Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.28613
Page: 511-522

IDX30 Portfolio Construction using K-Means Clustering with MAD Risk Optimization and Sortino Ratio Evaluation

[1]Rifki Pebriyandi
Department of Statistics, Tanjungpura University, Pontianak, Indonesia

[2]Evy Sulistianingsih
Department of Statistics, Tanjungpura University, Pontianak, Indonesia

[3]Hendra Perdana
Department of Statistics, Tanjungpura University, Pontianak, Indonesia

[4]Wirda Andani
Department of Statistics, Tanjungpura University, Pontianak, Indonesia

[5]Shantika Martha
Department of Statistics, Tanjungpura University, Pontianak, Indonesia

Article Info

Article history:

Accepted: 10 July 2026

Keywords:

Downside Risk;
K-Means Clustering;
MAD;
Simplex
Sortino Ratio.

ABSTRACT

A stock portfolio plays an important role in managing risk and achieving optimal returns in volatile markets. This study proposes an integrated framework that combines K-Means Clustering, Mean Absolute Deviation (MAD), and the Sortino Ratio. The main contribution lies in linking clustering-based asset selection with downside risk optimization and evaluation, enabling portfolio construction that accounts for asset similarity, risk measurement, and investor-oriented performance assessment. This approach addresses the limitation of previous studies that apply these methods separately by providing a more structured basis for downside risk-adjusted portfolio selection. Using daily IDX30 stock data from June 2024 to June 2025, samples were selected based on index consistency. The results indicate that a portfolio of ANTM and INDF achieved the highest Sortino Ratio of 0.2179. These findings suggest that combining high-return stocks with moderate volatility, supported by clustering and downside risk optimization, can improve downside risk-adjusted performance, providing practical guidance for investors.

Corresponding Author:

Evy Sulistianingsih,
Department of Statistics
Tanjungpura University, Pontianak, Indonesia
Email: evysulistianingsih@math.untan.ac.id

1. INTRODUCTION

Stocks are a primary financial instrument offering high return potential but also exposing investors to uncertainty in achieving expected returns [1]. In Indonesia, stock market participation has increased significantly, with more than 744,000 new investors recorded in 2024 [2], reflecting growing market awareness alongside more complex investment decisions under volatile price movements. The large number of traded stocks with heterogeneous risk–return characteristics further complicates asset selection [3], highlighting the need for systematic portfolio strategies. Despite these risks, stocks remain attractive due to their higher growth potential compared to other instruments [4]. Moreover, the Indonesian market exhibits relatively high volatility, as indicated by the IDX Composite Index (IHSG), which recorded a return standard deviation of approximately 0.01236, emphasizing the importance of effective portfolio management.

Portfolio formation is essential as it allows investors to combine multiple stocks and distribute risk through diversification [5], [6], thereby mitigating losses when individual assets underperform [7], [8]. In recent years, cluster analysis has been widely applied in portfolio construction to group stocks based on similarities in return behavior or volatility [9], [10], with K-Means Clustering being popular due to its computational efficiency and ability to produce relatively homogeneous asset groups [11]. However, clustering is not the final objective; rather, it serves as a basis for constructing representative portfolios that enhance inter-cluster diversification and achieve more balanced risk–return exposure.

Following cluster formation, portfolio construction requires optimal asset allocation to manage risk and enhance performance. The clustered assets are then used as inputs for portfolio optimization, ensuring diversified exposure across different risk–return characteristics. Mean Absolute Deviation (MAD) is widely used for this purpose because it measures risk through the average magnitude of return deviations and is less sensitive to extreme values than variance-based measures [12]. By treating deviations symmetrically without squaring returns, MAD provides a more robust representation of risk in volatile and non-normal market conditions, such as the Indonesian stock market [13]. Once optimal portfolio weights are obtained, performance evaluation becomes essential, with the Sortino Ratio offering a more relevant measure by focusing on downside risk and better reflecting investor preferences compared to symmetric measures such as the Sharpe Ratio [14], [15].

Previous studies have applied K-Means Clustering, MAD, and the Sortino Ratio in portfolio analysis, but mostly in separate stages. For example, study [16] employed MAD after stock grouping based on factor analysis, study [17] combined K-Means Clustering with mean–variance optimization, and study [18] focused solely on portfolio performance evaluation using the Sortino Ratio. However, the integrated application of these three approaches within the IDX30 context remains limited, indicating that portfolio construction often cannot simultaneously account for stock similarity, risk optimization, and downside-oriented performance evaluation within a unified framework, particularly under the relatively high volatility of IDX30 constituent stocks. From a theoretical perspective, integrating clustering-based asset grouping with risk optimization and downside risk evaluation aligns with the principles of diversification and risk–return efficiency in Modern Portfolio Theory, where clustering serves as a data-driven pre-selection mechanism, MAD provides a robust risk measure for volatile and potentially non-normal return distributions, and the Sortino Ratio emphasizes downside risk in line with investors' loss aversion. By combining these approaches, this study contributes to portfolio optimization research through: (1) cluster-based asset grouping using K-Means Clustering to identify stocks with similar return characteristics, (2) portfolio risk optimization using MAD to manage risk in volatile markets, and (3) portfolio performance evaluation using the Sortino Ratio to provide a more relevant assessment of downside risk, thereby supporting more structured and risk-aware portfolio construction.

2. RESEARCH METHOD

This study employs a descriptive quantitative approach to analyze financial data for IDX30 stock portfolio formation. Portfolio optimization is conducted through three stages: K-Means Clustering groups stocks based on return similarity, the Mean Absolute Deviation (MAD) method determines portfolio weights by minimizing risk, and the Sortino Ratio evaluates performance based on downside risk. The integration of these methods is justified by their complementary roles, where clustering reduces asset complexity and facilitates portfolio construction, MAD optimizes risk under volatile and non-normal conditions, and the Sortino Ratio evaluates downside performance in line with investor risk preferences. An equal-weight portfolio is included as a benchmark for comparison. Although the framework simplifies inter-stock relationships, it provides a practical and structured reference for investors operating under volatile market conditions.

2.1 Stock Returns and Expected Returns

Stock returns represent the rate of gain or loss obtained from an investment over a given period, whether in individual stocks or portfolios [19]. Returns indicate the profit or loss earned by investors, where positive values reflect gains and negative values indicate losses. In financial analysis, returns are commonly classified into realized returns, which are derived from historical data, and expected returns, which represent anticipated future performance. In this study, the actual return of the *i*-th stock at time *t*, denoted as $R_{i(t)}$, is calculated using logarithmic returns based on daily closing price changes, as presented in Equation (1) [20], [21]:

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

where $R_{i(t)}$ denotes the return of the *i*-th stock at time *t*, $P_{i(t)}$ is the closing price of the *i*-th stock at time *t*, and $P_{i(t-1)}$ is the closing price of the *i*-th stock in the previous period.

In contrast, the expected return is calculated as the average of all returns over the observation period, as shown in Equation (2) [22]:

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

where $E(R_i)$ is the expected return of stock *i*-th, $R_{i(t)}$ is the return of the *i*-th stock at time *t*, and *T* is the number of stock return observations.

2.2 Stock Risk and Expected Portfolio Return

Stock risk reflects the uncertainty of returns due to price fluctuations and the potential deviation between actual and expected outcomes [23], [24]. It is commonly measured by the standard deviation of returns, where higher values indicate greater volatility and risk. Mathematically, stock risk is defined as the standard deviation of returns, obtained by taking the square root of the return variance, as shown in Equation (3) [25]:

$$S_i = \sqrt{\frac{\sum_{t=1}^{T}\left(R_{i(t)} - E(R_i)\right)^2}{T-1}} \tag{3}$$

where $E(R_i)$ is the expected return of stock *i*-th, $R_{i(t)}$ is the return of the *i*-th stock at time *t*, and *T* is the number of stock return observations.

In portfolio construction, individual stock characteristics such as risk and return are aggregated to form portfolio-level measures. The expected portfolio return represents the weighted average return of all assets, where stocks with higher weights contribute more significantly to overall performance [26], [27]. This measure is essential for balancing return potential and associated risks in determining optimal asset allocation. Mathematically, the expected portfolio return is calculated using Equation (4) [16]:

$$E(R_p) = \sum_{i=1}^{n} w_i E(R_i) \tag{4}$$

where $E(R_p)$ is the expected return of the portfolio, w_i is the investment weight of the *i*-th stock, *n* is the number of stocks, and $E(R_i)$ is the expected return of stock *i*-th.

2.3 Data Standardization

Data standardization is applied to minimize clustering bias arising from distance-based calculations, which are highly sensitive to differences in data scales or measurement units among variables [28]. In practice, this step ensures that all return variables contribute proportionally to the distance calculation in K-Means Clustering, preventing stocks with larger numerical ranges or higher volatility from disproportionately dominating the Euclidean distance measure and distorting the resulting cluster structure. This process uses the Z-score method, which adjusts each value based on the mean and standard deviation, using Equation (5):

$$z_i = \frac{R_{i(t)} - E(R_i)}{S_i} \tag{5}$$

where z_i is the standardized Z-Score value of the *i*-th stock, $E(R_i)$ is the expected return of stock *i*, $R_{i(t)}$ is the return of the *i*-th stock at time *t*, and S_i is the standard deviation of the *i*-th stock.

2.4 Correlation and Multicollinearity Test

Correlation analysis is used to assess the relationships among stock returns in a portfolio [29]. While Pearson's correlation is commonly used, it assumes normally distributed data [30]. Since several stock return series in this study deviate from normality, Spearman's rank correlation, a nonparametric measure less sensitive to extremes, is employed to capture monotonic relationships under volatile conditions [31]. The mathematical formulation of the Spearman correlation coefficient is presented in Equation (6) [32]:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{r(r^2-1)} \tag{6}$$

where ρ_s is the Spearman's rank correlation coefficient, d_i is the rank difference between the two variables for the *i*-th observation, and *r* is the number of data pairs.

Correlation values near −1 indicate strong diversification potential, while values near +1 suggest stocks move together, reducing diversification benefits. Next, multicollinearity was examined using the Variance Inflation Factor (VIF) to detect excessive linear dependence among the return variables, which could interfere with risk estimation and portfolio construction. The VIF value was calculated using Equation (7) [25]:

$$\mathrm{VIF_i} = \frac{1}{1 - R_i^2} \tag{7}$$

where $\mathrm{VIF_i}$ is the Variance Inflation Factor value for the *i*-th stock, R_i^2 represents the coefficient of determination obtained from regressing the *i*-th stock return on the remaining stock returns and *i* is the number of independent variables, where *i* = 1, 2, …, *n*.

2.5 K-Means Clustering

The K-Means algorithm is a non-hierarchical clustering method that groups data based on similarity patterns among objects [33]. It is widely used due to its computational efficiency, interpretability, and suitability for grouping stocks based on return similarity in data-driven portfolio construction. In this context, clustering acts as a pre-selection mechanism that reduces the asset space by grouping stocks with similar characteristics, supporting diversification by capturing differences across clusters while minimizing redundancy within clusters. The procedure begins by specifying the number of clusters (k) and initializing centroids randomly. Given its sensitivity to initial centroids, different initializations may lead to varying clustering results. Distances between observations and centroids are then computed using the Euclidean distance, as shown in Equation (8) [17]:

$$d(x_i, v_i) = \sqrt{\sum_{u=1}^{U} (x_{iu} - v_{iu})^2} \tag{8}$$

where $d(x_i, v_i)$ is the Euclidean distance between x_i and v_i, v_i is the *i*-th centroid of the cluster, x_i is the *i*-th cluster object, and U is the number of variables used in the cluster.

Each object is assigned to the cluster with the nearest centroid, after which a new centroid is calculated as the average of the cluster members, as shown in Equation (9):

$$v_i = \frac{\sum_{j=1}^{k} x_j}{k} \tag{9}$$

where v_i is the *i*-th centroid of the cluster, x_j is the *j*-th cluster object, and k is the number of objects that are members of the cluster.

By reducing redundancy among similar stocks through clustering, the subsequent optimization stage becomes more stable and computationally efficient. This structured integration allows the optimization process to focus on representative assets, thereby improving the robustness of portfolio weight estimation under volatile market conditions.

2.6 Elbow and Silhouette Methods

The Elbow and Silhouette methods were applied to identify the most appropriate number of clusters in the K-Means Clustering analysis. The Elbow method utilizes a graph based on the Sum of Squared Error (SSE), which measures the dispersion of data within clusters or the closeness of each observation to its corresponding cluster centroid [34]. A smaller SSE indicates more compact and homogeneous clusters. As clusters increase, SSE generally decreases; the optimal number is at the point where SSE reduction slows ("elbow"), balancing cluster compactness and simplicity [35]. The SSE value is calculated using Equation (10) [34]:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{p} (x_{ij} - v_i)^2 \tag{10}$$

where p is the number of data points in cluster *i*, v_i is the *i*-th centroid of the cluster, x_{ij} is the *j*-th object in the *i*-th cluster, and k is the number of clusters.

Meanwhile, the Silhouette method assesses cluster quality based on the proximity of objects within and between clusters. The Silhouette value ranges from –1 to 1, with higher values indicating better-formed clusters [36]. The Silhouette value is calculated using Equation (11) [37]:

$$s(i) = \frac{b(m) - a(m)}{\max[a(m), b(m)]} \tag{11}$$

where $a(m)$ is the average distance from the *m*-th observation to all other points in the same cluster and $b(m)$ is the average distance from the *m*-th observation to all other points in different clusters.

Following the determination of the optimal number of clusters, representative stocks are selected from each cluster based on their risk–return characteristics. Three criteria are applied: (1) stocks with the highest expected return $(E(R_i) = \max\{E(R_1), \dots, E(R_k)\})$, (2) stocks with the lowest risk measured by standard deviation $(S_i = \min\{S_1, \dots, S_k\})$, and (3) stocks that balance high return and low risk within each cluster. These criteria capture extreme and trade-off positions in the risk–return space, representing maximum return, minimum risk, and their balance within each cluster.

2.7 Mean Absolute Deviation

The Mean Absolute Deviation (MAD) method is an extension of the Markowitz model that employs a linear programming approach to assess portfolio risk [13]. Unlike the Markowitz model, which uses variance, MAD measures risk based on the average absolute deviation between actual and expected returns [16]. In this study, portfolio risk is measured using the Mean Absolute Deviation, which is defined in Equation (12) as the average absolute difference between stock returns and their corresponding expected returns:

$$MAD_i = \frac{\sum_{t=1}^{T} |R_{i(t)} - E(R_i)|}{T} \quad (12)$$

where $E(R_i)$ is the expected return of stock *i*, $R_{i(t)}$ is the return of the *i*-th stock at time *t*, and *T* is the number of stock return observations.

MAD has the advantage of being robust to non-normal returns and independent of the covariance matrix [38]. However, it ignores inter-stock dependencies, which may slightly limit risk assessment accuracy compared to variance–covariance methods like Markowitz optimization. Despite this, MAD remains a practical and reliable tool for portfolio risk management in volatile markets. Portfolio optimization is done by minimizing the risk $\sigma(W)$, as expressed in Equation (13):

$$\text{Minimizing } \sigma(W) = \sum_{i=1}^{n}(MAD_i w_i) \quad (13)$$

where $\sigma(W)$ is the portfolio risk value and w_i is the size of the investment weight for the *i*-th stock.

Besides using the objective function, forming an optimal portfolio through the Mean Absolute Deviation (MAD) method also considers three constraint functions, as defined in Equations (14), (15), and (16) [39]:

$$E(R_1)w_1 + E(R_2)w_2 + \cdots + E(R_n)w_n \geq R_{min} \quad (14)$$

$$w_1 + w_2 + \cdots + w_n = 1 \quad (15)$$

$$0 \leq w_i \leq u_i, \text{with } i = 1,2, \ldots, n \quad (16)$$

where R_{min} is the minimum rate of return and u_i is the maximum weight of the *i*-th stock.

The value of R_{min} can be obtained using Equation (17):

$$R_{min} = \frac{\sum_{i=1}^{n} E(R_i)}{n} \quad (17)$$

where $E(R_i)$ is the expected return of stock *i* and *n* is the number of shares in the portfolio.

2.8 Sortino Ratio

The integration of clustering and MAD-based optimization with the Sortino Ratio provides a more comprehensive evaluation framework, as it not only considers asset similarity and risk minimization but also emphasizes downside risk, which is often overlooked in traditional variance-based approaches. The Sortino Ratio, introduced by Sortino in the early 1980s, evaluates portfolio performance based on excess return relative to the Minimum Acceptable Return (MAR), which is typically the risk-free rate such as the Bank Indonesia Certificate rate [40]. Unlike the Sharpe Ratio, it considers only downside deviation, where returns below MAR are treated as risk while returns above MAR are considered beneficial. The calculation of the Sortino Ratio is expressed in Equation (18) [41]:

$$S_{p_i} = \frac{E(Rp_i) - MAR}{Sd_i} \quad (18)$$

where S_{p_i} is the Sortino Ratio of portfolio *i*, R_{p_i} is the average return of portfolio *i*-th, MAR is the average risk−free rate, and S_{d_i} is the downside standard deviation of portfolio *i*-th.

The downside deviation value is calculated using the formula presented in Equation (19) [42]:

$$Sd_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[(\min\left(Rp_{i,t} - MAR, 0\right)^2]} \quad (19)$$

where only the negative values of $Rpi_t - MAR$(i.e., returns below MAR) are considered as risk.

3. RESULT AND ANALYSIS

This study uses quantitative secondary data consisting of daily closing prices of IDX30 constituent stocks from June 20, 2024, to June 20, 2025, obtained from Yahoo Finance, and the average BI Rate sourced from Statistics Indonesia (BPS) with a value of 2.52%. Daily stock prices were transformed into return series using Equation (1), which serve as the main input for the clustering and portfolio analysis.

The sample was selected using purposive sampling, including stocks that consistently remained in the IDX30 index during the observation period based on index evaluations conducted between January 2024 and April 2025. Based on these criteria, 25 out of 30 IDX30 stocks were included, while five stocks were excluded because they did not meet the index requirements related to transaction value, trading frequency, and market capitalization. This selection ensures data consistency and investability of the analyzed stocks, although the exclusion may introduce a degree of selection bias.

The dataset contains complete daily observations with no missing values. Extreme returns were retained as they reflect actual market movements, and the data were standardized using the Z-score method to ensure comparability during clustering. The one-year observation period captures short- to medium-term risk–return behavior; however, it may not fully represent long-term market cycles or extreme conditions. In addition, the results rely on in-sample evaluation, which may limit robustness and generalizability. Therefore, future research should incorporate out-of-sample validation and rolling-window analysis to better assess the stability and predictive performance of the proposed framework.

Descriptive statistics for the returns of the 25 selected stocks reveal notable differences in return behavior, volatility, and risk characteristics. Stocks such as MDKA, ADRO, and BRPT exhibit relatively high standard deviations, indicating substantial return fluctuations and elevated volatility. In particular, MDKA and ADRO also record negative average returns, suggesting an unfavorable risk–return profile and a higher likelihood of losses. Similar patterns are observed for GOTO, INKP, and UNVR, which show high variability combined with low or negative returns. In contrast, ICBP, BBCA, and ASII display lower standard deviations, indicating more stable return patterns. Meanwhile, ANTM and BRPT record the highest expected returns among the sampled stocks, although they are accompanied by relatively high volatility. These differences highlight the importance of considering both return and risk characteristics in portfolio construction and justify the need for portfolio optimization. The descriptive statistics for the returns of the 25 stocks are presented in Table 1.

Table 1. Descriptive Statistics

No	Stock Code	Min	Max	Average	Standard Deviation
1	ADRO	0.152908	-0.284961	-0.001480	0.038425
2	AKRA	0.200671	-0.121214	-0.000868	0.029334
3	AMRT	0.101003	-0.093434	-0.000729	0.025994
4	ANTM	0.099922	-0.155171	0.004272	0.030748
5	ASII	0.052565	-0.093685	0.000075	0.017504
6	BBCA	0.057331	-0.089153	-0.000303	0.017120
7	BBNI	0.085942	-0.079022	-0.000268	0.023841
8	BBRI	0.088251	-0.106733	-0.000494	0.023059
9	BMRI	0.082960	-0.107500	-0.000763	0.023892
10	BRPT	0.157467	-0.168228	0.002059	0.036614
11	CPIN	0.117261	-0.098047	-0.000063	0.021967
12	GOTO	0.126041	-0.156161	0.000843	0.033340
13	ICBP	0.046847	-0.075637	0.000113	0.016543
14	INCO	0.118920	-0.162260	-0.000886	0.031670
15	INDF	0.064539	-0.075712	0.001299	0.019115
16	INKP	0.096355	-0.170730	-0.001863	0.026515
17	KLBF	0.080043	-0.073092	-0.000124	0.024952
18	MDKA	0.159630	-0.159630	-0.000306	0.038736
19	MEDC	0.089612	-0.066249	0.000663	0.023502
20	PGAS	0.072539	-0.115134	0.000593	0.020767
21	PTBA	0.085655	-0.060625	0.000896	0.017036
22	SMGR	0.139356	-0.159195	-0.001053	0.030717
23	TLKM	0.070793	-0.059660	-0.000230	0.021418
24	UNTR	0.065454	-0.158406	-0.000117	0.021143
25	UNVR	0.157490	-0.167433	-0.003137	0.031158

3.1 Stock Returns and Expected Returns

Stock returns represent the profit or loss obtained by investors from stock price changes. Returns were calculated using logarithmic returns as defined in Equation (1). The expected return was calculated as the average of daily returns using Equation (2). Only stocks with positive expected return values were selected and are presented in Table 2.

Table 2. Positive Expected Return Stocks

No	Stock Code	Expected Return
1	ANTM	0.004272
2	ASII	0.000075
3	BRPT	0.002059
4	GOTO	0.000843
5	ICBP	0.000113
6	INDF	0.001299
7	MEDC	0.000663
8	PGAS	0.000593
9	PTBA	0.000896

Based on Table 2, nine stocks exhibit positive expected returns during the research period, indicating potential profitability for investors if the return trends persist. However, the magnitude of these returns varies across stocks, suggesting that not all IDX30 constituents offer the same return potential. Therefore, further analysis is required to determine the optimal combination of stocks for portfolio formation in order to maximize potential investor return.

3.2 Data Standardization

After identifying nine stocks with positive expected returns, the next step is data standardization. This step is necessary because the K-Means Clustering algorithm relies on distance-based calculations that are sensitive to differences in data scale. Without standardization, variables with larger numerical ranges may dominate the distance measurement and bias the clustering results. Therefore, the return data were standardized using the Z-score method, as presented in Equation (5).

3.3 Correlation and Multicollinearity Test

Before performing cluster analysis, two preliminary tests were conducted to ensure the suitability of the selected stocks. First, Spearman's correlation test was used to examine the relationships among stock returns, as excessively strong correlations may indicate similar risk characteristics and reduce the effectiveness of clustering. The results generally show weak correlations, suggesting that stock return movements are not closely aligned, which supports potential diversification. Second, the Variance Inflation Factor (VIF) was calculated to assess multicollinearity, with all values below 10, confirming that each stock provides independent information for clustering and portfolio formation. The combined results of Spearman's correlation coefficients and VIF values are presented in Table 3.

Table 3. Spearman's Correlation Coefficients and VIF Values of Selected IDX30 Stocks

Stock	ANTM	ASII	BRPT	GOTO	ICBP	INDF	MEDC	PGAS	PTBA	VIF
ANTM	1.000	0.226	0.195	0.164	0.057	0.137	0.215	0.234	0.411	1.2431
ASII	0.226	1.000	0.213	0.186	0.129	0.273	0.195	0.198	0.343	1.2226
BRPT	0.195	0.213	1.000	0.196	0.101	0.111	0.322	0.267	0.226	1.1470
GOTO	0.164	0.186	0.196	1.000	0.170	0.181	0.097	0.153	0.286	1.1958
ICBP	0.057	0.129	0.101	0.170	1.000	0.324	0.059	0.127	0.166	1.1657
INDF	0.137	0.273	0.111	0.181	0.324	1.000	0.145	0.164	0.212	1.1510
MEDC	0.215	0.195	0.322	0.097	0.059	0.145	1.000	0.384	0.281	1.2251
PGAS	0.234	0.198	0.267	0.153	0.127	0.164	0.384	1.000	0.366	1.2162
PTBA	0.411	0.343	0.226	0.286	0.166	0.212	0.281	0.366	1.000	1.4009

3.4 K-Means Clustering Analysis

The K-Means clustering technique was applied in this study by first determining the optimal number of clusters using the Elbow and Silhouette methods. The Elbow method evaluates clustering quality based on the Sum of Squared Errors (SSE), where a lower SSE indicates tighter within-cluster dispersion. In contrast, the Silhouette coefficient measures clustering quality by simultaneously capturing within-cluster cohesion and between-cluster separation, with higher values indicating more coherent and well-separated clusters.

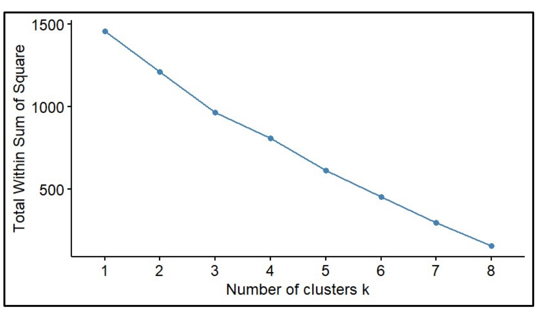

Figure 1. Elbow graph

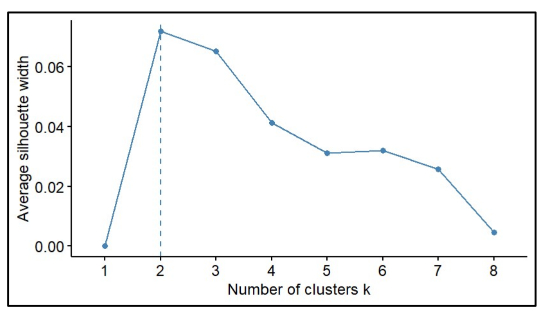

Figure 2. Silhouette graph

As shown in Figures 1 and 2, the SSE curve does not exhibit a clear inflection pointThe largest decrease occurs when the number of clusters increases from $k = 1$ to $k = 2$, after which the reduction becomes more gradual, indicating that the Elbow method alone does not provide a strong basis for selecting the optimal number of clusters. However, the highest average Silhouette value is obtained at $k = 2$, Therefore, this empirical evidence supports the selection of two clusters as the most meaningful partition for subsequent portfolio construction, as grouping stocks with similar risk–return behavior enhances the stability and interpretability of portfolio optimization.

After determining the optimal number of clusters (k = 2) using the Elbow and Silhouette methods, the K-Means clustering process was initiated by randomly selecting the initial centroids from the nine stocks with positive expected returns using the RANDOM function in Microsoft Excel. In this step, ANTM and PTBA were selected as the initial centroids. The Euclidean distance was then calculated to measure the similarity between each stock and the centroids, and stocks were assigned to clusters based on the shortest distance. The clustering process produced two groups: Cluster 1 consisting of ANTM and BRPT, and Cluster 2 comprising ASII, GOTO, ICBP, INDF, MEDC, PGAS, and PTBA. After updating the centroids, no changes in cluster membership were observed, indicating that the algorithm reached convergence after a single iteration and the clustering results were therefore considered stable. The final clustering results and the corresponding stock characteristics are presented in Table 5.

Table 5. Clustering Results and Stock Characteristics

Cluster	Stock	Expected Return	Risk	Range Ratio	Beta
Cluster 1	ANTM	0.00427	0.0306	126%	1.00
	BRPT	0.00205	0.0364	98%	1.67
Cluster 2	ASII	0.00007	0.0174	19%	0.79
	GOTO	0.00084	0.0331	54%	1.33
	ICBP	0.00011	0.0164	28%	0.41
	INDF	0.00129	0.0190	35%	0.50
	MEDC	0.00066	0.0234	43%	0.71
	PGAS	0.00059	0.0206	28%	0.60
	PTBA	0.00089	0.0169	30%	0.69

Based on the clustering results presented in Table 5, Cluster 1 is dominated by stocks with aggressive characteristics, which have relatively high levels of risk, volatility, and beta, making them more sensitive to the movement of the Jakarta Composite Index (JCI) and offering greater potential for returns. Therefore, this cluster is suitable for investors with a high-risk tolerance (risk-takers). Conversely, Cluster 2 consists of stocks with defensive-moderate characteristics, characterized by lower risk and volatility and more stable returns, making them more suitable for investors with a low-risk preference (risk-averse).

Referring to these cluster characteristics, representative stocks were selected to construct the portfolio using three complementary criteria: the highest expected return to capture growth potential, the lowest risk to represent defensive characteristics, and a combination of high expected return and low risk to reflect a balanced risk–return profile. These criteria were intended to ensure that the portfolio captures diverse investment profiles and enhances diversification. Since the sample consists of IDX30 constituents, which are generally liquid and representative of market movements, additional factors such as liquidity and market correlation were not used as primary selection criteria. Based on this process, ANTM, INDF, and PTBA were selected as representative stocks and subsequently used to construct the optimal diversified portfolio, as presented in Table 6.

Table 6. Formed Stock Portfolio

Portfolio	Source	Stock Code
1	Cluster 1	ANTM, BRPT
2	Cluster 2	ASII, GOTO, ICBP, INDF, MEDC, PGAS, PTBA
3	Representative Stock 1	ANTM, INDF
4	Representative Stock 2	ANTM, PTBA
5	Representative Stock 3	ANTM, INDF, PTBA

3.5 Mean Absolute Deviation and Sortino Ratio

Portfolio analysis is conducted using the Mean Absolute Deviation (MAD) approach to determine the optimal stock combination based on the risk–return trade-off. MAD is employed because it provides a linear and computationally efficient measure of risk. The portfolio optimization model is formulated through objective and constraint functions, as presented in Equations (11) to (15) in the methodology section. The MAD value for each stock is calculated using Equation (12), while the objective function in Equation (13) minimizes total portfolio risk. The constraint functions in Equations (14), (15), and (16) represent the minimum portfolio return requirement, the total investment weight, and the maximum allowable weight for each asset. Based on this formulation, the linear optimization model is solved using the simplex method to obtain the optimal portfolio weights, as expressed in the following simplex equation:

First Portfolio

Minimize: $$\sigma(W_1) = 0.022254w_1 + 0.024987w_2$$

With the constraint

$$0.00427w_1 + 0.00205w_2 \geq 0.00316$$
$$w_1 + w_2 = 1$$
$$w_i \leq 60\%, \qquad i = 1,2$$

Second Portfolio

Minimize:

$$\sigma(W_2) = 0.012847w_3 + 0.022892w_4 + 0.012458w_5 + 0.014289w_6 + 0.017646w_7 + 0.014422w_8 + 0.012082w_9$$

With the constraint

$$0.00007w_3 + 0.00084w_4 + 0.00011w_5 + 0.00129w_6 + 0.00066w_7 + 0.00059w_8 + 0.00089w_9 \geq 0.00064$$

$$w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9 = 1$$

$$w_i \leq 15\%, \quad i = 3,4,5,6,7,8,9$$

Third Portfolio

Minimize:

$$\sigma(W_3) = 0.022254w_{10} + 0.014289w_{11}$$

With the constraint

$$0.00427w_{10} + 0.00129w_{11} \geq 0.00278$$

$$w_{10} + w_{11} = 1$$

$$w_i \leq 60\%, \quad i = 10,11$$

Fourth Portfolio

Minimize:

$$\sigma(W_4) = 0.022254w_{12} + 0.012082w_{13}$$

With the constraint

$$0.00427w_{12} + 0.00089w_{13} \geq 0.00258$$

$$w_{12} + w_{13} = 1$$

$$w_i \leq 60\%, \quad i = 12,13$$

Fifth Portfolio

Minimize:

$$\sigma(W_5) = 0.022254w_{14} + 0.014289w_{15} + 0.014289w_{16}$$

With the constraint

$$0.00427w_{14} + 0.00089w_{15} + 0.00129w_{16} \geq 0.00215$$

$$w_{14} + w_{15} + w_{16} = 1$$

$$w_i \leq 40\%, \quad i = 14,15,16$$

Optimization was performed using the simplex method with the assistance of POM-QM and RStudio. Portfolios containing a larger number of stocks indicate a higher level of diversification, while portfolios with fewer stocks represent a more concentrated investment strategy. The results of the MAD optimization are presented in Table 7.

Table 7. Stock Investment Weights

Portfolio	Stock Code	Weights
1	ANTM	0.6000
	BRPT	0.4000
2	ASII	0.1368
	GOTO	0.1132
	ICBP	0.1500
	INDF	0.1500
	MEDC	0.1500
	PGAS	0.1500
	PTBA	0.1500
3	ANTM	0.5000
	INDF	0.5000
4	ANTM	0.5000
	PTBA	0.5000
5	ANTM	0.3424
	INDF	0.2576
	PTBA	0.4000

Portfolio performance was then evaluated using the Sortino Ratio, which measures returns relative to downside risk. This evaluation was conducted after determining the stock weights for each portfolio to assess how efficiently each portfolio generates returns while minimizing potential losses. A higher Sortino Ratio indicates better risk-adjusted performance. The results of the Sortino Ratio calculation are presented in Table 8.

Table 8. Portfolio Performance Values

Portfolio	Expected Portfolio Return	Downside Risk	Sortino Ratio
1	0.00338	0.01665	0.18814
2	0.00064	0.00852	0.04547
3	0.00278	0.01162	0.21792
4	0.00258	0.01302	0.17900
5	0.00215	0.01021	0.18862

Based on Table 8, the third portfolio achieves the highest Sortino Ratio (0.2179), indicating relatively superior performance in managing downside risk. This outcome is driven by the combination of ANTM's high return and INDF's relatively stable volatility, resulting in a balanced risk–return profile. These results are in line with previous studies [16], [17], which suggest that portfolios composed of assets with different risk characteristics, such as either a high-return stock or a mix of relatively stable stocks, can provide favorable risk–return outcomes. Practically, these findings suggest that investors and fund managers may benefit from combining high-return and relatively stable assets to enhance risk-adjusted performance, particularly in volatile markets such as IDX30.

Table 9. Comparison of Portfolio Performance Using Sortino Ratio

Portfolio	Sortino Ratio	
	MAD Optimization	Equal-Weight
1	0.18814	0.17196
2	0.04547	0.04460
3	0.21792	0.21792
4	0.17900	0.17900
5	0.18862	0.19135

To further evaluate the effectiveness of the proposed method, the Sortino Ratios of MAD-optimized portfolios were compared with an equal-weight benchmark, as shown in Table 9. The results indicate that MAD optimization generally produces comparable and, in several cases, higher Sortino Ratios, particularly in Portfolio 1 and Portfolio 2, reflecting improved risk-adjusted performance. Portfolio 3 and Portfolio 4 show identical results due to matching weight allocations, while Portfolio 5 performs slightly better under the equal-weight scheme. Overall, these findings confirm that MAD-based optimization is capable of delivering competitive and, in some cases, superior performance compared to the equal-weight benchmark. However, as the evaluation is based on in-sample data, the results may have limited generalizability. Future studies could incorporate out-of-sample or rolling-window approaches to enhance robustness.

4. CONCLUSION

This study demonstrates that integrating K-Means Clustering, Mean Absolute Deviation (MAD), and the Sortino Ratio provides a structured framework for constructing IDX30 stock portfolios within the studied period. Nine stocks were grouped into two clusters to form representative portfolios, with weights determined using MAD to control downside risk without strict distributional assumptions. The third portfolio achieved the highest Sortino Ratio (0.2179), indicating relatively better downside risk-adjusted performance within the sample. These findings illustrate the importance of diversification in balancing return and risk. From a methodological perspective, this study contributes to portfolio optimization and mathematical finance by providing a structured framework that integrates clustering-based asset selection with linear risk optimization and downside risk evaluation, offering a systematic approach for modeling and solving portfolio selection problems under volatile conditions.

However, this study has limitations, including a relatively short observation period, static portfolio weights, and reliance on in-sample evaluation, which may affect the robustness of the findings. Future research should consider longer observation periods, out-of-sample or rolling-window validation, and additional performance measures, such as the Treynor Ratio and Jensen's Alpha, to further enhance the robustness and practical relevance of the results for investors and portfolio managers.

5. REFERENCES

[1] M. Dahlquist and M. Ibert, “Equity Return Expectations and Portfolios: Evidence from Large Asset Managers,” *Rev. Financ. Stud.*, vol. 37, no. 6, pp. 1887–1928, May 2024, doi: 10.1093/rfs/hhae008.

[2] M. Su and C. Wang, “Policy Announcement, Investor Attention, and Stock Volatility: Evidence From the New Energy Vehicle Industry,” *Front. Psychol.*, vol. 13, pp. 1–11, 2022, doi: 10.3389/fpsyg.2022.838588.

[3] M. Rachwalski and Q. Wen, “Idiosyncratic Risk Innovations and the Idiosyncratic Risk-Return Relation,” *Rev. Asset Pricing Stud.*, vol. 6, no. 2, pp. 303–328, Dec. 2016, doi: 10.1093/rapstu/raw002.

[4] N. Li, C. Wei, and L. Zhang, “Risk factors in the Indonesian stock market,” *Pacific-Basin Finance Journal*, vol. 82, p. 102175, Dec. 2023, doi: 10.1016/j.pacfin.2023.102175.

[5] F. Cesarone, R. Giacometti, M. L. Martino, and F. Tardella, “A return-diversification approach to portfolio selection,” *Computational Management Science*, vol. 22, no. 2, pp. 1–31, 2025, doi: 10.1007/s10287-025-00538-1.

[6] D. Jayeola, Z. Ismail, and S. Firdaus Sufahani, “Effects of diversification of assets in optimizing risk of portfolio,” *Malaysian Journal of Fundamental and Applied Sciences*, vol. 13, no. 4, pp. 584–587, 2017, doi: 10.11113/mjfas.v0n0.567.

[7] F. Hua, R. Tang, and X. Leng, “Environmental Uncertainty and Idiosyncratic Risk: The Dual Role of Diversification Strategy,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 5, pp. 866–875, 2023, doi: 10.20965/jaciii.2023.p0866.

[8] R. Huang, D. Kambouroudis, and D. G. McMillan, “Is portfolio diversification still effective: evidence spanning three crises from the perspective of U.S. investors,” *Journal of Asset Management*, vol. 26, no. 2, pp. 115–135, 2025, doi: 10.1057/s41260-025-00398-z.

[9] T. Agarwal, H. Quelle, and C. Ryan, “Principal Component Analysis for Clustering Stock Portfolios,” *Arizona Journal of Interdisciplinary Studies* /, vol. 7, pp. 64–75, 2021.

[10] J. Korzeniewski, “Efficient Stock Portfolio Construction by Means of Clustering,” *Acta Universitatis Lodziensis. Folia Oeconomica*, vol. 1, no. 333, pp. 85–92, 2018, doi: 10.18778/0208-6018.333.06.

[11] M. I. Boloş, Ştefan Rusu, M. Leordeanu, C. D. Sabău-Popa, D. C. Perţicaş, and M. I. Crişan, “K-Means Clustering for Portfolio Optimization: Symmetry in Risk–Return Tradeoff, Liquidity, Profitability, and Solvency,” *Symmetry (Basel).*, vol. 17, no. 6, pp. 1–29, 2025, doi: 10.3390/sym17060847.

[12] Z. Hosseini-Nodeh, R. Khanjani-Shiraz, and P. M. Pardalos, “Portfolio optimization using robust mean absolute deviation model: Wasserstein metric approach,” *Financ. Res. Lett.*, vol. 54, p. 103735, Jun. 2023, doi: 10.1016/j.frl.2023.103735.

[13] W. S. Lam, W. H. Lam, and S. H. Jaaman, “Portfolio optimization with a mean–absolute deviation–entropy multi-objective model,” *Entropy*, vol. 23, no. 10, pp. 1–14, 2021, doi: 10.3390/e23101266.

[14] C. Spanaus and J. Wenzelburger, “Aggregation of downside risk and portfolio selection,” *J. Math. Econ.*, vol. 119, p. 103138, Aug. 2025, doi: 10.1016/j.jmateco.2025.103138.

[15] J. Sen and A. Dutta, “Design and Analysis of Optimized Portfolios for Selected Sectors of the Indian Stock Market,” *2022 International Conference on Decision Aid Sciences and Applications, DASA 2022*, pp. 567–573, 2022, doi: 10.1109/DASA54658.2022.9765289.

[16] 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, 2024, doi: 10.30598/barekengvol18iss3pp1753-1764.

[17] 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.

[18] A. Hilmy and L. V. Shafira, “Idx Bumn20 Performance Measurement With Sharpe, Treynor, and Sortino,” *JEM17: Jurnal Ekonomi Manajemen*, vol. 7, no. 1, pp. 119–128, 2022, doi: 10.30996/jem17.v7i1.6510.

[19] M. J. Brennan and A. P. Taylor, “Expected returns and risk in the stock market,” *J. Empir. Finance*, vol. 72, no. December 2020, pp. 276–300, 2023, doi: 10.1016/j.jempfin.2023.03.002.

[20] Y. Jiang, J. Olmo, and M. Atwi, “Dynamic robust portfolio selection under market distress,” *North American Journal of Economics and Finance*, vol. 69, no. PB, p. 102037, 2024, doi: 10.1016/j.najef.2023.102037.

[21] J. Ginting and S. Staenly, “Constructing Efficient Frontiers in Cryptocurrency Market Using Long-Run GARCH Volatility,” *ZERO: Jurnal Sains, Matematika dan Terapan*, vol. 9, no. 2, p. 546, Oct. 2025, doi: 10.30829/zero.v9i2.25928.

[22] S. Kozak, S. Nagel, and S. Santosh, “Shrinking the cross-section,” *J. financ. econ.*, vol. 135, no. 2, pp. 271–292, Feb. 2020, doi: 10.1016/j.jfineco.2019.06.008.

[23] Y. Zhao, B. Shi, F. Song, and L. Zhou, “Ambiguity — Risk — Return : Evidence from the Stock Market in China,” *American Journal of Applied Mathematics*, vol. 10, no. 4, pp. 118–124, 2022, doi: 10.11648/j.ajam.20221004.11.

[24] Sanghavi and S. M. Benedict, "Unlocking Investment Insights : Exploring Standard Deviation and Coefficient of Variation in Stock Analysis," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 4, pp. 311–318, 2024.
[25] 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*, vol. 19, no. 4, pp. 2609–2620, 2025, doi: 10.30598/barekengvol19iss4pp2609-2620.
[26] W. Sudiyono, "Risk and Return Management of Sharia Stocks in Jakarta Islamic Index for the Period of 2009 – 2019 Using Markowitz Portofolio Management," *International Journal of Economics, Business and Accounting Research (IJEBAR)*, vol. 5, no. 1, pp. 63–79, 2021, doi: 10.29040/ijebar.v5i1.1699.
[27] X. Zhang, "Portfolio theory in investment decision making: Evidence from NVIDIA and Costco," *Highlights in Business, Economics and Management*, vol. 39, pp. 186–194, 2024, doi: 10.54097/9htb3441.
[28] C. Ye, W. Li, Y. Ye, Z. He, and W. Zheng, "Construction of Valuation System with Chinese Characteristics Based on Entropy Weight Method and K-means," *Financial Engineering and Risk Management*, vol. 6, no. 11, 2023, doi: 10.23977/ferm.2023.061119.
[29] T. Millington and M. Niranjan, "Construction of minimum spanning trees from financial returns using rank correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 566, p. 125605, Mar. 2021, doi: 10.1016/j.physa.2020.125605.
[30] A. R. Kusumawardani, "Volatilitas Saham Pt Pertamina Geothermal Energy Tbk (Pgeo) Dalam Konteks Dinamika Inflasi Tahun 2023-2024," *Jurnal Ekonomi, Bisnis dan Pendidikan*, vol. 4, no. 10, p. 5, 2024, doi: 10.17977/um066v4i102024p5.
[31] R. Juwita, F. Rahayu, and M. A. Sari, "Comparison Analysis of ' Safe Investment ' (Gold , Mutual Funds , and Bonds)," *JIKA: Jurnal Ilmu Keuangan dan Perbankan*, vol. 12, no. 1, pp. 147–156, 2022.
[32] E. F. El-Hashash and R. H. A. Shiekh, "A Comparison of the Pearson, Spearman Rank and Kendall Tau Correlation Coefficients Using Quantitative Variables," *Asian Journal of Probability and Statistics*, vol. 20, no. 3, pp. 36–48, 2022, doi: 10.9734/ajpas/2022/v20i3425.
[33] M. Amelia, A. Faqih, and A. R. Rinaldi, "Penerapan Metode K-Means Clustering Dalam Pemetaan Kemiskinan Kabupaten/Kota Di Indonesia Untuk Perencanaan Kebijakan Yang Tepat," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, pp. 437–444, 2025, doi: 10.23960/jitet.v13i2.6231.
[34] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J : Multidisciplinary Scientific Journal*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
[35] S. Özarpacl, B. Klllç, O. C. Bayrak, A. Özdemir, Y. Yilmaz, and M. Floyd, "Comparative analysis of the optimum cluster number determination algorithms in clustering GPS velocities," *Geophys. J. Int.*, vol. 232, no. 1, pp. 70–80, 2023, doi: 10.1093/gji/ggac326.
[36] A. M. Harsha and V. V. S. Kesava Rao, "A Comprehensive Analysis of Clustering Methods for Portfolio Optimization," *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2024, no. 10s, pp. 27–36, 2023.
[37] F. Ali, R. Basheer, M. Kawas, and B. Alkhatib, "Towards Detecting Influential Members and Critical Topics from Dark Web Forums: A Data Mining Approach," *Journal of Information and Organizational Sciences*, vol. 47, no. 1, pp. 1–20, 2023, doi: 10.31341/jios.47.1.1.
[38] J. F. Bello, E. S. Taiwo, and I. Adinya, "Modified Models for Constrained Mean Absolute Deviation Portfolio Optimization," *International Journal of Mathematical Sciences and Optimization: Theory and Applications*, vol. 10, no. 1, pp. 12–24, 2024.
[39] E. Sulistianingsih *et al.*, "an Examination of the Green Stock Portfolio in Connection With the 2024 Indonesian Republic Presidential General Election," *Barekeng*, vol. 19, no. 4, pp. 2543–2556, 2025, doi: 10.30598/barekengvol19iss4pp2543-2556.
[40] Y. Kroll, A. Marchioni, and M. Ben-Horin, "Sortino(γ): A Modified Sortino Ratio With Adjusted Threshold," *Journal of Accounting and Finance*, vol. 23, no. 6, pp. 13–22, 2024, doi: 10.33423/jaf.v23i6.6699.
[41] D. K. Malhotra, T. Mooney, R. Poteau, and P. Russel, "Assessing the Performance and Risk-Adjusted Returns of Financial Mutual Funds," *International Journal of Financial Studies*, vol. 11, no. 4, p. 136, Nov. 2023, doi: 10.3390/ijfs11040136.
[42] B. A. Nugroho and R. Robiyanto, "Peran Cryptocurrency dalam Meningkatkan Kinerja Portofolio Indeks High Dividend dan Indeks SRI-KEHATI," *AFRE Accounting and Financial Review*, vol. 6, no. 1, pp. 1–14, 2023.