IMPLEMENTATION OF K-MEANS ALGORITHM USING PHYTON BASED ON PROFITABILITY AND SOLVENCY RATIO

NELSI WISNA[1], MAYSHAFIRA LISTIANINGSIH[2], SYALMA NURINDAH AL-AZHARY[3], KASTAMAN[4], RASWYSHNOE BOING KOTJOPRAYUDI[5]

[1,2,3,4,5]School of Applied Science, Telkom University, Bandung, Indonesia

E-mail: nelsie@telkomuniversity.ac.id[1], mayshfraa@student.telkomuniversity.ac.id[2], syalmanurindaha@student.telkomuniversity.ac.id[3], kastaman@telkomuniversity.ac.id[4], raswyshnoe@telkomuniversity.ac.id[5]

ABSTRACT

This research aims to calculate the profitability and solvency of companies and cluster them using the K-Means clustering analysis method. Profitability ratios are retrieved by Return on Assets (ROA) and Return on Equity (ROE). Solvency ratio is calculated based on Debt to Assets Ratio (DAR) and Debt to Equity Ratio (DER). The research was conducted at subcontractor companies registered on the Indonesia Stock Exchange for the 2019 period. The selection sample used purposive sampling method so that 16 companies were found. The research data is secondary data, namely audited financial reports and obtained by through the official website of the Indonesia Stock Exchange. The results of the calculation of ROA profitability show 7 companies in the very good category, while the results of ROE calculations show very good and good categories, namely 11 companies. Solvency calculations for the DAR ratio explain that 16 companies are in the healthy category while the DER ratio shows that 5 companies have a healthy category. Profitability cluster analysis shows that the optimum number of clusters is 3 clusters. Cluster 1 composed of 7 companies, Cluster 2 composed of 1 company and Cluster 3 composed of 8 companies. Solvency cluster analysis shows that the optimum number of clusters is 3 clusters. Cluster 1 composed of 3 companies, Cluster 2 composed of 7 companies and Cluster 3 composed of 6 companies.

Keywords: K-Means, ROA, ROE, DAR, DER

1. INTRODUCTION

The result of data processing that provide benefits to users called information, the results of data processing will be classified so that they can be used for decision making [1]. Accounting is an information system, which is a system used in providing financial information [2]. The quantitative information about economic entities that is useful in making economic decisions to determine choices among alternative of actions called accounting information [3]. Accounting systems have a role to provide added value to users in the form of providing various financial information that can be used for planning, controlling, and decision-making activities of the company [4].

The company's success in managing resources can be seen from its financial performance. Performance assessments that can be used are profitability and solvency. Profitability is a ratio that explains the company's capabilities to earn profit in a specific period and Solvability explains the company's capabilities to meet its debt obligations, both long-term debt and short-term debt. Profitability can be assessed using several ratios including Return on Investment, Return on Asset and Return on Equity [5]. The Solvability can be measured in several ratios including Debt to equity Ratio (DER), Debt to Asset Ratio (DAR),Times Interest Earned and Long-Term Debt to Equity Ratio [6].

The phenomenon that occurs in relation to the profitability and solvency of sub-construction companies is that the shares of construction issuers are declining because they are influenced by two

main factors, namely the sentiment of rising benchmark interest rates globally and the increase in Covid-19 cases [7]. The net profit of the issuer (ADHI) fell 96.38% in 2020 to IDR 23.98 billion compared to 2019 IDR 663.80 billion, due to the Covid-19 pandemic, the company's profit value per share also fell to IDR 7 from 2019 IDR 186 [8]. Covid has brought business activities to a relative standstill, including restrictions on the mobility of goods and people. During the Covid-19, the company's profit or loss was sharply corrected due to the reduction in project work progress terms that could be billed to customers, resulting in an average revenue decrease of up to 70% [9].

This study aims to calculate the profitability and solvability of the company in 2019. 2019 is a year when there has been no lockdown due to the Covid-19 pandemic in Indonesia so that this study is able to explain whether the company's profitability is in the very good, good, and less good categories. The Solvability calculation explains the calculation of whether the company is in a healthy condition, warning, or high risk. Clustering in this study uses k-means clustering, which is a non-hierarchical cluster analysis method that can classify a group of objects according to similar characteristics. When the required number of clusters is known then K-means clustering can be used. In cluster analysis, the level of similitude of the characteristics of objects in one group (cluster) is very high, while the level of similitude of characteristics between objects in a cluster with other clusters is very low [10].

2. LITERATURE REVIEW

2.1. Profitability Ratio
The ratio comparison of the company's capabilities to earn profits from revenue based on certain measurements. Profitability ratios are used to record financial transactions and to show the size of the profit earned by the company. The effectiveness and efficiency of management can be seen from the calculation of profitability ratios that exist in the elements of financial statements [11].

2.1.1. *Return on Asset*
Return on assets indicates the results of the total assets utilized by the company in its operational activities. ROA explains the ability of effectiveness in using assets to get income [6]. This ratio explains the company's ability to earn profit after tax using all the assets owned by the company.

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$$

2.1.2. Return on Equity
The relationship between annual profit after tax and shareholders' equity. ROE is used to show a measure of the effectiveness of funds for shareholders who have invested. The higher the ROE, the higher the amount of net profit generated [12].

$$\text{ROE} = \frac{\text{Net Income}}{\text{Shareholder's Equity}}$$

2.2. Solvability Ratio

The company's capabilities to pay loans or pay off long-term debt is called solvency. The solvency ratio indicates that the company's cash is sufficient to pay off long-term obligations, this ratio can also be a measure of the company's financial health.

2.2.1. Debt to Equity Ratio (DER)
Ratio that shows the extent to which own capital guarantees all debt. This ratio also shows the comparison between external party funds and company owner funds [13].

$$\text{DER} = \frac{\text{Total Liabilities}}{\text{Total Equity}}$$

2.2.2. Debt to Assets Ratio (DAR)
Ratio that measures the share of assets used to guarantee the overall liability [13]. Debt to Asset Ratio is able to measure how much the number of assets in the company where the assets are financed with debt or how much the company's debt affects the management of assets. The higher the DAR value, the greater the source of funds through loans to finance assets.

$$\text{DER} = \frac{\text{Total Liabilities}}{\text{Total Assets}}$$

2.3. Analysis Cluster

Analysis cluster is a multivariate statistical analysis that has the aim of knowing the structure of data by placing the similarity of observation objects into one data group so that they can be distinguished between one group and another or by separating objects into several groups that are different in nature between one group and another [14].

2.4. *K-Means Clustering*

K-means clustering is one of the non-hierarchical cluster analysis method. An analysis method that can be used to classify objects based on their proximity.

Cluster analysis will classify objects with similar characteristics or objects with different into one cluster [15].

K-means has the ability to classify large amounts of data with quick and efficient computing time. K-means is a clustering method with a centroid-based partitioning method. In its application, it requires three parameters, namely the number of clusters k, cluster initialization, and system distance, usually k-means is run independently with different initializations. This algorithm in principle only groups the data into a local minimum. The k-means algorithm iteratively increases the variation of values within each cluster, where objects are then placed in the closest group, calculated from the cluster centroid. A new centroid is determined when all data has been placed in the closest cluster. [16]

2.5. Python Programming Language

A high-level programming language who developed by Guido van Rossum in 1989 is called python that is designed to provide tremendous convenience to programmers in terms of time efficiency, as well as program development. Python can be used to create standalone programs and script programming [17].

3. RESEARCH METHODOLOGY

The data source used in this study case is secondary data which is obtained from the company annual financial statements published by the Indonesia Stock Exchange (IDX). The sample selection in this study was determined by criteria:

Table 1: Research Sample

No	Sampling Criteria	Amount
1	Sub-contractor companies who published annual financial reports on Indonesia Stock Exchange during 2019	18
2	Companies that do not consistently publish	(2)
Total companies in the research sample		16

In Table 1, there are 18 sub-contractor companies in Indonesia Stock Exchange but 2 (two) of them do not consistently publish their annual financial reports so that the total companies sampled in this study were 16 companies. The companies that became the sample on this research are listed in Table 2:

Table 2: Company Details

No	Company Codes	No	Company Codes	No	Company Codes	No	Company Codes
1	CSIS	5	SSIA	9	TAMA	13	JKON
2	NRCA	6	IDPR	10	DGIK	14	WKST
3	WEGE	7	PBSA	11	SRKN	15	WIKA
4	TOPS	8	ADHI	12	PTPP	16	TOTL

The variables used in this study are profitability and Solvability ratios consisting of:

Table 3: Research Variables

Variables	Description
ROA	Return on Asset
ROE	Return on Equity
DAR	Debt to Asset Ratio
DER	Debt to Equity Ratio

The methodology used in this analysis is the Knowledge Discovery in Database (KDD) methodology processed with the Python programming language with the following stages [18].
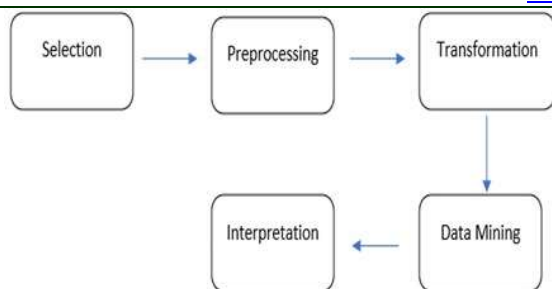
Figure 1: Data Analysis Phase

Based on Figure 1, the phase of data analysis are: (1) Selection stage: selecting the dataset to be used, namely data on return on assets, return on equity, debt to asset ratio, and debt to equity ratio of sub-contractor companies in 2019, (2) Preprocessing stage: performing data processing and data normalization using the min-max scaller feature, (3) Transformation stage: the process of transforming the selected data so that the data is suitable for the data mining process, (4) Data mining stage: the stage of data analysis using the k-means model to get clusters from each data. The command for the K-

Means method modeling process uses the Scikit Learn library, (5) Interpretation stage: the stage of translating the patterns generated from data mining.

4. RESULTS AND DISCUSSION

4.1. Ratio Calculation Result

The results of the ratio calculation show in Table 4, that ROA with good and excellent categories are those with values above 3%. ROA with this category is shown by NRCA, TOTL, WIKA, JKON and WEGE. The ROE ratio shows that companies with good categories are ADHI, NRCA, PTPP, TOTL, WIKA, JKON, WEGE, PBSA and SRKN. The calculation of the DAR ratio shows that all companies are able to show that all debts will be covered by the company's assets where the ratio value is below 1. The DER value is below or equal to 100% or 1, so the company's condition is in the healthy category, while based on the ratio calculation, the healthy companies are DGIK, JKON, IDPR, PBSA and SSIA.

Table 4: Ratio Calculation Results

No	Company	ROA	ROE	DAR	DER
1	ADHI	1.82%	9,73%	81.28%	434.30%
2	DGIK	0.09%	0,18%	49.77%	99.09%
3	NRCA	4.11%	8,28%	50.42%	101.68%
4	PTPP	2.04%	6,97%	70.72%	241.48%
5	TOTL	5.92%	16,30%	63.65%	175.14%
6	WIKA	4.22%	13,64%	69.06%	223.23%
7	WSKT	0.84%	3,53%	76.25%	321.00%
8	JKON	4.10%	7,50%	45.26%	82.67%
9	CSIS	-1.96%	-3,48%	56.32%	128.93%
10	IDPR	-0.18%	-0,45%	39.34%	64.86%
11	TOPS	-7.02%	-12,30%	57.02%	132.68%
12	WEGE	7.36%	12,21%	60.31%	151.98%
13	PBSA	1.84%	7,18%	25.60%	34.41%
14	SRKN	8.22%	14,70%	55.91%	126.81%
15	SSIA	1.68%	3,77%	44.66%	80.71%
16	TAMA	1.16%	1,39%	83.60%	509.67%

Source: Data processed by researchers (2023)

4.2. K-Means Algorithm Calculation
4.2.1. Profitablity Ratio Characteristics

The characteristics of the profitability ratio can be seen using descriptive statistics to see a general description of the data used. Descriptive statistical results of the profitability ratio are:

Table 5: Probability Ratio Characteristics

Index	ROA	ROE
mean	0.02131	0.05575
std	0.03687	0.07441
min	-0.07	-0.123
max	0.082	0.163

Based on Table 5, the information can be explained that the ROA variable, namely Return on Asset, has a mean index value of 0.02131, a standard deviation value of 0.03687, a minimum of -0.07 and a maximum value of 0.082. Standard deviation is used to see how far the data distribution is from the average or mean. A large standard deviation value indicates that the wider the variation of a data. From this index, it can show that the company that has the lowest ROA value is TOPS and the company that has the highest ROA value is SRKN. The ROE variable, namely Return on Equity, has a mean index value of 0.05575, a standard deviation value of 0.07441, a minimum of -0.123 and a maximum value of 0.163. Standard deviation is used to see how far the data distribution is from the average or mean. A large standard deviation value indicates that the wider the variation of the data. From this index, it can show that the company that has the lowest ROE value is TOPS and the company that has the highest ROE value is TOTL.

4.2.2 Clustering Profitability Ratio

The Scatter Plot diagram is used to see the data correlation between Return on Asset and Return on Equity.
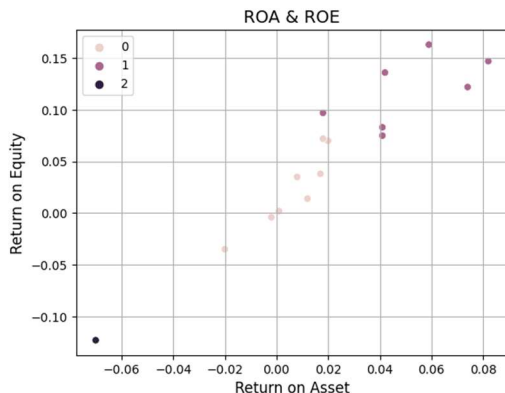

Figure 2: Scatter Plot Diagram of Profitability

Figure 2 shows the range of values for cluster 1 marked with purple dots is at a value of 0.02 – 0.16, for cluster 2 marked with black dots is between -0, and for cluster 3 marked with pink dots is between -0.02 until -0.06. shows the range of values where for cluster 1 marked with purple dots is in the range of ROA values between 0.02 - 0.08 and ROE values 0.06 – 0.08, cluster 2 marked with black dots is in the range of ROA values -0.05 and ROE values in the range -0.05, and cluster 3 marked with pink dots is in the range of ROA values between -0.02 - 0.02 and ROE values in the range -0.06 – 0.06.

4.2.3. Optimal Number of Clusters Determination

The method used to evaluate the optimum number of clusters is the elbow method. The following is an elbow image of the profitability ratio:
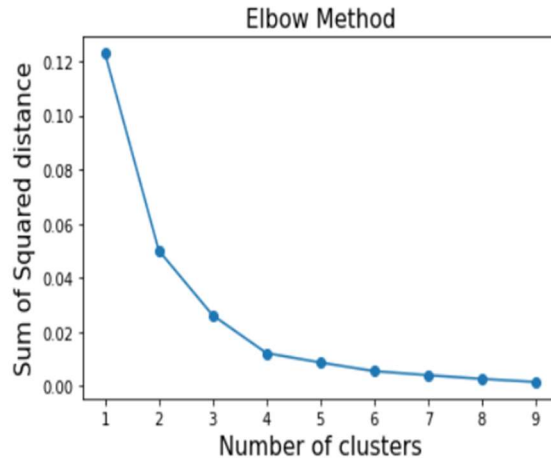

Figure 3: Elbow of profitability ratios

Based on Figure 3, the optimum number of clusters to be used in this clustering is 3 clusters. Table 6 explains that there are 7 (seven) companies in cluster 1, 1 (one) company in cluster 2 and 8 (eight) companies in cluster 3.

Table 6: Company Code Based on Clusters

Cluster	Company Code
1	ADHI, NRCA, TOTL, WIKA, JKON, WEGE, SRKN
2	TOPS
3	DGIK, PTPP, WSKT, CSIS, IDPR, PBSA, SSIA, TAMA

4.3.1 Clustering Solvability Ratio

The characteristics of the profitability ratio can be seen using descriptive statistics to see a general description of the data used. Descriptive statistical results of the solvency ratio are:

Table 7: Solvability Ratio Characteristics

Index	DAR	DER
mean	0.58075	1.81793
std	0.15753	1.35101
min	0.256	0.344
max	0.836	5.097

Based on Table 7, the information can be explained that the DAR variable, namely the Debt to Asset Ratio, has a mean index value of 0.58075, a standard deviation value of 0.15753, a minimum of 0.256 and a maximum value of 0.836. Standard deviation is used to see how far the data distribution is from the average or mean.

A large standard deviation value indicates that the wider the variation of the data. From this index, it can show that the company that has the lowest DAR value is PBSA and the company that has the highest DAR value is TAMA. The DER variable, has a mean index value of 1.81793, a standard deviation value of 1.35101, a minimum of 0.344 and a maximum value of 5.097. Standard deviation is used to see how far the data distribution is from the average or mean. A large standard deviation value indicates that the wider the variation of the data. From this index, it can show that the company that has the lowest DER value is PBSA and the company that has the highest DER value is TAMA.

4.3.2. Clustering Solvability Ratio

The Scatter Plot diagram is used to see the data correlation between Debt to Equity Ratio and Debt to Assets Ratio.
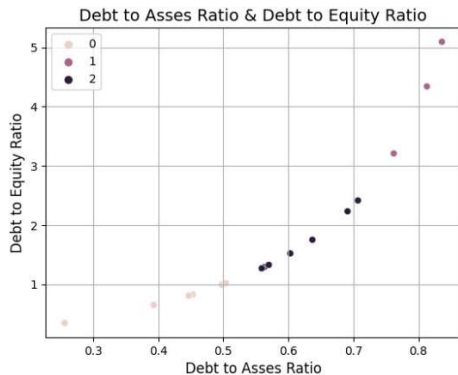

Figure 4: Scatter Plot Diagram of Solvability

From Figure 4, the value range for cluster 1 marked with purple dots is between 0.84 - 1, for cluster 2 marked with black dots is between 0.5 - 0.75, and for cluster 3 marked with pink dots is between 0 - 0.42. shows the range of values where for cluster 1 marked with purple dots is in the range of DAR values between 0.78 - 0.85 and DER values 3.1 - 6, cluster 2 marked with black dots is in the range of DAR values 0.55 - 0.71 and DER values in the range 1.2 - 2.5, and cluster 3 marked with pink dots is in the range of DAR values between 0 - 0.5 and DER values in the range 0 - 1.

4.3.3. Optimal Number of Clusters Determination

The method used to evaluate the optimum number of clusters is the elbow method. The following is an elbow image of the solvability ratio:
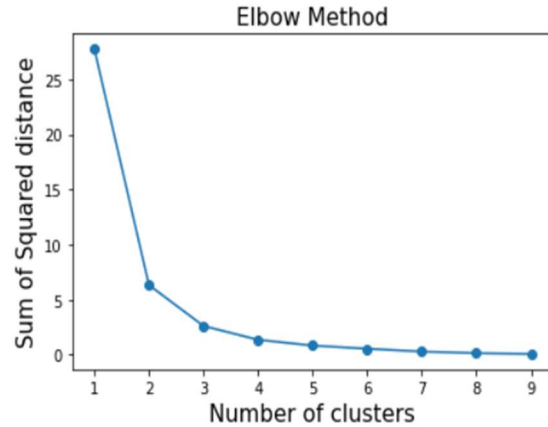

Figure 5: Elbow of Solvability Ratios

Based on Figure 5, the optimum number of clusters to be used in this clustering is 3 clusters. Table 8 explains that there are 3(three) companies in cluster 1, 7 (seven) company in cluster 2 and 6 (six) companies in cluster 3.

Table 8 Company Code Based on Clusters

Cluster	Company Code
1	ADHI, WSKT, TAMA
2	PTPP, TOTL, WIKA, CSIS, TOPS, WEGE, SRKN
3	DGIK, NRCA, JKON, IDPR, PBSA, SSIA

5.   CONCLUSION

Based on the calculation of the profitability ratio, it shows that companies with good and very good ratio levels are NRCA, TOTL, WIKA, JKON, WEGE, SRKN. This is in accordance with the company with Cluster Id 1 is a company with good and excellent ROA and ROE categories. The calculation of solvency shows that the DAR ratio of all companies shows that they are in a safe condition where the relationship between the amount of long-term debt and the amount of their capital and assets, the company can cover debt. The calculation of the DER ratio shows that companies with DER values below 100% or 1 are DGIK, NRCA, JKON, IDPR, PBSA and SSIA. This is in accordance with the company with Cluster Id 3 which is a company with DER category below 100%.

REFERENCES:

[1] E. Y. Anggraeni and R. Irviani, Pengantar Sistem Informasi, Yogyakarta: Penerbit ANDI, 2017.

[2] E. L. M. Sitinjak, B. Harcahyo, L. Kusumaningrum and L. Y. Nugraheni, Akuntansi Dasar 1 : Ringkasan Teori dan Soal, Jakarta: Grasindo, 2008.

[3] A. R. Belkaoui, Teori Akuntansi, Jakarta: Salemba Empat, 2000.

[4] M. Romney and P. Steinbart, Accounting Information System 11th Edition, 11 ed., Pearson Prentice Hall, 2019.

[5] F. I. Analisis Laporan Keuangan, Bandung: Alfabeta, 2016.

[6] Kashmir, Analisis Laporan Keuangan., Jakarta: Rajawali Pers, 2013.

[7] A. Fernando, "Kok Babak Belur? Beginilah Nasib Saham Konstruksi Hari Ini," 2022. [Online]. Available: https://www.cnbcindonesia.com/market/20220117102008-17-307848/kok-babak-belur-beginilah-nasib-saham-konstruksi-hari-ini.

[8] L. Olavia, "Laba Adhi Karya Anjlok 96,38% Akibat Pandemi," 2021. [Online]. Available: Investor.id: https://investor.id/market-and-corporate/243668/laba-adhi-karya-anjlok-9638-akibat-pandemi.

[9] T. Pranoto, "BUMN konstruksi mencatatkan kinerja yang rapuh sepanjang tahun lalu," 12 April 2021. [Online]. Available: https://newssetup.kontan.co.id/news/bumn-konstruksi-mencatatkan-kinerja-yang-rapuh-sepanjang-tahun-lalu-1?page=all.

[10] R. Johnson and D. Wichern, Applied Multivariate Statistical Analysis (6 ed.), New Jersey: Prentice Hall, 2007.

[11] Ompusunggu, Hermaya, Wage and Sunarto, Manajemen Keuangan, Batam: Batam Publisher, 2021.

[12] Hery, Analisis Kinerja Manajemen, Jakarta: Grasindo, 2015.

[13] Hantono, Konsep Analisa Laporan Keuangan dengan Pendekatan Rasio dan SPSS, Yogyakarta: Deepublish, 2018.

[14] R. R. Prakasa and A. P. Kurnianingtyas, SPSS untuk Analisis & Perencanaan Kota yang Lebih baik, Yogyakarta: Penerbit ANDI, 2022.

[15] A. Syukria, R. . Y. Putra, M. Larasati, T. M. Putri and A. Fadhlan, "Klasterisasi Perusahaan Yang Tergabung Dalam Jakarta Islamic Index 70 Berdasarkan Rasio Profitabilitas Menggunakan K-Means Clustering," *Jurnal Manajemen dan Kewirausahaan,* pp. 1-6, 2022.

[16] E. Irwansyah and M. Faisal, Advanced Clustering Teori dan Aplikasi, Yogyakarta: DeePublish, 2015.

[17] H. Wadi, Pemograman Phyton: untuk Pelajar dan Mahasiswa, TR Publisher, 2016.

[18] F. A. Hermawati, Data Mining, Yogyakarta: Andi, 2013.