Current Science Research Bulletin

Volume 03 (06), June 2026, pp. 168-175

CrossRef DOI: 10.55677/csrb/02-V03I06Y2026

Available on: https://csrbjournal.org/

e-ISSN: 3050-7103
p-ISSN: 3050-709X

Application of K-Medoids Clustering for The Formation of Mean-Semivariance Portfolios on The Idxshagrow Index

Aulia Kusuma Dewi[1], Agus Rusgiyono[2], Rahmila Dapa[3*]
[1,2,3] Department of Statistics, Faculty of Science and Mathematics, Diponegoro University

ABSTRACT: The number of investors in the Indonesian capital market continues to grow, underscoring the need for investment strategies that deliver optimal returns with controlled risk. Since investors prioritize downside risk over profit fluctuations, this study applies a downside-risk approach to portfolio formation. Stocks in the IDX Sharia Growth (IDXSHAGROW) index are grouped using K-Medoids Clustering based on Return on Assets (ROA) and Return on Equity (ROE). From each cluster, one stock with the highest expected return is selected to construct a Mean–Semivariance portfolio. Portfolio performance is evaluated using the Sharpe Ratio, while risk is measured using the Historical Simulation Value-at-Risk method. The results show that four optimal clusters were formed. The Mean–Semivariance portfolio consists of EMTK (25.16%), RAJA (2.47%), SSIA (9.72%), and TAPG (62.66%), with an expected return of 0.003446. A positive Sharpe Index value of 0.144198 indicates that the portfolio outperforms the risk-free return. VaR risk measurement at a 95% confidence level with a 1-day holding period of IDR 301,801.4, assuming investment capital of IDR 10,000,000.

KEYWORDS: IDX Sharia Growth; K-Medoids Clustering; Profitability Ratios; Mean-Semivariance; Sharpe Ratio; VaR Historical Simulation

1. INTRODUCTION

Investment in the capital market is currently experiencing rapid growth. This is evidenced by data from the Kustodian Sentral Efek Indonesia (2025), which shows that the number of capital market investors in Indonesia has grown significantly from 14.87 million at the end of 2024 to 19.2 million in October 2025. This increase is also accompanied by an expanding variety of investment instruments. One of the most well-known investment instruments is stocks, which are traded through the Indonesia Stock Exchange. The Indonesia Stock Exchange provides various stock indices as investment benchmarks, one of which is the IDX Sharia Growth (IDXSHAGROW), comprising 30 sharia-compliant stocks focused on profit growth and featuring strong trading liquidity.

When investing, investors not only expect returns but also face risks; therefore, it is essential to understand a company's financial performance before making investment decisions. One of the analyses used is profitability ratios, which measure a company's ability to generate profits (Desiyanti, 2017). Return on Equity (ROE) measures how effectively a company generates profits for its shareholders, while Return on Assets (ROA) measures the company's efficiency in utilizing its assets to generate profits. Information regarding these profitability ratios serves as the foundation for investors in the risk management phase. To minimize risk, investors need to diversify their portfolios by combining several assets, thereby reducing risk without sacrificing expected returns (Tandelilin, 2017). The main challenge in portfolio construction is selecting the right stocks from the many available. One method is cluster analysis, a multivariate technique for grouping stocks based on their characteristics and differences between groups.

K-Medoids Clustering is a clustering method that can be applied in the stock selection process. This method is a more robust alternative to K-Means when dealing with outliers. In K-Medoids Clustering, each cluster is represented by an object, known as a medoid, within the cluster (Gubu et al., 2021). The final step in cluster analysis is validating the clustering results to determine the optimal number of clusters. In this context, the Silhouette Coefficient can be used.

For each cluster generated, the top-performing stock is selected to form the portfolio. In 1952, Markowitz introduced the Mean–Variance method to determine portfolio weights based on average return and risk variance, assuming returns are normally

Cite the Article: *Dewi, A.K., Rusgiyono, A., Dapa, R. (2026). Application of K-Medoids Clustering for The Formation of Mean-Semivariance Portfolios on The Idxshagrow Index. Current Science Research Bulletin, 3(6), 168-175.* https://doi.org/10.55677/csrb/02-V03I06Y2026

Publication Date: *June 16, 2026*

Corresponding Author: Rahmila Dapa

distributed. However, in practice, stock return data often do not meet this assumption. Therefore, in 1959, Markowitz proposed the Mean–Semivariance approach, which measures downside risk—that is, deviations of returns below the expected level. This approach is more suitable for investors who tend to avoid losses and allows the optimization process to be carried out using techniques similar to the Mean–Variance method. The formed portfolio needs to be evaluated for performance and risk to ensure it has achieved the set investment objectives and to estimate potential future losses.

This study aims to cluster the 30 stocks in the IDX Sharia Growth (IDXSHAGROW) index based on profitability ratios using the K-Medoids Clustering algorithm. The optimal cluster is validated using the Silhouette Coefficient. Based on the formed clusters, a portfolio will be constructed using the Mean-Semivariance method, selecting the highest expected return from each cluster. Portfolio performance is measured using the Sharpe ratio to assess how well the portfolio performs, a consideration for investors in making future investment decisions. Value at Risk (VAR) using the Historical Simulation method is used to measure portfolio risk.

2. METHODOLOGY

a. Stock Investment

Investment is the act of allocating funds or other resources in the present with the expectation of generating a profit in the future (Tandelilin, 2017). Investments can be made in financial assets such as bonds, stocks, and deposits, or in real assets such as precious metals and real estate. One of the most widely recognized investment assets is stock. Stocks are certificates of ownership in a company organized as a Limited Liability Company (Maruddani, 2019). Shareholders have the opportunity to earn returns through capital gains, which result from the difference between the selling and buying prices, as well as dividends, which are a distribution of the company's profits. Returns are divided into two categories: realized returns, which have already occurred, and expected returns, which have not yet occurred but are anticipated to occur in the future. One type of realized return is the log return, calculated using Equation (1), while expected returns are determined by averaging all realized returns using Equation (2).

$$R_{i,t} = ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1)$$

$$E(R_i) = \mu_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t} \quad (2)$$

where $R_{i,t}$: the return on stock i in period t, P_t: the stock price in period t, $E(R_i)$: the expected return on stock I, and T: the number of return periods

Indonesian Stock Exchange has launched various stock indices that serve as benchmarks reflecting the performance of various stocks (Tandelilin, 2017). The Indonesia Stock Exchange launched the IDX Sharia Growth (IDXSHAGROW) as one of the sharia stock indices in the capital market. The IDX Sharia Growth (IDXSHAGROW) index measures the price performance of 30 sharia-compliant stocks that exhibit rising net profit and revenue, strong financial performance, and transaction liquidity (Bursa Efek Indonesia, 2025).

b. K-Medoids Clustering

Cluster analysis is a method for grouping a number of objects into several clusters based on their characteristics, so that objects in the same cluster are highly similar to one another but differ from objects in other clusters. Cluster analysis groups objects based on their degree of similarity, measured by the distance between them. In cluster analysis, there are two assumptions that must be met: the representative sample assumption and the absence of multicollinearity.

The K-Medoids method, also known as Partitioning Around Medoids (PAM), is a non-hierarchical clustering method that groups n objects into k clusters. K-Medoids clustering uses medoids as cluster centers. Medoids refer to objects within a cluster that have the minimum average difference between that point and all other cluster members (Gubu et al., 2021). According to Han et al. (2012)), the K-Medoids Clustering algorithm is as follows:

1. Determine the number of clusters (k) to be formed.
2. Select k objects at random from the data as initial medoids.
3. Assign each object to a specific cluster based on the shortest distance between that object and each of the selected medoids.
4. Randomly select an object as a candidate for a new medoid.
5. Calculate D (the difference in the total shortest distances) resulting from swapping the initial medoids with the new medoids.
6. If $D < 0$, swap the initial medoids with the new medoids to form a new set of medoid objects.
7. Repeat steps 3 through 6 until there are no further medoid changes or $D \geq 0$.

The silhouette coefficient is used to evaluate clustering results and determine the optimal number of clusters. The silhouette coefficient measures the closeness of relationships within a cluster and the extent to which it is separated from other clusters. The steps for validating clusters using the silhouette coefficient are as follows (Struyf et al., 1997):

1. Calculate the average distance between the i-th object and all objects in the same cluster using Equation (3).

$$a(i) = \frac{1}{n_A - 1} \sum_{j \epsilon A, j \neq i} d_{i,j} \quad (3)$$

where $a(i)$: the average distance of object I from all objects in cluster A, $d(i,j)$: the distance between object i and object j in cluster C, and n_A: the number of objects in cluster A.

2. Calculate the average distance of object i from all objects in other clusters using Equation (4).

$$d(i,C) = \frac{1}{n_C} \sum_{j \epsilon C} d_{i,j} \quad (4)$$

where $d(i,j)$: the distance between object i and object j in cluster C, where $C \neq A$; n_C: the number of objects in cluster C

3. Determine the minimum value of $d(i,C)$ for each C, where the minimum value, or $b(i)$ represents the difference in the average distance of the i-th object to its nearest neighbors using Equation (5).

$$b(i) = \min_{C \neq A} d\,(i,C) \quad (5)$$

where $b(i)$: the minimum value of the average distance of the i-th object to all objects in the other cluster C

4. Calculate the Silhouette Coefficient for each i-th object using Equation (6).

$$SC(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (6)$$

5. Calculate the average Silhouette Coefficient for all objects using Equation (7)

$$\overline{SC} = \frac{1}{n} \sum_{i=1}^{n} SC(i) \quad (7)$$

c. Portfolio Construction Using Mean-Semivariance

A portfolio is a combination of several securities selected by an investor as an investment target over a specific time period and under specific conditions (Maruddani, 2019). The Mean-Semivariance portfolio method was introduced by Harry Markowitz in 1959 as an extension of the Mean-Variance method, first introduced in 1952. Markowitz et al. (1993) noted that semivariance is a more appropriate measure of risk than variance, as investors are more concerned about portfolio underperformance (portfolio performance that falls short of set objectives) than overperformance. Semivariance is the mean of the squared differences between a specified threshold and observations that fall below that threshold. Portfolio semivariance provides a measure of downside risk that focuses on negative fluctuations in returns, thereby ignoring all values above the average or target return.

Markowitz (1959) estimated portfolio semivariance using the following formula:

$$S_p^2 = \frac{1}{T} \sum_{g=1}^{G} \left(w_i R_{ig}\right)^2 \quad (8)$$

The semivariance of a portfolio with benchmark B is expressed as follows:

$$S_{pB}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j S_{ijB} \quad (9)$$

$$S_{ijB} = \frac{1}{T} \sum_{g=1}^{G} \left(R_{ig} - B_g\right)\left(R_{jg} - B_g\right) \quad (10)$$

The mean-semivariance of a portfolio is more complex to solve than the mean-variance because the semicovariance matrix is endogenous and asymmetric. Estrada (2008) proposed a heuristic approach that yields a symmetric and exogenous semicovariance matrix by defining semicovariance and semivariance. The semivariance of the return on stock i relative to benchmark B is as follows:

$$S_{iB}^2 = E\left\{\left[Min\left(R_{i,t} - B_t, 0\right)\right]^2\right\} = \frac{1}{T} \sum_{t=1}^{T} \left[Min\left(R_{i,t} - B_t, 0\right)\right]^2 \quad (11)$$

The semicovariance between stock i and stock j relative to benchmark B is as follows:

$$S_{ijB} = E\left\{Min\left(R_{i,t} - B_t, 0\right).Min\left(R_{j,t} - B_t, 0\right)\right\} = \frac{1}{T} \sum_{t=1}^{T} \left[Min\left(R_{i,t} - B_t, 0\right).Min\left(R_{j,t} - B_t, 0\right)\right] \quad (12)$$

where S_{iB}^2: the semivariance of stock i relative to benchmark B, S_{ijB}: the semicovariance between stock i and stock j relative to benchmark B, B_t: the benchmark at period t, and T: the number of observation periods.

The Mean-Semivariance portfolio weighting is performed using the weight $\boldsymbol{w} = [w_1 \cdots w_N]^T$ (Entrisnasari, 2015). To minimize the semivariance $(\boldsymbol{w}^T \boldsymbol{\Sigma}_{sv} \boldsymbol{w})$, the weight vector w is determined by considering the following two constraints:

1. The portfolio's mean return (μ_p) must meet the target return, expressed as $\boldsymbol{w}^T\boldsymbol{\mu} = \mu_p$
2. The sum of the weights of the resulting portfolio must equal one, expressed as $\boldsymbol{w}^T\mathbf{1}_N = 1$

where $\boldsymbol{w}$: an $N \times 1$ stock weight vector, $\boldsymbol{w}^T$: the transpose of vector $\boldsymbol{w}$, $\boldsymbol{\Sigma}_{sv}$: an $N \times N$ semivariance-semicovariance matrix, and $\mathbf{1}_N$: a column vector of length N × 1.

The optimization problem can be solved using the Lagrange function (Maruddani, 2019).

$$L = \boldsymbol{w}^T\boldsymbol{\Sigma}_{sv}\boldsymbol{w} + \lambda_1(\mu_p - \boldsymbol{w}^T\boldsymbol{\mu}) + \lambda_2(1 - \boldsymbol{w}^T\mathbf{1}_N) \tag{13}$$

where L: Lagrange function and λ: Lagrange multiplier

The optimal value of the weight vector $\boldsymbol{w}$ is obtained by differentiating the Lagrange function with respect to $\boldsymbol{w}$. The Lagrange function is minimized if it satisfies the conditions $\frac{dL}{dw} = 0$ dan $\frac{d^2L}{dw^2} > 0$.

The first derivative,

$$\frac{dL}{d\boldsymbol{w}} = 0$$

$$\frac{d}{d\boldsymbol{w}}\left[\boldsymbol{w}^T\boldsymbol{\Sigma}_{sv}\boldsymbol{w} + \lambda_1(\mu_p - \boldsymbol{w}^T\boldsymbol{\mu}) + \lambda_2(1 - \boldsymbol{w}^T\mathbf{1}_N)\right] = 0$$

$$2\boldsymbol{\Sigma}_{sv}\boldsymbol{w} - \lambda_1\boldsymbol{\mu} - \lambda_2\mathbf{1}_N = 0 \tag{14}$$

$$2\boldsymbol{\Sigma}_{sv}\boldsymbol{w} = \lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N$$

$$\boldsymbol{\Sigma}_{sv}\boldsymbol{w} = \frac{1}{2}(\lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N)$$

$$\boldsymbol{w} = \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}(\lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N) \tag{15}$$

Equation (15), both sides are multiplied by $\mathbf{1}_N{}^T$

$$\mathbf{1}_N{}^T\boldsymbol{w} = \frac{1}{2}\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}(\lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N)$$

since $\mathbf{1}_N{}^T\boldsymbol{w} = 1$, then

$$1 = \frac{1}{2}\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}(\lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N)$$

$$2 = \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}(\lambda_1\boldsymbol{\mu} + \lambda_2\mathbf{1}_N)$$

$$2 = \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_1\boldsymbol{\mu} + \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_2\mathbf{1}_N$$

$$\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_2\mathbf{1}_N = 2 - \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_1\boldsymbol{\mu}$$

$$\lambda_2 = \frac{2 - \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_1\boldsymbol{\mu}}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}$$

By substitution, we obtain

$$\boldsymbol{w} = \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}\left(\lambda_1\mu + \left(\frac{2 - \mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_1\boldsymbol{\mu}}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}\right)\mathbf{1}_N\right)$$

$$\boldsymbol{w} = \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}\lambda_1\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}\frac{2\mathbf{1}_N - \lambda_1\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\boldsymbol{\mu}\mathbf{1}_N}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}$$

$$\boldsymbol{w} = \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}\boldsymbol{\lambda}_1\boldsymbol{\mu} + \frac{\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N} - \frac{1}{2}\boldsymbol{\Sigma}_{sv}{}^{-1}\left(\frac{\lambda_1\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\boldsymbol{\mu}\mathbf{1}_N}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}\right)$$

$$\boldsymbol{w} = \frac{1}{2}\lambda_1\left(\boldsymbol{\Sigma}_{sv}{}^{-1}\boldsymbol{\mu} - \frac{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\boldsymbol{\mu}}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N\right) + \frac{\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}$$

In a portfolio with efficient semivariance, there are no restrictions on the portfolio mean, so $\lambda_1 = 0$. Thus, the mean-semivariance weights can be calculated using the formula:

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N}{\mathbf{1}_N{}^T\boldsymbol{\Sigma}_{sv}{}^{-1}\mathbf{1}_N} \tag{16}$$

where $\boldsymbol{\Sigma}_{sv}{}^{-1}$: the inverse of the semivariance-semicovariance matrix and $\mathbf{1}_N{}^T$: the transpose of the first column vector, which has N elements transpose vektor kolom 1 sebanyak N

Taking the second derivative of Equation (14) yields a positive semidefinite matrix, which proves that the resulting Lagrange function has a minimum value.

$$\frac{d^2L}{d\boldsymbol{w}^2} = \frac{d}{d\boldsymbol{w}}[2\boldsymbol{\Sigma}_{sv}\boldsymbol{w} - \lambda_1\boldsymbol{\mu} - \lambda_2\mathbf{1}_N] = 2\boldsymbol{\Sigma}_{sv}$$
$$2\boldsymbol{\Sigma}_{sv} > 0$$

d. Portfolio Performance and Risk Evaluation

Portfolio performance can be evaluated using the Sharpe ratio. The Sharpe ratio indicates how much a portfolio's return exceeds the risk-free investment return per unit of portfolio risk. The Sharpe ratio is calculated using Equation (17).

$$SR = \frac{E(R_p) - R_f}{s_p} \tag{17}$$

where SR: the portfolio's Sharpe ratio, $E(R_p)$: the expected return of the portfolio, s_p: the standard deviation of the portfolio's return, and R_f: the average interest rate of the portfolio's risk-free investment.

Portfolio risk can be calculated using Value at Risk (VaR).Value at Risk (VaR) is a measure that calculates the maximum potential loss over a specific period at a specified confidence level. One nonparametric approach to measuring Value at Risk is Historical Simulation, which uses historical asset return data to estimate VaR. The VaR calculation is performed using the following equation:

$$VaR = V_0 \times kuantil\ \alpha \times \sqrt{t} \tag{18}$$

where VaR: maximum potential loss, V_0: investment amount, $kuantil\ \alpha$: return value at the $\alpha \times T$ data point from the historical data, and t: holding period.

3. DATA AND MATERIAL

The data used are secondary, obtained from official sources. Specifically, it consists of a list of the 30 stocks included in the IDX Sharia Growth Index (IDXSHAGROW) following the major evaluation in May 2025, obtained from idx.co.id. Additionally, the profitability ratios—Return on Assets (ROA) and Return on Equity (ROE)—for each stock as of October 2025 were used as clustering variables. Daily closing prices for the 30 IDXSHAGROW stocks and the Composite Stock Price Index (as a benchmark) for the period from October 17, 2024, to October 17, 2025, obtained from finance.yahoo.com, were used for portfolio construction.

The following are the data analysis steps in this study:

1. Prepare data on Return on Assets (ROA), Return on Equity (ROE), daily closing prices of stocks included in the IDXSHAGROW index, daily closing prices of the Composite Stock Price Index (IHSG), and the BI Rate for the specified period.
2. Standardizing the ROA and ROE data of the constituent stocks using the Maximum Absolute Scaler.
3. Conducting tests for the assumptions of a representative sample and non-multicollinearity on the standardized data.
4. Analyzing clusters using the K-Medoids Clustering method.
5. Determining the optimal number of clusters using the Silhouette Coefficient.
6. Calculate the return value for each stock.
7. Calculate the semivariance and semicovariance values of the selected stocks to form the semivariance-semicovariance matrix.
8. Calculate the weights of each stock comprising the Mean-Semivariance portfolio.
9. Evaluate the performance of the formed portfolio using the Sharpe ratio.
10. Calculate the Value at Risk of the portfolio using the Historical Simulation method.

4. RESULTS AND DISCUSSION

The clustering process uses the K-Medoids Clustering algorithm on standardized data. Before performing the cluster analysis, two assumptions must be met, namely:

1. Assumption of a representative sample
   The sample is considered representative of the population based on the KMO (Kaiser-Mayer-Olkin) test. A KMO value between 0.5 and 1 indicates that the sample meets the criteria for representativeness. The calculated KMO value of 0.5 indicates that the representativeness assumption is met.
2. Assumption of Non-Multicollinearity
   Multicollinearity can be detected using the Variance Inflation Factor (VIF) for each variable. If the VIF value is < 10, the assumption of non-multicollinearity is met. The VIF values for each variable are shown in Table 1.

Table 1. VIF Values for Each Variable

Variable	VIF
ROA	6.146467
ROE	6.146467

Based on Table 1, the VIF values for ROA and ROE are below 10; therefore, it can be concluded that there is no multicollinearity between the two variables, or the assumption of non-multicollinearity is met.

Next, clustering was performed using K-Medoids Clustering in RStudio for k = 2, 3, 4, 5, 6, 7, and 8. The clustering results for each k were evaluated using the Silhouette Coefficient to determine the optimal number of clusters. The Silhouette Coefficient values for k = 2, 3, 4, 5, 6, 7, and 8 are shown in Table 2.

Table 2. Silhouette Coefficient Values for Each k

Number of Clusters	Nilai
2	0.456588
3	0.528773
4	0.560517
5	0.470533
6	0.425532
7	0.394517
8	0.330656

The results in Table 2 indicate that four clusters constitute the optimal number, with the highest Silhouette Coefficient of 0.560517. Therefore, four stocks will be selected to form the portfolio. The results in Table 2 show that four clusters are optimal, with the highest silhouette coefficient of 0.560517. Therefore, four stocks will be used to construct the portfolio.

Table 3. Best Clustering Results Cluster

Cluster	Stock Members	Centroid Value	
		ROA	ROA
1	ADRO, AKRA, DSNG, ELSA, EMTK, ITMG, JPFA, MAPA, PTBA, TINS, TKIM, UNTR	9.373333	15.135000
2	BRIS, CTRA, ERAA, ESSA, ICBP, INDF, ISAT, MEDC, PGAS, RAJA	4.185000	9.657000
3	EXCL, INCO, INKP, KIJA, KPIG, RAJA	1.091667	1.578333
4	MARK, TAPG	28.205000	33.325000

The clustering results with 4 optimal clusters are shown in Table 3. Based on Table 3, Cluster 1 contains 12 stock members, Cluster 2 contains 10 stock members, Cluster 3 contains 6 stock members, and Cluster 4 contains 2 stock members. Information regarding the characteristics of the clustering results is based on the centroid value of each cluster. According to Table 3, Cluster 4 contains stocks with the highest ROA and ROE compared to the other clusters, and Cluster 2 contains stocks with the lowest ROA and ROE compared to the other clusters.

The portfolio selection is based on choosing one stock from each cluster with the highest expected return. The expected return values were calculated using stock return data from October 17, 2024, to October 17, 2025. Based on the expected return values, the EMTK stock represents Cluster 1, the RAJA stock represents Cluster 2, the SSIA stock represents Cluster 3, and the TAPG stock represents Cluster 4. The stocks with the highest expected returns in each cluster are shown in Table 4. These four stocks also come from different sectors: EMTK from the technology sector, RAJA from the energy sector, SSIA from the infrastructure sector, and TAPG from the consumer non-cyclicals sector. This sectoral diversity provides diversification benefits that can reduce industry-specific risk and help stabilize the portfolio against market volatility.

Table 4. Nilai *Expected Return* Tertinggi

Klaster	Saham	*Expected Return*
1	PT Mahkota Teknologi Tbk (EMTK)	0.004999
2	PT Rukun Raharja Tbk (RAJA)	0.004029
3	PT Surya Semesta Internusa Tbk (SSIA)	0.001674
4	PT Triputra Agro Persada Tbk (TAPG)	0.003098

The four representative stocks from each cluster form the Mean-Semivariance portfolio. The construction of the Mean-Semivariance portfolio begins with the calculation of the semivariance and semicovariance of the selected stocks to form the semivariance-semicovariance matrix. The results of the semivariance-semicovariance matrix calculations are presented in Table 5.

Corresponding Author: Rahmila Dapa

Table 5. Semivariance -Semicovariance Matrix

	EMTK	RAJA	SSIA	TAPG
EMTK	0.000462	0.000226	0.000184	0.000117
RAJA	0.000226	0.001187	0.000234	0.000165
SSIA	0.000184	0.000239	0.000667	0.000153
TAPG	0.000117	0.000165	0.000153	0.000264

The calculation of the Mean-Semivariance portfolio weights was performed using Equation (16), yielding the following weights for each stock.

Table 6. Weights of Each Stock

Stock Code	Weight	Weight Percentage
EMTK	0.251565	25.16%
RAJA	0.024688	2.47%
SSIA	0.097155	9.72%
TAPG	0.626592	62.66%

The resulting portfolio yields an expected return of 0.003446, indicating that the portfolio will generate a return of 0.3446% on the total invested capital in the future. Portfolio performance was measured using the Sharpe ratio (Equation (17), yielding a positive value of 0.144197. This indicates that the formed portfolio outperforms the average risk-free investment rate. Portfolio risk measurement using the Value at Risk (VaR) method via Historical Simulation, calculated using Equation (18), yields a portfolio VaR value of -301,853.5 at a 95% confidence level with a holding period of 1 day.

5. CONCLUSION

Based on the results and discussion, the following conclusions were drawn:

The clustering results for the IDX Sharia Growth (IDXSHAGROW) index using K-Medoids Clustering indicated an optimal number of 4 clusters, as determined by the silhouette coefficient. Cluster 1 consists of 12 members: ADRO, AKRA, DSNG, ELSA, EMTK, ITMG, JPFA, MAPA, PTBA, TINS, TKIM, and UNTR. Cluster 2 consists of 10 members: BRIS, CTRA, ERAA, ESSA, ICBP, INDF, ISAT, MEDC, PGAS, and RAJA. Cluster 3, with 6 members, consists of EXCL, INCO, INKP, KIJA, KPIG, and SSIA. Cluster 4, with 2 members, consists of MARK and TAPG.

The results of the portfolio formation from IDX Sharia Growth (IDXSHAGROW) stocks using the Mean-Semivariance method were derived from the representative stocks of each cluster with the highest expected return, yielding 4 portfolio constituent stocks with the following weights: EMTK at 25.16%, RAJA at 2.47%, SSIA at 9.72%, and TAPG at 62.66%. The expected return of the resulting portfolio is 0.003446.

Performance evaluation of the resulting portfolio using the Sharpe ratio yields a positive value of 0.144197. This indicates that the resulting portfolio outperforms the average risk-free investment rate. Portfolio risk was measured using the Value-at-Risk (VaR) method via Historical Simulation. With a capital of Rp10,000,000, a 95% confidence level, and a holding period of 1 day, the estimated maximum loss an investor would incur over the next 1 day is Rp301,853.5.

REFERENCES

1. Bursa Efek Indonesia. (2025). *Indeks Saham Syariah*. https://www.idx.co.id/id/idx-syariah/indeks-saham-syariah
2. Desiyanti, R. (2017). *Teori Investasi Dan Portofolio*. Bung Hatta University Press.
3. Entrisnasari, F. V. (2015). Analisis Portofolio Optimum Saham Syariah Menggunakan Mean Semivarian. *Jurnal Fourier*, *4*(1), 41–57.
4. Estrada, J. (2008). Mean-Semivariance Optimization: A Heuristic Approach. *Journal Of Applied Finance*, 57–72.
5. Gubu, L., Rosadi, D., & Abdurakhman. (2021). Pembentukan Portofolio Saham Menggunakan Klastering Time Series K-Medoid Dengan Ukuran Jarak Dynamic Time Warping. *Jurnal Aplikasi Statistika & Komputasi Statistik*, *13*(2), 35–46.
6. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd Edition). Morgan Kaufmann.
7. Kustodian Sentral Efek Indonesia. (2025). *Statistik Pasar Modal Indonesia: September 2025*. https://web.ksei.co.id/files/Statistik_Publik_September_2025_final.pdf
8. Markowitz, H. (1959). *Portofolio Selection: Efficient Diversification Of Investments*. John Wiley & Sons, Inc.

Corresponding Author: Rahmila Dapa

9. Markowitz, H., Todd, P., Xu, G., & Yamane, Y. (1993). Computation of Mean-Semivariance Efficient Sets by the Critical Line Algorithm. *Annals of Operations Research*, *45*, 307–317.
10. Maruddani, D. A. I. (2019). *Value at Risk Untuk Pengukuran Risiko Investasi Saham: Aplikasi dengan Program R*. Wade Group.
11. Struyf, A., Hubert, M., & Rousseeuw, P. J. (1997). Integrating Robust Clustering Techniques in S-PLUS. *Computational Statistics & Data Analysis*, *26*, 17–37.
12. Tandelilin, E. (2017). *Pasar Modal: Manajemen Portofolio & Investasi*. Kanisius.

Corresponding Author: Rahmila Dapa