. Then Gain (S,A), namely the information obtained from attribute A relative to the output data S. Furthermore, information is obtained from the output data or the dependent variable S, which is grouped based on attribute A and denoted by gain (S,A).

Meanwhile, to determine the results of Entropy (S) use the following formula:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \tag{8}$$

As for establishing a node, the steps and rules that need to be followed are as follows: 1) count the total number of cases; 2) find the gain value of each parameter; 3) calculate the gain for each remaining parameter; 4) select the gain value that is entered for the first time. Furthermore, after knowing the formula and steps for forming nodes, the data from the preprocessing results in Table 1 is processed using this formula. In this process, there are several calculation steps as follows:

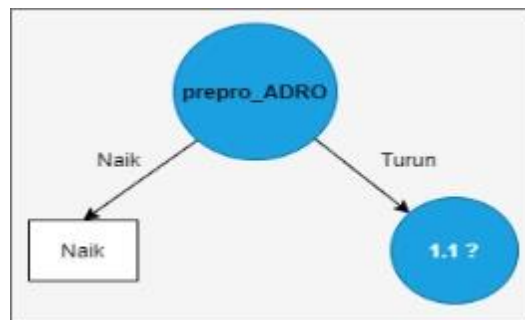
*Step 1*

*In this first calculation, the highest Gain values are in the two variables, namely prepro\_ADRO and prepro\_ASII.*

**Table 3.** Total Entropy and Information Gain Results

	Amount Case	Increase	Decrease	Entropy	Information Gain	
Total	10	6	4	0,97095059		
		prepro_ACES				0,170950594
Increase	2	2	0	0		
Decrease	6	3	3	1		
Fixed	2	1	1	1		
		prepro_ADRO				0,281290899
Increase	3	3	0	0		
Decrease	7	3	4	0,98522814		
		prepro_AKRA				0,085475297
Increase	4	2	2	1		
Decrease	5	3	2	0,97095059		
Fixed	1	1	0	0		
		prepro_ANTM				0
Increase	5	3	2	0,97095059		
Decrease	5	3	2	0,97095059		
		prepro_ASII				0,281290899
Increase	3	3	0	0		
Decrease	7	3	4	0,98522814		

From the results of these calculations, according to the decision tree rules, one attribute is selected as the root, namely the prepro\_ADRO variable. The prepro\_ADRO attribute has two values, namely "increase" and "decrease". The "increase" value in prepro\_ADRO has classified the cases into one, i.e., the decision is "increase" so no further calculations are needed. However, the "down" value still requires further calculations so that it can be classified into the next node, as shown in the figure 1.



**Fig. 1.** Decision Tree Step 1

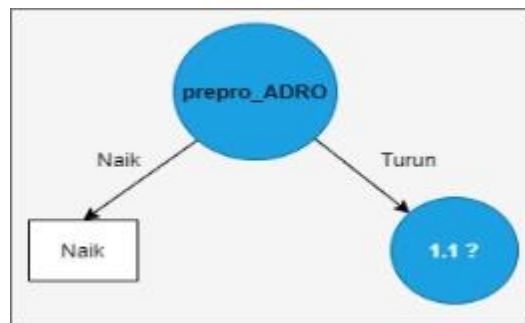
Step 2

To find out the next node, the remaining variables are calculated using the same method as in calculation 1, based on the *prepro\_ADRO* variable with a decreasing value. Here are the results of the second calculation:

**Table 4.** Entropy Results and Information Gain Node *prepro\_ADRO* Direction Down

	Amount Case	Increase	Decrease	Entropy	Information Gain	
<i>prepro_ADRO</i> , decrease	7	3	4	0,98522814		
		<i>prepro_ACES</i>				0,69951385
Increase	2	2	0	0		
Decrease	3	3	0	0		
Fixed	2	1	1	1		
		<i>prepro_AKRA</i>				0,413799565
Increase	2	0	2	0		
Decrease	4	2	2	1		
Fixed	1	1	0	0		
		<i>prepro_ANTM</i>				0,020244207
Increase	3	1	2	0,91829583		
Decrease	4	2	2	1		
		<i>prepro_ASII</i>				0,198117421
Increase	1	1	0	0		
Decrease	6	2	4	0,91829583		

The highest Gain value is in the *prepro\_ACES* variable with a value of 0.69951385. So according to the decision tree rules, *prepro\_ACES* is chosen as the next node. The *prepro\_ACES* variable has three values, namely "increase", "decrease", and "fixed". The "increase" and "decrease" values in *prepro\_ACES* have classified the cases into one, that is, both have an "increase" decision, so no further calculations are needed. However, the "fixed" value still requires more calculations so that it can be classified into the next node as shown in the following figure:



**Fig. 2.** Decision Tree Step 2

Step 3

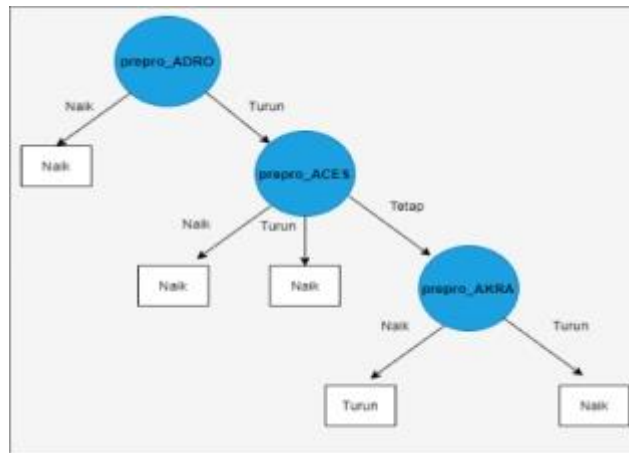
For the next selection of nodes, select the variable *prepro\_ADRO* with the value "decrease" and *prepro\_ACES* with the value "fixed". Calculations are performed on the remaining variables using the same method as those in steps 1 and 2, based on the 2 selected nodes. Following are the results of the entropy and Information Gain of the *prepro\_ADRO* node in the downward direction & *prepro\_ACES* in the fixed direction.

**Table 5.** Entropy Results and Information Gain Node prepro\_ADRO Down Direction & prepro\_ACES Fixed Direction

		Number of Cases	Increase	Decrease	Entropy	Information Gain
prepro_ADRO, decrease & prepro_ACES, fixed		2	1	1	1	
	prepro_AKRA					1
	Increase	1	0	1	0	
	Decrease	1	1	0	0	
	prepro_ANTM					0
	Increase	1	1	0	0	
	Decrease	1	0	1	0	
	prepro_ASII					0
	Increase	1	1	0	0	
	Decrease	1	0	1	0	

The highest Gain value is in the prepro\_AKRA variable with a value of 1, so according to the decision tree rules, prepro\_AKRA is chosen as the next node. The prepro\_AKRA variable has two values, namely "increase" and "decrease". The "increase" and "decrease" values in prepro\_AKRA have classified cases into 1, each of which is a down and up decision, so no further calculations are needed.

In this regard, the search for nodes was terminated because each of them had classified itself into 1 group without being further divided. So it is concluded that all cases have been classified, so other variables cannot be included in the decision tree. The following is the final decision tree figure that is formed:



**Fig. 3.** Decision Tree Step 3

3.2. Discussion

a. Implementation of C4.5 Algorithm

After finding the final results of all cases in the decision tree in Figure 3, the next step is to implement the C4.5 algorithm. However, previously the data was normalized using the Rapid Miner software. This normalization is to avoid data anomalies and data inconsistencies such as missing data or data that is out of the way of the proper value. First, you can do the Missing value in Statistics so that you can see whether there is missing data or not.

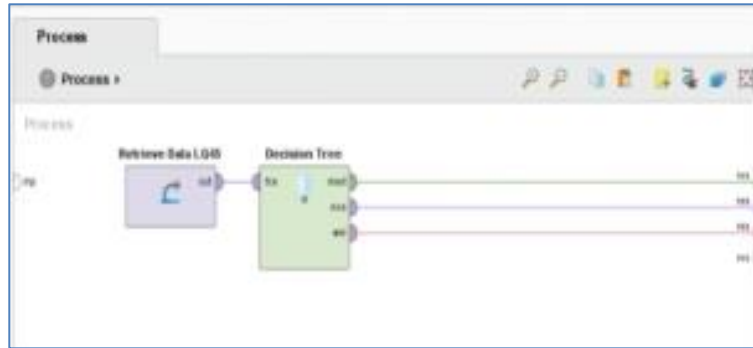


Fig. 4. Missing Value Process

Label	Frequency	Yes	No	Yes
BHIC	0	134 (1)	130	130 (78), 134 (42) [1 more]
ACES	0	148	137	137 (57), 148 (11) [1 more]
ADRO	0	100	107	107 (87), 100 (3) [1 more]
ARMA	0	106	102	102 (82), 106 (18) [1 more]
ANFM	0	144	174	174 (174), 144 (300) [1 more]
ASB	0	154	142	142 (84), 154 (302) [1 more]
BBCA	0	108	107	107 (87), 108 (10) [1 more]
BBM	0	144	107	107 (57), 144 (117) [1 more]

Fig. 5. Results of the Missing Value Process

Statistical results show that each variable does not have a missing value, so it is stated that the data is ready to be processed in feature selection towards C4.5 classification. Data ready to be processed using Rapid Miner with prior labeling. In this case there are 45 stocks, as the independent variable and the ICI as the dependent variable (Label).

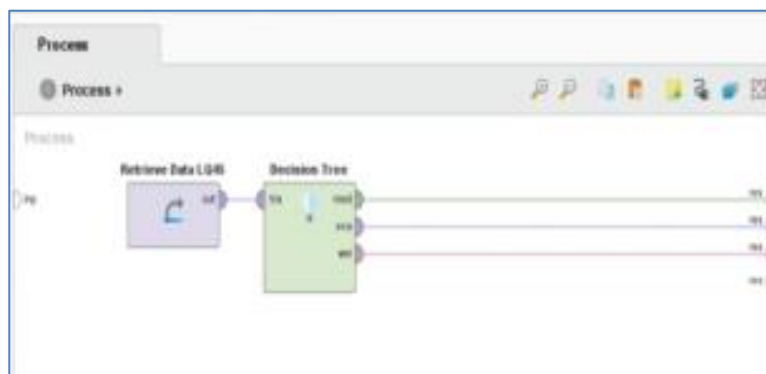


Fig. 6. Design of Rapid Miner – C4.5 (Decision Tree)

The data is retrieved, then the operator is combined with the C4.5 algorithm to form a decision tree. So as to produce the following decision tree:



**Fig. 7.** Decision Tree Results – C4.5

The results of C4.5 which are applied to LQ45 stock data, show that the BMRI variable is at the first node and TLKM, HMSP, and BBNI at the second node which have great weight in influencing the results of "Class". Furthermore, this data will be re-classified through the feature selection process.

*b. Feature Selection*

Feature selection is used to optimize which attributes have a significant and relevant influence on the data, and the selection is based on the Correlation Feature Selection (CFS) and Information Gain (IG) selection methods by taking the variable that has the 5 highest values. As for this study, an experiment was used with three selection methods, namely CFS, GI, and a combination of CFS and GI. This step is processed using Rapid Miner.

First, namely the selection method Correlation Feature Selection (CFS). This method works by minimizing a subset of the input data variables that will be used in the training and testing processes by selecting variables that are considered to have a high correlation with other variables (Chamakura and Saha 2019).

Data that is ready will be retrieved, then separated by labels through the "Set Role" operator so that the labels are not counted when categorizing data and do not affect the results. Next, the operator is combined with "Weight by Correlation" to see how much correlation each attribute has to the results given. So that the results show that the SMRA and BMRI variables have the greatest weight in influencing the results of "Class". More detail is presented in Figure 8.



**Fig. 8.** Results of Rapid Miner – Correlation Feature Selection

The second feature selection is using the Information Gain (IG) method. Where data that is ready will be retrieved, it will be separated by labels through the "Set Role" operator so that the labels are not counted when categorizing data and do not affect the results. Next, the operator is combined with the "Weight by Information Gain Ratio" to see how much weight each attribute has on the results given.

The results of the selection of the information gain feature show that the BMRI variable has the greatest weight in influencing the results of "Class". The complete data on the results of feature selection using the IG method are presented in Figure 9.

attribute	weight
BMRI	0.188
BBRI	0.173
BBRI	0.169
TLKM	0.141
AGI	0.139
BBCA	0.135
BBTH	0.131
PWON	0.124
HRSP	0.124
BBDE	0.119
WKA	0.113
BBRA	0.110
INTP	0.110
GGRM	0.106
INDP	0.106
JSMR	0.106
SMGR	0.106
PGAS	0.106

**Fig. 9.** Results of Rapid Miner – Information Gain

The third feature selection is a combination of Correlation Feature Selection (CFS) and Information Gain (IG). By combining two operators at once, namely "Weight by Correlation" and "Weight by Information Gain" to see how much weight each attribute has on the results given.

Then, the results of feature selection using the combination method of Correlation Feature Selection (CFS) and Information Gain (IG) show that the BMRI variable has the greatest weight in influencing the results of "Class".

attribute	weight
BMRI	0.188
BBRI	0.173
BBRI	0.169
TLKM	0.141
AGI	0.139
BBCA	0.135
BBTH	0.131
PWON	0.124
HRSP	0.124
BBDE	0.119
WKA	0.113
BBRA	0.110
INTP	0.110
GGRM	0.106
INDP	0.106
JSMR	0.106
SMGR	0.106
PGAS	0.106

**Fig. 10.** The Result of Combining Correlation Feature Selection and Information Gain

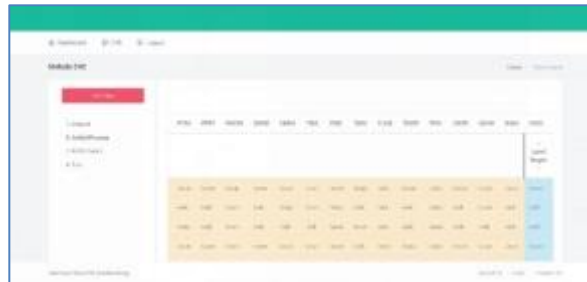
c. Stock Data Fluctuation Accuracy Level

Furthermore, this research produces an application program to make it easier to know the accuracy of stock fluctuations contained in the LQ45 index using Correlation-Based Feature Selection (CFS) and Information Gain (IG) techniques. In using this application, the first step is to fill out the login form, which will then automatically redirect to the dashboard page. Then, to see the level of accuracy of a dataset, select menu item C4.5 and input data files. In this regard, the first experiment was carried out by inputting the original data file from the LQ45 stock index.



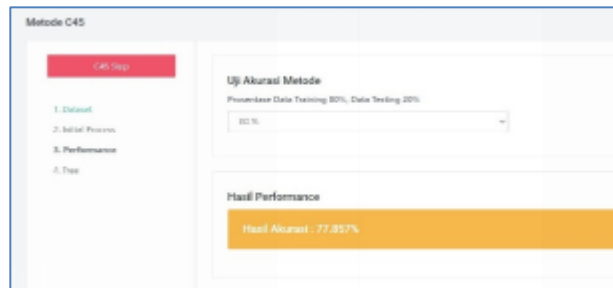
**Fig. 11.** Original Dataset Selection

In the next menu, the initialization of the input data process is shown, as shown in Figure 12, using the original LQ45 stock data as the first experiment.



**Fig. 12.** Original Data Process Initialization

Then find the results of the accuracy of the input data by selecting the "performance" menu. An example is seen in Figure 13 which shows the percentage of the original LQ45 stock data.



**Fig. 13.** Original Data Performance Results

In Figure 13, it is stated that the entire comparison of the distribution of training data and testing data that has the greatest accuracy value is in the proportion of 80% for training data and 20% for testing data, with an accuracy result of 77.857%. In addition, in the next menu, the shape of the decision tree can be seen from the input data.



Fig. 14. Original Tree Data Results

In the next stage, the use of this application is continued to find accurate results from fluctuations in LQ45 stock data using the C4.5 algorithm with the Correlation-Based Feature Selection (CFS) method, the Information Gain (IG) method, and the combination method between Correlation-Based Feature Selection (CFS) and Information Gain (IG).

In the section on using the CFS method, the data file that is input is the result of CFS feature selection using the rapid miner that was done previously. The performance results found, namely the greatest accuracy value, are training data at a proportion of 70% and testing data at a proportion of 30% with an accuracy value of 77.619%.

Furthermore, by using the IG method, the data file that is input is the data from the IG feature selection using the rapid miner that was done previously. The performance results show that training data has the highest accuracy value (78.333%) and testing data has the lowest accuracy value (70%).

Meanwhile, for the use of the CFS and IG combination method, the data file that is input is the data from the CFS and IG feature selection using the rapid miner that was done previously. The performance results state that the greatest accuracy value is training data at a proportion of 70% and testing data at a proportion of 30%, for an accuracy of 78.333%.

#### 4. Conclusion

The conclusion of this study is that the accuracy rate of stock fluctuations resulting from the original data (without selection features) is 77.857%. Then the addition of the Correlation-based Feature Selection (CFS) feature to the data actually reduced the accuracy to 77.619%. In contrast to the addition of the Information Gain feature, which has a major influence on increasing the results of data accuracy to 78.333%. Meanwhile, the addition of the Correlation-based Feature Selection (CFS) and Information Gain (IG) combination also increases the accuracy result to 78.333%. This means that the C4.5 calculation process with the combined feature selection technique of Correlation-based Feature Selection (CFS) and Information Gain affects the results of processing accuracy to be better (higher). Furthermore, in order for the results of this research to continue, it is necessary to develop the data by conducting experiments on other stock indices such as the IDX30 index, Sri Kehati, Kompas100, and so on. This can also be done by adding feature selection techniques and using other classification methods for comparison.

#### References

- Andrews, A., Alhaddad, A., & Boukhris, S. (2019). Black-Box Model-Based Regression Testing of Fail-Safe Behavior in Web Applications. *Journal of Systems and Software* 149:318–39. doi: 10.1016/j.jss.2018.11.020.
- Benhar, H., Idri, A., & Fernández-Alemán, J.L. (2020). Data Preprocessing for Heart Disease Classification: A Systematic Literature Review. *Computer Methods and Programs in Biomedicine* 195. doi: 10.1016/j.cmpb.2020.105635.
- Cabana, A., Charrier, C., & Louis, A. (2019). Mono and Multi-Modal Biometric Systems Assessment by a Common Black Box Testing Framework. *Future Generation Computer Systems* 101:293–303. doi: 10.1016/j.future.2019.04.053.

- Cahyani, N., & Muslim, M. A. (2020). Increasing Accuracy of C4.5 Algorithm by Applying Discretization and Correlation-Based Feature Selection for Chronic Kidney Disease Diagnosis. *Scientific Journal of Informatics* 2018(March):1–10.
- Chamakura, L., & Saha, G. 2019. An Instance Voting Approach to Feature Selection. *Information Sciences* 504:449–69. doi: 10.1016/j.ins.2019.07.018.
- Chirumalla, K. 2021. Building Digitally-Enabled Process Innovation in the Process Industries: A Dynamic Capabilities Approach. *Technovation* 105. doi: 10.1016/j.technovation.2021.102256.
- Junaedi, D., & Salistia, F. (2020). Dampak Pandemi Covid-19 Terhadap Pasar Modal Di Indonesia. *Al-Kharaj : Jurnal Ekonomi, Keuangan & Bisnis Syariah* 2(2):109–38. doi: 10.47467/alkharaj.v2i4.112.
- Karle, N. (2021). Debentures and Their Appeal to Investors and Companies. *International Journal of Law Management & Humanities* 4(1):1657–79.
- Lalo, Kristianus, A., Batarius, P., & Siki, Y. C. H. (2021). Implementasi Algoritma C4.5 Untuk Klasifikasi Penjualan Barang Di Swalayan Dutalia. *Jurnal Teknik Informatika UNIKA Santo Thomas* 1–12. doi: 10.54367/jtiust.v6i1.1089.
- Madyan, M., Sasikirono, N., Sumiati, & Meidiaswati, H. (2020). The Impact of Public Ownership and Share Warrants on Market Performance of IPOs: Evidence from the Indonesian Stock Exchange (IDX). *Asia Pacific Management Review* 25(4):226–34. doi: 10.1016/j.apmr.2020.05.002.
- Sari, A. P. (2019). Pengaruh Jumlah Uang Beredar Dan Pertumbuhan Ekonomi Terhadap Kinerja Reksadana Saham. *J-MAS (Jurnal Manajemen Dan Sains)* 4(2):362. doi: 10.33087/jmas.v4i2.120.
- Sucimanah, Istifani, Hariyanto, D., & Safitri, H. (2018). Pengaruh Debt to Equity Ratio, Return On Asset, Net Profit Margin Terhadap Harga Saham Perusahaan Yang Terdaftar Dalam Indeks Bisnis 27. *Jurnal Produktivitas* 5(1). doi: 10.29406/jpr.v5i1.1263.
- Suwandi, E. D. (2020). Analisis Perkembangan Bursa Efek Indonesia Dan Saham Syariah Di Indonesia Periode 2013-2018. *Jurnal Ilmiah Akuntansi Dan Keuangan* 9(2):160–69. doi: 10.32639/jiak.v9i2.626.
- Syukur, A. & Istiawan, D. (2020). Prediction of LQ45 Index in Indonesia Stock Exchange: A Comparative Study of Machine Learning Techniques. *International Journal of Intelligent Engineering and Systems* 14(1):453–63. doi: 10.22266/IJIES2021.0228.42.
- Tukino, T. (2019). Application of the C4.5 Algorithm to Predict Profits at PT SMOE Indonesia. *Jurnal Sistem Informasi Bisnis* 9(1):39.
- Widodo, W. (2018). Analisis Pengaruh Indeks Harga Saham Gabungan Regional Asia Terhadap Indeks Harga Saham Gabungan Indonesia. *EkBis: Jurnal Ekonomi Dan Bisnis* 1(2):148. doi: 10.14421/ekbis.2017.1.2.1016.