Computer Science and Information Technologies
Vol. 5, No. 3, November 2024, pp. 243~253
ISSN: 2722-3221, DOI: 10.11591/csit.v5i3.p243-253

Company clustering based on financial report data using k-means

Gusti Aditya Aromatica Firdaus, Lili Ayu Wulandhari
Department of Computer Science, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Apr 17, 2024
Revised Jul 27, 2024
Accepted Sep 8, 2024

Keywords:

Clustering
Financial report data
K-means
Machine learning
Persona analysis

ABSTRACT

Stock investment is the act of providing funds or assets to obtain future payments for gifts given. In its application, novice investors often make mistakes, one of which is not knowing the health condition of the company they want to target. By applying the machine learning clustering method based on company financial report data, it was found that 2 clusters were formed. This can show the current condition of the company so that it can be a consideration for investors, such as clusters of companies that have a profit trend that is always stable and increasing, or clusters of companies that are in the process of developing their business and groups of companies that have large amounts of debt from year to year.

Corresponding Author:

Gusti Aditya Aromatica Firdaus
Department of Computer Science, BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia
Email: gusti.firdaus001@binus.ac.id

1. INTRODUCTION

Stock investment is the act of providing funds or assets to obtain future payments for the gifts that have been made [1]. This investment states that the individual or business entity that makes their investment will obtain partial ownership of the company. In practice, novice investors often make mistakes, including not doing enough research before buying shares, not having a strategy when investing, and not understanding the risks of investing in shares. Therefore, several things need to be considered when buying shares, namely conducting research about the company, understanding the company's financial reports, determining the objectives of investing, and risk tolerance [2], [3]. However, it is still very difficult for ordinary people to find out the current condition of a company if they only pay attention to the financial reports. With the help of a model, it will be easier to identify whether the company's condition is good or not.

Nowadays, technology is developing very quickly, one of which is the emergence of the concept of artificial intelligence (AI) [4]-[6]. The existence of artificial intelligence in the current era has provided many benefits in various fields. One form of utilizing AI is stock analysis, which is important in making decisions when making investments. Knowing the condition of a company using machine learning (ML) has been done by researchers before. AI has an impact on the financial sector, one of which is stock analysis, which is important in making decisions when making investments.[7]. The development of ML has made it easier to analyze data and identify patterns or trends that can be used to make investment decisions. Apart from that, the application of ML can be implemented in stock analysis to predict company performance. Stock analysis involves evaluating specific trading instruments, investment sectors or the market as a whole [2]. Fundamental

analysis is a method often used in the stock analysis process, which determines the real or "fair market" value of shares by examining related economic and financial factors. This analysis considers the intrinsic value of an investment, which is based on the issuing company's finances and current market and economic conditions [3], [8]-[10].

Nowadays, technology is developing rapidly, one of which is the emergence of the concept of AI. The existence of AI in the current era has provided many benefits in various fields [11]. One of them is the use of the industrial and stock sectors. In some cases that have been applied, including decision making when making investments [12], [13], clusters of financial data and others. The cluster itself is a method that aims to identify the basic patterns or structures of data [14] one of the studies that has been carried out and developed to the stage of density-based clustering and ensemble data clustering [15].

The use of machine learning in clustering methods is very much in the industrial sector and other sectors. The research conducted is related to the analysis of hierarchical relationships and clustering that have occurred in the capital market over the past few years [16]. Apart from this, research shows the use of machine learning, namely the k-algorithm. Means has succeeded in helping the financial industry [17]. The use of the k-means clustering method is also supported for use as a clustering method based on a method review carried out by researchers from Australia [18]. In other studies, clusters have been conducted for the detection and optimization of financial data [19]. In other studies using k-means, it was also applied to cluster networks so that predictions could be made as a follow-up [20]. In addition, several further studies were also conducted related to the application of clustering such as for fraud detection clusters using k-means [21]. Build a clustering model for Heat Transfer Fluid control [22]. Cluster model for cancer disease experienced [23]. Build mode machine learning clustering for the basis of further research in terms of predicting stock prices [24]. Research to form clusters in terms of education [25]. There was also research conducted in 2018 to continue previous research to identify algorithm suitability to get better results than previous research [26].

The evaluation used to determine the optimal k value for the k-means model will be based on a silhouette score, as outlined in research conducted in 2020 by researchers from Baltimore [27]. In addition to the Baltimore study, there is also research conducted by Shutaywi and Kachouie [28]. Further studies have been conducted that are related to this area, including those that use the silhouette score to provide research recommendations for future studies [29]. These studies include research on assessing the quality of cluster results [27], building cluster models on existing data [30], clustering in the context of power generation capacity [31], finding the best cluster from high-level model clustering [32], and determining the optimal number of clusters [33], as well as identifying the best cluster with the formed number of clusters [34]. Based on this background and the related studies, this research will focus on clustering companies based on financial data reports published from 2018 to 2022 using the k-means method, aiming to identify the clusters formed and assess the condition of the companies based on the cluster analysis conducted. The remaining sections of this work are structured as follows: in section 2, we explain the formation of the dataset and the clustering method to produce the clusters that will be analyzed, in section 3 we explain the results of the analysis that has been carried out on the clusters that have been obtained using the method mentioned in the method. Previously and provide conclusions on the research that has been carried out in section 4.

2. METHOD

In this research, the observations made were collecting data from the internet by collecting manually from several sources, either from the Indonesian Stock Exchange website or the company's website. After obtaining data from the company's financial reports, of course the data needs to be processed first before being used as a dataset for developing the Clustering model. Figure 1 shows the flow of data processing carried out in this research. The company's financial report data collected will be checked, where if it is less than 3 years then the data will not be used, while the available financial reports are the same or more than 3 years to 5 years then it will proceed to the data processing. Examples of financial reports from companies whose data will be used are in Figure 2.

From the financial report as in the Figure 2, several variables will be taken, namely debt ratio, profit, book value per share (BVPS), and earning per share. Each variable selected is a variable that can indicate the quality of a company [35], [36]-[40]. Some of these parameters are not written in the financial statements so they need to be calculated first, including BVPS using (1).

$$BVPS = \frac{NE}{SB} \tag{1}$$

Where, BVPS is book value per share, NE is equity value, and SD is outstanding shares.

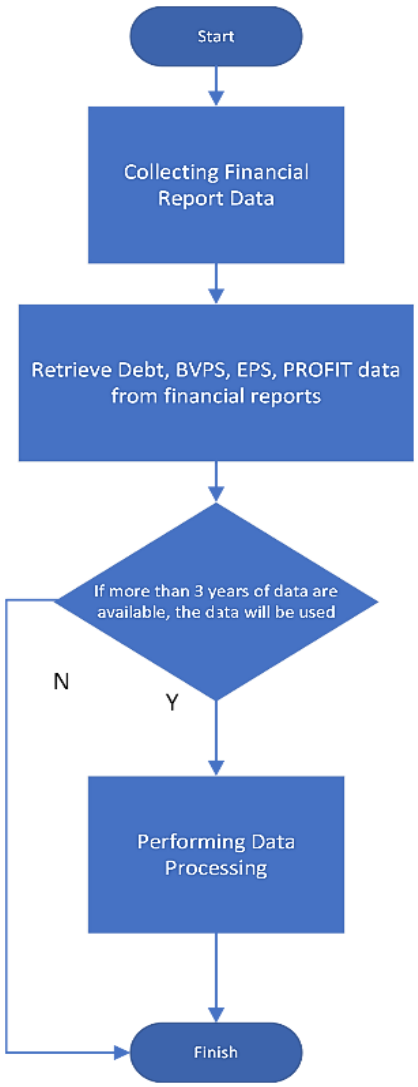


Figure 1. Data processing flow

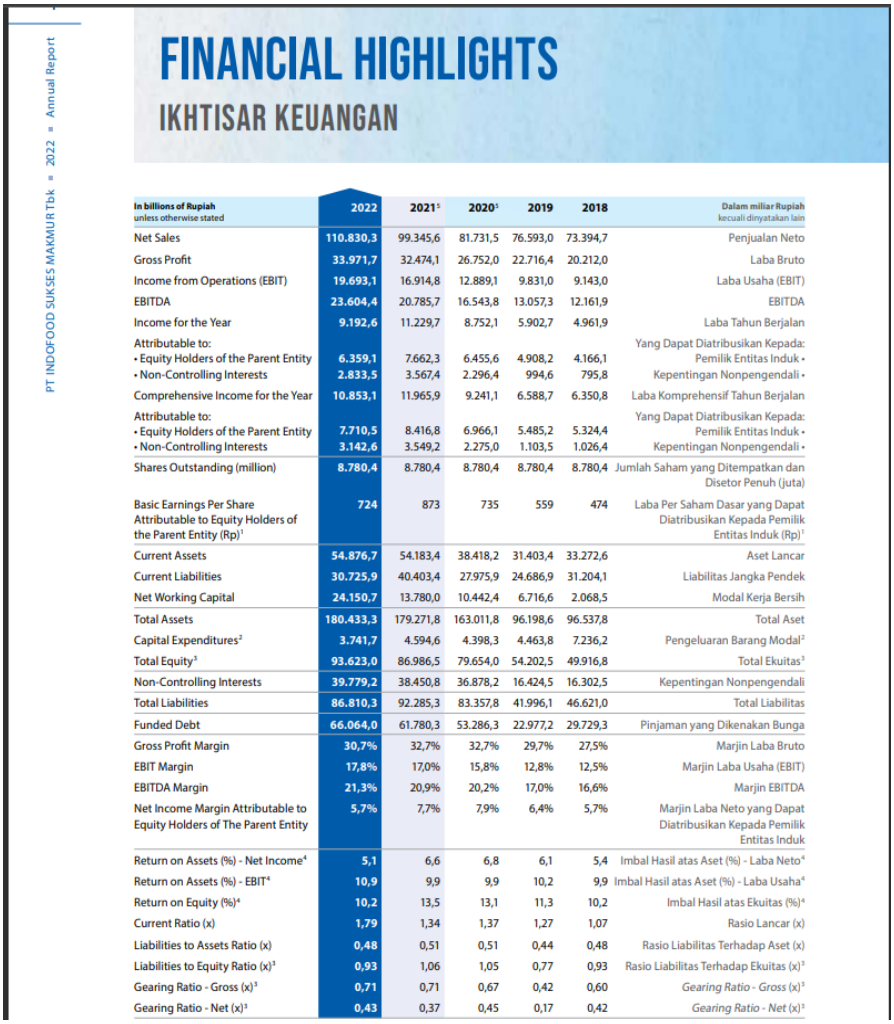
In Table 1 is a data sample that has been calculated after collecting data manually. After data collection, the first selection was carried out by paying attention to companies that had available data from at least 2020 to 2022 and a maximum from 2018. After selecting the data, a total of 35 companies were obtained in the Consumer Industry sector. The process continues with data processing so that all financial report data has the same nominal units.

The research carried out was to create an appropriate model for clustering companies based on their performance. This research certainly requires data from financial reports. Apart from the variables from the financial report data, this research will use the architecture, namely k-means. The use of this architectural model is to get the best classification model results and find out which algorithm is more suitable to get a model with the best evaluation performance. Of course, there are several things that will be done as a general description of the design of the research that will be carried out. The design of the clustering model in Figure 3 for the company is based on the company's performance. In the process of creating a data model that has been formed, processing is carried out by entering the data pre-processing stage. At this stage, data processing is carried out starting from checking for null data contained in the data and continuing with data normalization by changing the scale of the data contained in the dataset used. The first thing to do is check for null values, which from the results shows in Table 2 that many columns have null values or "0" in them.

As in the Figure 3, you can see that several columns have the status "true", which indicates that the column has a value of null or "0". After checking, continue with checking outlier values to determine the appropriate way to overcome "missing handling value". After checking for outliers, it was found that 7 variables had no outliers and 28 variables had outliers. After finding outliers, we proceed with handling missing

values using 2 methods. The first way uses the average value to fill in the null value or "0", the second way uses the medium value to fill in the null value or "0". The next stage in data preprocessing is to carry out standard scaling using the "Standardscaler()" method.

After preprocessing the data, the data is continued with application using the k-means method. In its implementation, apart from using the simple k-means method, it is also combined with another method, namely using the silhouette score method. This method is used to determine the number of clusters to be formed and the application of the silhouette score to determine the optimal number of clusters to be formed. In the process, we were given 10 cluster shape trials and obtained 2 clusters with the best silhouette value, namely 0.64 as in Table 3. After getting the optimal cluster, we continued with clustering data conversion to carry out persona analysis.



PT INDOFOOD SUKSES MAKMUR Tbk ▪ 2022 ▪ Annual Report

FINANCIAL HIGHLIGHTS

IKHTISAR KEUANGAN

In billions of Rupiah unless otherwise stated	2022	2021[5]	2020[5]	2019	2018	Dalam miliar Rupiah kecuali dinyatakan lain
Net Sales	110.830,3	99.345,6	81.731,5	76.593,0	73.394,7	Penjualan Neto
Gross Profit	33.971,7	32.474,1	26.752,0	22.716,4	20.212,0	Laba Bruto
Income from Operations (EBIT)	19.693,1	16.914,8	12.889,1	9.831,0	9.143,0	Laba Usaha (EBIT)
EBITDA	23.604,4	20.785,7	16.543,8	13.057,3	12.161,9	EBITDA
Income for the Year	9.192,6	11.229,7	8.752,1	5.902,7	4.961,9	Laba Tahun Berjalan
Attributable to:						Yang Dapat Diatribusikan Kepada:
• Equity Holders of the Parent Entity	6.359,1	7.662,3	6.455,6	4.908,2	4.166,1	Pemilik Entitas Induk •
• Non-Controlling Interests	2.833,5	3.567,4	2.296,4	994,6	795,8	Kepentingan Nonpengendali •
Comprehensive Income for the Year	10.853,1	11.965,9	9.241,1	6.588,7	6.350,8	Laba Komprehensif Tahun Berjalan
Attributable to:						Yang Dapat Diatribusikan Kepada:
• Equity Holders of the Parent Entity	7.710,5	8.416,8	6.966,1	5.485,2	5.324,4	Pemilik Entitas Induk •
• Non-Controlling Interests	3.142,6	3.549,2	2.275,0	1.103,5	1.026,4	Kepentingan Nonpengendali •
Shares Outstanding (million)	8.780,4	8.780,4	8.780,4	8.780,4	8.780,4	Jumlah Saham yang Ditempatkan dan Disetor Penuh (juta)
Basic Earnings Per Share Attributable to Equity Holders of the Parent Entity (Rp)[1]	724	873	735	559	474	Laba Per Saham Dasar yang Dapat Diatribusikan Kepada Pemilik Entitas Induk (Rp)[1]
Current Assets	54.876,7	54.183,4	38.418,2	31.403,4	33.272,6	Aset Lancar
Current Liabilities	30.725,9	40.403,4	27.975,9	24.686,9	31.204,1	Liabilitas Jangka Pendek
Net Working Capital	24.150,7	13.780,0	10.442,4	6.716,6	2.068,5	Modal Kerja Bersih
Total Assets	180.433,3	179.271,8	163.011,8	96.198,6	96.537,8	Total Aset
Capital Expenditures[2]	3.741,7	4.594,6	4.398,3	4.463,8	7.236,2	Pengeluaran Barang Modal[2]
Total Equity[3]	93.623,0	86.986,5	79.654,0	54.202,5	49.916,8	Total Ekuitas[3]
Non-Controlling Interests	39.779,2	38.450,8	36.878,2	16.424,5	16.302,5	Kepentingan Nonpengendali
Total Liabilities	86.810,3	92.285,3	83.357,8	41.996,1	46.621,0	Total Liabilitas
Funded Debt	66.064,0	61.780,3	53.286,3	22.977,2	29.729,3	Pinjaman yang Dikenakan Bunga
Gross Profit Margin	30,7%	32,7%	32,7%	29,7%	27,5%	Marjin Laba Bruto
EBIT Margin	17,8%	17,0%	15,8%	12,8%	12,5%	Marjin Laba Usaha (EBIT)
EBITDA Margin	21,3%	20,9%	20,2%	17,0%	16,6%	Marjin EBITDA
Net Income Margin Attributable to Equity Holders of The Parent Entity	5,7%	7,7%	7,9%	6,4%	5,7%	Marjin Laba Neto yang Dapat Diatribusikan Kepada Pemilik Entitas Induk
Return on Assets (%) - Net Income[4]	5,1	6,6	6,8	6,1	5,4	Imbal Hasil atas Aset (%) - Laba Neto[4]
Return on Assets (%) - EBIT[4]	10,9	9,9	9,9	10,2	9,9	Imbal Hasil atas Aset (%) - Laba Usaha[4]
Return on Equity (%)[4]	10,2	13,5	13,1	11,3	10,2	Imbal Hasil atas Ekuitas (%)[4]
Current Ratio (x)	1,79	1,34	1,37	1,27	1,07	Rasio Lancar (x)
Liabilities to Assets Ratio (x)	0,48	0,51	0,51	0,44	0,48	Rasio Liabilitas Terhadap Aset (x)
Liabilities to Equity Ratio (x)[3]	0,93	1,06	1,05	0,77	0,93	Rasio Liabilitas Terhadap Ekuitas (x)[3]
Gearing Ratio - Gross (x)[3]	0,71	0,71	0,67	0,42	0,60	*Gearing Ratio - Gross* (x)[3]
Gearing Ratio - Net (x)[3]	0,43	0,37	0,45	0,17	0,42	*Gearing Ratio - Net* (x)[3]

Figure 2 Example financial statement

Table 1. Sample dataset

	BVPS		EPS		Debt		Profit	
	2021	2022	2021	2022	2021	2022	2021	2022
GGRM	32.5	33.3	Rp. 2,913*	Rp. 1,445*	Rp. 30,676*	Rp. 30,706*	Rp. 5,605*	Rp. 277*
HARTADINATA	10.9	83.7	Rp. 4,212*	Rp. 5,505*	Rp. 1,962*	Rp. 2,126*	Rp. 194*	Rp. 1,475*
SAMPOERNA	88.3	95.4	Rp. 1,510*	Rp. 1,410*	Rp. 23,899*	Rp. 26,617*	Rp. 7,364*	Rp. 6,359*
INDOFOOD	100.1	98.3	Rp. 873*	Rp. 724*	Rp. 92,285*	Rp. 86,810*	Rp. 11,965*	Rp. 1,969*
KIMIAFARMA	76.8	59.5	Rp. 54*	Rp. 30*	Rp. 289*	Rp. 109*	Rp. 10,528*	Rp. 558*
UNILEVER	88.3	95.4	Rp. 151*	Rp. 141*	Rp. 14,747*	Rp. 14,321*	Rp. 5,758*	Rp. 706*

*= thousand units

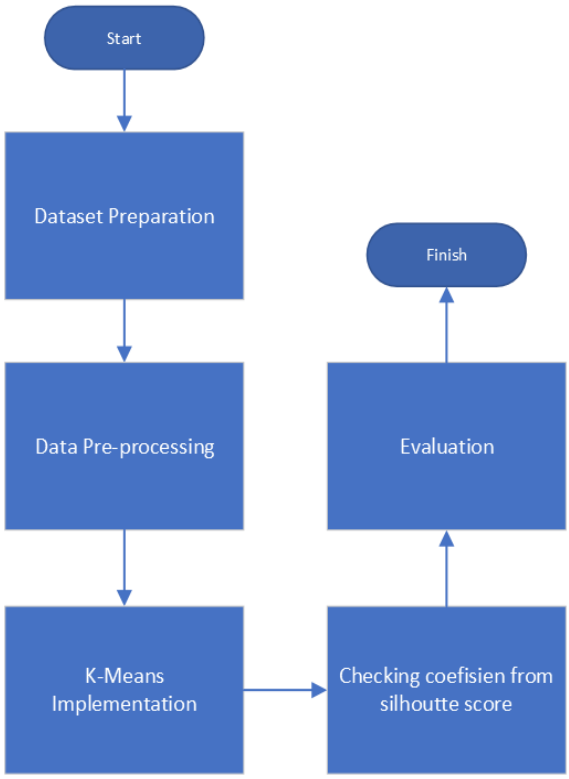


Figure 3. Implementation K-means

Table 2. Checking NULL value

Variable	Status
BVPS 2018	TRUE
EPS 2018	TRUE
DEBT 2018	TRUE
PROFIT 2018	TRUE
BVPS 2019	TRUE
EPS 2019	TRUE
DEBT 2019	TRUE
PROFIT 2019	TRUE
BVPS 2020	TRUE
EPS 2020	TRUE
DEBT 2020	FALSE
PROFIT 2020	FALSE
BVPS 2021	TRUE
EPS 2021	TRUE
DEBT 2021	FALSE
PROFIT 2021	FALSE
BVPS 2022	TRUE
EPS 2022	TRUE
DEBT 2022	FALSE
PROFIT 2022	FALSE

Table 3. Testing K value

K Value	Score
2	0.64
3	0.6
4	-0.32
5	-0.33
6	-0.27
7	-0.22
8	-0.38
9	-0.06

After getting results like the Table 3. Then proceed with calculating the number of companies in each cluster as well as a list of companies included in cluster 1 or 2. The results of this calculation show 29 companies in cluster 1 and 6 companies in cluster 2. After getting the clustering results, it will be continued with persona analysis of the results so that an explanation is obtained based on the results obtained.

3. RESULTS AND DISCUSSION

In the persona analysis process, several analyzes are carried out on the dataset which has been divided into 2 clusters. The following is a list of companies based on clusters that have been formed from clustering modeling using k-means. Cluster 1 consists of Hartadinata, Mayora, Cimory, Charoen, Alfamart, Multi Bintang, Ultra Jaya, Japfa, Segar Kumala Supra Boga, Wicaksana, Millennium, Cleo, Delta, Akasha, Jobobu Jarum, Garuda, Siantar Top, Sari Roti, Tiga Rasa, Indo Boga, Mulia Boga, Sekar Laut, Wilmar, Budi Starch, BPS, Panca Multi Perdana, Sentra Food and Kino. Meanwhile, cluster 2 consists of Kimia Farma, Gudang Garam, Indofood, Unilever, Sampoerna, and Campina.

Several things were carried out aimed at finding out the purpose of forming a cluster group, including the changes in BVPS, EPS, Profit and Debt values that occur every year as well as the Profit to Debt values that are experienced every year. In experiments such as in Figures 4 and 5, it is known that the bvps owned by cluster 1 companies in a positive context is the same as cluster 2, only in Cluster 1 the bvps value is greater. Based on these first results, it is known that companies in cluster 1 provide greater profits and are more profitable than cluster 2 if they experience liquidity. Every year since 2018 the average bvps in cluster 1 companies has had a fairly stable value.

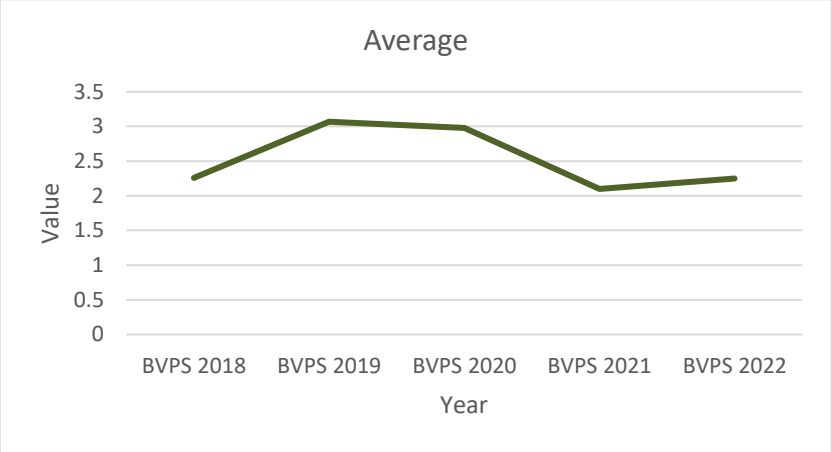


Figure 4. Average BVPS cluster 1

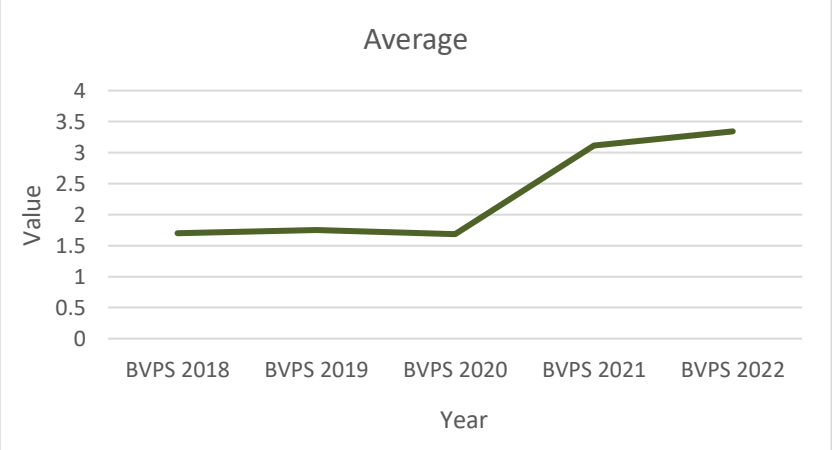


Figure 5. Average BVPS cluster 2

Meanwhile, the average bvps in cluster 2 companies had a lower value from 2018 to 2020 and increased in 2021 and 2022. These results are shown in Figures 4 and 5. The increase experienced also shows that the bvps value of companies in cluster 2 can be better than that in Cluster 1. This shows that the bvps value of companies in cluster 1 is more stable when experiencing change compared to companies in cluster 2.

In the next experiment, as in Figures 6 and 7, it is known that the EPS owned by cluster 1 companies is greater than that of cluster 2. Based on these second results, it is known that companies in cluster 1 get more profit from each share than companies in cluster 2. This is also supported by average data from EPS, where in cluster 1 the average EPS value is stable and tends to increase and even in 2022 there will be a more significant increase, whereas in cluster 2 the average EPS value has experienced a significant change. Every year. This is also supported by average data from EPS, where in cluster 1 the average EPS value is stable and tends to increase and even in 2022 there will be a more significant increase, whereas in Cluster 2 the average EPS value has experienced a significant change every year.

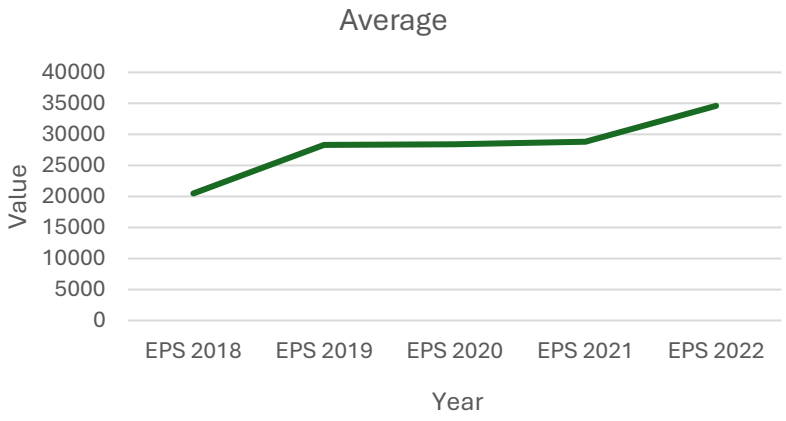


Figure 6. Average EPS cluster 1

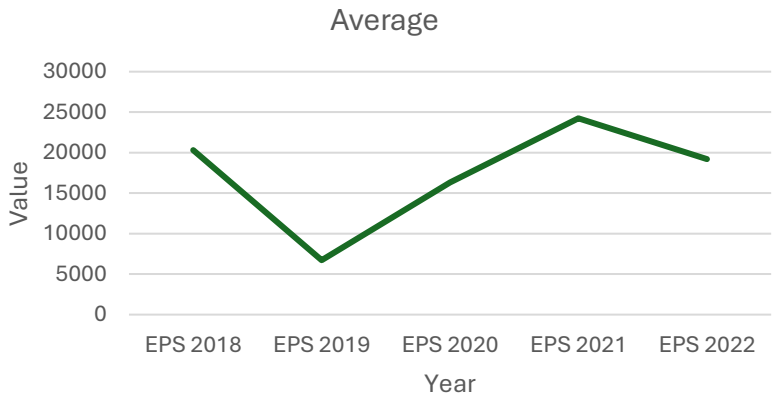


Figure 7. Average EPS cluster 2

In the next experiment, it is known in Figures 8 and 9 that changes in debt for cluster 1 companies have an average value of changes in debt that always increases compared to data from cluster 2 companies, where the average value of changes in debt tends to decrease. There is a result that is one of the differences that can be seen from the results of the data that has been carried out, namely that the average company profits in Figures 10 and 11 cluster 1 always increase, but the debt owned by cluster 1 companies also increases. Meanwhile, for cluster 2 companies, when relative profits decrease, this is accompanied by an increase in the value of debt. This shows that cluster 1 most likely consists of a group of mid-level companies that are still

trying to develop their companies and are in the process of company expansion. This has risks if the company fails.

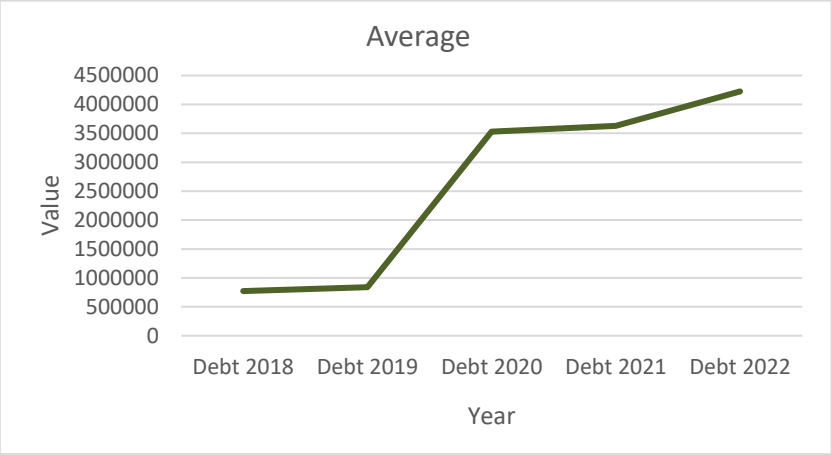


Figure 8. Average debt cluster 1

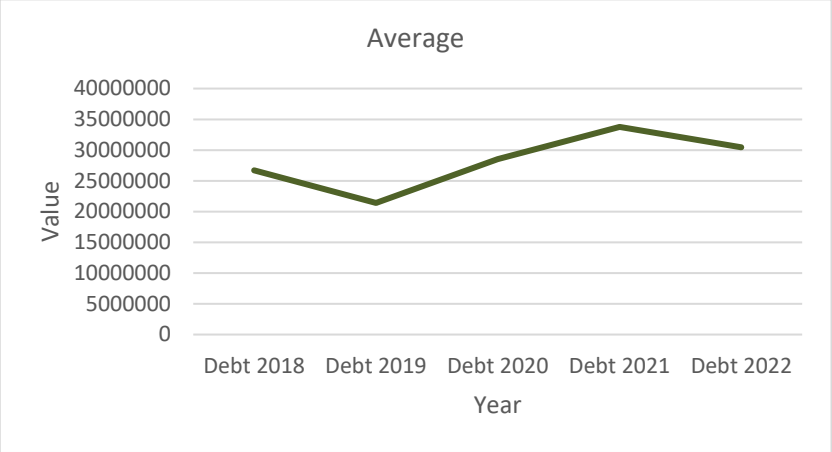


Figure 9. Average debt cluster 2

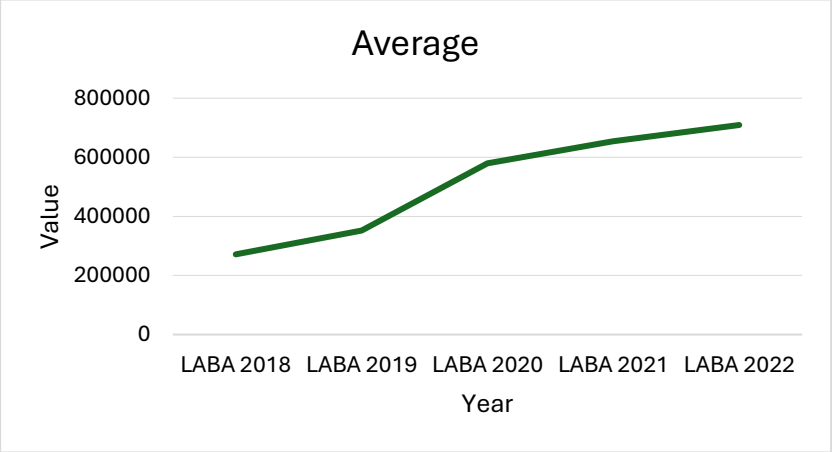


Figure 10. Average profit cluster 1

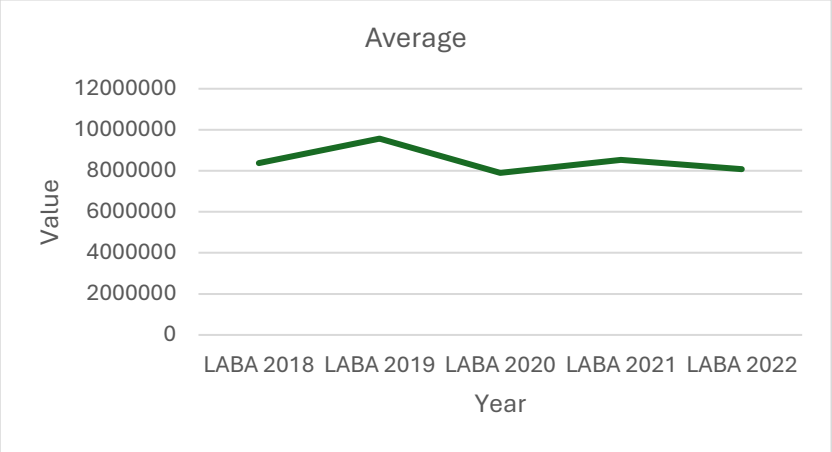


Figure 11. Average profit cluster 2

Meanwhile, cluster 2 companies are companies that have not expanded their company development and are more concerned with managing the condition of their own large companies. This shows that the company has experience and will be a challenge when conditions do not improve in the future. Meanwhile, an experiment was carried out regarding the percentage of profits on debt, where cluster 1 companies in Table 4 on average experienced very large ups and downs and also had large values. Meanwhile, cluster 2 companies in Table 5 have a relatively small average change with a lower percentage than cluster 1, but in Table 6 they experience relatively stable annual changes in their ability to pay debts using the profits obtained compared to Table 7. This of course affects the company's readiness to use profits to pay debts.

Table 4. Average percentage of profit earned towards debts owned by cluster 1

Average cluster 1	2018	2019	2020	2021	2022
	53.08654	120.1412	30.61343	62.15671	47.33334

Table 5. Average percentage of profit earned towards debts owned by cluster 2

Average cluster 2	2018	2019	2020	2021	2022
	50.51053	54.37762	37.65287	36.71564	45.69892

Table 6. Average percentage difference experienced by cluster 1

Average percentage difference experienced by cluster 1	2018-2019	2019-2020	2020-2021	2021-2022
	74.29378	-60.5282	31.54329	-14.8234

Table 7. Average percentage difference experienced by cluster 2

Average percentage difference experienced by cluster 2	2018-2019	2019-2020	2020-2021	2021-2022
	3.86709	-16.7247	-0.93723	8.983282

4. CONCLUSION

The results of this research showed that the clusters formed displayed varying patterns when data visualization was carried out. The results of this visualization can show the state of the company in each cluster. This will certainly help with the problem mentioned at the beginning, namely that many people do not understand how to determine whether a company is in good condition or not. It is hoped that when choosing when you want to invest in shares in a company, it will be easier to understand the condition of that company. With the conditions displayed based on research results, the condition of each cluster has positive and negative values so that in future determinations it is hoped that it can be adjusted to conditions that suit each investor, so that the investments made are not mistaken.

ACKNOWLEDGEMENTS

The authors thanks to Bina Nusantara University for the research grant and supporting this research.

REFERENCES

[1] G. Tang, F. Chiclana, and P. Liu, "A decision-theoretic rough set model with q-rung orthopair fuzzy information and its application in stock investment evaluation," *Applied Soft Computing Journal*, vol. 91, 2020, doi: 10.1016/j.asoc.2020.106212.
[2] B. G. Buchanan and D. Wright, "The impact of machine learning on UK financial services," *Oxford Review of Economic Policy*, vol. 37, no. 3, pp. 537–563, Sep. 2021, doi: 10.1093/oxrep/grab016.
[3] S. K. Sahu, A. Mokhade, and N. D. Bokde, "An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: recent progress and challenges," *Applied Sciences*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031956.
[4] S. Natale and A. Ballatore, "Imagining the thinking machine: Technological myths and the rise of artificial intelligence," *Convergence: The International Journal of Research into New Media Technologies*, vol. 26, no. 1, pp. 3–18, Feb. 2020, doi: 10.1177/1354856517715164.
[5] D. B. S. Manager and A. Analytics, "Neeltje Van Horen; Gosia Makarczuk; Verina Oxley; Candy Stephens; Juan Mateos-Garcia; Pete Thomas; Tangy Morgan; Andreas Viljoen." 2020.
[6] S. Russell, "The history and future of AI," *Oxford Review of Economic Policy*, vol. 37, no. 3, pp. 509–520, 2021, doi: 10.1093/oxrep/grab013.
[7] A. Mirestean *et al.*, "Powering the digital economy: opportunities and risks of artificial intelligence in finance," *Departmental Papers*, no. 24, 2021, doi: 10.5089/9781589063952.087.
[8] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications*, vol. 184, Dec. 2021, doi: 10.1016/j.eswa.2021.115537.
[9] G. Kumar, S. Jain, and U. P. Singh, "Stock market forecasting using computational intelligence: a survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1069–1101, 2021, doi: 10.1007/s11831-020-09413-5.
[10] Y. Wang and G. Yan, "Survey on the application of deep learning in algorithmic trading," *Data Science in Finance and Economics*, vol. 1, no. 4, pp. 345–361, 2021, doi: 10.3934/DSFE.2021019.
[11] J. Adlar, M. Alfter, A. To, And J. Bellis, "The impact of artificial intelligence on work," 2018.
[12] Q. Chen, "Stock market prediction using machine learning," pp. 458–465, 2023.
[13] J. Zou *et al.*, "Stock market prediction via deep learning techniques: a survey," *arXiv preprint arXiv: 2212.12717*, Dec. 2022.
[14] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, Jun. 2015, doi: 10.1007/s40745-015-0040-1.
[15] S. Bose *et al.*, "An ensemble machine learning model based on multiple filtering and supervised attribute clustering algorithm for classifying cancer samples," *PeerJ Computer Science*, vol. 7, Sep. 2021, doi: 10.7717/peerj-cs.671.
[16] G. Marti, F. Nielsen, M. Bińkowski, and P. Donnat, "A review of two decades of correlations, hierarchies, networks and clustering in financial markets," *Signals and Communication Technology*, pp. 245–274, 2021, doi: 10.1007/978-3-030-65459-7_10.
[17] J. Kumar, M. Debika, and B. Editors, "Advances in intelligent systems and computing 937 emerging technology in modelling and graphics proceedings of IEM graph 2018," pp. 99–111, 2020.
[18] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: a comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, Aug. 2020, doi: 10.3390/electronics9081295.
[19] T. Li, G. Kou, Y. Peng, and P. S. Yu, "An integrated cluster detection, optimization, and interpretation approach for financial data," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13848–13861, Dec. 2022, doi: 10.1109/TCYB.2021.3109066.
[20] S. Dong, Y. Xia, and T. Peng, "Network abnormal traffic detection model based on semi-supervised deep reinforcement learning," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4197–4212, Dec. 2021, doi: 10.1109/TNSM.2021.3120804.
[21] Z. Huang, H. Zheng, C. Li, and C. Che, "Application of machine learning-based k-means clustering for financial fraud detection," *Academic Journal of Science and Technology*, vol. 10, no. 1, pp. 33–39, 2024, doi: 10.54097/74414c90.
[22] P. Chanfreut, J. M. Maestre, A. J. Gallego, A. M. Annaswamy, and E. F. Camacho, "Clustering-based model predictive control of solar parabolic trough plants," *Renewable Energy*, vol. 216, 2023, doi: 10.1016/j.renene.2023.118978.
[23] A. Yaqoob, R. Musheer Aziz, and N. K. Verma, "Applications and techniques of machine learning in cancer classification: a systematic review," *Human-Centric Intelligent Systems*, vol. 3, no. 4, pp. 588–615, Sep. 2023, doi: 10.1007/s44230-023-00041-3.
[24] M. Li, Y. Zhu, Y. Shen, and M. Angelova, "Clustering-enhanced stock price prediction using deep learning," *World Wide Web*, vol. 26, no. 1, pp. 207–232, Jan. 2023, doi: 10.1007/s11280-021-01003-0.
[25] R. Liu, "Data analysis of educational evaluation using k-means clustering method," *Computational Intelligence and Neuroscience*, pp. 1–10, Jul. 2022, doi: 10.1155/2022/3762431.
[26] S. Nanjundan, S. Sankaran, C. R. Arjun, and G. P. Anand, "Identifying the number of clusters for K-Means: A hypersphere density-based approach," *arXiv preprint arXiv: 1912.00643*, 2019.
[27] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proceedings-2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 2020, pp. 747–748, doi: 10.1109/DSAA49011.2020.00096.
[28] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.
[29] T. Saheb and M. Dehghani, "Artificial intelligence for sustainability in energy industry: a contextual topic modeling and content analys," pp. 1–36, 2021.
[30] A. Vysala and D. J. Gomes, "Evaluating and validating cluster results," in *Computer Science and Information Technology*, Jul. 2020, pp. 37–47, doi: 10.5121/csit.2020.100904.
[31] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using k-means algorithm and silhouette coefficient," in *2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, Sep. 2020, pp. 258–262, doi: 10.1109/ICT-PEP50916.2020.9249773.
[32] A. Naghizadeh and D. N. Metaxas, "Condensed silhouette: an optimized filtering process for cluster selection in k-means," *Procedia Computer Science*, vol. 176, pp. 205–214, 2020, doi: 10.1016/j.procs.2020.08.022.

[33] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13638-021-01910-w.
[34] L. Lenssen and E. Schubert, "Medoid Silhouette clustering with automatic cluster number selection," *Information Systems*, vol. 120, 2024, doi: 10.1016/j.is.2023.102290.
[35] Nursiami And S. C. Simamora, "Pengaruh return on equity, current ratiodan debt to equity ratioterhadap dividend payout ratio pada perusahaan sub sektor konstruksi." 2021.
[36] A. M. Ahmed, N. A. Sharif, M. N. Ali, and I. Hágen, "Effect of firm size on the association between capital structure and profitability," *Sustainability*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151411196.
[37] B. Grabinska, M. Kedzior, D. Kedzior, and K. Grabinski, "The impact of corporate governance on the capital structure of companies from the energy industry. The case of Poland," *Energies*, vol. 14, no. 21, Nov. 2021, doi: 10.3390/en14217412.
[38] B. T. Rahmanto and S. Nurjanah, "Influence analysis of fundamental information, company size and sales growth toward share price (Studies in pharmacy industry company listed in Indonesia stock exchange)," *International Journal of Economic Research*, vol. 14, no. 17, pp. 11–25, 2017.
[39] Heliani, F. Mareta, A. Ulhaq, E. Resfitasari, I. Febriani, and S. Elisah, "Effect of debt to equity ratio, current ratio, total assets turnover, earning per share, price earning-ratio, sales growth, and net profit margin on return on equity," *Proceedings of the International Conference on Economics, Management and Accounting (ICEMAC 2021)*, vol. 207, 2022, doi: 10.2991/aebmr.k.220204.047.
[40] A. T. Gharaibeh, M. H. Saleh, O. Jawabreh, And B. J. A. Ali, "An empirical study of the relationship between earnings per share, net income and stock price," *Applied Mathematics and Information Sciences*, vol. 16, no. 5, pp. 673–679, Sep. 2022, doi: 10.18576/amis/160502.

BIOGRAPHIES OF AUTHORS

Lili Ayu Wulandhari is a computer science lecturer at Bina Nusantara University. She specializes in machine learning, data science, and text mining. She successfully obtained her master's and Ph.D. degrees in computer science from the Malaysian University of Technology. Lili possesses a decade of experience as an academic as well as expertise as a professional data scientist. She has successfully applied AI and machine learning techniques in several domains and data types, including text, image, and financial data. She is actively involved as both the chairman and a member of the campus's internal and national grants. In the past five years, the research has resulted in over 10 articles that have been distributed in respected national and international journals.

Gusti Aditya Aromatica Firdaus is a master's student at BINUS Graduate Program-Master of Computer Science, Bina Nusantara University with a focus on data science. His undergraduate education background is in Information Technology at Lambung Mangkurat University. He is also a data engineer. He can be contacted at email: gusti.firdaus001@binus.ac.id.