Stock Portfolio Optimization Based on Financial and Risk–Return Clustering and TOPSIS with MVEP–MAD Weighting

Wirda Andani[1*], Shantika Martha[1], Evy Sulistianingsih[1], Muhammad Fikri[1], Cinta Priscillia Maharani[1], Rifki Pebriyandi[1]
[1]Statistics Study Program, Universitas Tanjungpura, Pontianak, Indonesia
*Corresponding author: wirda.andani@math.untan.ac.id

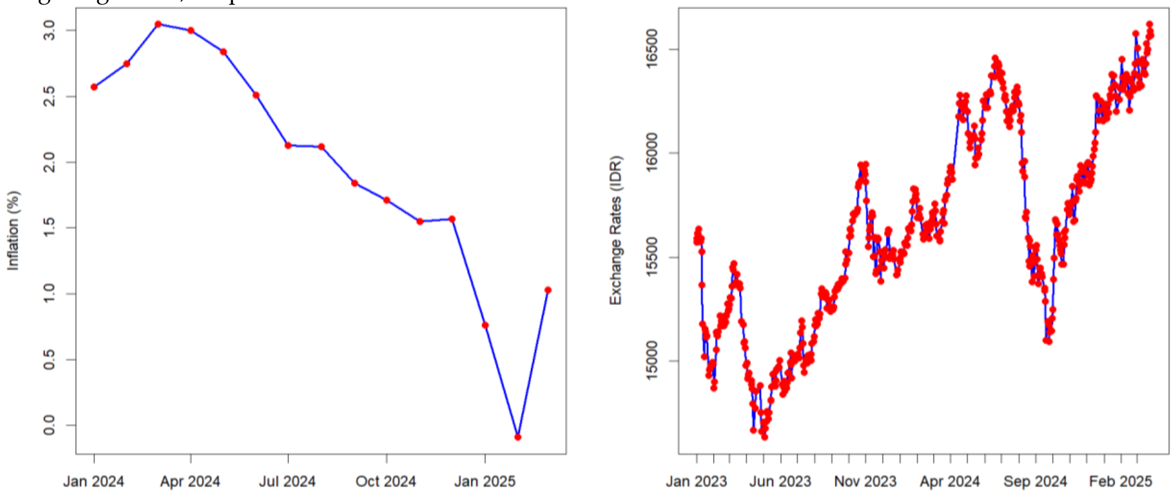
Received: 1 June 2026
Revised: 24 August 2026
Accepted: 26 August 2026

ABSTRACT – **This research aims to construct an optimal stock portfolio from the Kompas100 index using stock performance indicators, fundamental indicators, K-Means, TOPSIS, and portfolio optimization. Of the 100 stocks, only 22 were suitable as candidates for portfolio formation. From these 22 stocks, 4 portfolio candidates were identified through K-means analysis and 7 through TOPSIS analysis. The next step was to determine the investment proportion for each stock in the portfolio using MVEP and MAD. Performance evaluation results show that Portfolio 4, consisting of PTRO and WIFI stocks, consistently yields the highest Sharpe Ratio under both weighting methods: 0.19 using MVEP and 0.21 using MAD. Portfolio 4's performance was then re-evaluated using data from April through December 2025, resulting in a higher Sharpe ratio for both the MAD and MVEP. Overall, this study demonstrates that the combination of the K-Means Clustering, TOPSIS, MVEP, and MAD methods can be used to assist in the stock selection process and the formation of an optimal portfolio that is more efficient than investing in a single stock because it provides a better balance between return and risk through investment diversification, while remaining stable for the next nine months.**

Keywords – **Feature Selection, Maximum Drawdown, Portfolio Diversification, Portofolio Optimization, Sharpe Ratio.**

I. INTRODUCTION

Over the past few years, the global geopolitical landscape has become increasingly complex, marked by international conflicts, trade tensions, and global economic uncertainty. These events have had a direct impact on Indonesia's economic stability [1]. In addition, Indonesia's macroeconomic conditions from January 2024 to March 2025 were marked by significant volatility in both inflation and the exchange rate. Inflation, as shown in Figure 1.a, fluctuated quite sharply during this period, rising as high as 3.05% in March 2024 before eventually declining significantly into deflation by February 2025. Meanwhile, the rupiah exchange rate, as shown in Figure 1.b, fluctuated significantly and weakened, reaching a high of 16,622 per U.S. dollar at the end of March 2025.



(a) (b)

Figure 1 (a) Inflation and (b) the exchange rate of the rupiah against the U.S. dollar from January 2024 to March 2025
Source: Data obtained from www.bi.go.id, processed by the author.

Under these circumstances, people who rely solely on regular savings accounts as a means of storing assets face the risk of having no protection against economic shocks, since savings accounts do not provide a buffer when market conditions become volatile. Furthermore, people who rely solely on regular savings accounts actually miss out on opportunities to earn higher returns, given that savings account interest rates are much lower than the potential returns from investments. If this situation is not balanced by proper financial management, people could potentially experience a decline in their welfare or find themselves in a state of serious financial instability in the long term [2]. Therefore, in a situation of economic uncertainty such as the current one, an appropriate financial management strategy is needed, one of which is through investment activities.

One investment instrument with high profit potential is stocks. Stocks offer relatively greater profit opportunities compared to other investment products. However, stock investing also carries high risks [3]. Therefore, a sound strategy is needed when investing in stocks to achieve the goal of generating substantial profits while minimizing potential risks.

Building a stock portfolio is one strategy in stock investing. The goal is to diversify investments so that the risks faced can be minimized. Through diversification, losses on one stock can be compensated by gains on other stocks, making the total portfolio risk more manageable [4]. Thus, building a stock portfolio is a crucial step in achieving investment goals, that is, obtaining optimal returns with measurable risk.

In Indonesia, there are a huge number of stocks listed on the Indonesia Stock Exchange, including those included in the Kompas100 index. With so many stocks to choose from, investors need to be more selective in determining which ones are worth adding to their portfolios. Mistakes in stock selection can increase risk and diminish the potential return on investment [5]. Worthy or high-quality stocks are generally indicated by a company's good financial condition and strong market performance [6]. Financial performance can be analyzed through fundamental indicators, while stock performance in the market can be assessed using return-based and risk-based indicators (volatility) [7]. Additionally, stock performance is evaluated using the Sharpe Ratio and Maximum Drawdown. These two aspects—financial performance and stock performance—serve as the primary basis for this study in determining a stock's suitability for investment.

Based on what has been explained, an approach is needed to group stocks based on the two aspects described earlier. The selected group consists of stocks with the best financial performance and stock performance. In this research, stock clustering will be performed using K-Means Clustering, which is a simple non-hierarchical cluster analysis method with high computational efficiency, as its iteration process is relatively fast compared to other clustering methods [8]. However, clustering using K-means cannot simultaneously combine these two aspects due to the differing data periods between fundamental indicators and return-based indicators. The data period for fundamental indicators is quarterly, whereas the data period for stock prices or stock returns is daily. Therefore, stock clustering will be performed separately based on fundamental indicators and based on stock performance indicators. In addition to clustering methods, multi-criteria decision-making methods such as the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) can also be used to rank stocks. Using TOPSIS, stocks can be ranked based on their proximity to the ideal solution, thereby identifying the best stocks suitable for being combined into a portfolio [9]. In this study, K-Means clustering and TOPSIS were used as two independent screening approaches, with the aim of expanding the set of portfolio candidates, rather than as a single, mathematically integrated method. K-Means clusters stocks based on the similarity of their performance characteristics, while TOPSIS ranks stocks based on their proximity to the ideal solution in terms of fundamental indicators.

Selecting the best stocks as portfolio candidates is not enough to create an optimal portfolio. Determining the investment allocation for each stock is also a key factor in portfolio optimization. Commonly used optimization methods include the Mean Variance Efficient Portfolio (MVEP) and Mean Absolute Deviation (MAD). MVEP focuses on minimizing variance as a measure of risk [10], while MAD uses absolute deviation as an alternative risk measure that is more robust against outliers [11]. Both methods can be used to determine the optimal weight of each stock in the portfolio.

Furthermore, to evaluate the performance of the resulting portfolio, the Sharpe ratio is used as a performance metric. The Sharpe ratio measures portfolio return relative to total risk. A portfolio with the highest Sharpe ratio can be considered optimal because it provides maximum return with minimal risk. Based on the above description, this research aims to develop a stock portfolio selection and optimization model using the K-Means clustering method and TOPSIS as two complementary yet independently operating screening approaches, followed by the optimization of investment weights using MVEP and MAD, and the evaluation of portfolio performance using the Sharpe ratio for Kompas100 stocks. To ensure that the resulting portfolio performance does not merely reflect a bias toward the historical data used during the portfolio formation stage, this study also conducts out-of-sample testing, which involves evaluating the performance of the selected portfolio using data from periods outside the portfolio formation period.

II. LITERATURE REVIEW

A. K-Means Clustering

K-Means is a non-hierarchical clustering method that aims to maximize the similarity of data within a cluster and minimize the similarity of data across clusters. The measure of similarity used in K-Means is the calculation of distance [12]. The following is the K-Means algorithm [13]:

Determine the number of clusters, k

1. Initialize random initial centroids, $c_1, \ldots, c_k$
2. Calculate the distance of each data point to the nearest centroid (typically using Euclidean distance)

$$d(x_i, c_j) = \sqrt{\sum(x_i - c_j)^2} \tag{1}$$

4. Assign data points to the nearest centroid
5. Update $c_1, \ldots, c_k$ by calculating the average of all xi that are members of their respective clusters

$$dc_j = \frac{1}{n_j}\sum x_i \tag{2}$$

6. Repeat steps 3–5 until convergence

However, before performing k-means clustering, a multicollinearity test is first conducted on the variables. In this research, the multicollinearity test was performed using Variance Inflation Factor (VIF). The VIF formula is as follows [14]:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3}$$

where VIF_j is the VIF value for the j-th variable and R_j^2 is the coefficient of determination of the regression of the j-th variable on the other variables. A VIF value > 10 indicates high multicollinearity. In this research, variables with high VIF values may be considered for elimination to reduce information redundancy and improve the stability of the k-means model [14].

Once multicollinearity is no longer detected among the variables, these variables often have different units and scales. This can cause problems in clustering, particularly when calculating distances, because variables with a large data scale will dominate, leading to biased values used to assess similarity between objects. Therefore, the data must first be standardized so that all variables have comparable scales. In this research, data standardization uses the min-max normalization formula. The formula for min-max normalization is as follows [15]:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

B. TOPSIS

This method operates on the principle that the best alternative (Stock) is the one with the shortest distance to the positive ideal solution and the longest distance from the negative ideal solution [16]. The steps for using TOPSIS are as follows [9]:

1. Construct a decision matrix consisting of a number of alternatives (stocks) and criteria (indicators or variables). The decision matrix is denoted as matrix A, where rows represent alternatives, columns represent criteria, and x_{ij} indicates the value of the i-th alternative for the j-th criterion.

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

   with m being the number of alternatives and n the number of criteria.
2. Criteria values are not always on the same scale or in the same units, so a normalization process is required to ensure that all criteria fall within a comparable range and can be fairly compared.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum x_{ij}^2}} \tag{5}$$

   Once all criteria have been normalized, the result is a normalized decision matrix.
3. The assignment of weights w_j to each criterion in this study uses equal weights, meaning each criterion has the same weight. The sum of all weights across all criteria must equal one.
4. Weighting the normalized decision matrix is performed by multiplying each element in the column of the normalized matrix by the criterion's weight, resulting in a weighted normalized matrix. Mathematically, this is expressed as:

$$v_{ij} = r_{ij} \cdot w_j \tag{6}$$

   where v_{ij} is an element of the weighted decision matrix, r_{ij} is an element of the normalization matrix, and w_j is the weight of the j-th criterion.
5. The next step is to determine the positive ideal solution and the negative ideal solution. The positive ideal solution is the set of best values for each criterion, while the negative ideal solution is the set of worst values for each criterion. Mathematically, the positive ideal solution is expressed as $v^+ = (v_1^+, v_2^+, \ldots, v_n^+)$, while the negative ideal solution is expressed as $v^- = (v_1^-, v_2^-, \ldots, v_n^-)$. The determination of values for each criterion

depends on the type of criterion used, namely benefit or cost criteria. For benefit-type criteria (the larger, the better), the values of the positive ideal solution are obtained from the maximum values in each column of the weighted matrix, while the negative ideal solution is obtained from the minimum values.

$$v_j^+ = \begin{cases} \max(v_{ij}), & i = 1,2,\dots,m \text{ if the criterion is a benefit} - \text{type criterion} \\ \min(v_{ij}), & i = 1,2,\dots,m \text{ if the criterion is a cost} - \text{type criterion} \end{cases} \quad (7)$$

Conversely, for cost-type criteria (the smaller, the better), the positive ideal solution is obtained from the minimum values, while the negative ideal solution is obtained from the maximum values.

$$v_j^- = \begin{cases} \min(v_{ij}), & i = 1,2,\dots,m \text{ if the criterion is a benefit} - \text{type criterion} \\ \max(v_{ij}), & i = 1,2,\dots,m \text{ if the criterion is a cost} - \text{type criterion} \end{cases} \quad (8)$$

6. Calculate the Euclidean distance of each alternative from v_j^+ and v_j^-

$$D_i^+ = \sqrt{\sum(v_{ij} - v_j^+)^2}$$
$$D_i^- = \sqrt{\sum(v_{ij} - v_j^-)^2} \quad (9)$$

7. Finally, calculate the preference values (the proximity of the alternatives to the ideal solutions)

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (10)$$

C. Mean-Variance Efficient Portofolio (MVEP) dan Mean Absolut Devoation (MAD)

In portfolio formation, investors need to determine the allocation of funds to each selected stock. This is important because each stock has a different level of risk and return, so the right allocation can help create a balance between risk and return. Thus, setting investment weights for each stock is a strategic step in optimizing overall portfolio performance. In this research, the MVEP and MAD weighting methods will be used. The Mean Variance Efficient Portfolio (MVEP) model uses variance as a measure of risk, while the Mean Absolute Deviation (MAD) method uses absolute deviation. The optimal weights in the Mean-Variance Efficient Portfolio method can be expressed as [13]:

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{1}_N}{\boldsymbol{1}_N^T\boldsymbol{\Sigma}^{-1}\boldsymbol{1}_N} \quad (11)$$

where $\boldsymbol{w}$ is the vector of stock weights in the portfolio, $\boldsymbol{\Sigma}^{-1}$ denotes the inverse of the covariance matrix of stock returns in the portfolio, and $\boldsymbol{1}_N$ is a column vector whose elements are all one, and $\boldsymbol{1}_N^T$ is the transpose of the vector $\boldsymbol{1}_N$.
The portfolio risk is $\sqrt{\sigma_p^2}$ and the constraint is $\sum w_i = 1$ with

$$\sigma_p^2 = w^T \Sigma w \quad (12)$$

Mean Absolute Deviation (MAD) is a measure of risk obtained from the average of the absolute values of the deviations of actual returns from expected returns. Mathematically, MAD is formulated as follows [13]:

$$MAD_i = \frac{\sum_{i=1}^{T} |R_{i(t)} - E(R_i)|}{T} \quad (13)$$

where R_i denotes the return on the i-th stock in period t, and $E(R_i)$ denotes the expected return on the i-th stock. After the Mean Absolute Deviation (MAD) values are obtained, the next step is to formulate a linear portfolio optimization model before solving it using the simplex method. The objective function in the MAD method is used to minimize portfolio risk and is expressed as follows [13]:

$$\sigma(w) = MAD_1 w_1 + MAD_2 w_2 + \cdots + MAD_n w_n \quad (14)$$

where $\sigma(w)$ represents the portfolio risk derived from the combination of the MAD values of each asset in the portfolio. Furthermore, MAD_i is the Mean Absolute Deviation value of the i-th asset, while w_i denotes the investment weight or the proportion of funds allocated to the i-th asset. The MAD portfolio model has three main constraints [13]. The first constraint states that the portfolio return must be greater than or equal to the specified minimum return, namely:

$$E(R_1)w_1 + E(R_2)w_2 + \cdots + E(R_n)w_n \geq R_m \quad (15)$$

where R_m denotes the minimum portfolio return. The second constraint states that the total sum of the investment weights of all assets in the portfolio must equal one:

$$w_1 + w_2 + \cdots + w_n = 1 \tag{16}$$

The third constraint states that the investment weight of each asset must not be negative and must not exceed a certain maximum limit (u_i):

$$0 \leq w_i \leq u_i, i = 1,2, \dots, n \tag{17}$$

The minimum portfolio return is obtained from the average expected return of all stocks, which is formulated as:

$$R_m = \frac{E(R)_1 + E(R)_2 + \cdots + E(R)_n}{n} \tag{18}$$

Furthermore, the expected return of the stock portfolio for each portfolio weighting method—both MVEP and MAD—can be calculated using the following equation:

$$E(R_p) = \sum_{i=1}^{n} w_i \cdot E(R_i) \tag{19}$$

III. METHODOLOGY

In this research, the data used consists of fundamental indicators and stock price data obtained from stocks included in the Kompas100 index. The fundamental indicators consist of companies' financial ratios for the period from the first quarter of 2024 (Q1 2024) to the first quarter of 2025 (Q1 2025). The financial ratios used include Earnings Per Share (EPS), Price-to-Earnings Ratio (PER), Price-to-Book Value (PBV), Return on Equity (ROE), Enterprise Value to EBITDA (EV/EBITDA), Debt-to-Equity Ratio (DER), Net Profit Margin (NPM), Total Asset Turnover (TATO), natural logarithm of total assets (LNTA), and Total Asset Growth (TAG). This data was obtained from the indopremier.com website.

The stock price data used in this research consists of daily closing prices accessible via finance.yahoo.com. This data is divided into two periods with different functions. The first period, from January 2, 2024, to March 31, 2025, is used as the training (in-sample) period. Data from this period is used to calculate stock returns, risk, the Sharpe ratio, maximum drawdown, and average drawdown per stock, which subsequently form the basis for the initial screening process. The second period, from April 1, 2025, to December 31, 2025, is used as the evaluation (testing/out-of-sample) period to test the consistency of the performance of the portfolio that has been formed.

The methods used in this research to identify portfolio candidates include the K-Means Clustering method and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method. After obtaining several portfolio candidates from these two methods, portfolio optimization was performed using the Mean Variance Efficient Portfolio (MVEP) and Mean Absolute Deviation (MAD) methods. Finally, a portfolio evaluation was conducted using the Sharpe Ratio to determine the best portfolio based on both MVEP and MAD weightings. The research analysis stages for obtaining the optimal stock portfolio are as follows:

1. Stock Selection

   Before performing clustering stocks using K-means Clustering and ranking stocks using TOPSIS, the stocks included in the Kompas100 were first selected based on their expected return, Sharpe Ratio, and maximum drawdown. The selected stocks are those with a positive expected return, a Sharpe Ratio greater than 0%, and a maximum drawdown of less than 44%. Expected Return, Sharpe Ratio, and Maximum Drawdown are calculated using the following formulas:

$$E(R_i) = \frac{1}{\text{n}} \sum_{t=1}^{n} R_{it} \tag{20}$$

$$S = \frac{E(R_i) - R_f}{\sigma_p} \tag{21}$$

$$MDD = \max\left(\frac{Peak - Trough}{Peak}\right) \tag{22}$$

   where $E(R_i)$ is the expected or average return of the i-th stock, R_{it} is the return of the i-th stock in period t, and n denotes the number of observation periods. The higher the expected return value, the greater the average expected profit from that stock. Furthermore, S is the Sharpe Ratio, R_f is the risk-free rate of return, and σ_i

denotes the standard deviation of the return of the i-th stock. The higher the Sharpe Ratio, the better the stock's performance, as it generates a higher return relative to the risk incurred.

Finally, *MDD* stands for Maximum Drawdown, *Peak* refers to the highest price or value of a stock before a downward movement, while *Trough* refers to the lowest price or value of a stock after the peak, before the price begins to rise again. The smaller the Maximum Drawdown value, the lower the risk of extreme price falls for that stock, meaning the stock's price movements are considered more stable.

2. Stock Clustering Using K-Means Clustering
   If the variables used are free of multicollinearity and have been standardized, the process can continue with clustering using K-Means.
3. Optimization of K-Means with Feature Selection Using ANOVA
   In this study, feature selection was performed using the Analysis of Variance (ANOVA) test to identify variables with the most significant differences in means across clusters. Generally, ANOVA compares between-group variation with within-group variation. If the between-group variation is significantly greater than the within-group variation, the variable is considered to have good ability to discriminate between clusters. The ANOVA test statistic is calculated using the F-ratio (F-statistic) [17]:

$$F = \frac{\dfrac{\sum_{i=1}^{k} n_i\,(\bar{x}_i - \bar{x})^2}{k-1}}{\dfrac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{N-k}} \tag{23}$$

   where k is the number of clusters, N is the total number of observations, n_i is the number of observations or members in the i-th cluster, $\bar{x}_i$ denotes the mean of the i-th cluster, $\bar{x}$ denotes the overall mean, and $\boldsymbol{x_{ij}}$ denotes the j-th observation in the i-th cluster. The larger the F-statistic value and the smaller the p-value, the more capable the variable is of distinguishing characteristics between clusters, and thus it is considered more important (feature important) in the clustering process. Variables with a p-value smaller than the significance level α=0.05 in this study were retained because they were considered to have a significant contribution to cluster formation.
4. Stock Ranking Using TOPSIS
   The formation of potential portfolios is not only based on k-means clustering but also on rankings using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method, based on several criteria or variables identified as significant by the ANOVA results. Furthermore, the data does not use the standardization results previously applied in the K-Means method but instead uses the raw data. This is because TOPSIS has its own normalization step in the calculation process; thus, using pre-standardized data could result in redundant data transformations. In the TOPSIS method, normalization is performed internally using Equation (5).
5. Portfolio Formation
   After obtaining several portfolio candidates using K-Means Clustering and TOPSIS, the next step is to calculate the optimal portfolio weights using the Mean Variance Efficient Portfolio (MVEP) method and the Mean Absolute Deviation (MAD) method.
6. Comparison of Portfolio Results and choose best portofolio
   Compare the portfolio results obtained from the MVEP and MAD methods based on the Sharpe Ratio. The portfolio with the highest Sharpe ratio among the portfolios generated by each weighting method is the best portfolio.
7. Selected Portfolio's Performance Evaluation
   The next step is to test the consistency of the selected portfolio's performance using evaluation data periode (testing/out of sample). The evaluation at this stage is conducted by applying the investment weight proportions (w_i) obtained from the MVEP and MAD optimization results during the training period. These weights are not recalculated or re-optimized at this stage to ensure that the evaluation truly reflects the portfolio's performance. The performance of the selected portfolio is calculated using portfolio return, portfolio risk, and the Sharpe ratio.

IV. RESULTS AND DISCUSSIONS

A. Stock Selection

In the first step, stocks included in the Kompas100 Index were selected based on their expected return. Stocks with an expected return below zero (negative) were eliminated from the analysis process. This was done because a negative expected return indicates that, on average, the stock tends to decline in value or presents a potential for loss during the observation period. Therefore, stocks with a negative expected return were not deemed suitable for consideration in portfolio construction.

Based on the calculation results shown in Table 1, out of a total of 100 stocks included in the Kompas100 index, 59 stocks have an expected return of less than zero. Consequently, these 59 stocks are eliminated from the further stages of analysis. Meanwhile, the remaining 41 stocks have a positive expected return and are thus deemed to still hold profit potential, so they are retained for the following stages of analysis.

Table 1 Expected Return of all stocks in the Kompas100 Index

Stock Name	$E(R_i)$	Stock Name	$E(R_i)$	Stock Name	$E(R_i)$	Stock Name	$E(R_i)$
ACES	-0.00089	BTPS	-0.00177	INTP	-0.00183	PNLF	0.00159
ADMR	-0.00116	CMRY	0.00069	ISAT	-0.00123	PTBA	0.00016
ADRO	-0.00049	CPIN	-0.00025	ITMG	-0.00037	PTPP	-0.00059
AKRA	-0.00082	CTRA	-0.00122	JPFA	0.00211	PTRO	0.00637
AMMN	-0.00020	DEWA	0.00220	JSMR	-0.00047	PWON	-0.00072
AMRT	-0.00092	DSNG	0.00164	KIJA	0.00116	RAJA	0.00194
ANTM	0.00010	ELSA	0.00041	KLBF	-0.00101	SCMA	0.00135
ARTO	-0.00185	EMTK	0.00029	KPIG	0.00319	SIDO	0.00048
ASII	-0.00034	ENRG	-0.00049	LSIP	0.00098	SMGR	-0.00272
AUTO	-0.00043	ERAA	0.00022	MAPA	-0.00018	SMIL	0.00605
AVIA	-0.00050	ESSA	0.00065	MAPI	-0.00074	SMRA	-0.00111
BBCA	-0.00023	EXCL	0.00061	MARK	0.00164	SRTG	0.00053
BBNI	-0.00055	FILM	-0.00069	MBMA	-0.00160	SSIA	0.00299
BBRI	-0.00092	GGRM	-0.00228	MDKA	-0.00165	SSMS	0.00242
BBTN	-0.00096	GJTL	0.00025	MEDC	-0.00016	SURI	-0.00131
BBYB	-0.00234	GOTO	0.00039	MIDI	-0.00050	TINS	0.00199
BDKR	-0.00333	HEAL	-0.00084	MIKA	-0.00044	TKIM	-0.00106
BFIN	-0.00092	HMSP	-0.00143	MNCN	-0.00112	TLKM	-0.00151
BMRI	-0.00032	HRUM	-0.00195	MTEL	-0.00050	TOBA	0.00079
BMTR	-0.00199	ICBP	0.00000	MYOR	-0.00049	TOWR	-0.00212
BNGA	0.00009	INCO	-0.00189	NCKL	-0.00104	TPIA	0.00201
BRIS	0.00140	INDF	0.00044	NISP	0.00047	UNIQ	0.00380
BRMS	0.00292	INDY	-0.00053	PANI	0.00321	UNTR	0.00017
BRPT	-0.00153	INET	0.00194	PGAS	0.00126	UNVR	-0.00309
BSDE	-0.00083	INKP	-0.00152	PGEO	-0.00128	WIFI	0.00994

Next, the 41 stocks that passed the screening were analyzed further using the Sharpe Ratio and Maximum Drawdown. These two indicators were used to evaluate stock performance not only in terms of returns but also by considering the level of investment risk. The Sharpe Ratio measures a stock's ability to generate returns relative to its total risk, while Maximum Drawdown measures the potential maximum decline in a stock's price during the observation period. Based on Table 2, 22 stocks remain that have a Sharpe Index greater than 0% and a Maximum Drawdown less than 44%. The stocks that were not eliminated will proceed to the clustering stage using K-means Clustering and the ranking stage using the TOPSIS method.

Table 2 Sharpe Ratio and Maximum Drawdown for 41 stocks with positive expected returns

Stock Name	Sharpe	MDD	Stock Name	Sharpe	MDD	Stock Name	Sharpe	MDD
ANTM	0%	33%	ICBP	-1%	21%	RAJA	4%	58%
BNGA	-1%	27%	INDF	2%	17%	SCMA	4%	29%
BRIS	4%	34%	INET	3%	45%	SIDO	2%	30%
BRMS	7%	36%	JPFA	8%	16%	SMIL	10%	71%
CMRY	2%	30%	KIJA	6%	17%	SRTG	1%	47%
DEWA	5%	33%	KPIG	7%	55%	SSIA	7%	43%
DSNG	5%	47%	LSIP	4%	27%	SSMS	5%	37%
ELSA	1%	30%	MARK	5%	26%	TINS	5%	42%
EMTK	0%	40%	NISP	2%	17%	TOBA	1%	58%
ERAA	0%	35%	PANI	7%	57%	TPIA	4%	49%
ESSA	2%	44%	PGAS	5%	15%	UNIQ	10%	40%
EXCL	2%	20%	PNLF	4%	43%	UNTR	0%	22%
GJTL	0%	28%	PTBA	0%	27%	WIFI	17%	38%
GOTO	1%	45%	PTRO	12%	43%			

B. Multicollinearity Test

Before conducting the K-means clustering analysis, a multicollinearity test was first performed on the variables used in the study. As previously explained in the background section, clustering was conducted separately based on fundamental indicators and stock performance indicators. Therefore, each indicator will be tested for multicollinearity.

Since the fundamental indicators are quarterly data and the study period runs from January 2, 2024, to March 31, 2025, the VIF test was conducted separately for each quarter. This approach was taken because the characteristics of the relationships between financial ratios can change in each company reporting period. The results of the VIF tests for each quarter are presented in Table 3. Based on these results, it can be seen that most financial ratio variables in each quarter have a VIF value below 10, so it can be said that they do not exhibit high multicollinearity. However, there are several variables with VIF values exceeding 10, namely the ROE variable in Q2 2024, Q4 2024, and Q1 2025, as well as the PBV variable in Q1 2025. High VIF values for the ROE variable indicate a fairly strong relationship between ROE and other financial ratio variables, particularly those that also represent corporate profitability. Therefore, the ROE variables for Q4 2024 and Q1 2025, which have the highest VIF values, were eliminated first.

Table 3 VIF values in the fundamental indicators for each quarter of 2024 and 2025

Variable	Q1 2024	Q2 2024	Q3 2024	Q4 2024	Q1 2025
EPS	2.00	1.98	1.93	1.91	1.82
PER	1.26	2.08	3.09	4.42	7.82
PBV	3.65	4.24	3.75	7.14	11.87
ROE	6.72	10.38	9.33	12.06	14.51
EV/EBITDA	2.50	3.68	3.17	3.73	6.47
Debt/Equity	2.64	3.38	4.25	3.32	6.24
NPM	2.50	4.40	4.00	5.75	4.97
TATO	2.79	4.64	4.10	4.61	3.82
LN.TA	3.84	4.73	6.82	5.26	4.62
TAG	1.65	1.32	1.95	2.92	2.00

After the elimination was performed, the results of the VIF re-test shown in Table 4 indicate that there are no longer any variables with a VIF value greater than 10. Thus, the PBV variable for Q1 2025 does not need to be eliminated because its multicollinearity has decreased. Meanwhile, the ROE variable for Q2 2024 was retained in the analysis because the VIF value obtained was not significantly different from the tolerance limit of 10. Additionally, the ROE variable was deemed to still have good ability to represent the company's profitability level and was therefore considered in the clustering process. Thus, the variables used in the subsequent analysis stage met the assumption of adequate multicollinearity.

Table 4 The VIF value for the fundamental indicator after eliminating the ROE variable

Variable	Q1 2024	Q2 2024	Q3 2024	Q4 2024	Q1 2025
EPS	2.00	1.98	1.93	1.91	1.67
PER	1.26	2.08	3.09	3.17	3.58
PBV	3.65	4.24	3.75	3.49	4.14
ROE	6.72	10.38	9.33	eliminated	
EV/EBITDA	2.50	3.68	3.17	2.51	3.89
Debt/Equity	2.64	3.38	4.25	2.05	3.43
NPM	2.50	4.40	4.00	2.78	2.41
TATO	2.79	4.64	4.10	5.25	2.06
LN.TA	3.84	4.73	6.82	2.80	4.43
TAG	1.65	1.32	1.95	2.51	1.92

The results of the VIF test on the stock performance indicator data are presented in Table 5. Based on these results, it can be seen that return, risk, and the Sharpe ratio have VIF values above 10, which means that these three variables exhibit high multicollinearity. Therefore, to address this multicollinearity, this study chose to eliminate the Sharpe Ratio variable. This decision was based on the consideration that Return and Risk are variables that represent a stock's performance and volatility, whereas the Sharpe Ratio is a derived variable that represents risk-adjusted efficiency. However, even though the Sharpe Ratio has been eliminated, multicollinearity between Return and Risk still exists. Consequently, the risk variable is eliminated. Afterwards, the VIF test was performed again and it is observed that there are no longer any VIF values above 10.

Table 5 VIF Values Before and After Variable Elimination in Stock Performance Indicators

Descriptions	Return	Risk	Sharpe Ratio	Maximum Drawdown	Average Drawdown
VIF Before Variable Elimination	17.64	14.54	12.56	4.91	3.06
VIF after the first round of variable elimination	8.67	13.98	Eliminated	4.66	2.97
VIF after the second round of variable elimination	1.79	Eliminated		3.37	2.49

C. KMeans Clustering

After confirming that there was no multicollinearity among the variables, they were standardized using min-max normalization. Then, K-means clustering analysis was performed on the fundamental indicators and stock performance indicators. K-means analysis on fundamental indicators was also performed separately for each quarterly period using financial ratio variables that had passed the multicollinearity test. Since K-Means is a non-hierarchical clustering method where the number of clusters formed is determined not by a dendrogram but subjectively by the researcher, a method is required to determine the optimal number of clusters (k) to ensure greater validity. In this study, the number k was determined using the Silhouette coefficient method.

Table 6 Silhouette coefficient values for various numbers of clusters in each fundamental indicators period

Period	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Q1 2024	0.259	0.285	0.234	0.222	0.244	0.246	0.250	0.266	0.252
Q2 2024	0.239	0.282	0.265	0.232	0.261	0.253	0.264	0.266	0.277
Q3 2024	0.304	0.313	0.290	0.293	0.312	0.311	0.275	0.272	0.252
Q4 2024	0.284	0.252	0.280	0.289	0.303	0.310	0.243	0.234	0.258
Q1 2025	0.353	0.310	0.313	0.305	0.283	0.267	0.273	0.269	0.268

The Silhouette method is used because it evaluates the quality of clustering results based on the average distance of objects from members within the same cluster (within-cluster distance) and the average distance to the nearest other cluster (between-cluster distance). The smaller the within-cluster distance and the larger the between-cluster distance, the better the cluster quality. Alternatively, in terms of the Silhouette coefficient value, the closer the value is to 1, the better the cluster quality. As shown in Table 6, the optimal number of clusters in the k-means algorithm is mostly 3. In Q4 2024, the optimal number of clusters is 7, and in Q1 2025, it is 2.

After obtained k optimal for each quarter, the clustering results were then analyzed using ANOVA to identify the important variables (features) which are able to differentiate the characteristics between clusters. Variables with significant mean differences between clusters are considered to make an important contribution to the cluster formation process.

Table 7 P-value in ANOVA Test

Variable	Q1 2024	Q2 2024	Q3 2024	Q4 2024	Q1 2025
EPS	0.914	0.833	0.627	0.000***	0.996
PER	0.891	0.898	0.774	0.000***	0.625
PBV	0.000***	0.000***	0.000***	0.000***	0.285
ROE	0.000***	0.000***	0.000***	eliminated	
EV/EBITDA	0.000***	0.000***	0.000***	0.000***	0.000***
Debt/Equity	0.014*	0.009**	0.007**	0.000***	0.000***
NPM	0.009**	0.039*	0.009**	0.000***	0.152
TATO	0.002**	0.001**	0.005**	0.012*	0.009**
LN.TA	0.008**	0.006**	0.000***	0.000***	0.001**
TAG	0.530	0.210	0.023*	0.037*	0.167

* significant at $\alpha = 0.05$
** significant at $\alpha = 0.01$
*** significant at $\alpha = 0.005$

The ANOVA test results are presented in Table 7. Based on these results, insignificant variables were eliminated from the clustering model, while significant variables were retained for further analysis. K-Means clustering was re-run using the selected important variables. For the Q1 and Q2 2024 data, the variables used were PBV, ROE, EV/EBITDA, DER, NPM, TATO, and LN TA. For the Q3 2024 data, the variables used were PBV, ROE, EV/EBITDA, DER, NPM, TATO, LN TA, and TAG. Finally, for the Q1 2025 data period, the variables used were EV/EBITDA, DER, TATO, and LN TA. The Q4 2024 data was not re-modeled using k-means because all variables were significant. The clustering model resulting from the variable selection was then re-evaluated using the silhouette coefficient.

The silhouette coefficient results in Table 8 show that the clustering model using feature-important variables has a higher silhouette value compared to the previous clustering model. This indicates that the variable selection process using ANOVA improves clustering quality by producing clusters that are more homogeneous within groups and more distinct across groups. Additionally, the optimal number of clusters for the Q2 2024 data changed from 3 to 7. Next, k-means clustering was performed based on the optimal k value obtained. The results of k-means clustering for fundamental indicators are shown in Figures 2a, 2b, 2c, 2d, and 2e.

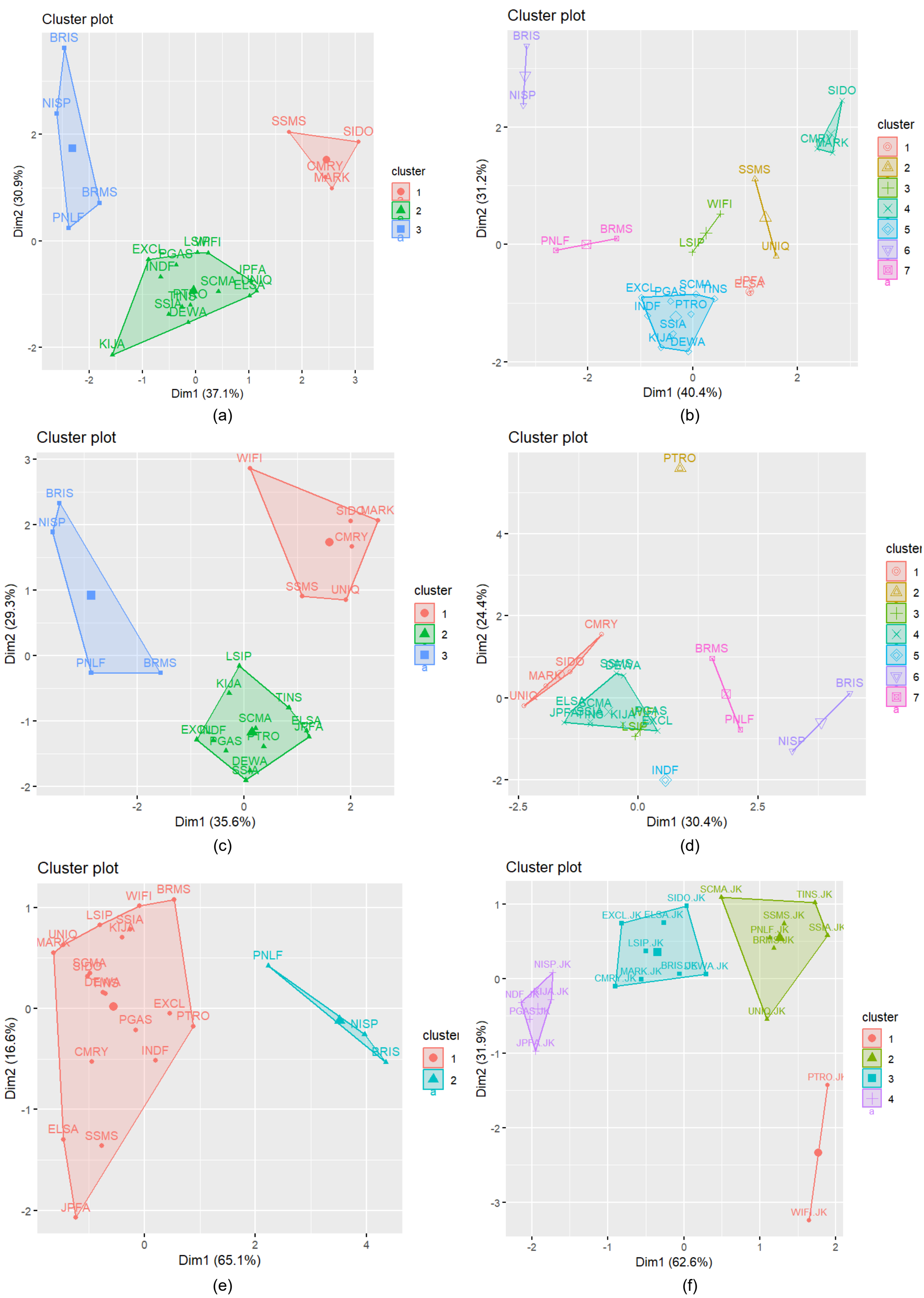


Figure 2. Clustering results based on data (a) Q1 2024, (b) Q2 2024, (c) Q3 2024, (d) Q4 2024, (e) Q1 2025, and (f) stock performance indicators

The stock performance indicator data was also analyzed using the optimal k. Based on the silhouette coefficient results shown in Table 9, the optimal number of clusters for the stock performance data is 4. The results of the K-means clustering of the stock performance indicators are shown in Figure 2f.

Table 8 Silhouette coefficient values before and after variable selection using ANOVA

Period	Optimal K Before Variable Selection	Silhouette Coefficient Value Before Variable Selection	Optimal K After Variable Selection	Silhouette Coefficient Value After Variable Selection
Q1 2024	3	0.285	3	0.373
Q2 2024	3	0.282	7	0.363
Q3 2024	3	0.313	3	0.367
Q4 2024	7	0.310		
Q1 2025	2	0.353	2	0.551

Table 9 Silhouette coefficient values for various numbers of clusters using stock performance indicator data

k	Silhouette coefficient values
2	0.4084
3	0.4063
4	0.4391
5	0.4244
6	0.3940
7	0.3592
8	0.3306
9	0.3034
10	0.3357

D. Stock Cluster Characteristics

The next step is to interpret the characteristics of each formed cluster. The interpretation of cluster results for fundamental indicator data is conducted by analyzing the average financial ratios within each cluster to identify groups of stocks with better fundamental conditions compared to other clusters. Clusters with high profitability values, such as EPS, ROE, and NPM, indicate a company's ability to generate higher profits [18]. A higher natural logarithm of total assets (LNTA) indicates that a company has larger total assets, thereby representing a larger company size. Furthermore, a higher total asset growth (TAG) indicates better company growth, while a negative value indicates a decline in the company's total assets [19]. Additionally, a high TATO value indicates the company's efficiency in utilizing assets to generate revenue [20].

From a risk perspective, clusters with a low Debt-to-Equity Ratio (DER) indicate a healthier leverage level and more controlled debt risk [20]. Meanwhile, generally, investors tend to seek stocks with relatively low valuation ratios such as PER, PBV, and EV/EBITDA, as they are considered to be undervalued or have the potential to be undervalued [21]. Based on these characteristics, clusters exhibiting a combination of high profitability, good asset efficiency, low leverage, and relatively reasonable valuations are categorized as clusters with healthier financial ratios. The stocks included in these clusters are then selected as potential candidates for the next stage of analysis.

Table 10 Average Fundamental Indicator Values Across Different Numbers of Clusters

Period	Cluster	EPS	PER	PBV	ROE	EV/EBITDA	DER	NPM	TATO	LN TA	TAG
Q1 2024	1	27.52	56.33	4.67	0.09	43.66	1.19	0.25	0.24	29.12	0.06
	2	38.18	157.75	0.92	0.02	28.16	1.12	0.06	0.17	30.30	0.03
	3	25.57	123.97	1.35	0.02	167.11	4.14	0.21	0.02	32.57	0.02
Q2 2024	1	93.56	9.30	0.89	0.09	10.84	1.29	0.06	0.68	30.60	0.03
	2	23.65	52.45	4.19	0.11	19.80	2.29	0.08	0.42	28.69	-0.01
	3	62.72	7.20	0.51	0.08	6.77	0.85	0.31	0.14	29.29	0.04
	4	53.53	36.59	5.68	0.16	28.31	0.16	0.28	0.52	28.73	-0.05
	5	83.69	75.25	0.89	0.03	12.85	1.11	0.05	0.30	30.85	0.05
	6	89.81	23.67	1.82	0.07	85.09	6.84	0.28	0.03	33.37	0.03
	7	12.98	76.33	0.72	0.01	61.45	1.41	0.13	0.04	31.82	0.01
Q3 2024	1	62.62	23.00	4.34	0.20	15.75	1.14	0.22	0.61	28.74	0.09
	2	156.60	50.65	1.04	0.07	8.33	1.00	0.09	0.54	30.73	-0.02
	3	80.54	45.85	1.58	0.06	59.55	4.16	0.21	0.06	32.62	0.03
Q4 2024	1	82.14	19.32	4.79	0.26	13.50	0.27	0.22	0.99	28.41	-0.02
	2	15.61	1769.70	6.86	0.04	20.87	2.47	0.01	0.79	30.28	0.23
	3	156.76	4.38	0.77	0.18	4.62	1.05	0.33	0.28	29.48	0.04
	4	104.53	26.33	1.27	0.11	6.79	1.22	0.06	0.79	30.58	0.02
	5	982.00	7.84	0.62	0.08	5.11	0.85	0.07	0.57	32.94	0.03
	6	183.85	11.97	1.75	0.14	45.28	6.99	0.28	0.06	33.46	0.04
	7	27.40	65.92	1.63	0.03	44.35	1.56	0.14	0.09	31.87	0.06
Q1 2025	1	37.03	143.57	2.31	0.03	37.13	1.00	0.12	0.18	30.16	0.05
	2	37.26	35.75	1.08	0.03	144.12	5.53	0.23	0.02	33.35	-0.01

Based on the clustering results shown in Table 10, in Q1 2024, the cluster with the healthiest financials was Cluster 1. The stocks classified in Cluster 1 are CMRY, MARK, SSMS, and SIDO. This cluster has a relatively better combination of fundamental indicators compared to other clusters. This is evident from its fairly good profitability metrics, with the highest ROE and NPM. Additionally, the highest TATO value of 0.24 indicates better asset utilization efficiency compared to other clusters, as well as the highest TAG.

In Q2 2024, the healthiest cluster is Cluster 3. The stocks classified in Cluster 3 are WIFI and LSIP. This cluster has the best combination of valuation, profitability, and risk compared to the other clusters. This is evident from a PER of 7.2, a PBV of 0.51, and an EV/EBITDA of 6.77, indicating that the stocks are undervalued. Additionally, Cluster 3 has the highest NPM at 0.31 and the lowest DER at 0.85 compared to most other clusters. Although the TATO is not particularly high, the combination of strong profitability and low leverage risk makes Cluster 3 the cluster with the healthiest fundamentals in Q2 2024.

Based on the results for Q3 2024, the healthiest cluster is Cluster 1. The stocks classified in Cluster 1 are CMRY, MARK, SSMS, SIDO, UNIQ, and WIFI. This cluster has the highest profitability, with an ROE of 0.20 and an NPM of 0.22. Additionally, a TATO of 0.61 indicates good asset efficiency. This cluster has a low EV/EBITDA ratio and a debt-to-equity ratio (DER) that remains under control at 1.14.

In Q4 2024, the healthiest cluster is Cluster 3. Stocks classified in Cluster 3 are WIFI and LSIP. This cluster has an excellent combination of fundamental indicators with a PER of 4.38, a PBV of 0.77, and an EV/EBITDA of 4.62, indicating relatively cheap stock valuations. Additionally, Cluster 3 also has the highest profitability with an ROE of 0.18 and an NPM of 0.33. The companies' leverage levels are also still considered healthy, with a DER of 1.05. Compared to other clusters, Cluster 3 demonstrates the best balance between low valuation, high profitability, and relatively low risk.

In Q1 2025, the healthiest cluster is Cluster 1. This cluster has a DER of 1, which is significantly lower than Cluster 2's. Additionally, the EV/EBITDA ratio for Cluster 1 is significantly lower at 37.13 compared to Cluster 2's 144.12. Although Cluster 1's profitability is not particularly high, it still offers a better combination of valuation and risk compared to Cluster 2.

Based on the results of the K-means clustering of the fundamental indicator data, three potential portfolios were identified. The first candidate portfolio consists of CMRY, MARK, SSMS, and SIDO. The second candidate portfolio consists of WIFI and LSIP. Finally, the third candidate portfolio consists of CMRY, MARK, SSMS, SIDO, UNIQ, and WIFI. Members of Cluster 1 for the Q1 2025 period were not included in the candidate portfolios because there were too many stocks, and they were already represented by the three previous candidate portfolios.

Table 11 Average Stock Performance Indicator Values for Different Numbers of Clusters

Cluster	Return	Risiko	Sharpe	MDD	Mean DD
1	0.008	0.055	0.142	0.409	0.092
2	0.002	0.037	0.060	0.385	0.137
3	0.001	0.024	0.032	0.288	0.100
4	0.001	0.017	0.046	0.165	0.060

The interpretation of the cluster results for stock performance indicator data was conducted by considering the balance between the level of return and investment risk in each cluster to identify groups of stocks that performed better than other clusters. Clusters with high returns and Sharpe ratios, as well as low risk, maximum drawdown, and average drawdown, indicate the clusters with the best stock performance.

Table 11 shows that each cluster exhibits a distinct investment profile. Cluster 1 has the highest return value of 0.008, significantly higher than the other clusters, and is the riskiest. Consequently, Cluster 1 has the highest Sharpe ratio, significantly higher than the other three clusters. However, this cluster has a relatively higher maximum drawdown compared to some of the other clusters. This indicates that the stocks in Cluster 1 exhibit a high-risk, high-return profile. Therefore, Cluster 1 is more suitable for aggressive investors with a high risk tolerance who are focused on achieving higher returns. Conversely, Cluster 4 has the lowest risk level, Maximum Drawdown, and Mean Drawdown compared to the other clusters. However, the returns generated are not as high as those of Cluster 1; the stocks in Cluster 4 are more suitable for conservative investors who tend to prioritize investment stability and avoid the risk of significant losses.

Based on this combination of return and risk indicators, in this study, the stocks classified in Cluster 1 were selected as the best cluster and designated as candidate stocks for portfolio construction. Therefore, the fourth portfolio candidate consists of PTRO and WIFI.

E. Stock Ranking Using TOPSIS

A TOPSIS score that approaches 1 indicates that the stock is getting closer to the positive ideal solution; in other words, the higher the TOPSIS score, the better the stock performs compared to others. Based on the ranking results using the TOPSIS method in Table 12, it is obvious that each period has different candidates for the best stocks. In the Q1 2024 period, nearly all stocks in the top 10 rankings had very high scores that were relatively close to one another, specifically above 0.97. Therefore, the top five stocks were selected as candidates for the fifth portfolio: SCMA, LSIP, JPFA, PTRO, and WIFI. These five stocks possess excellent fundamental characteristics and are close to ideal conditions.

In the Q2 2024 period, there were only three stocks with scores above 0.9: PGAS, EXCL, and TINS. These three stocks were selected as candidates for the sixth portfolio because they possess significantly better fundamental quality compared to other stocks during that period. Furthermore, in Q3 2024, there were five stocks with scores above 0.9: PGAS, EXCL, ELSA, INDF, and LSIP. These five stocks were selected as candidates for the seventh portfolio. In Q4 2024, there was a significant drop in scores after the top-ranked stock. The INDF stock received a very high score of 0.9944, while the second-ranked stock, JPFA, only scored 0.5970. Since the portfolio must contain at least two stocks, only the top two stocks were selected as candidates for the eighth portfolio, namely INDF and JPFA. In Q1 2025, several stocks had very high scores that were close to one another, specifically above 0.9. These stocks are EXCL, BRMS, PNLF, LSIP, SSIA, and WIFI. These six stocks were selected as candidates for the ninth portfolio.

Meanwhile, based on stock performance indicators, only the three stocks with the highest scores were selected as portfolio candidates: WIFI, PGAS, and DEWA. These three stocks became candidates for the tenth portfolio. Overall, the TOPSIS results indicate that stocks such as PGAS, EXCL, LSIP, INDF, and WIFI tended to consistently rank at the top during the study period. Consequently, these five stocks were also selected as the eleventh portfolio candidates.

Table 12 TOPSIS Ranking Results for the Top 10 Stocks Based on Fundamental and Stock Performance Indicators

Ranking	Q1 2024	Q2 2024	Q3 2024	Q4 2024	Q1 2025	Stock Performance
1	SCMA (0.9862)	PGAS (0.9527)	PGAS (0.9360)	INDF (0.9944)	EXCL (0.9847)	WIFI.JK (0.8791)
2	LSIP (0.9837)	EXCL (0.9416)	EXCL (0.9167)	JPFA (0.5970)	BRMS (0.9847)	PGAS.JK (0.7839)
3	JPFA (0.9789)	TINS (0.9351)	ELSA (0.9132)	ELSA (0.4983)	PNLF (0.9840)	DEWA.JK (0.7456)
4	PTRO (0.9777)	LSIP (0.8072)	INDF (0.9117)	PNLF (0.4761)	LSIP (0.9831)	LSIP.JK (0.7356)
5	WIFI (0.9763)	CMRY (0.8009)	LSIP (0.9011)	TINS (0.4749)	SSIA (0.9813)	BRIS.JK (0.7355)
6	INDF (0.9760)	SIDO (0.7991)	TINS (0.8969)	SIDO (0.4704)	WIFI (0.9719)	INDF.JK (0.7331)
7	BRIS (0.9756)	SSIA (0.6987)	JPFA (0.8939)	EXCL (0.4672)	INDF (0.8418)	SSMS.JK (0.7174)
8	PGAS (0.9751)	DEWA (0.6986)	SCMA (0.8927)	MARK (0.4610)	SSMS (0.8408)	SCMA.JK (0.7169)
9	DEWA (0.9729)	INDF (0.6978)	SSIA (0.8915)	LSIP (0.4596)	PTRO (0.8407)	JPFA.JK (0.6801)
10	MARK (0.9629)	MARK (0.6976)	DEWA (0.8905)	SCMA (0.4576)	CMRY (0.8406)	KIJA.JK (0.6209)

F. *Weighting Method Using MVEP and MAD*

Table 13 presents the results of stock weighting using the Mean-Variance Efficient Portfolio and Mean Absolute Deviation portfolio optimization methods. These two methods generate different investment weight allocations across the selected stocks because they employ different risk measurement approaches. In the MVEP method, weighting is determined based on the variance and covariance structure among stocks, as this weighting aims to minimize portfolio risk [22]. Thus, it is evident that stocks with low risk tend to receive a larger allocation of funds.

Table 13 Stock Weight Allocations Based on the MVEP and MAD Methods

Portofolio	Stock	MVEP weighting	MAD weighting
1	SSMS	0.054	0.185
	MARK	0.242	0.400
	SIDO	0.440	0.400
	CMRY	0.264	0.015
2	WIFI	0.089	0.500
	LSIP	0.911	0.500
3	WIFI	0.051	0.180
	UNIQ	0.115	0.200
	SSMS	0.036	0.020
	MARK	0.211	0.200
	SIDO	0.347	0.200
	CMRY	0.240	0.200
4	PTRO	0.541	0.500
	WIFI	0.459	0.500
5	SCMA	0.174	0.139
	LSIP	0.405	0.250
	JPFA	0.338	0.250
	PTRO	0.051	0.111
	WIFI	0.032	0.250
6	PGAS	0.411	0.400
	EXCL	0.534	0.298
	TINS	0.055	0.302
7	PGAS	0.145	0.250
	EXCL	0.246	0.250
	ELSA	0.194	0.144
	INDF	0.343	0.250
	LSIP	0.072	0.106
8	JPFA	0.246	0.500
	INDF	0.754	0.500
9	EXCL	0.457	0.200
	BRMS	0.101	0.007
	PNLF	0.073	0.200
	LSIP	0.252	0.200
	SSIA	0.086	0.200
	WIFI	0.031	0.193
10	WIFI	0.074	0.400
	PGAS	0.757	0.340
	DEWA	0.169	0.260
11	PGAS	0.206	0.250
	EXCL	0.278	0.250
	LSIP	0.103	0.047
	INDF	0.393	0.250
	WIFI	0.020	0.203

On the other hand, the MAD method uses mean absolute deviation as a measure of risk; thus, stocks with a combination of high returns and low MAD values will receive a larger investment proportion in the portfolio. The difference in weighting results between MVEP and MAD subsequently affects the expected return, risk, and Sharpe Index of the resulting portfolio. Therefore, portfolio performance evaluation is necessary to determine which weighting method can produce the most efficient portfolio.

G. Best Portfolio Selection

In terms of portfolio risk, Table 14 shows that the portfolio with MVEP weighting has lower risk than the portfolio with MAD weighting. This aligns with the theory, which aims to form a portfolio with the minimum variance among all possible portfolios that can be formed [22], [23]. This Table also shows that the best portfolio with highest Sharpe Ratio for both weighting methods is Portfolio 4, consisting of the stocks WIFI and PTRO. A higher Sharpe Ratio indicates that the portfolio is capable of delivering a more optimal return relative to the risk faced by investors.

When analyzed individually, WIFI stock had the highest expected return and Sharpe Index among Kompas100 stocks during the study period. In addition to its excellent stock performance, the results of K-Means clustering on fundamental indicator data from Q2 2024 to Q1 2025 show that WIFI consistently falls into the group of stocks with strong fundamentals. Ranking results using TOPSIS also show that WIFI ranked in the top five during the Q1 2024 and Q1 2025 periods. This indicates that WIFI possesses an excellent combination of stock performance and fundamental indicators, making it a worthy core component of an optimal portfolio.

Table 14 Portfolio Performance Based on the MVEP and MAD Weighting Methods

Portofolio	Stocks	Weighting Methods	Expected Return	Risk	Sharpe Ratio
1	SSMS, MARK, SIDO, CMRY	MAD	0.00131	0.01767	0.06272
2	WIFI, LSIP		0.00546	0.02597	0.20255
3	WIFI, UNIQ, SSMS, MARK, SIDO, CMRY		0.00316	0.02124	0.13948
4	PTRO, WIFI		0.00816	0.03694	0.21540
5	SCMA, LSIP, JPFA, PTRO, WIFI		0.00415	0.02428	0.16269
6	PGAS, EXCL, TINS		0.00129	0.01758	0.06172
7	PGAS, EXCL, ELSA, INDF, LSIP		0.00074	0.01305	0.04138
8	JPFA, INDF		0.00127	0.01431	0.07510
9	EXCL, BRMS, PNLF, LSIP, SSIA, WIFI		0.00317	0.02233	0.13311
10	WIFI, PGAS, DEWA		0.00498	0.02646	0.18047
11	PGAS, EXCL, LSIP, INDF, WIFI		0.00265	0.01768	0.13834
1	SSMS, MARK, SIDO, CMRY	MVEP	0.00092	0.01356	0.05330
2	WIFI, LSIP		0.00177	0.02086	0.07549
3	WIFI, UNIQ, SSMS, MARK, SIDO, CMRY		0.00171	0.01234	0.12219
4	PTRO, WIFI		0.00801	0.03991	0.19564
5	SCMA, LSIP, JPFA, PTRO, WIFI		0.00199	0.01616	0.11069
6	PGAS, EXCL, TINS		0.00095	0.01411	0.05327
7	PGAS, EXCL, ELSA, INDF, LSIP		0.00063	0.01059	0.04103
8	JPFA, INDF		0.00085	0.01392	0.04680
9	EXCL, BRMS, PNLF, LSIP, SSIA, WIFI		0.00150	0.01345	0.09666
10	WIFI, PGAS, DEWA		0.00206	0.01801	0.10331
11	PGAS, EXCL, LSIP, INDF, WIFI		0.00090	0.01116	0.06247

Meanwhile, PTRO is the stock with the second-highest expected return and Sharpe Ratio among the Kompas100 stocks. In terms of fundamental indicators based on the TOPSIS ranking results, PTRO ranked 4th in Q1 2024 and re-entered the top 10 in Q1 2025, which is ranking 9th. This indicates that PTRO possesses solid fundamental quality and consistently strong stock performance throughout the study period. Thus, this study successfully screened Kompas100 index by combining stock performance analysis, fundamental indicators, K-Means Clustering, TOPSIS, and also weighting methods resulting in the optimal portfolio comprising WIFI and PTRO. This combination of the two stocks results in a well-diversified portfolio, as it provides a balance between high returns and more manageable risk compared to investing in a single stock.

Because when compared to an investment in a single stock, such as WIFI stock, it is evident that Portfolio 4's expected return of 0.008 is slightly lower than the expected return of WIFI stock individually, which is 0.009. However, portfolio construction significantly reduced investment risk. The risk of Portfolio 4 is only 0.04, whereas the risk of a single investment in WIFI stock reaches 0.06. Consequently, the portfolio's Sharpe Ratio is higher than the Sharpe Ratios of the stocks individually. This indicates that the portfolio delivers more efficient investment performance because it generates a good return while simultaneously mitigating the risks investors might face through diversification across several stocks.

H. Selected Portfolio's Performance Evaluation

Next, the performance of the selected portfolio, Portfolio 4, which consists of PTRO and WIFI stocks, was tested to determine whether its performance remained stable outside the time period during which it was formed. This follow-up performance evaluation used data from the testing period (April 1–December 31, 2025), which was not used during the candidate selection or investment weighting stages. The evaluation results shown in Table 15 indicate a Sharpe Ratio of 0.24 for the MAD weighting method and 0.21 for the MVEP weighting method—both higher than the results from the training period, which were 0.21 (MAD) and 0.19 (MVEP). It can be seen that the performance of Portfolio 4 during the testing period did not decline compared to the training period; in fact, it showed a slight improvement under both weighting methods.

Table 15 Out-of-Sample Sharpe Ratio Performance of Portfolio 4 Using MVEP and MAD Weighting

Portofolio	Weighting Methods	Expected Return	Risk	Sharpe Ratio
Portofolio 4 (PTRO, WIFI)	MAD	0.00788	0.03223	0.23823
	MVEP	0.00810	0.03731	0.21169

To further examine the movement patterns of these two stocks in Portofolio 4 during the testing period, the closing prices were visualized using time-series charts, as shown in Figure 3. At first glance, Figure 3 indicates that both stocks exhibit the same upward trend. However, upon closer inspection, these two stocks have distinct characteristics, allowing them to complement each other in the formation of a stock portfolio. These differing price movement patterns are one of the key benefits of portfolio diversification, as price fluctuations between stocks do not move in lockstep.

WIFI stock experienced fairly aggressive price increases in early April, mid-July, and mid-October 2025, offering high return potential for the portfolio. However, WIFI's price movements are relatively more volatile, as evidenced by several sharp price spikes and drops, which also carry the potential for high risk. The risk associated with WIFI stock can be mitigated by PTRO stock, which is relatively stable yet still shows a gradual upward trend from April through early September 2025. Subsequently, PTRO stock continues to rise through the end of the year.

Interestingly, during certain periods when WIFI experienced a price decline, PTRO actually continued to show an upward trend or was able to maintain its price level. This difference in price movement patterns indicates the potential benefits of diversification when both stocks are included in a single portfolio. Thus, the risk resulting from fluctuations in one stock can be better managed because some of its movements can be offset by the other stock.

Combining these two stocks in a single portfolio creates a balance between potential returns and investment risk. Aggressive-growth stocks like WIFI can boost the portfolio's expected return, while PTRO helps maintain the stability of the portfolio's overall performance. Therefore, combining these two stocks can result in a more optimal portfolio compared to investing in just one stock.

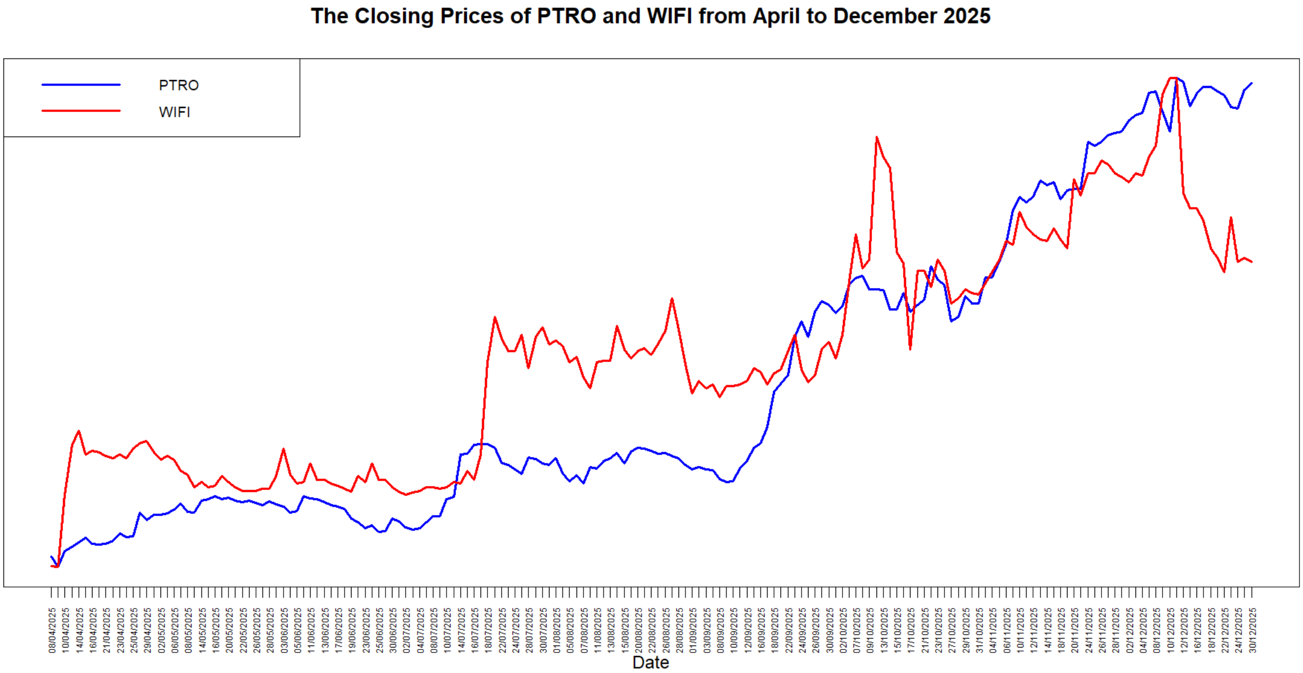


Figure 3. Closing Price Movements of the Stocks in the Optimal Portfolio (PTRO and WIFI) for the Period April - December 2025

V. CONCLUSIONS AND SUGGESTIONS

In the process of building a portfolio using stocks included in the Kompas100 Index, it turned out that of the 100 stocks, only 22 were suitable based on expected return, risk, Sharpe ratio, maximum drawdown, and average drawdown. Next, these 22 stocks were clustered to identify those with healthy financial characteristics and strong stock performance in the market using K-Means clustering. From the K-Means clustering results, 4 potential portfolios were selected. In addition to clustering analysis, stock selection for portfolio construction also utilized a ranking method, namely TOPSIS. Based on the output from the TOPSIS method, 7 portfolio candidates were obtained, bringing the total to 11 portfolio candidates. For these 11 portfolio candidates, investment allocation proportions were also calculated using the MVEP and MAD weighting methods, and portfolio performance was evaluated using the Sharpe Ratio. Based on the evaluation results, the best portfolio is Portfolio 4, consisting of WIFI and PTRO stocks. Next, the portfolio's performance was retested to determine whether it remained stable outside the time period during which it was formed. It turned out that Portfolio 4's performance during the testing period, April through December 2025, did not decline compared to the training period; in fact, it showed a slight improvement under both weighting methods.

Thus, this study successfully constructed a stock portfolio using the K-Means clustering approach and the TOPSIS method as two candidate screening methods, followed by the determination of investment proportions using the MVEP and MAD weighting methods. The combination of these two stocks results in a well-diversified portfolio, as it strikes a balance between high returns and more manageable risk compared to investing in a single stock. Therefore, a portfolio investment strategy is more recommended, especially for investors seeking more stable and efficient investment performance in the long term.

This study still has several limitations that could serve as opportunities for development in future research. First, future studies could use a longer observation period to better capture diverse market conditions and generate a portfolio more robust against economic changes.

Second, this study analyzed separate data periods. Future research could employ clustering methods on panel data to analyze the data as a whole. Third, portfolio construction in this study utilized the MVEP and MAD methods. Future research could develop optimization models using alternative methods, such as Mean-CVaR, Genetic Algorithms, and Particle Swarm Optimization.

Fourth, this study considers only fundamental indicators and stock performance indicators. Future research could considet macroeconomic factors and market sentiment derived from news and social media to improve the quality of the stock selection process. With these enhancements, it is hoped that future research will yield stock selection and portfolio optimization models that are more comprehensive, accurate, and adaptable to capital market dynamics, as well as provide better evaluation results.

ACKNOWLEDGEMET

The authors would like to express their sincere appreciation to the Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Tanjungpura, for providing financial support that made this research possible. The authors declare that there are no conflicts of interest related to this work.

REFERENCES

[1] Lentera Today, "BI Ungkap Geopolitik Global Picu Rupiah Melemah hingga ke Rp17.300," Lentera Today. Accessed: May 03, 2026. [Online]. Available: https://lenteratoday.com/post/item/232529/BI-Ungkap-Geopolitik-Global-Picu-Rupiah-Melemah-hingga-ke-Rp17300

[2] M. A. Santoso and I. W. Wenagama, "DAMPAK DEPRESIASI NILAI TUKAR RUPIAH TERHADAP DOLAR, PERTUMBUHAN UANG BEREDAR DAN LAJU PDRB TERHADAP INFLASI | E-Jurnal Ekonomi Pembangunan Universitas Udayana," *E-Jurnal Ekonomi Pembangunan Universitas Udayana*, vol. 7, no. 12, pp. 2698–2731, Dec. 2018.

[3] S. Syamsudin and M. Khaddafi, "Analisis Komparatif Risiko dan Return Pada Saham dan Obligasi di Bursa Efek Indonesia: Implikasi Bagi Investor," *Jurnal Ekonomi, Bisnis dan Manajemen*, vol. 3, no. 2, pp. 61–76, Jun. 2024, doi: 10.58192/ebismen.v3i2.2209.

[4] V. T. Winarty and S. Safarina, "Portfolio Optimization of Indonesian Banking Stocks Using Robust Optimization," Oct. 17, 2025, *arXiv*: arXiv:2510.15288. doi: 10.48550/arXiv.2510.15288.

[5] V. E. Hidayat and A. S. Pamungkas, "Pengaruh Literasi Keuangan, Persepsi Risiko, dan Toleransi Risiko terhadap Keputusan Investasi pada Saham," *JMK*, vol. 4, no. 3, pp. 767–776, Aug. 2022, doi: 10.24912/jmk.v4i3.19771.

[6] A. A. D. Purwanti, D. S. P. Koesoemasari, and E. S. Wahyuningsih, "Return Saham, Faktor Fundamental, Risiko, dan Suku Bunga di LQ 45," *Jurnal Ilmiah Akuntansi dan Keuangan*, vol. 13, no. 1, pp. 1–11, Jun. 2024, doi: 10.32639/jiak.v13i1.690.

[7] G. N. Sabatini, O. R. Putri, and N. K. Wardhany, "Impact Dengan Analisis Fundamental Dab Analisis Teknikal Terhadap Harga Saham Pada PT. Telkom Indonesia Tbk," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 11175–11192, Jan. 2026, doi: 10.31004/riggs.v4i4.4815.

[8] D. Tohendry and D. Jollyta, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKKAN SAHAM BERDASARKAN PRICE EARNING RATIO DAN PRICE TO BOOK VALUE," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi (JMApTeKsi)*, vol. 5, no. 1, pp. 1–7, Apr. 2023.

[9] E. Ebermam, H. Knidel, and R. A. Krohling, "Development of a hybrid method for stock trading based on TOPSIS, EMD and ELM," Jun. 14, 2022, *arXiv*: arXiv:2206.06723. doi: 10.48550/arXiv.2206.06723.

[10] I. A. Majidah, A. Rahim, and M. Bahri, "Mean Variance Complex-Based Portfolio Optimization," *Statistics, Optimization & Information Computing*, vol. 12, no. 5, pp. 1382–1396, Jun. 2024, doi: 10.19139/soic-2310-5070-2023.

[11] D. Chen, Y. Wu, J. Li, X. Ding, and C. Chen, "Distributionally robust mean-absolute deviation portfolio optimization using wasserstein metric," *J Glob Optim*, pp. 1–23, May 2022, doi: 10.1007/s10898-022-01171-x.

[12] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, Eighth edition. Australia Brazil Mexico South Africa Singapore United Kingdom United States: Cengage, 2019.

[13] E. Sulistianingsih *et al.*, "AN EXAMINATION OF THE GREEN STOCK PORTFOLIO IN CONNECTION WITH THE 2024 INDONESIAN REPUBLIC PRESIDENTIAL GENERAL ELECTION," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 4, pp. 2543–2556, Sep. 2025, doi: 10.30598/barekengvol19iss4pp2543-2556.

[14] M. Sriningsih, D. Hatidja, and J. D. Prang, "PENANGANAN MULTIKOLINEARITAS DENGAN MENGGUNAKAN ANALISIS REGRESI KOMPONEN UTAMA PADA KASUS IMPOR BERAS DI PROVINSI SULUT," *JIS*, vol. 18, no. 1, p. 18, Jul. 2018, doi: 10.35799/jis.18.1.2018.19396.

[15] S. Martha, W. Andani, and S. W. Rizki, "Perbandingan Metode k-Nearest Neighbor, Regresi Logistik Biner, dan Pohon Klasifikasi pada Analisis Kelayakan Pemberian Kredit," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 10, no. 2, pp. 262–273, Nov. 2022, doi: 10.34312/euler.v10i2.16751.

[16] D. Ilham, N. E. Putri, N. Patricia, N. F. Nst, and S. Safrizal, "Penerapan Metode TOPSIS untuk Memilih Laptop Terbaik Sesuai Kebutuhan Konsumen," *Journal of Manufacturing and Enterprise Information System*, vol. 3, no. 1, pp. 32–40, Apr. 2025, doi: 10.52330/jmeis.v3i1.415.

[17] T. N. Annisa, J. Jasmir, and N. Nurhadi, "Comparison of ANOVA and Chi-Square Feature Selection Methods to Improve Machine Learning Performance in Anemia Classification," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 1925–1940, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.5017.

[18] A. Santoso, "Analisis Price to Book Value Ratio melalui Net Profit Margin, Return on Asset, and Return on Equity," *Jurnal Akutansi dan Bisnis*, vol. 15, no. 2, pp. 88–94, Agustus 2015.

[19] D. A. Vemalasari and M. E. Puspita, "Pengaruh Total Aset Terhadap Harga Saham pada Perusahaan Sektor Energi di Bursa Efek Indonesia Periode 2018-2022," 2024.

[20] D. Fadilah, M. Rimawan, and O. Ovriyadin, "Analisis Pengaruh Total Asset Turnover (TATO) dan Debt to Equity Ratio (DER) Terhadap Harga Saham Pada PT Unilever Indonesia Tbk," *EBISMEN*, vol. 4, no. 3, pp. 207–224, Sep. 2025, doi: 10.58192/ebismen.v4i3.3687.

[21] T. Li, "Assessing Stock Value: P/E, PEG, and P/B Ratios Explained," Investopedia. Accessed: May 22, 2026. [Online]. Available: https://www.investopedia.com/articles/investing/101316/how-tell-if-stock-overvalued-or-undervalued.asp

[22] E. Fitriani, Y. Hikmah, and I. R. Hikmah, "COMPARISON OF VALUE AT RISK USING HISTORICAL AND MONTE CARLO METHODS ON PT XYZ STOCK PORTOFOLIO," *JABT*, vol. 6, no. 2, pp. 102–111, Jun. 2024, doi: 10.7454/jabt.v6i2.1105.

[23] F. Farkhati, A. Hoyyi, and Y. Wilandari, "ANALISIS PEMBENTUKAN PORTOFOLIO OPTIMAL SAHAM DENGAN PENDEKATAN OPTIMISASI MULTIOBJEKTIF UNTUK PENGUKURAN VALUE AT RISK," *Jurnal Gaussian*, vol. 3, no. 3, pp. 371–380, 2014, doi: 10.14710/j.gauss.3.3.371-380.