K-MEANS CLUSTERING OF INDONESIAN BANKING STOCKS USING FINANCIAL RATIOS

Fajar Rizki Aditiya[1]; Nina Sulistiyowati[2]; Siska[3]
Information System[1,2,3]
Universitas Singaperbangsa Karawang, Karawang, Indonesia[1,2,3]
https://www.unsika.ac.id/[1,2,3]
2210631250081@student.unsika.ac.id[1*], nina.sulistio@unsika.ac.id[2],
siska@staff.unsika.ac.id[3*]
(*) Corresponding Author

Abstract— *This study aims to classify banking sector issuers listed on the Indonesia Stock Exchange based on financial ratios, namely Return on Assets (ROA), Return on Equity (ROE), and Loan to Deposit Ratio (LDR), to assist investors in analyzing financial performance and making more objective investment decisions. The method used in this study is Knowledge Discovery in Databases (KDD) with the K-Means clustering algorithm. The dataset was obtained from the annual financial reports of 25 banking issuers for the period 2022–2024. The research stages consist of data selection, data cleaning, data transformation, data mining, and interpretation/evaluation. In the transformation stage, the average values of ROA, ROE, and LDR were calculated and normalized using the Min-Max method. The optimal number of clusters was determined using the Elbow method, while cluster quality was evaluated using the Silhouette Coefficient. The results indicate that the optimal number of clusters is three with a Silhouette Coefficient value of 0.5946. The clustering results consist of a dominant cluster containing 22 issuers with relatively stable financial performance, a second cluster consisting of two issuers characterized by higher lending activity and relatively higher risk, and a third cluster containing one issuer with significantly lower financial performance. These findings reveal latent patterns among banking issuers that may not be easily identified through conventional ratio analysis and provide a clearer structural overview of banking sector performance to support investment evaluation. These insights provide a clearer structural overview of the banking sector and may assist investors in identifying banks with comparable financial risk and performance profiles.*

Keywords: *Banking Issuers, Clustering, Data Mining, Financial Ratios, K-Means*

Intisari— *Penelitian ini bertujuan untuk mengklasifikasikan emiten sektor perbankan yang terdaftar di Bursa Efek Indonesia berdasarkan rasio keuangan, yaitu Return on Assets (ROA), Return on Equity (ROE), dan Loan to Deposit Ratio (LDR), untuk membantu investor dalam menganalisis kinerja keuangan dan membuat keputusan investasi yang lebih objektif. Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD) dengan algoritma clustering K-Means. Dataset diperoleh dari laporan keuangan tahunan 25 emiten perbankan untuk periode 2022–2024. Tahapan penelitian terdiri dari pemilihan data, pembersihan data, transformasi data, penambangan data, dan interpretasi/evaluasi. Pada tahap transformasi, nilai rata-rata ROA, ROE, dan LDR dihitung dan dinormalisasi menggunakan metode Min-Max. Jumlah cluster optimal ditentukan menggunakan metode Elbow, sedangkan kualitas cluster dievaluasi menggunakan Koefisien Silhouette. Hasil menunjukkan bahwa jumlah cluster optimal adalah tiga dengan nilai Koefisien Silhouette sebesar 0,5946. Hasil pengelompokan terdiri dari kelompok dominan yang berisi 22 emiten dengan kinerja keuangan yang relatif stabil, kelompok kedua yang terdiri dari dua emiten yang dicirikan oleh aktivitas pinjaman yang lebih tinggi dan risiko yang relatif lebih tinggi, dan kelompok ketiga yang berisi satu emiten dengan kinerja keuangan yang jauh lebih rendah. Temuan ini mengungkapkan pola laten di antara emiten perbankan yang mungkin tidak mudah diidentifikasi melalui analisis rasio konvensional dan memberikan gambaran struktural yang lebih jelas tentang kinerja sektor perbankan untuk mendukung evaluasi investasi. Wawasan ini memberikan gambaran struktural yang lebih jelas tentang sektor perbankan dan dapat membantu investor dalam mengidentifikasi bank dengan profil risiko dan kinerja keuangan yang sebanding.*

Kata Kunci: *Emiten Perbankan, Clustering, Penambangan Data, Rasio Keuangan, K-Means*

INTRODUCTION

The capital market plays a vital role in supporting national economic growth by facilitating capital allocation between surplus and deficit economic units. In Indonesia, this function is carried out by the Bursa Efek Indonesia (Indonesia Stock Exchange/IDX), which serves as the official marketplace for trading long-term financial instruments such as stocks and bonds. As one of the pillars of the national financial system, the IDX contributes to economic stability by improving market liquidity and encouraging public investment participation (Agustina et al. 2022). Furthermore, the performance of the capital market is often reflected through the movement of the Composite Stock Price Index (IHSG), which represents overall investor confidence and macroeconomic conditions (Rasyid and Sosrowidigdo 2022). Within this ecosystem, issuers play an essential role as entities that offer securities to the public to obtain long-term funding for business expansion and capital strengthening (Effendi, Siwi, and Silalahi 2024).

Among the sectors listed on the IDX, the banking sector holds a strategic position due to its dual function as a financial intermediary and as a public investment instrument. Publicly listed banks not only collect and distribute funds but also serve as indicators of financial system stability (Muzahid, Wijaya, and Isra 2023). The financial performance of banking issuers directly affects investor decisions and market confidence. However, differences in profitability, efficiency, and liquidity among banks create varying levels of competitiveness and risk exposure (Almunawwaroh 2022). These disparities highlight the importance of conducting systematic performance analysis to better understand the relative positioning of banking issuers within the capital market.

Financial ratios are widely used to measure banking performance and stability. Return on Assets (ROA) reflects the ability of banks to generate profit from total assets, while Return on Equity (ROE) measures efficiency in utilizing shareholders' equity to produce returns. Loan to Deposit Ratio (LDR), on the other hand, indicates liquidity risk by comparing total loans to total deposits (Lusiana, Tripermata, and Putri 2025). Previous studies indicate that profitability ratios such as ROA and ROE experienced pressure during periods of economic uncertainty, including the COVID-19 pandemic, particularly in banks with weaker asset and capital management (Saerang, Tommy, and Christiano 2014). Moreover, inadequate liquidity management reflected in high LDR values can increase financial vulnerability, whereas controlled LDR levels contribute to operational stability (Rasyid and Sosrowidigdo 2022). These findings demonstrate that variations in ROA, ROE, and LDR reflect differences in management effectiveness, capital adequacy, and risk mitigation strategies among banking institutions.

Traditional approaches in evaluating banking performance commonly rely on comparative ratio analysis. However, such methods often analyze each ratio separately and fail to capture multidimensional relationships among financial indicators (Dalimunthe and Lubis 2023). As a result, hidden patterns and structural similarities among banks may remain unidentified, leading to less objective performance assessment. In recent years, data-driven analytical approaches have been increasingly adopted to overcome these limitations. The Knowledge Discovery in Databases (KDD) framework provides systematic stages including data selection, preprocessing, transformation, data mining, and interpretation to extract meaningful insights from large datasets (Srimurdianti et al. 2023).

Several previous studies have applied data mining techniques in financial and classification contexts. (Jurnal and Mea 2024) applied the Naïve Bayes algorithm to classify nutritional status data, achieving high accuracy performance. (Trees, Forest, and Neighbours 2024) compared Decision Tree and Support Vector Machine (SVM) methods in classification tasks, while (Sahrul 2025) implemented Random Forest combined with cross-validation for prediction purposes. Although these studies demonstrate the effectiveness of data mining algorithms in classification problems, clustering-based analysis for grouping financial performance in the Indonesian banking sector remains limited. In particular, empirical clustering studies focusing on banking issuers listed on the IDX using recent financial ratio data are still scarce.

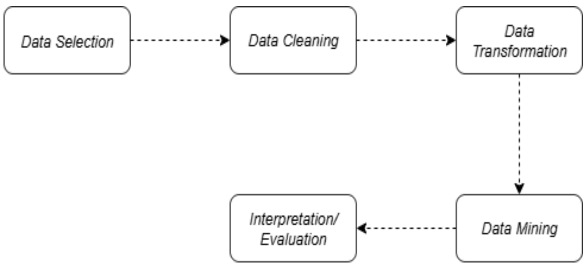
To address this gap, this study applies the K-Means clustering algorithm within the KDD framework to group banking sector issuers listed on the Indonesia Stock Exchange based on financial ratios (Wibowo and Sasongko 2022). K-Means is widely recognized for its effectiveness in grouping numerical data by minimizing intra-cluster distance and maximizing inter-cluster separation (Tohendry and Jollyta 2023). Compared to alternative clustering techniques such as K-Medoids, K-Means offers computational efficiency and suitability for medium-sized numerical datasets (Salman et al. 2025). The novelty of this research lies in the implementation of K-Means clustering on averaged financial ratios (ROA, ROE, and LDR) of banking issuers for the 2022–2024 period, providing an objective segmentation of banking performance using a structured KDD process.

Therefore, this study aims to cluster banking sector issuers listed on the Indonesia Stock Exchange based on ROA, ROE, and LDR ratios using the K-Means algorithm to identify homogeneous performance groups. The results are expected to provide objective insights into banking competitiveness, assist investors in evaluating financial health and risk levels, and contribute to the development of data-driven financial performance analysis in the Indonesian capital market context.

MATERIALS AND METHODS

This study employed a quantitative research design to analyze and cluster the financial performance of banking sector issuers listed on the Bursa Efek Indonesia during the 2022–2024 period. The research object consisted of annual financial statement data obtained from banking companies that met the criteria of data completeness and availability throughout the observation period. Issuers with incomplete financial reports were excluded to ensure dataset consistency and reliability of the clustering results.

The research methodology applied in this study was Knowledge Discovery in Databases (KDD), which consists of five main stages: data selection, data cleaning, data transformation, data mining, and interpretation/evaluation. These stages were conducted sequentially to ensure systematic data processing and valid clustering outcomes. The overall stages of the research methodology are illustrated in Figure 1.



Source: (Algoritma et al. 2023)

Figure methodology Knowledge Discovery in Databases (KDD)

In the data selection stage, annual financial reports of banking issuers for the 2022–2024 period were collected from officially published financial statements. The dataset included financial ratios representing key aspects of banking performance, particularly profitability and liquidity. In the data cleaning stage, the dataset was examined to remove missing values and eliminate duplicate records to ensure data accuracy and consistency before further analysis.

After the data cleaning process, three financial ratio features were retained as clustering variables, namely Return on Assets (ROA), Return on Equity (ROE), and Loan to Deposit Ratio (LDR). These ratios were selected because they represent key dimensions of banking financial performance, including profitability, capital efficiency, and liquidity management. ROA reflects the bank's ability to generate profit from its total assets, while ROE measures the effectiveness of equity utilization in generating returns for shareholders. LDR indicates the proportion of loans provided relative to the deposit base and represents the bank's liquidity management and credit distribution capacity.

Although other indicators such as Non-Performing Loans (NPL) and Capital Adequacy Ratio (CAR) are commonly used in banking performance analysis, they were not included in this study to maintain model simplicity and to focus on core indicators directly related to profitability and lending activities. In addition, preliminary data exploration indicated that the selected ratios were sufficient to represent variations in financial performance among the observed banking issuers.

The data mining stage involved implementing the K-Means clustering algorithm to group banking issuers based on similarities in their financial characteristics. Prior to clustering, the optimal number of clusters was determined using the Elbow method. The Euclidean Distance metric was used to measure similarity between data points and cluster centroids. The analysis was conducted using Google Colaboratory with the Python programming language and relevant machine learning libraries.

Although alternative clustering algorithms such as K-Medoids and Hierarchical Clustering are commonly used in clustering analysis, this study focuses on the K-Means algorithm due to its computational efficiency and effectiveness in handling numerical datasets. K-Means is widely applied in financial data clustering because it performs well in identifying patterns in structured numerical variables such as financial ratios. In this study, the dataset consists of normalized numerical variables (ROA, ROE, and LDR), making K-Means suitable for grouping issuers based on similarity in financial characteristics. Furthermore, the evaluation using the Silhouette Coefficient indicated that the clustering structure obtained in this study demonstrates a satisfactory level of cluster separation and cohesion.

Finally, in the interpretation and evaluation stage, the clustering results were assessed using the Silhouette Coefficient to measure the quality of the formed clusters. A higher silhouette value indicates better separation and cohesion between clusters,

Vol. 23, No. 1 March 2026 | DOI https://doi.org/10.33480/techno.v23i1.8334

demonstrating the effectiveness of the clustering model in identifying homogeneous groups of banking issuers.

RESULTS AND DISCUSSION

This study applies the Knowledge Discovery in Databases (KDD) methodology combined with the K-Means clustering algorithm to classify banking sector issuers listed on the Indonesia Stock Exchange based on financial performance. The variables used in this study consist of Return on Assets (ROA), Return on Equity (ROE), and Loan to Deposit Ratio (LDR) during the 2022–2024 period.

The dataset used in this study was obtained from the annual financial reports of 25 banking issuers for the 2022–2024 period. Each issuer was analyzed using three financial ratios: ROA, ROE, and LDR. With 25 issuers, three variables, and three years of observation, a total of 225 numerical data points were processed in this study.

The number of issuers included in this study reflects banking companies listed on the Indonesia Stock Exchange that met the criteria of data completeness during the observation period. Although the dataset consists of 25 issuers, this sample represents a substantial portion of the banking sector population listed on the IDX, making it relevant for sector-level analysis. In the context of financial data analysis, clustering is often applied to relatively small datasets when the objective is to identify structural patterns within a specific industry group rather than to develop large-scale predictive models. Furthermore, the use of normalized financial ratios and cluster validation using the Silhouette Coefficient helps ensure that the resulting clusters maintain an acceptable level of separation and cohesion despite the limited sample size.

The dataset used in this study was obtained from the annual financial reports of 25 banking issuers for the 2022–2024 period. Each issuer was analyzed using three financial ratios: ROA, ROE, and LDR. With 25 issuers, 3 variables, and 3 years of observation, a total of 225 numerical data points were processed in this study.

Table 1. Financial Ratio Dataset of Banking Issuers in 2022

Emiten (Kode)	ROA	ROE	LDR
BBCA	0.031	0.1842	0.676
BBRI	0.0274	0.1687	1.0972
BBHI	0.0244	0.0421	1.4927
BBNI	0.0179	0.1318	1.0516
BBTN	0.0076	0.1176	2.2356
BDMN	0.0167	0.0696	0.9172
PNBN	0.0154	0.0645	0.8832
AGRO	0.00082	0.00338	0.7521
BABP	0.0031	0.0194	0.7754
BBKP	-0.0559	-0.4487	0.9735
BMRI	0.0226	0.1782	0.8552
BJBR	0.0124	0.1523	1.025
BJTM	0.015	0.1348	0.7146
BNBA	0.0047	0.0127	0.9471
BNGA	0.0166	0.1126	0.8413
BRIS	0.0642	0.5857	0.8905
BVIC	0.0087	0.0611	0.8335
INPC	0.0022	0.0137	0.5328
MEGA	0.0286	0.1964	0.6826
NISP	0.0139	0.0972	0.7597
AMAR	-0.0345	-0.0489	2.2489
BGTG	0.0051	0.0147	0.5269
BBMD	0.0315	0.1149	0.8326
BNLI	0.0079	0.0535	0.7455
BSIM	0.0047	0.0304	0.4236

Source: (Research Results, 2026)

Table 2. Financial Ratio Dataset of Banking Issuers in 2023

Emiten (Kode)	ROA	ROE	LDR
BBCA	0.0346	0.23	0.7
BBRI	0.0308	0.1909	1.1667
BBHI	0.0349	0.134	1.9387
BBNI	0.0194	0.1364	0.8574
BBTN	0.008	0.1149	0.1107
BDMN	0.0165	0.0732	
PNBN	0.0117	0.0511	0.883
AGRO	0.00196	0.00712	0.7797
BABP	0.0217	0.0043	
BBKP	-0.0718	-0.4291	0.9394
BMRI	0.0276	0.2089	1.0062
BJBR	0.0089	0.1088	0.9887
BJTM	0.0142	0.1209	0.7989
BNBA	0.0056	0.0142	0.9502
BNGA	0.0196	0.1328	0.874
BRIS	0.0143	0.2473	0.8732
BVIC	0.0034	0.0266	0.8816
INPC	0.0056	0.0352	0.6151
MEGA	0.0266	0.598	0.741
NISP	0.0164	0.1096	0.8478
AMAR	0.054	0.0406	2.536
BGTG	0.0111	0.0321	0.732
BBMD	0.026	0.0851	0.8966
BNLI	0.01	0.0646	0.7464
BSIM	0.0014	0.0097	0.3853

Source: (Research Results, 2026)

Table 3. Financial Ratio Dataset of Banking Issuers in 2024

Emiten (Kode)	ROA	ROE	LDR
BBCA	0.0378	0.2087	0.7767
BBRI	0.0304	0.1876	0.8945
BBHI	0.0334	0.0643	1.2155
BBNI	0.0192	0.1296	0.9151
BBTN	0.0064	0.0923	0.9063
BDMN	0.0136	0.0635	1.1734

Emiten (Kode)	ROA	ROE	LDR
PNBN	0.0117	0.0511	0.883
AGRO	0.0039	0.0147	0.7475
BABP	0.0036	0.0204	0.7611
BBKP	-0.0762	-0.7948	0.8855
BMRI	0.0304	0.1876	0.8945
BJBR	0.0142	0.1035	0.7581
BJTM	0.0171	0.1297	0.5501
BNBA	0.0009	0.0041	1.1636
BNGA	0.0157	0.1262	0.5899
BRIS	0.0149	0.1698	1.0315
BVIC	0.004	0.0259	0.4861
INPC	0.0023	0.0125	0.8529
MEGA	0.0266	0.1395	0.5346
NISP	0.0155	0.1054	0.6748
AMAR	0.0442	0.0644	2.3525
BGTG	0.0195	0.0585	0.7139
BBMD	0.0243	0.0793	1.0528
BNLI	0.0138	0.0837	0.889
BSIM	0.0067	0.0431	0.3582

Source: (Research Results, 2026)

The dataset shows variations in profitability and liquidity among issuers. Several issuers exhibit negative ROA and ROE values, indicating suboptimal financial performance during certain periods. At the data cleaning stage, missing value and duplicate data checks were conducted.

Table 4. Missing Value And Duplcate Data Checking Results

Year	Missing Values (Emiten)	Missing Values (ROA)	Missing Values (ROE)	Missing Values (LDR)	Duplicate Data
2022	0	0	0	0	0
2023	0	0	0	0	0
2024	0	0	0	0	0

Source: (Research Results, 2026)

The results indicate that two missing values were found in the LDR variable for the 2023 dataset, while no missing values were identified in 2022 and 2024. The missing values were retained in their original condition to maintain data authenticity and avoid bias caused by imputation. Next, duplicate data checking was performed. The duplicate checking process resulted in a value of 0, indicating that no duplicate records were found in the dataset. The transformation stage involved calculating the average value of ROA, ROE, and LDR for each issuer across the 2022–2024 period. If a value was unavailable in one period, the average was calculated based on the available data.

Table 5. Average ROA, ROE, and LDR Results

No	Emiten (Kode)	ROA	ROE	LDR
0	AGRO	0.002227	0.0084	0.75977
1	AMAR	0.021233	0.0187	2.37913
2	BABP	0.009467	0.0147	0.76825
3	BBCA	0.034467	0.207633	0.71757
4	BBHI	0.0309	0.080133	1.54897
5	BBKP	-0.06797	-0.55753	0.9328
6	BBMD	0.027267	0.0931	0.92733
7	BBNI	0.018833	0.1326	0.94137
8	BBRI	0.029533	0.1824	1.0528
9	BBTN	0.007333	0.108267	1.0842
10	BDMN	0.0156	0.068767	1.0453
11	BGTG	0.0119	0.0351	0.6576
12	BJBR	0.011833	0.121533	0.92393
13	BJTM	0.015433	0.128467	0.68787
14	BMRI	0.026867	0.191567	0.91863
15	BNBA	0.003733	0.010333	1.0203
16	BNGA	0.0173	0.123867	0.7684
17	BNLI	0.010567	0.067267	0.79363
18	BRIS	0.031133	0.334267	0.93173
19	BSIM	0.004267	0.027733	0.38903
20	BVIC	0.005367	0.037867	0.73373
21	INPC	0.003367	0.020467	0.66693
22	MEGA	0.027267	0.3113	0.65273
23	NISP	0.015267	0.104067	0.76077
24	PNBN	0.012933	0.055567	0.88307

Source: (Research Results, 2026)

After obtaining the average values, Min-Max Normalization was applied to transform all variables into a range between 0 and 1. This normalization ensures that no variable dominates the distance calculation in the K-Means algorithm.

Table 6. Min-Max Normalization Results

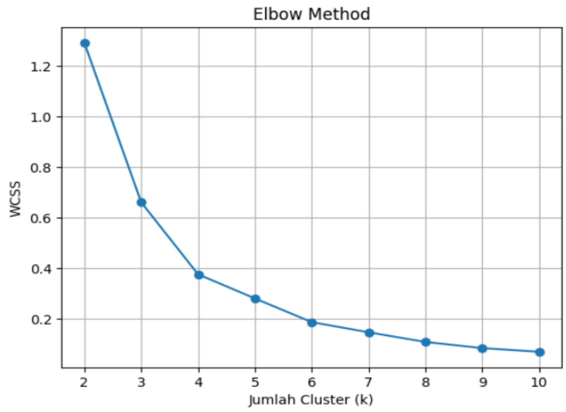
No	Emiten (Kode)	ROA	ROE	LDR
0	AGRO	0.68526	0.6346	0.18629
1	AMAR	0.87081	0.64615	1
2	BABP	0.75594	0.64166	0.19055
3	BBCA	1	0.858	0.16508
4	BBHI	0.96518	0.71503	0.58285
5	BBKP	0	0	0.27324
6	BBMD	0.92971	0.72957	0.27049
7	BBNI	0.84738	0.77387	0.27754
8	BBRI	0.95184	0.82971	0.33353
9	BBTN	0.73511	0.74658	0.34931
10	BDMN	0.81582	0.70229	0.32977
11	BGTG	0.77969	0.66454	0.13495
12	BJBR	0.77904	0.76146	0.26878
13	BJTM	0.81419	0.76923	0.15016
14	BMRI	0.92581	0.83999	0.26612
15	BNBA	0.69997	0.63677	0.3172
16	BNGA	0.83241	0.76407	0.19063
17	BNLI	0.76668	0.70061	0.20331
18	BRIS	0.96746	1	0.2727
19	BSIM	0.70517	0.65628	0
20	BVIC	0.71591	0.66764	0.17321
21	INPC	0.69639	0.64813	0.13964
22	MEGA	0.92971	0.97425	0.13251
23	NISP	0.81256	0.74187	0.18679
24	PNBN	0.78978	0.68749	0.24825

Source: (Research Results, 2026)

The determination of the optimal number of clusters was carried out using the Elbow Method by analyzing the Within Cluster Sum of Squares (WCSS). Mathematically, WCSS is defined as:

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2 \qquad (1)$$

where k represents the number of clusters, C_i is cluster i, x is a data point, and μ_i is the centroid of cluster i. The Elbow graph shows a significant decrease in WCSS from k = 2 to k = 3, after which the reduction becomes relatively stable.



Source: (Research Results, 2026)
Figure 2. Elbow Method Graph

Based on the elbow point observed at k = 3, three clusters were selected as the optimal number for classification. The K-Means algorithm with k = 3 successfully grouped 25 banking issuers into three clusters based on the similarity of their financial performance.

Table 7. K-Means Clustering Results

No	Emiten (Kode)	ROA	ROE	LDR
0	AGRO	0.68526	0.6346	0.18629
1	AMAR	0.87081	0.64615	1
2	BABP	0.75594	0.64166	0.19055
3	BBCA	1	0.858	0.16508
4	BBHI	0.96518	0.71503	0.58285
5	BBKP	0	0	0.27324
6	BBMD	0.92971	0.72957	0.27049
7	BBNI	0.84738	0.77387	0.27754
8	BBRI	0.95184	0.82971	0.33353
9	BBTN	0.73511	0.74658	0.34931
10	BDMN	0.81582	0.70229	0.32977
11	BGTG	0.77969	0.66454	0.13495
12	BJBR	0.77904	0.76146	0.26878
13	BJTM	0.81419	0.76923	0.15016
14	BMRI	0.92581	0.83999	0.26612
15	BNBA	0.69997	0.63677	0.3172
16	BNGA	0.83241	0.76407	0.19063
17	BNLI	0.76668	0.70061	0.20331
18	BRIS	0.96746	1	0.2727
19	BSIM	0.70517	0.65628	0
20	BVIC	0.71591	0.66764	0.17321
21	INPC	0.69639	0.64813	0.13964
22	MEGA	0.92971	0.97425	0.13251
23	NISP	0.81256	0.74187	0.18679
24	PNBN	0.78978	0.68749	0.24825

Source: (Research Results, 2026)

Table 8 presents the distribution of issuers in each cluster

Table 8. Distribution of Banking Issuers in Each Cluster

Cluster	Issuers (Emiten)
0	AMAR, BBHI
1	AGRO, BABP, BBCA, BBMD, BBNI, BBRI, BBTN, BDMN, BGTG, BJBR, BJTM, BMRI, BNBA, BNGA, BNLI, BRIS, BSIM, BVIC, INPC, MEGA, NISP, PNBN
2	BBKP

Source: (Research Results, 2026)

The centroid values of each cluster are shown in Table 9.

Table 9. Centroid Values of Each Cluster

Cluster	ROA	ROE	LDR
0	0.917995	0.68059	0.791426
1	0.815265	0.746753	0.217582
2	0	0	0.273236

Source: (Research Results, 2026)

Cluster 1 contains the majority of issuers, indicating that most banking companies share relatively similar financial characteristics. Cluster 0 represents issuers with relatively aggressive lending behavior combined with strong profitability. Cluster 2 consists of a single issuer with significantly lower financial performance compared to others. To evaluate clustering quality, the Silhouette Coefficient method was applied.

The evaluation results using the Silhouette Coefficient show a value of **0.5946**, indicating a reasonably good clustering structure, with sufficient cohesion within clusters and clear separation between clusters. This suggests that the K-Means algorithm is able to group banking issuers effectively based on the similarity of their financial characteristics.

Furthermore, the results demonstrate that the integration of the KDD (Knowledge Discovery in Databases) methodology and the K-Means algorithm successfully classifies banking issuers based on fundamental financial ratios. The formation of three clusters indicates the presence of distinct financial performance patterns among banking companies. These findings are in line with previous studies, which state that K-Means is effective in grouping numerical financial data based on similarity measures.

The use of financial ratios such as Return on Assets (ROA), Return on Equity (ROE), and Loan to Deposit Ratio (LDR) provides an objective and fundamental basis for issuer classification, beyond relying solely on technical market indicators. From a practical perspective, this clustering approach can assist investors in identifying issuers with similar financial performance profiles, thereby supporting more informed and data-driven investment decision-making. However, this study is limited to a three-year observation period and only considers three financial variables. Future research is recommended to incorporate additional financial indicators and extend the observation period in order to produce more comprehensive and robust clustering results.

CONCLUSION

This study concludes that the implementation of the K-Means algorithm using financial ratios Return on Assets (ROA), Return on Equity (ROE), and Loan to Deposit Ratio (LDR) through the Knowledge Discovery in Databases (KDD) stages successfully classified 25 banking issuers listed on the Indonesia Stock Exchange into three clusters. The clustering results consist of a dominant cluster containing 22 issuers with relatively stable financial performance, a cluster of two issuers characterized by higher lending distribution and relatively higher risk, and a single-issuer cluster reflecting the lowest financial performance. These findings indicate that financial ratio-based clustering can provide a clearer overview of issuer characteristics and support more informed investment decision-making.

The evaluation using the Silhouette Coefficient produced a value of 0.5946, indicating that the clustering structure has a reasonably good level of cohesion and separation. This result demonstrates that the K-Means algorithm is capable of generating meaningful groupings and can serve as an analytical tool for assessing banking sector performance. However, this study is limited to a three-year observation period and three financial variables. Future research is recommended to incorporate additional financial indicators and longer observation periods to enhance the robustness of the clustering results.

REFERENCE

Agustina, Dewi, Alean Kistiani, Hegy Suryana, And Asih Triatmaja. 2022. “ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI HARGA SAHAM PADA PERUSAHAAN SEKTOR PERBANKAN YANG TERDAFTAR DI BURSA EFEK INDONESIA (BEI) TAHUN 2016-2020.” 10(2): 185–94.

Algoritma, C, Untuk Prediksi, Evaluasi Anggota, Satuan Pengamanan, Studi Kasus, And P T Yimm. 2023. “Jurnal Indonesia : Manajemen Informatika Dan Komunikasi IMPLEMENTASI DATA MINING MENGGUNAKAN Abstrak Jurnal Indonesia : Manajemen Informatika Dan Komunikasi.” 4(3): 1566–80.

Almunawwaroh, Medina. 2022. “Studi Faktor-Faktor Yang Mempengaruhi Profitabilitas Bank Umum Syariah Di Indonesia.” 4(10): 4518–22.

Dalimunthe, N, And N K Lubis. 2023. “Peran Lembaga Perbankan Terhadap Pembangunan Ekonomi: Fungsi Dan Tujuannya Dalam Menyokong Ketenagakerjaan.” *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah* 8(30): 956–63. Https://Journal.Um-Surabaya.Ac.Id/Mas/Article/View/20997%0Ahttps://Journal.Um-Surabaya.Ac.Id/Mas/Article/Download/20997/7132.

Effendi, Meredith Fiona, Celina Tri Siwi, And Hermanto Silalahi. 2024. “Informasi Yang Menyesatkan Dalam Prospektus : Bagaimanakah Tanggung Jawab Hukum Emiten ? Masyarakat . Karena Pasar Modal Yakni Sarana Yang Dapat Menunjang Keuangan Dalam Waktu.” 2(1).

Jurnal, Jimea, And Ilmiah Mea. 2024. “ANALISIS TEKNIKAL SAHAM ENERGI MENGGUNAKAN JIMEA | Jurnal Ilmiah MEA (Manajemen , Ekonomi , Dan Akuntansi).” 8(2): 187–206.

Lusiana, Sindi Ayu, Lukita Tripermata, And Andini Utari Putri. 2025. “Analisis Fundamental Dan Teknikal Terhadap Return Saham Pada Perusahaan Multisektor IDX 30 Yang Terdaftar Di BEI.” 4(4): 873–79. Doi:10.54259/Akua.V4i4.5646.

Muzahid, Mukhlisul, Aryati Ismet Wijaya, And Mauli Isra. 2023. “ANALISA GOOD CORPORATE GOVERNANCE TERHADAP KINERJA PERUSAHAAN PADA EMITEN JAKARTA ISLAMIC INDEX 70.” 25(1): 9–18.

Rasyid, Harun Al, And Suryanto Sosrowidigdo. 2022. “Pengaruh Return On Assets (ROA) Dan Return On Equity (ROE) Terhadap Kinerja Kesehatan Capital Adequacy Ratio (CAR) Pada Bank BTPN.” 6(6): 620–31.

Saerang, Ivonne, Parengkuan Tommy, And Mario Christiano. 2014. “Analisis Terhadap Rasio-Rasio Keuangan Untuk Mengukur Profitabilitas Pada Bank-Bank Swasta Yang Go Public Di Bursa Efek Indonesia.” *Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*

2(4): 817–30.

Sahrul, Muhammad. 2025. “Sentimen Analisis Pada Ulasan Aplikasi Indodana Di Google Play Store Menggunakan Algoritma Logistic Regression , Naïve Bayes Dan Support Vector Machine Guna Memperoleh Informasi Mengenai Sentimen Yang Ada Dalam Suatu Pendapat . Penelitian.” 5: 136–47.

Salman, Muhammad Dzaki, Rahmaddeni Rahmaddeni, Nanda Rizki Pratama, M Nakhlah Farid A, Ahmad Agung Setiawan, Fenisya Zalianti, And Isra Bil Huda. 2025. “Perbandingan Kinerja Algoritma Clustering K-Means Dan K-Medoids Dalam Pengelompokan Sekolah Di Provinsi Riau Berdasarkan Ketersediaan Sarana Dan Prasarana: Comparison Of K-Means And K-Medoids Clustering Algorithm Performance In Grouping Schools In Riau Pr.” *MALCOM: Indonesian Journal Of Machine Learning And Computer Science* 5(3): 797–806. Https://Www.Journal.Irpi.Or.Id/Index.Php/Malcom/Article/View/1950.

Srimurdianti, Anggi, Sukamto, Wawan Setiawan, Enda Esyudha, And Pratama. 2023. “JEPIN (Jurnal Edukasi Dan Penelitian Informatika) Data Mining Untuk Pengelompokan Saham Pada Sektor Energi Dengan Metode K-Means.” *Jurnal Edukasi Dan Penelitian Informatika* 9(1): 76–81. Https://Jurnal.Untan.Ac.Id/Index.Php/Jepin/Article/Viewfile/62509/75676597024.

Tohendry, Derrick, And Deny Jollyta. 2023. “Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Saham Berdasarkan Price Earning Ratio Dan Price To Book Value.” *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi* 5(1): 3–9.

Trees, Decision, Random Forest, And D A N K-Nearest Neighbours. 2024. “Journal Of Social And Economics Research ANALYSIS OF FACTORS INFLUENCING THE DECISION TO PURCHASE TOUR PACKAGES USING THE DECISION TREES , RANDOM FOREST AND K-NEAREST.” 6(2): 25–39.

Wibowo, Arief, And Raden Sasongko. 2022. “PENERAPAN DATA MINING PADA SUKU BUNGA INVESTASI DEPOSITO DI INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PROFITABILITAS (Application Of Data Mining On Deposit Investment Rate In Indonesia Using K-Means Clustering Method For Profitab.” *Ascarya* 2(1): 70–80. Https://Doi.Org/10.53754/Iscs.V2i1.369.