Available online at https://journal.rescollacomm.com/index.php/ijqrm/index

International Journal of Quantitative Research and Modeling

e-ISSN 2721-477X
p-ISSN 2722-5046

Vol. 7, No. 2, pp. 174-181, 2026

Mean-VaR Portfolio Diversification Based on K-Medoids Clustering

Deva Putra Setyawan[1], Alim Jaizul Wahid [2*], Riza Andrian Ibrahim[3]

[1] *Doctoral of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*

[2] *Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia*

[3] *Doctoral of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*

**Corresponding author email: alimjaizul47@gmail.com*

Abstract

This study develops a diversified stock portfolio by integrating the Mean-Value at Risk (Mean-VaR) model with K-Medoids clustering. The approach groups stocks according to similar risk-return characteristics before the portfolio optimization stage. The data consist of daily closing prices of LQ45 index constituents from 3 February to 31 July 2025, obtained from the Indonesia Stock Exchange and Yahoo Finance. Of the 45 LQ45 stocks, 18 stocks satisfied the criteria of data completeness, liquidity, market capitalization stability, and sector representation. Clustering was performed using expected return and 95% Value at Risk (VaR) as input variables. The best clustering structure was obtained for two clusters, with a Silhouette Index of 0.6882. The first cluster represents aggressive stocks with relatively high expected returns and higher downside risk, including ANTM, BRPT, AMMN, and MDKA. The second cluster represents defensive stocks with lower risk and more stable returns, including INDF, ASII, ICBP, BBCA, and TLKM. The optimal Mean-VaR portfolio was constructed with minimum inter-cluster allocation constraints of 30% for the aggressive cluster and 70% for the defensive cluster. The resulting portfolio produced a daily expected return of 0.003272 and a 95% VaR of -0.029053. These results indicate that K-Medoids clustering can support portfolio diversification by identifying distinct risk-return groups and improving risk control in investment allocation.

Keywords: Investment portfolio, Mean-VaR, K-Medoids, Value at Risk, stock diversification

1. Introduction

The capital market is an important indicator of economic development because it channels funds from investors to firms that require capital for productive activities. Among available investment instruments, stocks are attractive because they provide opportunities for high returns (Wahid et al., 2025). However, stock investment also contains substantial risk due to price fluctuations that cannot be predicted with certainty. Therefore, quantitative approaches are required to help investors manage the trade-off between return and risk objectively (Hidayat et al., 2023).

Value at Risk (VaR) is a widely used risk measure in financial risk management. VaR estimates the maximum potential loss that may occur at a specified confidence level and investment horizon (Sarykalin et al., 2008). Compared with the classical Mean-Variance model, the Mean-VaR framework is more suitable for financial return data because it can account for asymmetric return distributions and extreme loss behavior (Parrak & Seidler, 2010).

Diversification is a key strategy in portfolio construction because it reduces total investment risk without necessarily sacrificing expected return (Wahid & Saputra, 2025). In this study, diversification is supported by clustering analysis. Clustering groups stocks based on similar return and risk characteristics, allowing investors to combine assets from different groups to create a more stable portfolio (Rudianto & Wijayanto, 2024).

K-Medoids is a partition-based clustering algorithm that uses actual data objects as cluster centers. Unlike centroid-based approaches, K-Medoids is more robust to outliers because the medoid is an observed object that best represents its cluster (Pramesti et al., 2017). This characteristic is useful for financial data, which often contain noise and extreme observations.

Although K-Medoids has been applied in various classification and grouping problems, its integration with Mean-VaR portfolio optimization remains limited. This study addresses this gap by developing a portfolio diversification model that combines K-Medoids clustering with Mean-VaR optimization. The objective is to identify risk-return stock groups and construct a diversified portfolio that balances return opportunity and downside risk control in the Indonesian stock market.

2. Literature Review

Modern portfolio theory explains that portfolio selection involves a trade-off between expected return and risk. Markowitz (1952) formulated portfolio optimization by minimizing variance for a given expected return. In practice, however, variance does not distinguish between upside and downside volatility. For investors who focus on potential losses, downside-risk measures such as VaR provide a more direct risk interpretation.

The Mean-VaR model extends portfolio optimization by measuring risk through the quantile of the loss distribution. This model is suitable for financial market data that may be non-normal, skewed, or affected by extreme fluctuations. Parrak and Seidler (2010) noted that Mean-VaR can provide a more realistic risk-return framework in crisis or volatile market conditions.

Clustering analysis can improve diversification by identifying assets with similar behavior. K-Medoids is suitable for this purpose because it selects representative observations as medoids and is less affected by extreme values than K-Means. Previous applications have shown that K-Medoids produces interpretable clusters in heterogeneous datasets (Bau et al., 2023; Pramesti et al., 2017). Rudianto and Wijayanto (2024) also showed that K-Medoids may provide a better within-between cluster ratio than K-Means in certain grouping problems.

Based on these findings, the integration of K-Medoids and Mean-VaR is expected to provide two advantages. First, the clustering stage identifies the intrinsic risk-return structure among stocks. Second, the optimization stage uses this structure to form a portfolio allocation that remains diversified across aggressive and defensive stock groups.

3. Materials and Methods

3.1. Data and Research Objects

This study uses a quantitative descriptive-computational approach. The research objects are stocks listed in the LQ45 index of the Indonesia Stock Exchange. The observation period covers two LQ45 composition cycles, namely February-April 2025 and May-July 2025. Daily closing price data were collected from 3 February to 31 July 2025.

The data were obtained from the official Indonesia Stock Exchange website and Yahoo Finance. Not all LQ45 constituents were included in the analysis. Stock selection was based on data completeness, trading liquidity, market capitalization stability, and sector representation. A maximum sector concentration criterion was also considered to maintain portfolio diversification.

The stock must be included in the LQ45 index during the February-April 2025 and May-July 2025 periods.
The stock must have complete daily closing price data and no trading suspension during the observation period.
The stock must have adequate transaction liquidity and relatively stable market capitalization.
Sector representation must be maintained so that the portfolio is not concentrated in one sector.

Table 1: Selected LQ45 stocks used in the analysis

No.	Ticker	Sector
1	INDF.JK	Consumer Defensive
2	CPIN.JK	Consumer Defensive
3	BBRI.JK	Financial Services
4	UNTR.JK	Basic Materials
5	BBCA.JK	Financial Services
6	BMRI.JK	Financial Services
7	ANTM.JK	Basic Materials
8	ASII.JK	Industrials
9	BBNI.JK	Financial Services
10	GOTO.JK	Technology
11	BRPT.JK	Basic Materials
12	AMMN.JK	Basic Materials
13	ICBP.JK	Consumer Defensive
14	ADRO.JK	Energy
15	TLKM.JK	Communication Services
16	BRIS.JK	Financial Services
17	MDKA.JK	Basic Materials
18	AMRT.JK	Consumer Defensive

3.2. Return and Risk Measurement

The daily logarithmic return of stock i at time t is calculated from the closing price as follows:

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

where $P_{i,t}$ denotes the closing price of stock i at time t. The expected return is computed as the average daily return:

$$E(R_{\mathrm{i}}) = \left(\frac{1}{T}\right)\sum_{t=1}^{T} R_{i,t} \quad (2)$$

The 95% Value at Risk is calculated using the historical return distribution. In this study, VaR is expressed as a negative value to indicate the potential downside loss at the 95% confidence level:

$$VaR_{0.95} = Quantile_{0.05}(R_{\mathrm{i}}) \quad (3)$$

3.3. K-Medoids Clustering

K-Medoids clustering is used to group stocks based on expected return and 95% VaR. The algorithm starts by selecting k medoids, assigning each stock to the nearest medoid, updating the medoids to minimize within-cluster dissimilarity, and repeating the process until the medoids no longer change (Puspaningsih et al, 2024). Euclidean distance is used after standardizing the variables so that expected return and VaR have comparable scales.

The steps of the K-Medoids algorithm (Bau et al.,2023; Rudianto & Wijayanto, 2024) are as follows:

a) Initialization: Choose k random objects as initial medoids.
b) Calculate distance: Use the Euclidean distance between each object and each medoid.

$$d(i,j) = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \quad (4)$$

c) Classification: Place each object into a cluster with the closest medoid.
d) Medoid update: For each cluster, select the object that minimizes the total distance to all cluster members.
e) Iteration: Repeat steps 2–4 until there are no new medoids (convergence).

This method then generates k clusters of stocks with similar risk and return characteristics. Each cluster is represented by a medoid stock that can be used as a candidate for portfolio composition.

3.4. Cluster Evaluation Using the Silhouette Index

The quality of clustering is evaluated using the Silhouette Index. For object i, the silhouette coefficient is defined as:

$$s(i) = \frac{[b(i) - a(i)]}{\max\{a(i), b(i)\}} \quad (5)$$

where $a(i)$ is the average distance between object i and other objects in the same cluster, and $b(i)$ is the smallest average distance between object i and objects in another cluster. Values close to 1 indicate appropriate clustering, values close to 0 indicate overlapping clusters, and negative values indicate possible misclassification.

3.5. Mean-VaR Portfolio Optimization

After the clusters are obtained, the portfolio is constructed using Mean-VaR optimization with non-negative weights. Inter-cluster allocation constraints are applied to maintain diversification: at least 30% of the total weight is allocated to the aggressive cluster and 70% to the defensive cluster. The portfolio expected return is expressed as:

$$E(R_{\mathrm{p}}) = \textstyle\sum_i w_{\mathrm{i}}\, E(R_{\mathrm{i}}), \quad \sum_i w_i = 1, \quad w_{\mathrm{i}} \geq 0 \quad (6)$$

The resulting portfolio is then interpreted based on its expected return, VaR, and allocation structure across clusters.

3.6. Research Stages

The research procedure was conducted through several systematic stages, starting from data collection to portfolio performance interpretation. These stages were designed to ensure that the portfolio formation process was based on measurable risk-return characteristics and supported by clustering analysis. The research stages are described as follows.

a) Data Collection
The first stage involved collecting daily closing price data of LQ45 index stocks listed on the Indonesia Stock Exchange. The observation period covered 3 February 2025 to 31 July 2025. The data were obtained from the official website of the Indonesia Stock Exchange and Yahoo Finance.

b) Parameter Estimation
The daily closing prices were transformed into daily logarithmic returns. Furthermore, two main parameters were calculated for each stock: expected return and Value at Risk (VaR) at the 95% confidence level. These parameters were used to represent the return and risk characteristics of each stock.

c) Stock Clustering Using K-Medoids
The expected return and VaR values were used as input variables in the K-Medoids clustering algorithm. This stage aimed to classify stocks into several groups based on similar risk-return characteristics. The use of K-Medoids was considered appropriate because it is more robust to outliers and uses actual objects as cluster centers.

d) Cluster Evaluation
The quality of the clustering results was evaluated using the Silhouette Index. This evaluation was used to determine the optimal number of clusters by measuring the degree of cohesion within clusters and separation between clusters. The cluster configuration with the highest Silhouette Index value was selected as the optimal clustering structure.

e) Mean–VaR Portfolio Formation
After the optimal clusters were obtained, the portfolio was constructed using the Mean–VaR approach. The portfolio formation process considered the expected return, VaR, and cluster-based diversification constraints to ensure that the selected portfolio did not concentrate only on one stock group.

f) Efficiency Analysis and Interpretation
The final stage involved analyzing and interpreting the efficiency of the formed portfolio. The portfolio performance was evaluated based on its expected return, VaR value, and diversification structure. This stage provided insight into whether the integration of K-Medoids clustering and the Mean–VaR model could produce a more diversified and risk-controlled portfolio.

The overall research procedure is illustrated in Figure 1.

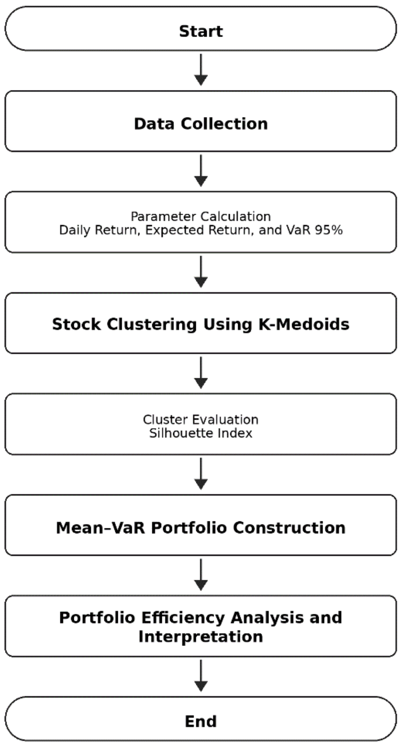

Figure 1: Research Flowchart

4. Results and Discussion

4.1. Selected Stocks and Return-Risk Parameters

Based on the selection criteria, 18 of the 45 LQ45 constituents were eligible for analysis. The selected stocks represent financial services, consumer defensive, basic materials, industrials, energy, technology, and communication services sectors. This composition provides a reasonable basis for diversification because it prevents the portfolio from being fully concentrated in a single sector.

Table 2: Expected return and 95% VaR of selected stocks

Ticker	Expected Return	VaR 95%
ADRO	-0.0005899	-0.04828
AMMN	0.0023589	-0.07286
AMRT	-0.0018719	-0.05113
ANTM	0.0081961	-0.07268
ASII	0.0014701	-0.03181
BBCA	-0.0005446	-0.03022
BBNI	-0.0002008	-0.04453
BBRI	-0.0001860	-0.04418
BMRI	-0.0008004	-0.04497
BRIS	0.0001366	-0.05436
BRPT	0.0106020	-0.08739
CPIN	0.0009659	-0.04598
GOTO	-0.0015259	-0.05546
ICBP	-0.0007068	-0.03119
INDF	0.0014022	-0.03783
MDKA	0.0064116	-0.09138
TLKM	0.0017673	-0.03980
UNTR	0.0006607	-0.03941

Table 2 shows substantial variation across stocks. BRPT, ANTM, and MDKA have relatively high expected returns but also high VaR values in absolute terms. In contrast, BBCA, ICBP, ASII, INDF, and TLKM show lower downside risk and more stable return characteristics. This contrast supports the use of clustering before portfolio optimization.

4.2. Preliminary Assumption Checks

Before clustering, preliminary checks were conducted using the Kaiser-Meyer-Olkin (KMO) measure and variance inflation factor (VIF). Although K-Medoids does not require strong correlation structures or normality assumptions, these checks help confirm that the variables are usable and not affected by excessive multicollinearity.

Table 3: KMO and VIF results

Variable	KMO	VIF
Expected Return	0.50	2.37505
VaR 95%	0.50	2.37505
Overall	0.50	-

The overall KMO value of 0.50 is at the marginal threshold. This value may be considered limited for factor analysis, but it is not a major limitation for K-Medoids because the algorithm is distance-based and non-parametric. The VIF value of 2.37505 is below 5, indicating that no serious multicollinearity exists between expected return and VaR.

4.3. K-Medoids Clustering Results

K-Medoids was applied to the standardized expected return and VaR variables. The evaluation of different k values indicated that $k = 2$ produced the highest Silhouette Index, with a value of 0.6882. This value indicates that the two-cluster solution has clear separation and good internal cohesion. The medoids are ANTM.JK for Cluster 1 and BBRI.JK for Cluster 2.

Table 4: K-Medoids clustering results

Ticker	Expected Return	VaR 95%	Cluster	Medoid
ADRO.JK	-0.000590	-0.04828	2	False
AMMN.JK	0.002359	-0.07286	1	False
AMRT.JK	-0.001870	-0.05113	2	False
ANTM.JK	0.008196	-0.07268	1	True
ASII.JK	0.001470	-0.03181	2	False
BBCA.JK	-0.000540	-0.03022	2	False
BBNI.JK	-0.000200	-0.04453	2	False
BBRI.JK	-0.000190	-0.04418	2	True
BMRI.JK	-0.000800	-0.04497	2	False
BRIS.JK	0.000137	-0.05436	2	False
BRPT.JK	0.010602	-0.08739	1	False
CPIN.JK	0.000966	-0.04598	2	False
GOTO.JK	-0.001530	-0.05546	2	False
ICBP.JK	-0.000710	-0.03119	2	False
INDF.JK	0.001402	-0.03783	2	False
MDKA.JK	0.006412	-0.09138	1	False
TLKM.JK	0.001767	-0.03980	2	False
UNTR.JK	0.000661	-0.03941	2	False

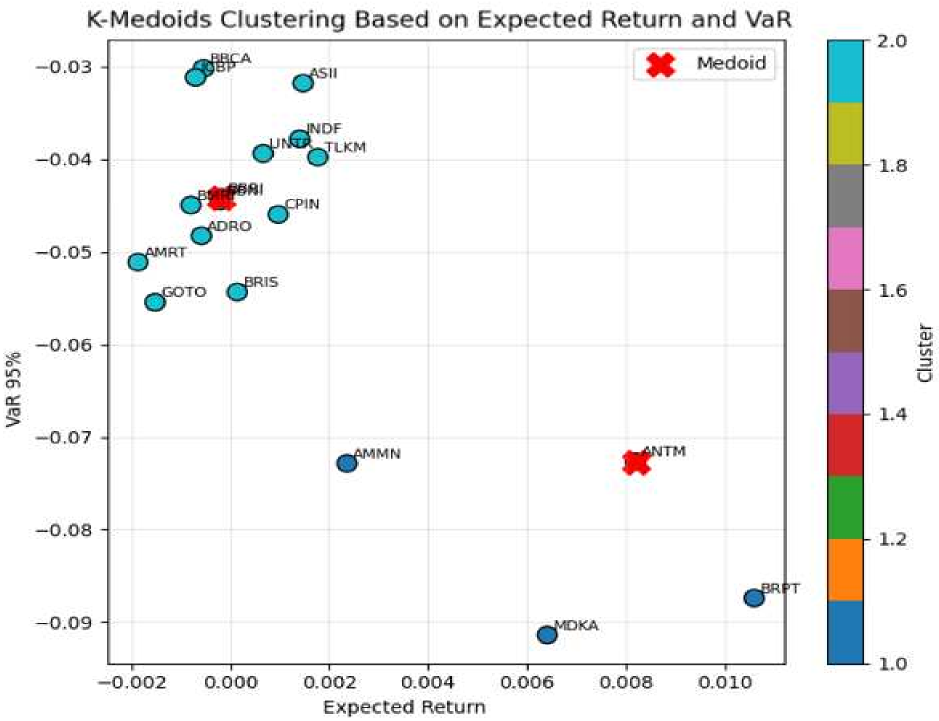

Figure 2: Distribution of stocks based on K-Medoids clustering

Cluster 1 consists of aggressive stocks with relatively high returns and higher potential losses, namely AMMN, ANTM, BRPT, and MDKA. These stocks are mainly associated with basic materials and commodity-related sectors, which tend to be more volatile. ANTM is selected as the medoid, indicating that it is the most representative member of the aggressive cluster.

Cluster 2 contains defensive stocks with lower risk and more moderate returns, including INDF, ASII, ICBP, BBCA, TLKM, and other selected stocks. This cluster is dominated by financial services, consumer defensive, communication services, and industrial stocks. BBRI is selected as the medoid, representing the central behavior of the defensive cluster.

4.4. Mean-VaR Portfolio Formation

After the clustering stage, portfolio weights were optimized using the Mean-VaR framework with non-negative weights and inter-cluster diversification constraints. The final allocation assigns 30% of the portfolio to the aggressive cluster and 70% to the defensive cluster.

Table 5: Mean-VaR portfolio weights

Ticker	Cluster	Weight
ANTM.JK	1	0.17003
BRPT.JK	1	0.12632
AMMN.JK	1	0.00365
MDKA.JK	1	0.00000
INDF.JK	2	0.19999
ASII.JK	2	0.19997
ICBP.JK	2	0.17207
BBCA.JK	2	0.06405
TLKM.JK	2	0.06392
UNTR.JK	2	0.00000
BMRI.JK	2	0.00000
BRIS.JK	2	0.00000
GOTO.JK	2	0.00000
ADRO.JK	2	0.00000
AMRT.JK	2	0.00000
BBNI.JK	2	0.00000
BBRI.JK	2	0.00000
CPIN.JK	2	0.00000

The largest weights in the aggressive cluster are assigned to ANTM and BRPT, with weights of 0.17003 and 0.12632, respectively. In the defensive cluster, the largest allocations are assigned to INDF, ASII, and ICBP, followed by BBCA and TLKM. Several stocks receive zero weights, indicating that their contribution to the optimized risk-return profile is less favorable under the applied constraints.

The optimized portfolio produces a daily expected return of 0.003272 and a 95% VaR of -0.029053. The VaR value indicates an estimated maximum daily loss of approximately 2.91% at the 95% confidence level. The dominance of defensive stocks helps stabilize the portfolio, while the 30% aggressive allocation preserves the opportunity for higher return.

4.5. Discussion

The results show that integrating K-Medoids clustering with Mean-VaR optimization provides a clearer understanding of the risk-return structure of LQ45 stocks. The clustering stage separates stocks into aggressive and defensive groups, while the optimization stage translates this structure into portfolio weights.

The two-cluster solution is consistent with the observed financial characteristics of the stocks. Commodity-related stocks such as BRPT, ANTM, AMMN, and MDKA tend to produce higher expected returns but are exposed to larger downside risk. In contrast, consumer defensive, banking, telecommunication, and industrial stocks tend to provide lower risk and more stable return behavior.

The portfolio allocation also confirms the importance of diversification across different risk profiles. A purely aggressive portfolio may increase expected return but also increase VaR substantially. A purely defensive portfolio may reduce risk but limit return potential. The 30%-70% allocation structure provides a balance between growth opportunity and risk control.

From a practical perspective, this model can be used by moderate or conservative investors who seek structured diversification based on quantitative evidence. More aggressive investors may increase allocation to Cluster 1, but they should be aware that higher expected return is accompanied by greater VaR exposure. Periodic portfolio rebalancing is also recommended because stock risk-return characteristics may change over time.

5. Conclussion

This study developed a Mean-VaR portfolio diversification model based on K-Medoids clustering using LQ45 stock data from 3 February to 31 July 2025. Of the 45 LQ45 stocks, 18 stocks met the eligibility criteria and were included in the analysis. Expected return and 95% VaR were used as the main variables for clustering and portfolio construction.

The K-Medoids analysis produced an optimal two-cluster structure with a Silhouette Index of 0.6882. Cluster 1 consists of aggressive stocks with high return potential and higher downside risk, while Cluster 2 consists of defensive stocks with lower risk and more stable returns. The medoids selected were ANTM.JK for Cluster 1 and BBRI.JK for Cluster 2.

The Mean-VaR optimization with inter-cluster allocation constraints resulted in a diversified portfolio with 30% allocation to the aggressive cluster and 70% allocation to the defensive cluster. The optimized portfolio generated a daily expected return of 0.003272 and a 95% VaR of -0.029053. These results indicate that the integration of K-Medoids and Mean-VaR can produce a diversified and risk-controlled portfolio.

Future research may extend this model by incorporating additional variables such as market beta, trading volume, fundamental indicators, or market sentiment. The model may also be tested under different market conditions, including crisis periods, to evaluate the stability of the cluster structure and portfolio performance.

References

Bau, M. K. D., Setyawan, Y., & Jatipaningrum, M. T. (2023). Perbandingan metode algoritma K-Means dan K-Medoids pada pengelompokan kabupaten/kota di NTT berdasarkan IPM 2020. Jurnal Statistika Industri dan Komputasi, 8(1), 48-57.

Hidayat, Y., Purwandari, T., Sukono, et al. (2023). Mean-Value-at-Risk portfolio optimization based on risk tolerance preferences and asymmetric volatility. Mathematics, 11, 4761.

Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77-91.

Parrak, R., & Seidler, J. (2010). Mean-Variance & Mean-VaR portfolio selection: A simulation based comparison in the Czech crisis environment. IES Working Paper 27/2010.

Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi metode K-Medoids clustering untuk pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(9), 723-732.

Puspaningsih, E. S., Di Asih, I. M., & Tarno, T. (2024). K-Medoids Clustering dan Mean-Value at Risk untuk Optimasi Portofolio Saham Jakarta Islamic Index. Indonesian Journal of Applied Statistics, 6(1), 85-95.

Rudianto, R. D., & Wijayanto, A. W. (2024). Analisis perbandingan K-Means dan K-Medoids dalam pengelompokan provinsi berdasarkan Indeks Demokrasi Indonesia. Komputika: Jurnal Sistem Komputer, 13(1), 19-27.

Sarykalin, S., Serraino, G., & Uryasev, S. (2008). Value-at-Risk vs. Conditional Value-at-Risk in risk management and optimization. INFORMS Tutorials in Operations Research.

Wahid, A. J., & Saputra, J. (2025). Portfolio Performance Analysis with Jensen's Alpha Using Single Index Model and CAPM on IDX30 Stocks. International Journal of Quantitative Research and Modeling, 6(2), 274-283.

Wahid, A. J., Riaman, R., & Sukono, S. (2025). A Systematic Literature Review on Mean-CVaR Based Financial Asset Portfolio Weight Allocation Using K-Means Clustering. CAUCHY: Jurnal Matematika Murni dan Aplikasi, 10(2), 1069-1091.