Journal of Information Systems and Informatics

Vol. 6, No. 3, September 2024 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v6i3.809 Published By DRPM-UBD

Stock Grouping Based on Price Earnings Ratio and Price Book Value Using K-Medoids Algorithm

Muhammad Oemar Abdillah[1], Raissa Amanda Putri[2]

[1,2]Information Systems, State Islamic University of North Sumatra, Medan, Indonesian
Email: [1]muhammadoemarabdillah@gmail.com, [2]raissa.ap@uinsu.ac.id

Abstract

Investing involves allocating funds to achieve optimal returns by evaluating opportunities and managing risks in asset acquisition. Recently, many news reports have highlighted issues in the Indonesian capital market, such as stock investors using online loan funds for trading, which often leads to debt. This research aims to apply the K-Medoids algorithm for stock clustering, enabling investors to select fundamentally sound stocks based on the Price Earnings Ratio (PER) and Price-Book Value (PBV). The K-Medoids method results show that Cluster 1 includes 93 stocks with moderate PER and PBV values. Cluster 2 comprises 91 stocks with the lowest PER and PBV values. Cluster 3 contains 113 stocks with the highest PER and PBV values. Developing a web-based system that classifies stocks based on Price Earnings Ratio (PER) and Price Book Value (PBV) can help investors analyze and make investment decisions more effectively.

Keywords: Investment, K-Medoids, Clustering, Stock

1. INTRODUCTION

Investment refers to the allocation of funds with the objective of achieving optimal returns through the evaluation of opportunities and risk management associated with asset purchases. Currently, the investment environment is no longer dominated by conventional options such as deposits, savings, or gold. One attractive form of investment is through the purchase of company shares. When an individual buys shares of a company, they are investing capital with the expectation of earning profits through the sale of shares and dividends provided by the company. Investing in stocks has become a popular alternative for the public.

With the increasing number of new investors, many remain confused about how to start investing in stocks, which sectors to choose, and which companies are worthy of allocating some of their funds [1]. Recently, numerous news reports have highlighted issues related to the Indonesian capital market, such as the behavior of stock investors using online loan funds to trade, ultimately leading to debt. Additionally, there are cases of stock investors buying shares based on hearsay, which subsequently results in financial losses[2]. In making stock

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

investment decisions, novice investors tend to select stocks randomly without conducting proper analysis. This often results in the stock prices decreasing when purchased and increasing when sold [3]. One approach to address this issue is by clustering stocks based on their value, distinguishing between cheap, moderate, and expensive stocks using clustering techniques. The outcome of this data clustering process yields several groups, including clusters of stocks categorized as cheap, moderate, and expensive.

Price Book Value (PBV) measures the relationship between the market value of a stock and its book value, or the value of assets as reported in the company's financial statements. The value of a company reflects the long-term prosperity of its shareholders, as the primary goal of the company is to enhance its value [4]. This ratio can be utilized to determine a stock's value, as the market value of a stock should reflect its book value [5].

Price Earnings Ratio (PER) is a ratio that describes the relationship between the market price of a stock and its earnings per share (EPS). PER is calculated by determining the intrinsic value of a stock and then comparing it to the market price of the stock to ascertain whether the stock is overvalued, correctly valued, or undervalued. [6]. The higher the Price-to-Earnings Ratio (PER), the more expensive the company's valuation. Conversely, the lower the PER, the cheaper the company's valuation. The PER also reflects the growth of the stock [7].

Data mining is the process of collecting and analyzing historical data to discover patterns, relationships, and regularities in large datasets. Prediction, comparison, classification, clustering, and estimation are some of the approaches used in data mining [8]. Clustering is the process of grouping data into various categories such that the data within a single category have a high degree of similarity. This grouping process is unsupervised, meaning that it lacks predefined labels, and therefore, all attributes within a category are considered identical [9].

The K-Medoids data mining algorithm, also known as PAM (Partitioning Around Medoids), uses medoids as cluster representatives. The K-Medoids algorithm is a clustering technique that can be used to group objects into clusters with similar or identical characteristics [10][11]. This is different from the K-Means algorithm, which uses the mean value as the cluster center [12]. K-Medoids can also address the shortcomings of the K-Means method, which is sensitive to outliers [13].

Research utilizing the k-medoids algorithm in the Clustering of Capture Fisheries Production was conducted by Fajrian [14]. In this study, the potential production of Albakora fish catches was classified into the moderate cluster. The catches of Alu-alu, Tongkol Krai, Bigeye Tuna, and Southern Bluefin Tuna were categorized into the low cluster. Meanwhile, the catches of Banyar, Black Pomfret, and White

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Pomfret were classified into the high cluster. Another study by Mahesa [15] applied the k-medoids algorithm in the Clustering of Vaccine and Serum Sales. This research employed the k-medoids method to aid in clustering vaccine and serum sales, enabling the process to be conducted without the need to analyze the specific requirements for vaccines and serums at each hospital.

Based on the issues outlined in the background, the researcher decided to perform stock clustering based on Price-Earnings Ratio (PER) and Price-Book Value (PBV) using a web-based approach. The objective of this study is to apply the K-Medoids algorithm for clustering stocks based on PER and PBV to provide fundamentally sound stock recommendations and assist investors in making investment decisions. Clustering is essential for grouping stocks into categories of cheap, moderate, and expensive, thereby facilitating analysis and investment decision-making.

2. METHODS

This research employs the Research and Development (R&D) methodology to design and build an information system for clustering stocks based on PER and PBV using the K-Medoids algorithm. Research and Development (R&D) is a series of processes or stages aimed at developing or improving a product, whether it is an existing one or a new one [16].

2.1. Data Collection

In this study, the researcher collected data from financial sources through the Indonesia Stock Exchange website, www.idx.co.id. Once the data was gathered, it was essential to understand the data by considering the available and required data attributes, as well as the quality of the data.

2.2. Waterfall Development Method

The Waterfall method is one of the most commonly used approaches in system development. This method excels because it provides a structured and orderly development process. Moreover, its straightforward structure facilitates good system documentation. The steps of waterfall method as shown in Figure 1. Base on Figure 1 can be explained as follows.

1) Requirement Analysis: This initial stage aims to determine the necessary system design, including the menus required by users to build the information system. During this phase, the requirements of users and other stakeholders are collected and analyzed to ensure that all system needs are identified and documented in detail.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

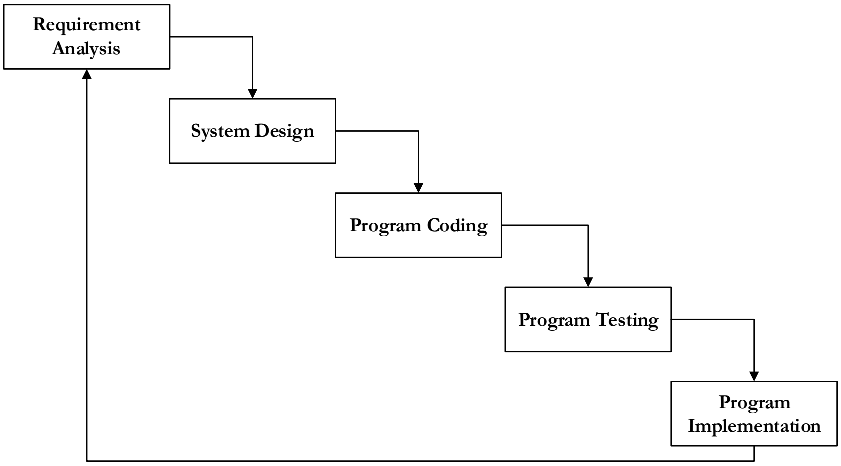

Figure 1. Waterfall Method [17]

2) System Design: The design stage is the phase of creating the blueprint for the system to be developed. The tool used in this system design process is UML (Unified Modeling Language), particularly the Use Case Diagram. The Use Case Diagram is utilized to illustrate the interaction between actors (users) and the system, as well as to define the main functionalities that the system must provide.
3) Program Coding: Program coding is the phase of translating the system design into instructions that can be understood by a computer. In this stage, the researcher writes program code using the PHP programming language and Phpmyadmin as the database, based on the specifications and design previously created. This process includes developing algorithms, data structures, and the necessary program modules.
4) Program Testing: In this stage, all input and output processes are tested to detect errors and bugs in the program code. Testing is conducted to ensure that the system operates according to the specified requirements and is free from errors that could impact the system's performance.
5) Program Implementation: Program implementation is the final phase in which the researcher deploys the fully developed and tested system. After implementation, the system is ready for use by the users.

2.3. K-Medoids Algorithm

The K-Medoids algorithm is a clustering technique used to partition a set of data into several groups (clusters) based on the similarity among the data points. K-Medoids utilizes medoids, which are data points located at the center of a cluster and have the minimum total distance to all other points within that cluster. The

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

process of clustering data using the K-Medoids algorithm involves several stages. The following outlines the steps of the K-Medoids algorithm:

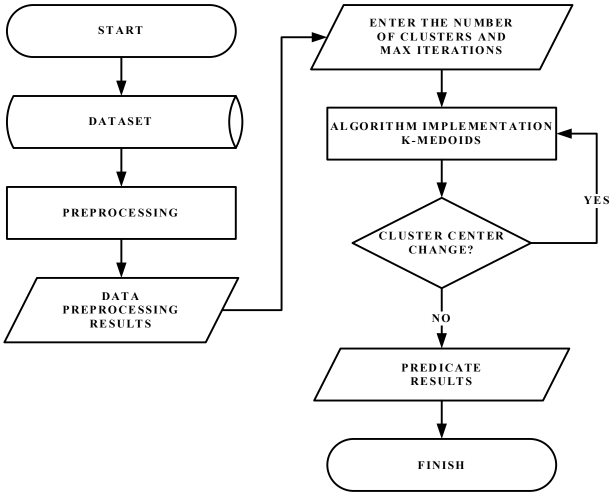

Figure 2. K-Medoids Clustering Algorithm

1) Start: The process begins with preparing the dataset to be used.
2) Data Preprocessing: During the data preprocessing stage, normalization of the dataset is performed to ensure that the dataset has consistent values and a reasonable range [18]. One commonly used formula for data normalization is Min-Max Normalization. The formula is as shown in Equation 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where:
(X') is the normalized data value.
(X) is the original data value.
($X_{\min}$) is the minimum value in the dataset.
($X_{\max}$) is the maximum value in the dataset.
3) Specify the Number of Clusters and Maximum Iterations: Determine the desired number of clusters and the maximum number of iterations.
4) Algorithm Implementation K-Medoids: This stage involves applying the K-Medoids algorithm to data that has undergone preprocessing, specifically min-max normalization, and determining the number of clusters and the maximum number of iterations. The K-Medoids algorithm will produce data clusters that will be analyzed in the subsequent stage.

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

5) The following formula is used in the K-Medoids algorithm for clustering: Selection of Initial Medoids: Randomly select k objects from the dataset as the initial medoids. Cluster Assignment: Assign each object in the dataset to the cluster with the nearest medoid, based on a specific distance metric (e.g., Euclidean distance). Calculation of New Medoids: For each cluster, compute a new medoid by minimizing the total distance between objects within the cluster and the medoid. Iteration: Repeat steps 2 and 3 until the medoids no longer change or the maximum number of iterations is reached. The Equation 2 for the Euclidean distance between two points (x) and (y) is:

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{2}$$

6) Has the Cluster Center Changed? Check whether the cluster centers have changed after the iteration. Yes: If the cluster centers have changed, return to the algorithm implementation step. No: If the cluster centers have not changed, proceed to the next step.
7) Result Prediction: Display the final outcome of the data clustering.
8) Completion: The process is complete.

3. RESULTS AND DISCUSSION

3.1. System Design

Use case diagram is a model for the behavior of the information system to be made, use case diagrams are used to find out what functions are in the system and who are entitled to use these functions [19]. The use case diagram in this section explains the interactions between the actors, namely the admin and the user, in utilizing the system as shown in Figure 3.

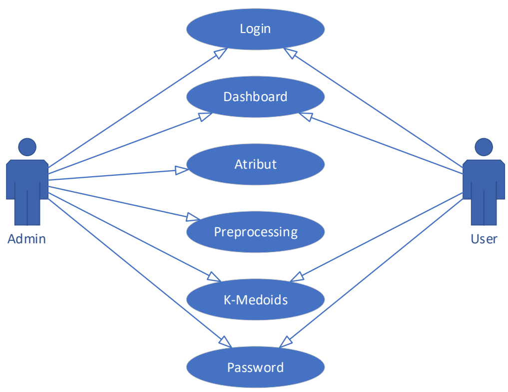

Figure 3. Use Case Diagram

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.2. Data Collection

The dataset utilized in this research is sourced from the Indonesian Stock Exchange website, specifically at www.idx.co.id. The stock data was retrieved on July 22, 2024.

3.3. Understanding Data

The data used consists of 295 rows with attributes: Symbol, Price, PER, PBV, Market Cap. In this study, clustering will only utilize two attributes, namely PER and PBV. Table 1 is an some of the data for analysis, presented in Table 2.

Table 1. Dataset

No	Symbol	Price	PER	PBV	Market Cap
1	AALI	5,800.00	10.51	0.5	11,163.19 B
2	ABMM	3,910.00	2.93	0.84	10,764.88 B
3	ACES	815	17.26	2.2	13,977.25 B
4	ACST	113	-5.07	4.32	1,432.29 B
5	ADCP	50	10.32	0.43	1,111.11 B
...	...	...	...	...	...
293	WINS	525	16.93	1.05	2,291.54 B
294	WIRG	90	13.6	1.56	1,074.48 B
295	WOMF	362	5.33	0.75	1,260.30 B
296	WOOD	224	12.73	0.34	1,442.00 B
297	WSKT	202	-1.34	1.31	5,818.98 B

The data in Table 1 is not processed immediately because the dataset is too large and contains significant numerical disparities. Therefore, data normalization is necessary. Table 2 displays the results of the normalization.

Table 2. Normalization Dataset

No	PER	PBV
1	0.5483	0.0154
2	0.3715	0.0309
3	0.7057	0.0926
4	0.1849	0.1888
5	0.5438	0.0123

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.4. Implementation K-Medoids

When the K-Medoids algorithm is applied, the data obtained with the condition that the desired clustering is 3 is used to generate the midpoint or centroid value. The following medoids were selected and used in the initial iterations in Table 3.

Table 3. Iteration 1 Medoids Center

Name	PER	PBV
C1	0.6203	0.0291
C2	0.75	0.1443
C3	0.8505	0.1039

To perform clustering on each data that has been obtained, the next step is to calculate the distance equation using the Euclidean Distance method. After calculating the distance for each data point, the data is grouped according to its cluster. The cluster group of a data point is determined by the nearest distance from the data to a cluster. The nearest distance of each data point to each cluster will be obtained in Table 4.

Table 4. Distance to Center Medoids 1

Code	PER	PBV	C1	C2	C3	Cluster
D001	0.5483	0.0154	0.0733	0.2394	0.3149	C1
D002	0.3715	0.0309	0.2488	0.3951	0.4846	C1
D003	0.7057	0.0926	0.1064	0.0681	0.1453	C2
D004	0.1849	0.1888	0.4638	0.5668	0.671	C1
D005	0.5438	0.0123	0.0783	0.2448	0.3201	C1
.....	.....	.....	.....	.....	.....	.....
D293	0.698	0.0404	0.0785	0.1162	0.1652	C1
D294	0.6203	0.0635	0.0345	0.1528	0.2337	C1
D295	0.4275	0.0268	0.1929	0.3433	0.43	C1
D296	0.6	0.0082	0.0291	0.2026	0.2682	C1
D297	0.2719	0.0522	0.3492	0.4869	0.5809	C1

Calculating the total deviation involves finding the difference between the new total cost and the old total cost. Previous Cost is 0, Cost Iteration 1 is 45.2466. Since the cost is now smaller than before, the iteration continues. After calculating the cost, next determine the value of new medoids (non-medoids). To determine the new medoids, new medoids are selected randomly, with the condition that once a medoid has been chosen, it cannot be selected again as a new medoid (non-medoids). Determination of New Medoid Values (Non-Medoid), the medoids selected and used are in the following Table 5.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Table 5. Iteration 2 Medoids Center

Name	PER	PBV
C1	0.4286	0.0554
C2	0.8144	0.0368
C3	0.7048	0.1307

To cluster the obtained data, the next step involves calculating the distance equation using the Euclidean Distance method. Once the distances for each data point are calculated, the data is then grouped into clusters. The cluster assignment for each data point is based on the shortest distance to a cluster. Table 6 display the nearest distances of each data point to the respective clusters.

Table 6. Distance to Center Medoids 2

Code	PER	PBV	C1	C2	C3	Cluster
D001	0.5483	0.0154	0.1261	0.2669	0.1944	C1
D002	0.3715	0.0309	0.0622	0.4429	0.3479	C1
D003	0.7057	0.0926	0.2795	0.1222	0.0381	C3
D004	0.1849	0.1888	0.2779	0.6475	0.5231	C1
D005	0.5438	0.0123	0.123	0.2716	0.1998	C1
.....	.....	.....	.....	.....	.....	.....
D293	0.698	0.0404	0.2698	0.1164	0.0906	C3
D294	0.6203	0.0635	0.1919	0.1959	0.1079	C3
D295	0.4275	0.0268	0.0286	0.387	0.2961	C1
D296	0.6	0.0082	0.1778	0.2162	0.1612	C3
D297	0.2719	0.0522	0.1567	0.5427	0.4399	C1

Calculating the total deviation involves finding the difference between the new total cost and the old total cost. Previous Cost: 45.2466, Cost Iteration 2: 33.4714. Since the cost is now smaller than before, the iteration continues. After calculating the cost, next determine the value of new medoids (non-medoids). Determining New Medoids Values (Non-Medoids) the medoids selected and used are in the following Table 7.

Table 7 Iteration 3 Medoids Center

Name	PER	PBV
C1	0.5222	0.0354
C2	0.7162	0.094
C3	0.4671	0.0418

The subsequent step involves computing the distance equation using the Euclidean Distance method during the 3rd iteration. After calculating the distance for each data point, the data is grouped into clusters. Each data point is assigned

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

to a cluster based on the shortest distance. Table 8 display the nearest distance of each data point to its respective cluster.

Table 8 Distance to Center Medoids 3

Code	PER	PBV	C1	C2	C3	Cluster
D001	0.5483	0.0154	0.0329	0.1854	0.0853	C1
D002	0.3715	0.0309	0.1507	0.3504	0.0962	C3
D003	0.7057	0.0926	0.1922	0.0106	0.2439	C2
D004	0.1849	0.1888	0.3705	0.5397	0.3182	C3
D005	0.5438	0.0123	0.0317	0.1907	0.0822	C1
.....	.....	.....	.....	.....	.....	.....
D292	0.2757	0.0831	0.2511	0.4407	0.1959	C3
D293	0.698	0.0404	0.1759	0.0566	0.2309	C2
D294	0.6203	0.0635	0.1021	0.1006	0.1548	C2
D295	0.4275	0.0268	0.0951	0.2964	0.0424	C3
D296	0.6	0.0082	0.0825	0.1444	0.1371	C1

Calculating the total deviation involves finding the difference between the new total cost and the old total cost. Previous Cost: 33.4714, Cost Iteration 3: 30.7988. Since the cost is now smaller than before, the iteration continues. After calculating the cost, next determine the value of new medoids (non-medoids). Determining New Medoids Values (Non-Medoids) the medoids selected and used are in the following Table 9.

Table 9 Iteration 4 Medoids Center

Name	PER	PBV
C1	0.5828	0.1262
C2	0.5147	0.0554
C3	0.6896	0.2778

After determining the New Medoid Value (Non-Medoid), The next step involves calculating the distance equation using the Euclidean Distance method. Once the distances for each data point are calculated, the data is then grouped into clusters. The cluster assignment for each data point is based on the shortest distance to a cluster. Table 10 will display the nearest distances of each data point to the respective clusters.

Table 10 Distance to Center Medoids 4

Code	PER	PBV	C1	C2	C3	Cluster
D001	0.5483	0.0154	0.116	0.0522	0.298	C2
D002	0.3715	0.0309	0.2318	0.1453	0.4027	C2
D003	0.7057	0.0926	0.1274	0.1946	0.1859	C1
D004	0.1849	0.1888	0.4028	0.3557	0.5124	C2
D005	0.5438	0.0123	0.1204	0.0521	0.3029	C2
.....	.....	.....	.....	.....	.....	.....
D293	0.698	0.0404	0.1436	0.1839	0.2376	C1

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Code	PER	PBV	C1	C2	C3	Cluster
D294	0.6203	0.0635	0.073	0.106	0.2252	C1
D295	0.4275	0.0268	0.1844	0.0918	0.3629	C2
D296	0.6	0.0082	0.1193	0.0975	0.2841	C2
D297	0.2719	0.0522	0.3196	0.2428	0.4747	C2

Calculating the total deviation involves finding the difference between the new total cost and the old total cost. Previous Cost: 30.7988, Cost Iteration 4: 38.3772. Because the cost is now greater than before, the iteration is stopped. The final iteration's outcome will serve as the clustering parameter. In this context, researchers identify which cluster members fall into the categories of expensive, medium, and cheap based on the centroid values: C1 represents medium, C2 represents cheap, and C3 represents high. Consequently, members of C1 belong to the medium cluster, members of C2 to the cheap cluster, and members of C3 to the high cluster. The cluster results using the k-medoids algorithm are shown in Table 11.

Table 11. Clustering Results

Code	Name	Medoids
D001	AALI	C1
D002	ABMM	C3
D003	ACES	C2
D004	ACST	C3
D005	ADCP	C1
D006	ADES	C2
D007	ADHI	C1
D008	ADMF	C3
D009	ADMR	C3

In the Figure 4, Cluster 1 consists of 93 stocks (31.3% of the total data) with moderate PER and PBV values. Then in Cluster 2, it includes 91 stocks (30.6% of the total data) which have the lowest PER and PBV values among the three clusters. And the last one In Cluster 3 contains 113 stocks (38.0% of the total data) with the highest PER and PBV values.

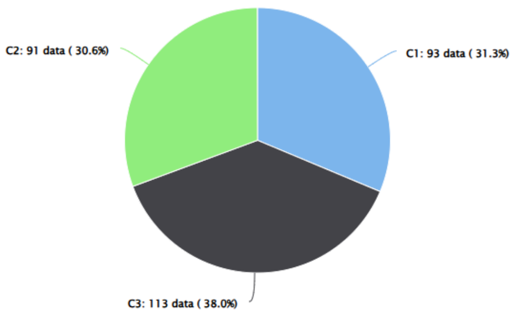

Figure 4 Clustering Graph

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.5. Implementation

The system is implemented according to a well-defined design, ensuring a seamless and intuitive experience for users navigating its various functionalities. The user interface, depicted in the accompanying images, is designed to facilitate easy interaction and efficient workflow management. The system begins with the login page as shown in Figure 5, where users are presented with straightforward options to either log in or register for an account. This initial step ensures secure access and helps manage different user roles, distinguishing between administrators and general users. Upon successful login, users are directed to the dashboard page (Figure 6). The dashboard serves as the central hub of the system, offering a comprehensive overview of its capabilities. It provides a brief explanation of the k-medoids algorithm, a core component of the system's analytical functions. Additionally, the dashboard includes several navigational menus—such as Attributes, Dataset, Preprocessing, K-Medoids, Password, and Logout—allowing users to easily move between different sections and perform various tasks.

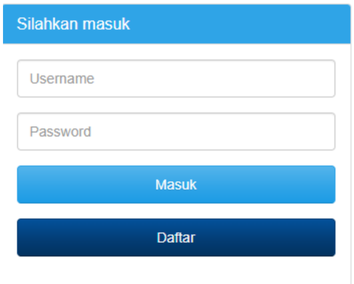

Figure 5. Login page

Figure 7 illustrates the attribute page, a key feature where administrators can add new attributes that define the criteria or parameters for stock grouping. This functionality provides flexibility, enabling administrators to customize the system according to specific needs or objectives. Moving forward, Figure 8 depicts the dataset input page, which is designed to streamline the data entry process. Here, administrators can import datasets in CSV format, which is a common and versatile data format. This capability allows the system to accommodate large datasets efficiently, ensuring that data is correctly formatted for subsequent analysis.

The preprocessing page**,** shown in Figure 9, is crucial for preparing the data for analysis. On this page, the system performs normalization of the dataset, adjusting the data so that different features are on a comparable scale, which is vital for accurate clustering and analysis. Next, Figure 10 presents the k-medoids algorithm calculation page, where users or administrators can specify the parameters for the algorithm, such as the number of clusters and iterations. This page is designed to

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

be user-friendly, requiring minimal input before the clustering process can be initiated with a single button click. The flexibility to choose the number of clusters and iterations allows for a tailored analysis, accommodating different data characteristics and analytical needs.

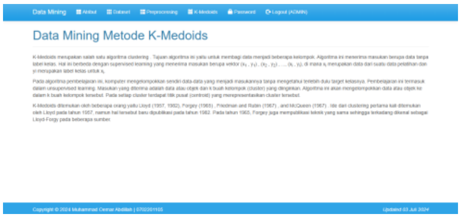

Figure 6. Dashboard

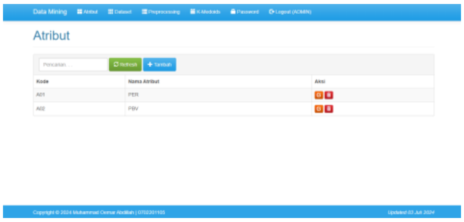

Figure 7. Attribute page

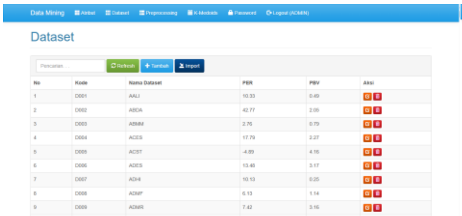

Figure 8. Dataset input page

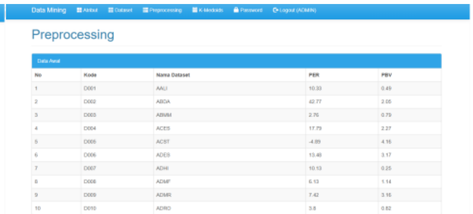

Figure 9. Preprocessing pages

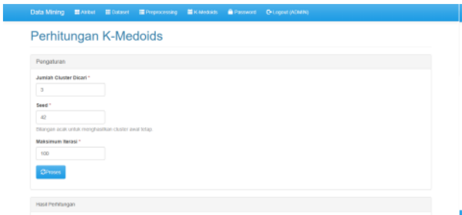

Figure 10. K-Medoids Algorithm

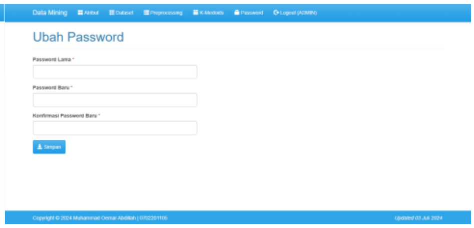

Figure 11. Change Password Page

Figure 11 shows the change password page, a critical feature for maintaining the security and integrity of the system. Both administrators and users have the ability to update their passwords, ensuring that account security can be managed independently and regularly. This page enhances the overall security protocol of the system by promoting best practices in password management. Overall, the system's design integrates a range of features that support effective data management and analysis while maintaining user-friendliness. The clear layout and intuitive navigation help users perform complex tasks with ease, from data entry and preprocessing to clustering and security management. By providing both functionality and flexibility, the system is well-suited to meet diverse user requirements and adapt to various data-driven applications.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.6. Testing

Testing is a crucial final stage in this research, aimed at ensuring that the developed system is ready for practical use. To validate its functionality and effectiveness, rigorous testing is conducted on two different user groups: admin and user as shown in Table 12 and Table 13.

Table 12. Admin Page Test

No	Pages Tested	Input	Expected Results	Conclusion
1	Login Page	Click the "Register button"	Displaying the Account Registration page	Succes
2	Login Page	Click the "sign in" button	Displaying the Dashboard	Succes
3	Attribute Menu	Click "Add"	Adding Attributes	Succes
4	Dataset Menu	Click the "Import" button	Inserting a dataset	Succes
5	Preprocessing Menu	Click the "preprocessing" menu	Displaying data preprocessing results	Succes
6	Menu K-Medoids	Click the "Process" button	Running the K-Medoids algorithm	Succes
7	Menu Password	Click the "Save" button	Changing Old Passwords	Succes
8	Logout Menu	Click the "Logout" button	Log out of the page and enter the login page	Succes

Table 13. User Page Test

No	Pages Tested	Input	Expected Results	Conclusion
1	Login Page	Click the "Register button"	Displaying the Account Registration page	Succes
2	Login Page	Click the "sign in" button	Displaying the Dashboard	Succes
3	Menu K-Medoids	Click the "Process" button	Running the K-Medoids algorithm	Succes
4	Menu Password	Click the "Save" button	Changing Old Passwords	Succes
5	Logout Menu	Click the "Logout" button	Log out of the page and enter the login page	Succes

3.7. Discussion

Based on the results obtained in this study, the K-Medoids algorithm has proven effective in clustering stocks using the Price Earnings Ratio (PER) and Price Book

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Value (PBV) as the primary attributes. The clustering process generated three distinct groups: Cluster 1 consists of 93 stocks (31.3% of the total dataset) with moderate PER and PBV values; Cluster 2 includes 91 stocks (30.6% of the total dataset) with the lowest PER and PBV values; and Cluster 3 comprises 113 stocks (38.0% of the total dataset) with the highest PER and PBV values. These clusters provide valuable insights for investors, allowing them to categorize stocks according to risk levels: low-risk stocks (Cluster 2), moderate-risk stocks (Cluster 1), and high-risk stocks (Cluster 3). This differentiation is crucial for investment strategies, enabling investors to align their portfolios with their risk tolerance and financial goals.

The system developed in this study facilitates the analysis of stock data by leveraging the K-Medoids algorithm, providing a user-friendly platform for investors to make informed decisions. By implementing the clustering algorithm, the system identifies groups of stocks with similar characteristics, helping investors minimize risk and optimize investment strategies. For instance, stocks grouped in Cluster 2 (low PER and PBV values) generally exhibit lower risk, making them suitable for conservative investors looking to avoid significant losses. Conversely, Cluster 3 stocks (high PER and PBV values) are indicative of higher risk but potentially higher returns, appealing to more aggressive investment strategies.

To ensure the system's reliability and readiness for practical use, extensive testing was conducted with two different user groups: admin and regular users. The testing results demonstrate that all pages and functionalities performed as expected, confirming the system's robustness. For the admin group, the testing covered a range of functionalities including login, attribute management, dataset input, preprocessing, algorithm execution (K-Medoids), password management, and logout procedures. Similarly, for the user group, the testing focused on login, clustering (K-Medoids), password management, and logout processes. In both cases, the system met all functional requirements, confirming that it is well-prepared for deployment.

The system design, as outlined in the use case diagram (Figure 3), illustrates the interactions between the different actors (admin and user) and the system. The admin is responsible for managing the data input and attributes, ensuring that the dataset is ready for clustering, while users primarily focus on using the algorithm to perform clustering analysis. The dataset, sourced from the Indonesian Stock Exchange and collected on July 22, 2024, contains 295 rows with attributes such as Symbol, Price, PER, PBV, and Market Cap. However, only the PER and PBV attributes were utilized for clustering purposes.

Given the significant numerical disparities in the dataset, normalization was necessary to ensure that the values of PER and PBV were on a comparable scale

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

before applying the K-Medoids algorithm. This normalization process allowed the algorithm to accurately calculate the Euclidean distances between data points and assign them to the appropriate clusters. The iterative process of the K-Medoids algorithm involved multiple iterations of distance calculations, cost assessments, and adjustments to the medoids (Table 3 to Table 10). The algorithm continued to refine the clusters until the cost stopped decreasing, ensuring optimal clustering results.

Through this approach, the study not only demonstrated the effectiveness of the K-Medoids algorithm in clustering stock data but also highlighted the system's ability to support investors in analyzing stock characteristics and making data-driven decisions. The clustering results (Figure 4) clearly define which stocks fall into the categories of low, medium, and high-risk, providing actionable insights for portfolio management.

Furthermore, the implementation of the system aligns with a well-defined design, featuring an intuitive interface that guides users through each step of the process—from logging in (Figure 5) to navigating the dashboard (Figure 6), managing attributes (Figure 7), inputting datasets (Figure 8), preprocessing data (Figure 9), running the K-Medoids algorithm (Figure 10), and managing security through password updates (Figure 11). The clear layout and logical flow of the interface ensure that both novice and experienced users can effectively utilize the system's capabilities.

The study effectively combines data analysis and system development to achieve a comprehensive solution for stock clustering. By using the K-Medoids algorithm, the study demonstrates a practical application of clustering techniques in the financial domain, specifically for stock analysis based on PER and PBV values. The analysis reveals that the system can effectively group stocks into distinct clusters, which helps investors understand the risk profiles associated with each group.

The decision to utilize the K-Medoids algorithm is well-founded, given its ability to handle large datasets and its suitability for clustering tasks where the number of clusters is predefined. The iterative refinement of the clusters based on cost reduction ensures that the algorithm arrives at the most optimal grouping, thereby enhancing the reliability of the results. The synthesis of data normalization, Euclidean distance calculations, and iterative medoid adjustments exemplifies a robust methodological approach, ensuring that the clustering outcomes are both accurate and meaningful.

The study's integration of system design with data analysis is a significant strength, as it translates complex analytical processes into a user-friendly platform that

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

supports decision-making. The use case diagram clarifies the roles of different users and their interactions with the system, while the comprehensive testing ensures that all functionalities are correctly implemented and ready for practical use. This study contributes to the field of financial data analysis by providing a validated tool that leverages clustering algorithms to assist investors in making informed decisions. It demonstrates how data science techniques can be effectively integrated into financial decision-making processes, offering a valuable resource for both novice and experienced investors.

4. CONCLUSION

This study effectively addresses the challenges faced by novice investors by applying the K-Medoids algorithm to cluster stocks based on Price Earnings Ratio (PER) and Price Book Value (PBV), resulting in three distinct clusters: Cluster 1, comprising 93 stocks (31.3% of the dataset) with moderate PER and PBV values; Cluster 2, including 91 stocks (30.6% of the dataset) with the lowest PER and PBV values; and Cluster 3, containing 113 stocks (38.0% of the dataset) with the highest PER and PBV values. By categorizing stocks into low, moderate, and high-risk groups, the system provides valuable guidance, recommending investors focus on stocks with low to moderate risk levels to minimize potential losses. The web-based clustering system demonstrates its utility in supporting investment decisions, with rigorous testing confirming that all functionalities for both admin and user groups operate as expected, ensuring its reliability and readiness for practical use. Future enhancements could include integrating additional financial indicators or refining the clustering method to further assist investors in navigating the complexities of stock investments.

REFERENCES

[1] M. S. Munir and P. Rahardiyanto, "Sistem Pendukung Keputusan Pemilihan Investasi Saham Syariah Dalam Indeks JII 70 Menggunakan Metode Ahp Beserta Perhitungan," *SPIRIT*, vol. 15, no. 2, 2023.
[2] R. P. Sari and M. R. Maulana, "Sistem Pendukung Keputusan Rekomendasi Emiten Saham Menggunakan Metode Simple Additive Weighting," *J. Sist. Komput. dan Inform.*, vol. 2, no. 3, p. 321, 2021.
[3] D. Tohendry and D. Jollyta, "Penerapan Algoritma K-Means Cluster Ing Untuk Pengelompokkan Saham Berdasarkan Price Earning Ratio Dan Price To Book Value," *J. Mhs. Apl. Teknol. Komput. dan Inf.*, vol. 5, no. 1, pp. 3–9, 2023.
[4] M. Mikrad and A. Budi, "Pengaruh Struktur Modal, Likuiditas, Dan Keputusan Investasi Terhadap Nilai Perusahaan Pada Perusahaan Jasa Sub Sektor Pariwisata, Hotel, Dan Restoran Yang Terdaftar Di Bei Tahun 2014-2018," *Dyn. Manag. J.*, vol. 4, no. 1, 2020.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[5] Tandelilin, "Pasar Modal Manajemen Portofolio dan Investasi," *Kanisius Daerah Istimewa Yogyakarta*, pp. 302, 2017.

[6] A. M. Peranginangin, "Keputusan Investasi Saham Dengan Analisis Fundamental Melalui Pendekatan Price Earning Ratio (Per)(Studi Pada Saham-Saham," *Jurakiman (Jurnal Akunt. dan Manajemen*, vol. 14, no. 2, pp. 91-110, 2021.

[7] S. Sukamulja, "Pengantar Pemodelan Keuangan Dan Analisis Pasar Modal (Edisi 1), Yogyakarta, Andi Offset," *Ed. 1, Yogyakarta, Andi Offset*, pp. 1–158, 2017.

[8] A. Ikhwan and N. Aslami, "Implementasi Data Mining untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 208–217, 2020.

[9] Z. Mustofa and Iman Saufik Suasana, "Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi," *J. Teknol. Inf. Dan Komun.*, vol. 9, no. 1, pp. 1–10, 2020.

[10] A. A. D. Sulistyawati and M. Sadikin, "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan, " *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 3. pp.516-526, 2021.

[11] H. Pohan, M. Zarlis, E. Irawan, H. Okprana, and Y. Pranayama, "Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia," *JUKI J. Komput. dan Inform.*, vol. 3, no. 2, pp. 97–104, 2021.

[12] E. Tasia and M. Afdal, "Comparison Of K-Means And K-Medoid Algorithms For Clustering Of Flood-Prone Areas In Rokan Hilir District Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," *Indones. J. Inform. Res.*, vol. 3, no. 1, pp. 65–73, 2023.

[13] G. Dwilestari, Mulyawan, Martanto, and I. Ali, "Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia," *JURSIMA J. Sist. Inf. dan Manaj.*, vol. 9, no. 3, pp. 282–290, 2021.

[14] F. Fajriana, "Analisis Algoritma K-Medoids pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 2, p. 263, 2021.

[15] G. Mahesa, Y. H. Chrisnanto, "Sistem Pengelompokkan Penjualan Vaksin Dan Serum di Pt Bio Farma Menggunakan Metode K-Medoids," in Prog *Prosiding Sains Nasional dan Teknologi*, 2019.

[16] M. S. Haq, M. Samani, Karwanto, and N. Hariyati, "Android-Based Digital Library Application Development," *Int. J. Interact. Mob. Technol.*, vol. 16, no. 11, pp. 224–237, 2022.

[17] Muhamad Alda, "Sistem Informasi Laundry Menggunakan Metode Waterfall Berbasis Android Pada Simply Fresh Laundry," *J. Teknol. Inf.*, vol. 3, no. 2, pp. 122–129, 2019.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[18] R. Kurniawan, A. Halim, and H. Melisa, "Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 2, pp. 903–912, 2023.

[19] R. A. Putri and A. M. Dewi, "Mobile-Based Document and Salary Application Case Study : Kemenag Labuhan Batu," *Infokum*, vol. 10, no. 1, pp. 722–730, 2021.