Advances in Computer Science Research, volume 96
Proceedings of the International Conference on Mathematics, Geometry, Statistics, and Computation (IC-MaGeStiC 2021)

Correlation Analysis Between the Number of Confirmed Cases of COVID-19 and Stock Trading in Indonesia

Dinagusti Sianturi* Alvida Rukmi

Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia.
Email: alvidamr@gmail.com

ABSTRACT
The COVID-19 pandemic has impact in every sector of life. Studies of the impact of the COVID-19 pandemic on stock trading are also being developed in Indonesia regarding to the number of industries affected by the pandemic. This research aims to provide information about the results of the correlation analysis between the number of confirmed cases of COVID-19 in Indonesia and the volume of stock transactions in Indonesia. From 600 stocks in Indonesia, all of them can be clustered into three cluster based on their transaction volume using K-Means clustering. Then correlation test is done between confirmed case of COVID-19 in Indonesia and the transaction volume of stocks in Indonesia synchronously. The result of this research indicate which stock transaction volumes have a positive correlation and which have a negative correlation with confirmed cases of COVID-19 in Indonesia in the appropriate time. Or it can be said that the number of confirmed cases of COVID-19 in Indonesia is increasing, does not causing stock transactions in Indonesia decrease, but stock transactions in Indonesia is also increasing.

Keywords: *Impact of COVID-19, K-Means, Pearson correlation test, Stock clustering.*

1. PRELIMINARY

Stock market is influenced by several factors, such as political, social and economy activities, public sentiment, and supply and demand from stock market participants[1] so that the stock market conditions are volatile and unpredictable. We can see public interest in a stock from the transaction volume of the stock. Transactions of a stock describe the interest of investors to buy or sell the stock[2].

COVID-19 pandemic certainly has an impact on the economic sector in Indonesia, one of which is the state of the stock market. Jakarta Composite Index (JCI) decreased in March 2020[3]. This is because many investors are nervous because of COVID-19, they sell their shares and causing stock prices to decrease[2]. However, the JCI is slowly starting to rise, which means the conditions of the Indonesian stock market is getting better.

Based on those problems, this study analyzes the impact between confirmed cases of COVID-19 and stock transaction volume in Indonesia. Based on the transaction volume of stock in Indonesia, a clustering process was carried out for these companies using the K-Means clustering method. The results of the clustering are divided into three clusters, namely high, medium, and low volume clusters. Then to find out which clusters are quite affected by confirmed cases of COVID-19, a correlation test is carried out to see the relationship between confirmed cases of COVID-19 and the volume of stock of those companies. It is hoped that this research can provide information regarding stock trading patterns during the COVID-19 pandemic.

2. RESEARCH METHODS

In this chapter, the method used in this research will be explained.

2.1. Data Source

The data used are daily data of confirmed cases of COVID-19 that taken from the https://covid19.go.id and daily data of stock transactions in Indonesia taken from https://finance.yahoo.com/. The period used for both data is from March to December 2020.

2.2. Model Analysis

At this stage, a series of methods are carried out to determine the correlation between the number of confirmed cases of COVID-19 and the volume of stock transactions in Indonesia. The first step is to do data preprocessing to get the initial data so that it is ready for use and able to get the best results. Preprocessing is done by removing unnecessary data and replacing missing values. After obtaining data that is ready to be used, the stocks in Indonesia are clustered based on the volume of stocks transactions. Before doing the clustering process, the silhouette score will be calculated to determine whether the number of clusters is optimal. After obtaining the maximum number of clusters, the clustering process was carried out using the K-Means clustering method. Stocks in Indonesia will be clustered into three clusters, namely low-volume, medium-volume, and high-volume clusters. The next step is to test the correlation between the number of confirmed cases of COVID-19 cases and the volume of stock transactions, which is to calculate the Pearson correlation coefficient and then perform a t-test.

3. RESULT AND DISCUSSION

3.1. Data Analysis

From the data that has been collected, in the data of confirmed cases of COVID-19 we have 305 rows of data with attribute date that contain the date and cases that contain the number of confirmed cases on that date. In this research, what is meant by the number of confirmed cases is the number of people who have tested positive for the corona virus based on the results of medical check.

Meanwhile, the stock transaction data contains 200 lines of data with the attribute date, open, close, high, low, adj close, and volume. Date contains the date of the stock transaction. Open, close, high, low each contains opening, closing, highest, and lowest prices of the stock. Volume contains the number of stocks traded on that date.

3.2. Data Pre-Processing

There is a difference in the number of data lines in the confirmed case data of COVID-19 cases and stock transaction data, where data of confirmed cases of COVID-19 has 305 lines of data while stock transaction data has 200 lines of data. The number of data lines in stock transaction data is less than the number of data lines in confirmed cases of COVID-19 due to stock trading not being carried out every day, while in Indonesia there are confirmed cases of COVID-19 in every day. Stock market in Indonesia is open from Monday to Friday[3] and the stock market is closed on Saturdays, national holidays, and public holidays. Due to the different number of rows of data, 105 rows of data were deleted from the data of confirmed cases of COVID-19. Deleted data is data on the date on which there was no stock transaction, that is data on confirmed cases of COVID-19 recorded on Saturdays, national holidays, and public holidays. Then the attributes of open, close, high, low, and adj, close on stock transaction data is removed. This is done because in the next stage, the attributes used from the stock transaction data are the date and volume attributes.

The next step is the handling the missing values. Missing values is the number 0 contained in the stock transaction data. Existing missing values will be replaced with mean[4] of the variable where the missing value is located. For example, in the stock transaction volume of AALI there is a number 0 then it is replaced with the average value of the stock transaction volume of AALI. Then in the ABBA stock transaction volume there is a number 0 and it is replaced with the average value of the ABBA stock transaction volume. In the same way, the number 0 is replaced in the transaction volume of other stocks. The process of replacing missing values is only carried out on stock transaction volume data, because in the COVID-19 case data the number 0 means there are no confirmed cases of COVID-19 or in other words the number 0 is not missing value.

After deleting 105 rows of confirmed cases of COVID-19 data, deleting unused attributes, and replacing existing missing values, the confirmed COVID-19 case data and stock transaction data both have the same number of data lines, that is 200 data lines and there are no missing values so that the data is already ready to use. The data that is ready for use after we do the preprocessing step is shown as in Table 1.

Table 1. Data after pre-processing

Date	Case	Stock Transaction Volume			
		AALI	ABBA	...	ZONE
2/3/2020	2	885500	1066400	...	3300
3/3/2020	0	1109600	59650800	...	100
4/3/2020	0	1050800	13327400	...	147732.5
5/3/2020	0	1220200	22350700	...	23100
6/3/2020	2	373600	2189300	...	75900
...	...	...	...	...	...
12/29/2020	7903	3184200	473192900	...	109800
12/30/2020	8002	3630900	282844000	...	109800

In Table 1, *Date* contains date, *Case* contains the number of confirmed cases of COVID-19, and *AALI*, *ABBA*, and so on until *ZONE* contains the transaction volume of stocks of AALI, ABBA, and so on until ZONE.

3.3. Stock Clustering Based on Transaction Volume

By using transaction volume data that has passed the previous preprocessing stage, 600 stocks in Indonesia are clustered into three clusters. Silhouette score is calculated to determine whether the number of clusters is optimum to obtain the optimum cluster as well. Using the jupyter notebook software and the Python programming language, the silhouette score obtained is 0.32. Based on the interval table for the interpretation of silhouette score[5], the value cannot be said to be good and requires additional methods. Therefore, the authors look for a measure of data concentration from each stock to then carry out the clustering process.

The stock transaction data obtained are very diverse. For example, on AALI stock, the transaction volume is in the range of 373,600 to 11,900,200. Then on ABBA stock, the transaction volume is in the range of 500 to 473,192,900. Because the data varies greatly, the measure of data concentration that is more suitable to be used is the median[6]. To find the median value, the data needs to be sorted from lowest to highest. The median lies in the middle data. In this study, because the number of data rows is 200, the median value is obtained by adding up the values in the 100th and 101st data rows and then dividing by two.

After getting the median value of each stock transaction volume and by ignoring the curse of dimensionality phenomenon where this phenomenon makes the information search process difficult[7] because there will be a lot of missing data, in this study a clustering process was done using the K-Means clustering method. Before clustering, the silhouette score will be calculated using the median data that has been obtained. For the number of clusters is three, the silhouette value obtained is 0.886. By consider the silhouette score interpretation table[5], the value obtained is good and the number of clusters as much as 3 can be used. This value is also much better when compared to the silhouette value obtained when performing the clustering process using the entire stock transaction data.

Furthermore, the stock clustering process is done into three clusters using the K-Means clustering method. The results of clustering are shown in Table 2.

From Table 2 it can be seen that stocks in cluster 0 have a lower median value than stocks in cluster 1, and stocks in cluster 2 have a higher median value than stocks in cluster 2 cluster 1. It can be concluded that cluster 0 is a low-volume cluster, cluster 1 is a medium-volume cluster, and cluster 2 is a high-volume cluster.

Table 2. Stock clustering results

Cluster	Median of Transaction Volume	Stock Name	Number of Stocks
0	117	MORE	557
	207,5	FMII	
	...	...	
	31.104.000	SSIA	
1	33.328.550	PTBA	36
	34.496.200	CTRA	
	...	...	
	107.833.450	TOWR	
2	133.706.400	PGAS	7
	139.172.900	TLKM	
	...	...	
	185.899.150	ZINC	

The stocks in the medium-volume cluster are ADRO, AGRO, ANTM, ASII, ASRI, BBKP, BBNI, BBTN, BEST, BMRI, BMTR, BOGA, BRIS, BRPT, BSDE, BWPT, CTRA, DMAS, DOID, ELSA, HMSP, HOKI, KLBF, MDKA, MEDC, MNCN, NATO, PTBA, PTPP, PURA, PWON, SMRA, TBIG, TOWR, WSBP, and WSKT. Meanwhile, stocks in high-volume clusters are BBRI, BULL, CARE, FREN, PGAS, TLKM, and ZINC.

3.4. Calculation of the Correlation Coefficient between the Number of Confirmed COVID-19 Cases and the Volume of Stock Transactions

At this stage, the correlation coefficient is calculated between the number of confirmed cases of COVID-19 and the transaction volume of stocks in medium-volume and high-volume clusters. The method used to calculate the correlation coefficient between the number of confirmed cases of COVID-19 and the volume of stock transactions is the Pearson test method. The formula to get the correlation coefficient is

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

In this study, the x variable is the number of confirmed cases of COVID-19 while the y variable is the volume of stock transactions. Take, for example, to calculate the correlation coefficient between the number of confirmed cases of COVID-19 and the transaction volume of BULL shares, we get

$$r_{xy} = \frac{25.707.249.712.449}{\sqrt{839.386.594}\sqrt{1.214.124.770.293.180.000}} \quad (2)$$

$$r_{xy} = 0{,}805273 \quad (3)$$

In the same way, a correlation coefficient is obtained between the number of confirmed cases of COVID-19 and the transaction volume of other stocks in the medium-volume and high-volume clusters. The results of the calculation of the correlation coefficient are shown in Table 3.

Table 3. Correlation coefficient between the number of confirmed cases of covid-19 and the volume of stock transactions in medium-volume and high-volume clusters

Stock Name	Correlation Coefficient	Stock Name	Correlation Coefficient
ADRO	0.467318	ELSA	0.378482
AGRO	0.425012	FREN	0.504953
ANTM	0.558946	HMSP	0.433405
ASII	-0.006715	HOKI	0.203953
ASRI	0.457105	KLBF	-0.079204
BBKP	0.465698	MDKA	0.055044
BBNI	-0.144073	MEDC	0.423462
BBRI	-0.331552	MNCN	0.173531
BBTN	-0.104203	NATO	0.762665
BEST	0.494568	PGAS	0.236263
BMRI	-0.249394	PTBA	0.046584
BMTR	0.410713	PTPP	0.106434
BOGA	0.554404	PURA	0.753115
BRIS	0.327478	PWON	0.004885
BRPT	0.125576	SMRA	0.246827
BSDE	0.101013	TBIG	-0.164439
BULL	0.805273	TLKM	0.367285
BWPT	0.341469	TOWR	0.169186
CARE	0.784804	WSBP	0.486141
CTRA	0.043490	WSKT	0.462392
DMAS	0.243897	ZINC	0.835581
DOID	0.394845		

3.5. Significance Test of Correlation Test Results

To test whether the correlation coefficient obtained is significant or can represent the population, a t-test is carried out with the following steps:

Determining the hypothesis:

H_0 : there is no correlation between the number of confirmed cases and the volume of stock transactions.

H_1 : there is a correlation between the number of confirmed cases and the volume of stock transactions

Test statistics:

$$t_{count} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}{}^2}} \quad (4)$$

$$t_{table} = t_{\frac{\alpha}{2},198} = 1{,}9712 \quad (5)$$

Test criteria:

If $|t_{count}| < t_{table}$ so H_0 is accepted, meanwhile if $|t_{count}| > t_{table}$ so H_0 is rejected. Take, for example, to test the significance of the correlation coefficient between confirmed cases of COVID-19 and the stock transaction volume of BULL, we get:

$$t_{count} = \frac{0{,}805273\sqrt{198}}{\sqrt{1-0{,}805273^2}} = 19{,}111 \quad (6)$$

Because the score of $|t_{count}| > t_{table}$ so H_0 is rejected, or it can be said that there is a correlation between confirmed cases of COVID-19 and the transaction volume of BULL shares. In the same way, the value of $|t_{count}|$ from the correlation coefficient that has been obtained previously is obtained. The results of the t-test of the correlation coefficient between confirmed cases of COVID-19 and the transaction volume of stocks in the medium and high volume clusters are shown in Table 4.

In Table 4, r_{ADRO} is the correlation between confirmed cases of COVID-19 and stock transaction volume of ADRO, r_{AGRO} is the correlation between confirmed cases of COVID-19 and stock transaction volume of AGRO, and so on until r_{ZINC} is the correlation between confirmed cases of COVID-19 and stock transaction volume of ZINC. It can be seen that after conducting a significance test, as many as 10 stocks in the medium volume cluster have no correlation with the number of confirmed cases of COVID-19 (H0 is accepted).

3.6. Analysis of Correlation Test Results

The results of the calculation of the correlation coefficient and the significance test in the previous stage show that the number of confirmed cases of COVID-19 in Indonesia has a correlation with the transaction volume of all stocks in the high-volume cluster and most of the stocks in the medium-volume cluster. The number of confirmed cases of COVID-19 has no correlation with the transaction volume of 10 stocks in the medium volume cluster.

The correlation coefficient is in the range -1 to 1. If the correlation coefficient is getting closer to the number -1 or 1 then the correlation is getting stronger and if the correlation coefficient is getting closer to 0 then the correlation is getting weaker. If the correlation coefficient is negative, then the correlation obtained is a negative correlation. However, if the correlation coefficient is positive, then the correlation obtained is a positive correlation. The meaning of negative correlation is that when confirmed cases of COVID-19 decrease,

transaction volume increases as well as when confirmed cases increase, transaction volume decreases. While the positive correlation means that when confirmed cases increase, the transaction volume also increases and vice versa.

Table 4. Correlation coefficient t-test results

Correlation	$\|t_{count}\|$	t_{table}	Testing Criteria
r_{ADRO}	7.437877	1,9712	H_0 is rejected
r_{AGRO}	6.606862	1,9712	H_0 is rejected
r_{ANTM}	9.485085	1,9712	H_0 is rejected
r_{ASII}	0.094492	1,9712	H_0 is accepted
r_{ASRI}	7.231777	1,9712	H_0 is rejected
r_{BBKP}	7.404931	1,9712	H_0 is rejected
r_{BBNI}	2.048662	1,9712	H_0 is rejected
r_{BBRI}	1.474295	1,9712	H_0 is rejected
r_{BBTN}	8.006989	1,9712	H_0 is accepted
r_{BEST}	3.623791	1,9712	H_0 is rejected
r_{BMRI}	6.338518	1,9712	H_0 is rejected
r_{BMTR}	9.373611	1,9712	H_0 is rejected
r_{BOGA}	4.876947	1,9712	H_0 is rejected
r_{BRIS}	1.781108	1,9712	H_0 is rejected
r_{BRPT}	1.428691	1,9712	H_0 is accepted
r_{BSDE}	5.112178	1,9712	H_0 is accepted
r_{BULL}	0.612535	1,9712	H_0 is rejected
r_{BWPT}	3.538801	1,9712	H_0 is rejected
r_{CARE}	6.047317	1,9712	H_0 is rejected
r_{CTRA}	5.753739	1,9712	H_0 is accepted
r_{DMAS}	6.767157	1,9712	H_0 is rejected
r_{DOID}	2.931488	1,9712	H_0 is rejected
r_{ELSA}	1.118013	1,9712	H_0 is rejected
r_{FREN}	0.775720	1,9712	H_0 is rejected
r_{HMSP}	6.577485	1,9712	H_0 is rejected
r_{HOKI}	2.479407	1,9712	H_0 is rejected
r_{KLBF}	16.592087	1,9712	H_0 is accepted
r_{MDKA}	0.656209	1,9712	H_0 is accepted
r_{MEDC}	1.506221	1,9712	H_0 is rejected
r_{MNCN}	16.108013	1,9712	H_0 is rejected
r_{NATO}	0.068741	1,9712	H_0 is rejected
r_{PGAS}	3.584053	1,9712	H_0 is rejected
r_{PTBA}	2.345787	1,9712	H_0 is accepted
r_{PTPP}	2.415485	1,9712	H_0 is accepted
r_{PURA}	7.827861	1,9712	H_0 is rejected
r_{PWON}	7.338009	1,9712	H_0 is accepted
r_{SMRA}	4.945065	1,9712	H_0 is rejected
r_{TBIG}	19.111318	1,9712	H_0 is rejected
r_{TLKM}	17.818920	1,9712	H_0 is rejected
r_{TOWR}	8.231886	1,9712	H_0 is rejected
r_{WSBP}	3.421373	1,9712	H_0 is rejected
r_{WSKT}	5.556507	1,9712	H_0 is rejected
r_{ZINC}	21.402176	1,9712	H_0 is rejected

Next, it will be seen how strong the correlation between the number of confirmed cases of COVID-19 and the transaction volume of stocks in the medium and high volume clusters is by paying attention to the correlation coefficient interpretation table[8]. The type and strength of the correlation between the number of confirmed cases of COVID-19 in Indonesia and the volume of stock transactions in the medium-volume and high-volume clusters are shown in Table 5.

In this study, stock clustering was carried out to see how the correlation between confirmed cases of COVID-19 and the volume of stock transactions in Indonesia was. From the results of the study, it was found that the confirmed cases of COVID-19 had a correlation with all stocks in the high-volume cluster and most of the stocks in the medium-volume cluster. From Table 5, it can be seen that from the forty-three stocks included in the medium-volume and high-volume cluster, twenty-nine stocks have a positive correlation with confirmed cases of COVID-19, four stocks have a negative correlation with confirmed cases of COVID-19, and ten stocks has no correlation with confirmed cases of COVID-19.

In daily report, the number of confirmed cases of COVID-19 tends to increase as can be seen in Figure 1. According to the results of the correlation test, it was found that several stocks in the medium- and high-volume clusters had a positive correlation with confirmed cases of COVID-19 in Indonesia, then this also means that the transaction volume of stocks that have a positive correlation with confirmed cases of COVID-19 also tends to increase. However, the correlation between confirmed cases of COVID-19 and the transaction volume of several stocks in Indonesia does not always mean that there is cause and effect.

Even though at the beginning of the COVID-19 pandemic in Indonesia, investor sentiment towards the stock market was low, causing the stock market to tend to fluctuate in a negative direction, but since early April 2020, stock market conditions began to improve

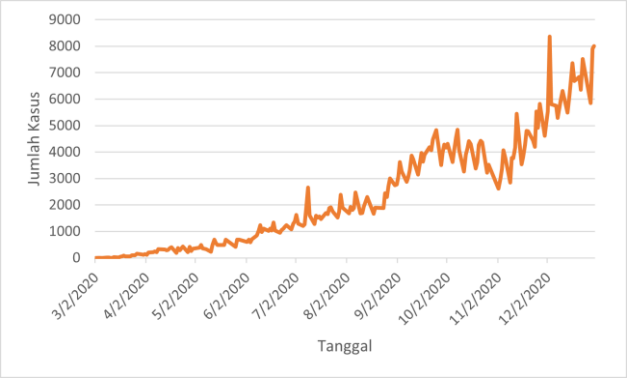

Figure 1 Growth chart for COVID-19 confirmed cases in Indonesia.

Table 5. Type and strengthens of correlation between number of confirmed cases of covid-19 and transaction volume of shares in medium-volume and high-volume clusters

Stock Name	Correlation Type	Correlation Strength
ADRO	positive correlation	moderate
AGRO	positive correlation	moderate
ANTM	positive correlation	moderate
ASII	no correlation	
ASRI	positive correlation	moderate
BBKP	positive correlation	moderate
BBNI	negative correlation	very weak
BBRI	negative correlation	weak
BBTN	no correlation	
BEST	positive correlation	moderate
BMRI	negative correlation	weak
BMTR	positive correlation	moderate
BOGA	positive correlation	moderate
BRIS	positive correlation	weak
BRPT	no correlation	
BSDE	no correlation	
BULL	positive correlation	very strong
BWPT	positive correlation	weak
CARE	positive correlation	strong
CTRA	no correlation	
DMAS	positive correlation	weak
DOID	positive correlation	weak
ELSA	positive correlation	weak
REN	Positive correlation	moderate
HMSP	positive correlation	moderate
HOKI	positive correlation	weak
KLBF	no correlation	
MDKA	no correlation	
MEDC	positive correlation	moderate
MNCN	positive correlation	very weak
NATO	positive correlation	strong
PGAS	Positive correlation	weak
PTBA	no correlation	
PTPP	no correlation	
PURA	positive correlation	strong
PWON	no correlation	
SMRA	positive correlation	weak
TBIG	negative correlation	very weak
TLKM	Positive correlation	weak
TOWR	positive correlation	very weak
WSBP	positive correlation	moderate
WSKT	positive correlation	moderate
ZINC	very strong	very strong

4. CONCLUSIONS AND SUGGESTIONS

Based on the results of this study, it can be concluded as follows:

1. Based on the silhouette score, the number of good clusters to cluster stocks based on transaction volume is three. And in this study, stock clusters based on transaction volume are low volume clusters, medium volume clusters and high volume clusters..
2. The Pearson correlation test shows a correlation or relationship between positive confirmed cases of COVID-19 and the transaction volume of several stocks in Indonesia.

As for suggestions that can be made in further related research is to use data with a longer period of time. Then it would be better if in further research a correlation test was carried out on stocks belonging to the small stock volume cluster. In further related research, dimension reduction can also be carried out using principal component analysis to avoid the phenomenon of the curse of dimensionality.

AUTHORS' CONTRIBUTIONS

Dinagusti, collected and contributed data from the resident cluter, proccesed data using K-Means to make clustering, analyzed correlation between covid-19 data and stock transaction data.

Alvida, conceived and designed the analysis result of clustering, supported references and edited writing in paper

ACKNOWLEDGMENTS

The research is supported by the Departement of Mathematics, Institute Teknologi Sepuluh Nopember, Ministry of Education and Culture.

REFERENCES

[1] S. Tabar, S. Sharma, D. Volkman, A new method for predicting stock market crashes using classification and artificial neural networks, International Journal of Business and Data Analytics, 1(3), 2020, pp 203-217.

[2] K.A.P. Kasuma, Y.D. Nugroho, Tinjauan kasus Ttrkonfirmasi positif COVID-19 terhadap iklim investasi di Indonesia : Peramalan dan Korelasi, (in Indonesian), Seminar Nasional Official Statistics, Vol. 1, 2020, pp. 190-195.

[3] Bursa Efek Infonesia, Jam Perdagangan, https://www.idx.co.id/investor/jam-perdagangan/, (accesed June, 26. 2021).

[4] A.C. Acock, What to do about missing values, 2012

[5] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of computational and applied mathematics, 20, 1987, pp 53-65.

[6] L.D. Martias, Statistika deskriptif sebagai kumpulan informasi, (in Indonesian), Fihris: Jurnal Ilmu Perpustakaan dan Informasi, 16(1), 2021, pp 40-59.

[7] M. Sarmilah, A. Adiwijaya, A. Atiqi, Analisis seleksi fitur genetic algorithm dan ekstraksi fitur wavelet pada klasifikasi microarray data menggunakan naïve bayes, (in Indonesian), eProceedings of Engineering, 5(1), 2018.

[8] S. Sugiyono, Metode penelitian kuantitatif dan kualitatif, (in Indonesian), Alfabeta Bandung, 2010.