BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2609–2620
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19no4pp2609-2620

INTEGRATION OF DAVIES-BOULDIN INDEX VALIDATION AND MEAN-VARIANCE EFFICIENT PORTFOLIO IN K-MEANS++ CLUSTERING FOR OPTIMIZATION OF THE LQ45 STOCK PORTFOLIO

David Jordy Dhandio[1], **Evy Sulistianingsih**[2*], **Neva Satyahadewi**[3]

[1,2,3]*Mathematics Department, Faculty of Mathematics and Natural Science, Universitas Tanjungpura Jln. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia*

*Corresponding author's e-mail: *evysulistianingsih@math.untan.ac.id*

Article History:

Received: 30th January 2025
Revised: 8th April 2025
Accepted: 2nd June 2025
Available online: 1st September 2025

Keywords:

Expected Return;
Financial Ratios;
Investment Risk;
K-Means++ Clustering;
Optimal Portfolio;
Sharpe Index.

ABSTRACT

Stock investment involves allocating funds to get returns based on the associated risks. In stock investments, returns and risks exhibit a linear correlation, meaning higher expected returns come with higher risks. Risk in stock investments can be minimized by forming portfolios using a cluster analysis approach, where the groups of stocks generated from the analysis represent the resulting portfolios. This research aims to form an optimal stock portfolio using K-Means++ Clustering, validated by the Davies Bouldin Index (DBI), the weighting of stocks in a portfolio using the Mean-Variance Efficient Portfolio (MVEP), and evaluated based on the Sharpe Index. The data used include stocks indexed in LQ45 from February 2020 to August 2024, stock closing prices from August 1, 2023, to August 1, 2024, company financial ratios as of June 2024, and the average Bank Indonesia interest rate from August 2023 to August 2024. Based on the financial ratios, K-Means++ Clustering and DBI validation identified three optimal clusters. Clusters 1 and 2, consisting of single stocks, cannot be directly utilized as portfolios due to the requirement for diversification. Each cluster's stocks with the highest expected return were selected to form a new portfolio. According to the MVEP analysis, the investment proportion of each stock in portfolio 1 is 44.10% (BBCA.JK), 15.40% (BBNI.JK), 2.89% (BMRI.JK), 15.02% (CPIN.JK), and 22.60% (PGAS.JK). In portfolio 2, the weights are 27.68% (BBTN.JK), 36.00% (ADRO.JK), and 36.33% (BMRI.JK). Based on the Sharpe Index, portfolio 2 achieved the highest value (0.048404) compared to portfolio 1 (0.034465), indicating that portfolio 2 shows a better risk-adjusted return than portfolio 1.

How to cite this article:

D. J. Dhandio, E. Sulistianingsih, and N. Satyahadewi, "INTEGRATION OF DAVIES-BOULDIN INDEX VALIDATION AND MEAN-VARIANCE EFFICIENT PORTFOLIO IN K-MEANS++ CLUSTERING FOR OPTIMIZATION OF THE LQ45 STOCK PORTFOLIO," *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2609-2620, December, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Investment is one of the actions in the capital market, which involves investing a certain amount of funds to obtain certain benefits in the future [1]. According to the data provided by Indonesia Central Securities Depository, the number of investors in Indonesia's capital market increased by 1.30% in January 2024, with the most significant increase being stock investors of 1.76% compared to December 2023 [2]. The high rise in stock investors indicates that this investment is currently in great demand by the public, so an understanding of investing is needed to provide benefits.

The goal of stock investing is to provide returns proportional to the risks taken. There is a linear correlation between return and risk, which means that the greater the expected return, the higher the risk that must be borne [3]. To minimize the risk of stock investment, the approach that can be taken is to form a stock portfolio [4]. Portfolio formation is allocating funds to select assets that minimize the investment risk of the portfolio. A stock portfolio is a type of investment that involves forming a collection of stocks (diversification) to reduce the risk of stock investment [5]. One method that can be used to create a stock portfolio is the Mean-Variance Efficient Portfolio (MVEP).

Investors need to be able to select the appropriate stocks when they are putting together a stock portfolio to ensure that the investment will yield a profit. A cluster analysis approach can be utilized to identify the stocks that will be included in a portfolio. Cluster analysis divides data into several groups with similar characteristics in the same group and different characteristics between groups [6]. Hierarchical clustering and non-hierarchical clustering are the two major classifications used in cluster analysis. There are two types of clusters, and the difference between them is based on determining the number of groups or clusters to be formed [7]. K-Means++ Clustering is a non-hierarchical clustering approach that develops the K-Means algorithm, which gives initial centroid selection based on simple probability [8]. The K-Means++ algorithm requires a method for determining the optimal number of clusters, one of which is the Davies-Bouldin Index validation.

The capital market is a place for buying and selling various financial instruments, including stocks [9]. The Indonesia Stock Exchange (IDX) is one of the institutions associated with the capital market in Indonesia. The large number of stock choices available on the capital market often makes it difficult for investors to choose the right stocks to invest in. Therefore, the IDX issues various stock indices based on performance, liquidity, and stability, including the LQ45 index. The LQ45 index comprises 45 stocks, which must accomplish the following criteria: the stocks must have been listed on the IDX for a minimum of three years, have the highest average market capitalization over the previous twelve months, and have significant transaction volume activity [10]. The LQ45 index is updated by the IDX every six months in February and August. Being listed on the LQ45 index is a source of pride for a company, so many investors dare to invest in stocks in the index. This can be used as a form of motivation for companies to maintain and improve their stock performance so that the stocks are always consistently indexed in the LQ45 [11].

Many researchers have applied cluster analysis to form stock portfolios. One of them is the formation of the LQ45 stock portfolio with K-Means Clustering and MVEP [12]. This research shows that the K-Means algorithm groups 11 stocks into 2 portfolios, with portfolio 1 having the highest expected return (0.000657) compared to portfolio 2 (0.000133) and portfolio 1 having the lowest risk (0.008953) compared to portfolio 2 (0.011706).

Another research conducted by applying cluster analysis in forming a stock portfolio was to optimize the Jakarta Islamic Index stock portfolio using Mean-Value at Risk (Mean-VaR) by creating a combination of stocks in the portfolio based on financial ratios using the K-Medoids Clustering algorithm with Davies Bouldin Index (DBI) validation [13]. Based on the findings of this research, it was determined that the DBI value offered the best possible number of clusters, which were six clusters, each representing the total number of stocks in the portfolio. The portfolio was formed based on one stock with the highest expected return from each cluster. Based on Mean-VaR, with a 95% confidence level, the estimated maximum loss of the portfolio is 5.06% of the initial investment capital.

In stock portfolio investment, investors want an optimal investment, meaning that the investment provides a return on the risk faced. The strategy that can be applied to create an optimal stock portfolio is to choose stocks with high stability, provide positive returns, and broad diversification [14]. This strategy can be applied by selecting stocks consistently indexed by LQ45 with a positive expected return in current market conditions. Similar research has been conducted previously, but none has optimized the LQ45 stock portfolio based on cluster analysis with this strategy approach. This strategy is applied in this research by forming an optimal stock portfolio using a combination of stocks formed with the K-Means++ algorithm based on the company's financial ratio, determining the number of optimal clusters with Davies Bouldin Index validation, and weighting stocks in the portfolio with MVEP. The clusters formed represent the portfolios formed, with the addition of one new portfolio based on stocks with the highest expected return in each cluster. All of the portfolios built will be assessed using the Sharpe index value to identify the portfolio with the optimum performance.

2. RESEARCH METHODS

2.1 Stock Return and Expected Return

The return on stock is the risk-adjusted rate of return that is achieved by buying and selling stocks over a specified period. **Equation (1)** can be used to calculate the return on investment for a stock **[15]**.

$$R_{i(u)} = \ln\left(\frac{P_{i(u)}}{P_{i(u-1)}}\right) \tag{1}$$

where:

$R_{i(u)}$: Return of the i-th stock in period u.

$P_{i(u)}$: Closing price of the i-th stock in period u.

$P_{i(u-1)}$: Closing price of the i-th stock in period $u-1$.

Then, an expected return of stock is the return that is anticipated to occur in the future. The expected return for a stock can be quantified by **Equation (2)** **[15]**.

$$E(R_i) = \frac{\sum_{u=1}^{U} R_{i(u)}}{U} \tag{2}$$

where:

$E(R_i)$: Expected return of the i-th stock.

U : Number of stock returns.

2.2 Variance, Covariance, and Risk of Stock

The variance of stock return is a measure of risk in stock investment. The greater the variance value, the greater the risk of stock investment. Stock variance can be calculated using **Equation (3)** **[16]**.

$$s_i^2 = \frac{\sum_{u=1}^{U}\left(R_{i(u)} - E(R_i)\right)^2}{U-1} \tag{3}$$

where s_i^2 is the variance of the i-th stock.

Stock covariance measures the extent to which two stock returns move together. Positive stock covariance means that the returns of both stocks move in the same direction **[17]**. Stock covariance can be calculated using **Equation (4)** **[16]**.

$$s_{i;j} = \frac{\sum_{u=1}^{U}\left[\left(R_{i(u)} - E(R_i)\right)\left(R_{j(u)} - E(R_j)\right)\right]}{U-1} \tag{4}$$

where $s_{i;j}$ is the covariance of the i-th stock and the j-th stock.

Stock risk is the deviation between the received and the expected returns. **Equation (5)** can be utilized to calculate the risk of stock investment **[16]**.

$$s_i = \sqrt{s_i^2} = \sqrt{\frac{\sum_{u=1}^{U}\left(R_{i(u)} - E(R_i)\right)^2}{U-1}} \tag{5}$$

where s_i is the risk of the i-th stock.

2.3 Data Standardization

Data standardization is changing data into a form suitable for analysis, such as cluster analysis. Data standardization aims to reduce data variation between variables with different units. Data standardization can be derived using **Equation (6)** to calculate the Z-Score value **[18]**.

$$Z_{i;x} = \frac{a_{i;x} - \bar{a}_x}{s_x} \tag{6}$$

where:

$Z_{i;x}$: Z-Score of the i-th object (stock) on the x-th variable
$a_{i;x}$: Data of the i-th object (stock) on the x-th variable
$\bar{a}_x$: Average of the x-th variable
s_x : Standard deviation of the x-th variable

2.4 Multicollinearity Test

Multicollinearity is the possible existence of a significant correlation between variables. Examining the Variance Inflation Factor (VIF) value can determine whether or not the multicollinearity assumption is valid. If the VIF value is ≥ 10, then multicollinearity occurs. The VIF value can be calculated using **Equation (7) [18]**.

$$VIF_x = \frac{1}{1 - R_x^2} \tag{7}$$

where:

VIF_x : Variance Inflation Factor Value of the x-th variable
R_x^2 : Determination coefficient of the x-th variable

2.5 Euclidean Distance

Cluster analysis is a method that involves grouping data or objects in such a way that there is a high degree of homogeneity between the objects in one group and a significant degree of variability between the groups **[19]**. By measuring the distance between things, it is possible to assess the homogeneity of the objects; the closer the distance between the objects, the more homogeneous the objects are **[20]**. The Euclidean distance is one method used to measure the distance between two objects. It is calculated by taking the root of the sum of the squares of the differences in values between objects. Euclidean distance can be calculated using **Equation (8) [18]**.

$$d_{i;j} = \sqrt{\sum_{x=1}^{X} \left(a_{i;x} - a_{j;x}\right)^2} \tag{8}$$

where:

$d_{i;j}$: Euclidean distances between i-th object and j-th object
X : Number of variables
$a_{j;x}$: Data of the j-th object on the x-th variable

2.6 K-Means++ Clustering

Hierarchical and non-hierarchical cluster analysis are the two methods that comprise cluster analysis. One of the non-hierarchical cluster algorithms that was developed within the K-Means algorithm is known as K-Means++. The K-Means has a weakness, namely that the initial centroid is determined randomly, thus increasing the number of iterations and increasing the variation of the cluster results formed **[21]**. The K-Means++ overcomes this weakness by determining the initial centroid using a simple probability calculation. Namely, the object with the highest chance will be selected as the initial centroid. The stages of cluster analysis with the K-Means++ start from determining the number of clusters (K), calculating the distance between objects, and determining the initial centroid. The number of initial centroids is the number of clusters that will be formed. The probability of an object being selected as the initial centroid can be calculated using **Equation (9) [18]**.

$$P_k(a_i) = \frac{d(a_{i;c})^2}{\sum_{i=1}^{I} d(a_{i;c})^2} \tag{9}$$

where:
$P_k(a_i)$: Probability of object a_i being selected as the new centroid of the cluster k
$d(a_{i;c})$: Distance of the i-th object to the initial centroid c
I : Number of research objects

After determining the initial centroid with the K-Means++ algorithm, the next stage is the same as the K-Means algorithm, namely, placing objects based on proximity and calculating new centroids. The new centroid in the K-Means algorithm is determined based on the average of objects in each cluster, which can be calculated using **Equation (10) [18]**.

$$c'_{k;x} = \frac{1}{I_k}\sum_{i=1}^{I_k} a_{i;x} \tag{10}$$

where:
$c'_{k;x}$: New centroid of cluster k from the x-th variable
I_k : Number of objects in cluster k

2.7 Davies Bouldin Index

A non-hierarchical cluster analysis is a type of cluster analysis in which the researcher determines the number of clusters depending on their specific requirements. Therefore, an approach is needed to determine the optimal number of clusters, meaning that the cluster has high similarity of characteristics between objects in the same cluster and high differences in characteristics between clusters **[19]**. One method used to determine the optimal number of clusters is the Davies-Bouldin Index (DBI) validation.

Davies-Bouldin Index (DBI) validation is started by calculating the Sum of Squares Within (SSW) value, namely the cohesion or proximity of objects to the centroid in the same cluster. The SSW value can be calculated using **Equation (11) [22]**.

$$SSW_k = \frac{1}{I_k}\sum_{i=1}^{I_k} d_{i;k} \tag{11}$$

where:
SSW_k : Distance of object proximity to the centroid of cluster k
$d_{i;k}$: Distance of the i-th object to the centroid of cluster k

The next step is calculating the Sum of Squares Between (SSB) value, which is the separation or distance between centroids in each cluster. The SSB value can be calculated using **Equation (12) [22]**.

$$SSB_{k;l} = d(c_k, c_l) \tag{12}$$

where:
$SSB_{k;l}$: Distance between the centroid of cluster k and the centroid of cluster l
c_k : Centroid of cluster k

After getting the SSW and SSB values, the next step is calculating the ratio value or comparing the two clusters. The ratio value can be calculated using **Equation (13) [22]**.

$$R_{k;l} = \frac{SSW_k + SSW_l}{SSB_{k;l}} \tag{13}$$

where $R_{k;l}$ is the ratio of cluster k and cluster l.

The ratio value that has been obtained will be used to calculate the DBI value. The DBI value can be calculated using **Equation (14) [22]**.

$$DBI = \frac{1}{K}\sum_{k=1}^{K} \max_{k\neq l}(R_{k;l}) \tag{14}$$

where K is number of clusters.

The Davies-Bouldin Index validation aims to maximize separation and minimize cohesion. The optimal number of clusters is the cluster with the minimum DBI value with a DBI value ≥ 0 [22].

2.8 Mean-Variance Efficient Portfolio

The optimal portfolio is an efficient portfolio chosen by investors based on the investor's attitude towards risk, whether the investor likes risk or avoids risk. An efficient portfolio is a portfolio that provides minimum risk with an inevitable return or a particular risk with maximum return [16]. One method for forming a portfolio by minimizing risk is the Mean-Variance Efficient Portfolio (MVEP). MVEP is a method of forming a portfolio with minimum variance based on its average return. Minimum variance is obtained by optimizing portfolio weighting or the proportion of funds allocated to each asset in the portfolio [23]. The limitation of the MVEP portfolio is that the total of all asset weights is equal to one. The asset weight in the portfolio is calculated using **Equation (15) [16]**.

$$\boldsymbol{w} = \frac{\Sigma^{-1}\,\mathbf{1}_N}{{\mathbf{1}_N}^T\,\Sigma^{-1}\,\mathbf{1}_N} \tag{15}$$

where:

$\boldsymbol{w}$: Asset weight in the portfolio, with $\boldsymbol{w}^T\mathbf{1}_N = 1$
Σ^{-1} : Inverse of the variance-covariance of asset returns in the portfolio
$\mathbf{1}_N$: Vector one with $N \times 1$ dimension
${\mathbf{1}_N}^T$: Transpose of $\mathbf{1}_N$
N : Number of assets in the portfolio

The weights obtained are used to calculate the portfolio's expected return, variance, and risk. Expected return is the expected rate of return in a portfolio investment. Expected return can be calculated using **Equation (16) [15]**.

$$E(R_p) = \sum_{i=1}^{N} w_i E(R_i) \tag{16}$$

where:

$E(R_p)$: Expected return of the portfolio
w_i : Weight of i-th asset in portfolio

Portfolio variance is a measure used to measure portfolio risk. Portfolio variance can be calculated using **Equation (17) [16]**.

$$s_p^2 = \boldsymbol{w}^T\Sigma\boldsymbol{w} \tag{17}$$

where:

s_p^2 : Variance of portfolio
Σ : Variance-covariance of asset returns in the portfolio

Portfolio risk is calculated by taking the square root of portfolio variance. Portfolio risk can be calculated using **Equation (18) [16]**.

$$s_p = \sqrt{s_p^2} = \sqrt{\boldsymbol{w}^T\Sigma\boldsymbol{w}} \tag{18}$$

where s_p is a portfolio risk.

2.9 Sharpe Index

The Sharpe index is a method used to measure portfolio performance by calculating the average difference between portfolio returns and risk-free asset returns, with portfolio risk. This index measures the suitability between the level of return and risk in portfolio investment. The higher the Sharpe index value, the better the portfolio performance. The Sharpe index can be calculated using **Equation (19) [17]**.

$$Sp = \frac{E(R_p) - R_f}{s_p} \tag{19}$$

where:

Sp : Sharpe index of portfolio

R_f : Return on risk-free assets

3. RESULTS AND DISCUSSION

This research uses four types of secondary data. The first data is a list of LQ45-indexed stocks from February 2020 to August 2024 obtained from the Indonesia Stock Exchange website. The data is used to select stocks with high stability criteria, and 24 stocks are obtained that are consistently indexed in the LQ45 during that period. The second data is the closing price of stocks from August 1, 2023, to August 1, 2024, on 24 stocks consistently indexed in the LQ45. The data is obtained from the Yahoo Finance website. The data is used to select stocks with positive expected returns and to calculate the weight of the portfolio investment using Microsoft Excel and RStudio software. In determining stocks with positive expected returns, the first step is to calculate the return on the 24 stocks using **Equation (1)**. After obtaining the stock return, the next step is to calculate the expected return on the stocks using **Equation (2)**. Based on the calculation of the expected return, 7 out of 24 stocks have a positive expected return. The stocks that pass the selection are the stocks that will be used to form a portfolio. The list of these stocks can be seen in **Table 1**.

Table 1. List of 7 Selected Stocks

No	Stock Code	Company Name	Expected Return
1	ADRO.JK	Adaro Energy Indonesia Tbk.	0.001291
2	BBCA.JK	Bank Central Indonesia Tbk.	0.000539
3	BBNI.JK	Bank Negara Indonesia (Persero) Tbk.	0.000660
4	BBTN.JK	Bank Tabungan Negara (Persero) Tbk.	0.000127
5	BMRI.JK	Bank Mandiri (Persero) Tbk.	0.000747
6	CPIN.JK	Charoen Pokphand Indonesia Tbk.	0.000040
7	PGAS.JK	Perusahaan Gas Negara Tbk.	0.000580

The third data is the financial ratio for June 2024 for seven selected stock companies. The data was obtained from the Indo Premier Sekuritas page, which includes Return on Asset (ROA), Return on Equity (ROE), Debt to Asset Ratio (DAR), and Debt to Equity Ratio (DER). The data is used to form a combination of stocks in the portfolio with the K-Means++ Clustering algorithm and Davies-Bouldin Index validation. Because the data have different units, data standardization is required using **Equation (6)**. The fourth data is the average daily conversion data of the Bank Indonesia interest rate (BI Rate) from August 2023 to August 2024, which is 0.02%. The data is used as a risk-free asset return to measure portfolio performance using the Sharpe index.

The main steps in this research are as follows.

1. Exploring and describing the general overview of the data.
2. Standardize the data and testing for multicollinearity.
3. Calculate the distance between the data using Euclidean distance.
4. Determine the number of clusters and group the data using K-Means++ Clustering.
5. Determine the optimal number of clusters using the Davies-Bouldin Index validation.
6. Forming a stock portfolio based on the highest expected return of each stock cluster in the optimal number of clusters.
7. Calculate the proportion of portfolio investment using the Mean-Variance Efficient Portfolio.
8. Calculate the expected return and the risk of the portfolio.
9. Calculate the portfolio performance using the Sharpe index, and interpret the performance of each portfolio.

3.1 Data Exploration

Descriptive statistics of the financial ratio data of 7 selected stocks are presented in **Table 2**.

Table 2. Descriptive Statistics

Variable	*I*	Mean	Minimum	Maximum
ROA	7	2.74	0.33	7.59
ROE	7	7.73	4.84	11.16
DAR	7	0.65	0.25	0.93
DER	7	4.74	0.33	13.68

Based on **Table 2**, the lowest ROA is recorded by BBTN.JK at 0.33%, while the highest ROA is recorded by ADRO.JK at 7.59%. The lowest ROE is also held by BBTN.JK at 4.84%, whereas the highest ROE belongs to BBCA.JK at 11.16%. For the DAR ratio, ADRO.JK has the lowest value at 0.25, and BBTN.JK has the highest at 0.93. Similarly, ADRO.JK records the lowest DER at 0.33, while BBTN.JK has the highest DER at 13.68. These results indicate that ADRO.JK demonstrates the highest ROA and ROE as well as the lowest DAR and DER, whereas BBTN.JK exhibits the opposite pattern. This suggests that the financial performance of ADRO.JK is superior compared to that of BBTN.JK.

3.2 Multicollinearity Test

Financial ratio data from data standardization is used to conduct multicollinearity tests. Multicollinearity tests measure the correlation between variables. In cluster analysis, the proper data is data that is not correlated with each other. Multicollinearity tests are measured by the Variance Inflation Factor (VIF) value. Using **Equation (7)**, the VIF value for each variable can be seen in **Table 3**.

Table 3. VIF Value

Variable	VIF Value
ROA	36.22
ROE	7.72
DAR	65.24
DER	12.00

Table 3 shows that the ROA, DAR, and DER variables are highly correlated, so it is necessary to eliminate the variable with the highest VIF value and re-check the VIF value. After eliminating the DAR variable, the VIF value is obtained, as seen in **Table 4**.

Table 4. VIF Value After Removing Variable DAR

Variable	VIF Value
ROA	2.57
ROE	1.11
DER	2.47

Based on **Table 4**, it is obtained that all variables are free from multicollinearity. After conducting a multicollinearity test, the next step is to calculate the distance between objects using the Euclidean distance in **Equation (8)**.

3.3 Formation of Stock Combinations in a Portfolio

The combination of stocks in a portfolio based on financial ratios using the K-Means++ algorithm, with the number of clusters formed being 2 and 3. This algorithm selects one object randomly as the first initial centroid, which is obtained, namely, BBTN.JK stock. The following initial centroid is obtained by calculating the probability using **Equation (9)**. After obtaining the initial centroid, the analysis is continued by placing objects based on the nearest Euclidean distance and calculating the new centroid using **Equation (10)**. The results of the stock clusters in the portfolio can be seen in **Table 5**.

Table 5. Cluster of Stocks

k	Initial Centroid	Cluster	Number of Members	Stocks
$k = 2$	BBTN.JK and ADRO.JK	1	3	BBNI.JK, BBTN.JK, BMRI.JK
		2	4	ADRO.JK, BBCA.JK, CPIN.JK, PGAS.JK
$k = 3$	BBTN.JK, ADRO.JK, and BBNI.JK	1	1	BBTN.JK
		2	1	ADRO.JK
		3	5	BBCA.JK, BBNI.JK, BMRI.JK, CPIN.JK, PGAS.JK

After getting the results of the stock clusters in the portfolio, the next step is to determine the optimal number of clusters with Davies-Bouldin Index (DBI) validation. The cluster with the smallest DBI value is the optimal number of clusters. The DBI value for each cluster can be seen in **Table 6**.

Table 6. DBI Value

k	DBI Value
2	1.128
3	0.452

Based on **Table 6**, k = 3 has the smallest DBI value of 0.452. Therefore, three clusters are the optimal number of clusters for forming a combination of stocks in a portfolio based on financial ratios.

A stock portfolio is a collection of various stocks, so clusters 1 and 2 are not used directly as a portfolio because the clusters only consist of 1 stock. Cluster 3 can be used directly as the first portfolio. This research selects representative stocks in each cluster to be formed into a second portfolio, namely, based on the highest expected return value. The flowchart in **Figure 1** below illustrates the portfolio formation process based on these clustering results and selection criteria.

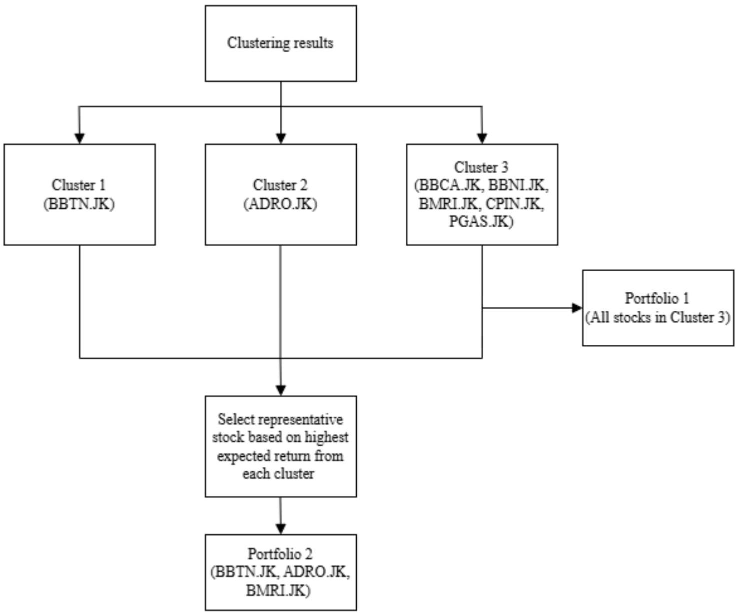

Figure 1. Formation of Portfolios Based on The Highest Expected Return

Figure 1 shows the clustering results of selected stocks and the formation of two portfolios. Portfolio 1 consists of all stocks in Cluster 3, and Portfolio 2 consists of the top-performing stock (based on expected return) from each cluster, namely BBTN.JK, ADRO.JK, and BMRI.JK stock.

3.4 Portfolio Weighting and Performance Calculation

The initial step in calculating portfolio weights with a Mean-Variance Efficient Portfolio (MVEP) is to calculate stock variance and covariance. Stock variance can be calculated using **Equation (3)**, and stock covariance can be calculated using **Equation (4)**. The obtained stock variance and covariance are presented as a variance-covariance matrix for each portfolio for the MVEP calculation process. Using **Equation (15)**, the weights for each portfolio are presented in **Table 7**.

Table 7. Stock Weight in the Portfolio

Portfolio	Stock	Weight
1	BBCA.JK	0.440975
	BBNI.JK	0.153958
	BMRI.JK	0.028926
	CPIN.JK	0.150158
	PGAS.JK	0.225983
2	BBTN.JK	0.276787
	ADRO.JK	0.359956
	BMRI.JK	0.363256

The weights obtained are used to calculate the portfolio's expected return, variance, and risk. The expected return can be calculated using **Equation (16)**, the variance can be calculated using **Equation (17)**, and the risk can be calculated using **Equation (18)**. The calculation results show that portfolio 1 has an expected return of 0.000498 with a risk of 0.000093, while portfolio 2 has an expected return of 0.000772 with a risk of 0.012522.

The expected return and risk of the portfolio obtained are used to measure portfolio performance with the Sharpe index. The portfolio with the highest Sharpe index value is the optimal portfolio, meaning that the portfolio provides a return according to the risk obtained. The Sharpe index of the portfolio can be calculated using **Equation (19)**, and it is obtained that the Sharpe index of portfolio 1 is 0.034465, and the Sharpe index of portfolio 2 is 0.048404. The Sharpe index value of portfolio 2 is greater than that of portfolio 1, so portfolio 2 shows better performance.

4. CONCLUSION

This study aims to form an optimal stock portfolio by selecting stocks consistently indexed by LQ45 and with a positive expected return. The combination of stocks in the portfolio is formed using the K-Means++ algorithm and Davies-Bouldin Index (DBI) validation. Based on the DBI value, it is obtained that the optimal number of clusters formed by the K-Means++ algorithm is three clusters with the smallest DBI value of 0.452.

There are two portfolios formed, with portfolio 1 consisting of stocks in cluster 3 and portfolio 2 consisting of stocks with the highest expected return in each cluster. Based on the analysis of optimal portfolio formation with a Mean-Variance Efficient Portfolio, the weights in portfolio 1 are BBCA.JK shares of 44.0975%, BBNI.JK shares of 15.3958%, BMRI.JK shares of 2.8926%, CPIN.JK shares of 15.0158%, and PGAS.JK shares of 22.5983%. Meanwhile, weights in portfolio 2 are BBTN.JK shares of 27.6787%, ADRO.JK shares of 35.9956%, and BMRI.JK shares of 36.3256%. Based on the assessment of portfolio performance with the Sharpe index, it is obtained that portfolio 2 has a higher Sharpe index value than portfolio 1. This shows that portfolio 2 has a more optimal performance than portfolio 1. Therefore, investors seeking a better risk-adjusted return may consider allocating their capital to Portfolio 2.

AUTHOR CONTRIBUTIONS

David Jordy Dhandio: Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Software, Validation, Visualization, Writing - Original Draft, Writing - Review and Editing. Evy Sulistianingsih: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing - Review and Editing. Neva Satyahadewi: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Validation, Writing - Review and Editing. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

The authors received no financial support for the research, authorship, and/or publication of this article.

ACKNOWLEDGMENT

The authors would like to express their gratitude and appreciation to all those who have already supported the completion of this work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest to report study.

REFERENCES

[1] I. S. Tanjung, A. Soemitra, and B. Dharma, "THE EFFECT OF INVESTMENT KNOWLEDGE, MINIMUM CAPITAL, MOTIVATION ON STUDENT INTEREST IN INVESTING IN THE SHARIA CAPITAL MARKET," *IIJSE: Indonesian Interdisciplinary Journal of Sharia Economics*, vol. 6, no. 3, pp. 1533–1552, 2023.

[2] Indonesia Central Securities Depository, "STATISTIK PASAR MODAL INDONESIA," PT Kustodian Sentral Efek Indonesia. [Online]. Available: www.ksei.co.id [access: 10 November 2024].

[3] E. du Toit, "REVISITING THE RELATIONSHIP BETWEEN DIFFERENT FINANCIAL RISK MEASURES AND THE MARKET RETURN ON ORDINARY SHARES IN SOUTH AFRICA," *SAJEMS: South African Journal of Economic and Management Sciences*, vol. 18, no. 2, pp. 218–231, Jul. 2015, doi: https://doi.org/10.4102/sajems.v18i2.790.

[4] V. Ivanyuk and D. Berzin, "AN ALGORITHM FOR CONSTRUCTING AN EFFICIENT INVESTMENT PORTFOLIO," in *Advances in Intelligent Systems and Computing*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 482–490. doi: https://doi.org/10.1007/978-3-030-63322-6_39.

[5] D. Jayeola, Z. Ismail, and S. F. Sufahani, "EFFECTS OF DIVERSIFICATION OF ASSETS IN OPTIMIZING RISK OF PORTFOLIO," *MJFAS: Malaysian Journal of Fundamental and Applied Sciences*, vol. 13, no. 4, pp. 584–587, 2017.doi: https://doi.org/10.11113/mjfas.v0n0.567

[6] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING CORN PLANTING FEASIBILITY AREA IN SOUTH LAMPUNG REGENCY," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021, pp. 1–9. doi: https://doi.org/10.1088/1742-6596/1751/1/012038.

[7] A. A. Ismael, Z. A. A. Al-Bairmani, and I. F. F. Al-Masferi, "COMPARISON OF THE TWO METHODS OF CLUSTER ANALYSIS (NON-HIERARCHICAL AND HIERARCHICAL) IN THE CLASSIFICATION OF LABORATORY QUALITY (GLP) AT THE UNIVERSITY OF BABYLON," *JUBPAS: Journal of University of Babylon for Pure and Applied Sciences*, vol. 29, no. 3, pp. 34–43, 2021.

[8] J. L. A. Colmenero and J. P. G. Miguel, "A REGIONALIZATION APPROACH BASED ON THE COMPARISON OF DIFFERENT CLUSTERING TECHNIQUES," *Applied Sciences*, vol. 14, no. 22, pp. 1–27, Nov. 2024, doi: https://doi.org/10.3390/app142210563.

[9] Tarmizi, "FORMS OF LEGAL PROTECTION FOR INVESTORS IN BUYING AND SELLING SHARES," *IJRS: International Journal Reglement & Society*, vol. 4, no. 3, pp. 254–260, 2023.

[10] A. Syukur and D. Istiawan, "PREDICTION OF LQ45 INDEX IN INDONESIA STOCK EXCHANGE: A COMPARATIVE STUDY OF MACHINE LEARNING TECHNIQUES," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 453–463, 2021, doi: https://doi.org/10.22266/ijies2021.0228.42.

[11] S. Aulia and S. Avriyanti, "PENGARUH PROFITABILITAS TERHADAP HARGA SAHAM PADA PERBANKAN YANG TERINDEKS LQ45 PERIODE TAHUN 2016-2021," *JAPB*, vol. 6, no. 3, pp. 733–748, 2023.

[12] Y. Pratama, E. Sulistianingsih, N. N. Debataraja, and N. Imro'ah, "K-MEANS CLUSTERING DAN MEAN VARIANCE EFFICIENT PORTFOLIO DALAM PORTOFOLIO SAHAM," *Jambura Journal of Probability and Statistics*, vol. 5, no. 1, pp. 24–30, 2024, doi: https://doi.org/10.37905/jjps.v5i1.20298.

[13] E. S. Puspaningsih, D. A. I. Maruddani, and Tarno, "K-MEDOIDS CLUSTERING DAN MEAN-VALUE AT RISK UNTUK OPTIMASI PORTOFOLIO SAHAM JAKARTA ISLAMIC INDEX," *Indonesian Journal of Applied Statistics*, vol. 6, no. 1, pp. 85–95, Jan. 2024, doi: https://doi.org/10.13057/ijas.v6i1.79231.

[14] U. Kalsum, B. Efendi, Z. Fikri, and N. Astuti, *OPTIMALISASI ALOKASI SUMBER DAYA KEUANGAN STRATEGI MAKSIMALKAN RETURN*. Malang: PT Literasi Nusantara Abadi Grup, 2023.

[15] A. N. Pratama, N. Satyahadewi, and E. Sulistianingsih, "ANALYSIS OF OPTIMAL PORTFOLIO FORMATION ON IDX30 INDEXED STOCK WITH THE MEAN ABSOLUTE DEVIATION METHOD," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1753–1764, Jul. 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1753-1764.

[16] V. P. Balqis, Subiyanto, and S. Supian, "OPTIMIZING STOCK PORTFOLIO WITH MARKOWITZ METHOD AS A REFERENCE FOR INVESTMENT COMMUNITY DECISIONS," *International Journal of Research in Community Service*, vol. 2, no. 2, pp. 71–76, 2021.doi: https://doi.org/10.46336/ijrcs.v2i2.213

[17] R. Hidayat, D. Anggraini, Y. Riyani, and E. Endri, "OPTIMAL SHARIA PORTFOLIO SELECTION TO OUTPERFORM THE STOCK MARKET IN THE POST-PANDEMIC ERA," *Quality - Access to Success*, vol. 23, no. 187, pp. 287–294, Apr. 2022, doi: https://doi.org/10.47750/QAS/23.187.35.

[18] R. A. Sary, N. Satyahadewi, and W. Andani, "APPLICATION OF K-MEANS++ WITH DUNN INDEX VALIDATION OF GROUPING WEST KALIMANTAN REGION BASED ON CRIME VULNERABILITY," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 4, pp. 2283–2292, Dec. 2024, doi: https://doi.org/10.30598/barekengvol18iss4pp2283-2292.

[19] A. S. Rostami, M. Badkoobe, F. Mohanna, H. Keshavarz, A. A. R. Hosseinabadi, and A. K. Sangaiah, "SURVEY ON CLUSTERING IN HETEROGENEOUS AND HOMOGENEOUS WIRELESS SENSOR NETWORKS," *Journal of Supercomputing*, vol. 74, no. 1, pp. 277–323, Jan. 2018, doi: https://doi.org/10.1007/s11227-017-2128-1.

[20] S. T. Wierzchoń and M. A. Kłopotek, "CLUSTER ANALYSIS," in *Algorithms of Cluster Analysis*, Institute of Computer Science, Polish Academy of Sciences, 2015, pp. 21–68.

[21] M. Mughnyanti, S. Efendi, and M. Zarlis, "ANALYSIS OF DETERMINING CENTROID CLUSTERING X-MEANS ALGORITHM WITH DAVIES-BOULDIN INDEX EVALUATION," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jan. 2020, pp. 1–6. doi: https://doi.org/10.1088/1757-899X/725/1/012128.

[22] F. Tempola and A. F. Assagaf, "CLUSTERING OF POTENCY OF SHRIMP IN INDONESIA WITH K-MEANS ALGORITHM AND VALIDATION OF DAVIES-BOULDIN INDEX," in *International Conference on Science and Technology*, 2018, pp. 730–733.doi: https://doi.org/10.2991/icst-18.2018.148

[23] W. Febryanti, E. Sulistianingsih, and D. Kusnandar, "ANALISIS VALUE AT RISK PADA PORTOFOLIO OPTIMAL SAHAM BLUE CHIP DENGAN METODE RESAMPLED EFFICIENT FRONTIER," *Epsilon: Jurnal Matematika Murni dan Terapan*, vol. 18, no. 1, pp. 30–40, 2024.doi: https://doi.org/10.20527/epsilon.v18i1.10170