JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 22, No. 1, September 2025, pp. 187-198
DOI: 10.20956/j.v22i1.43875

Optimization of Stock Portfolio Investment based on K-Means Clustering using Markowitz Method (A case study: IDX-MES BUMN17 Index)

Muhammad Rifki Nisardi[1*], Restu Ananda Putra[2], Sri Muslihah Bakhtiar[3], Hartina Husain[1], Wahyuni Ekasasmita[1], Kusnaeni[1]
[1)]Institut Teknologi Bacharuddin Jusuf Habibie, Parepare, South Sulawesi, Indonesia
[2)]Universitas Gadjah Mada, Yogyakarta, Indonesia
[3)] Universitas Muhammadiyah Kolaka Utara Kolaka Utara, South East Sulawesi, Indonesia
***Email**: muhammadrifkinisardi@ith.ac.id[*], restuanandaputra@ugm.ac.id, srimuslihahb@umkota.ac.id, hartinahusain@ith.ac.id, wahyuni.ekasasmita@ith.ac.id, kusnaeni25@ith.ac.id*
******Correspondence Author*

Abstract

A stock portfolio is a combination of two or more equity securities invested over a specific period and under certain conditions. This research analyzes stock combinations that can be formed into an optimal portfolio using the Markowitz method. The Markowitz method is employed to maximize returns and minimize the risk of a portfolio. The data used in this study consists of daily closing prices from the IDX-MES BUMN17 index, one of the indices in Indonesian Stock Exchange, between January 2023 and December 2023. Based on the results obtained, two recommended portfolios are identified, known as the Minimum Variance Portfolio (MVP) and Tangency Portfolio. The optimal portfolio can serve as an option depending on the investor's risk profile.

Keywords: IDX-MES BUMN17, Markowitz method, Minimum Variance Portfolio (MVP), Stock Portfolio, Tangency Portfolio.

1. INTRODUCTION

Investment is the activity of allocating a certain amount of capital with the aim of obtaining future returns. In making investments, each investor has individual preferences and specific objectives that guide their investment choices, including the type of investment instrument selected. There are various types of investment instruments available, one of which is financial assets in the form of stocks. Stocks represent ownership in a company and grant shareholders rights to dividends as well as participation in corporate decisions. Investing in stocks can be profitable when the company's value increases, but it also carries the risk of value depreciation when the company underperforms. This duality compels investors to carefully consider two key aspects when investing in stocks: how to maximize returns and, simultaneously, how to minimize potential risks. In the stock market, there is generally a positive correlation between expected return and risk—the higher the expected return, the higher the potential risk that must be borne by the investor, and vice versa [4], [15].

One strategy that investors can employ to reduce investment risk in stocks is diversification. Diversification involves constructing a stock portfolio by allocating investments across several companies, thereby mitigating overall portfolio risk [1], [2], [12]. Diversification is essential for investors seeking to reduce risk without significantly diminishing the expected rate of return. Portfolio diversification can be implemented either randomly or through the use of portfolio optimization methods. One such method is the Markowitz approach [10], [14].

The Markowitz method is a quantitative technique used to determine the optimal allocation of stocks within a portfolio by considering multiple constraints. The method is based on the principle of maximizing expected returns for a given level of risk, or minimizing risk for a given level of expected return. These dual objectives result in a set of optimal portfolios from which an investor may choose, depending on their preferred balance of return and risk. This collection of optimal portfolios is commonly referred to as the Efficient Frontier [3], [6].

Before applying the Markowitz optimization, it is often beneficial to perform a preliminary grouping of stocks that exhibit similar behavioural patterns. One effective approach to achieve this is by employing K-Means clustering, an unsupervised machine learning algorithm widely used for pattern recognition and data segmentation. K-Means clustering partitions stock data—typically based on historical returns, volatility, or other market-related features—into “k” clusters, where each cluster contains stocks with similar statistical characteristics. This process involves iteratively assigning stocks to the nearest cluster centroid and updating the centroids based on the average of assigned data points until the algorithm converges [11], [13]. By grouping stocks in this manner, investors can reduce the dimensionality of the optimization problem and avoid the overrepresentation of highly correlated assets, which could distort the risk estimation in the Markowitz model. Furthermore, by selecting representative stocks from different clusters or applying optimization within each cluster, investors can construct a more diversified and robust portfolio. This hybrid approach effectively integrates data-driven clustering with classical portfolio theory, enhancing both the analytical rigor and practical effectiveness of investment decision-making.

However, several previous studies have applied the Markowitz method without considering optimal diversification decisions. Diversification is often conducted subjectively, based on the preferences of investors or researchers. This study focuses on the application of the Markowitz method to one of the stock indices listed on the Indonesia Stock Exchange (IDX), namely the IDX-MES BUMN 17. This index comprises up to 17 state-owned enterprise (SOE) stocks that are included in the list of sharia-compliant securities and are among the largest in terms of market capitalization and liquidity. From these 17 stocks, a portfolio will be constructed consisting of selected stocks. The stock selection (diversification) will be based on the K-Means Clustering method, in which stocks are classified according to their return and volatility characteristics.

2. THEORETICAL REVIEW

Stock Returns

According to Maruddani and Purbowati (2009) as cited in [8], return is one of the key factors that motivates individuals to invest, as it reflects the outcome of the investment. The higher the return, the greater the profit an investor can potentially earn. The formula used to calculate stock return is:

$$R_{i,t} = \ln\left(\frac{P_{it}}{P_{i(t-1)}}\right) \tag{2.1}$$

$R_{i,t}$ = Return of stock *i* at time *t*
P_{it} = Closing price of stock *i* at time *t*
$P_{i(t-1)}$ = Closing price of stock *i* at time *t-1*

Expected Stock Return

The expected return represents the anticipated profit that an investor aims to receive in the future. It can be estimated by calculating the mean of historical returns of an asset [7]:

$$E(R_i) = \frac{\sum_{i=1}^{n} R_{i,t}}{n} \tag{2.2}$$

Variance and Standard Deviation of Stock Returns

Variance and standard deviation are statistical measures used to assess the dispersion or deviation of data. In the context of stock investment, risk can be interpreted as the degree of deviation between realized returns and expected returns. The greater the deviation, the higher the risk associated with the asset. Variance and standard deviation can therefore be used to quantify the investment risk of an asset. The mathematical formulas are as follows [5]:

$$\sigma_i^2 = \frac{1}{n}\sum\left(R_{i,t} - E(R_i)\right)^2 \tag{2.3}$$

$$\sigma_i = \sqrt{\sigma_i^2}$$

Where:

σ_i^2 : Variance of stock *i*
σ_i : Standard deviation of stock *i*

Covariance Between Two Stocks

Covariance is a measure used to assess the directional relationship between two variables. The formula to calculate the covariance is as follows [16]:

$$Cov(R_A, R_B) = \sigma_{R_A,R_B} = \sum_{i=1}^{n} \frac{[(R_{Ai} - E(R_A))\cdot(R_{Bi} - E(R_B))]}{n} \tag{2.4}$$

Where:

$Cov(R_A, R_B)$: Covariance of returns between stocks A and B
R_{Ai} : Realized return of stock A in the *i*-th observation
R_{Bi} : Realized return of stock B in the *i*-th observation
$E(R_A)$: Expected return of stock A
$E(R_B)$: Expected return of stock B

Correlation Between Two Stocks

The correlation coefficient measures the strength and direction of the linear relationship between two variables, relative to their individual standard deviations. The formula to compute the correlation coefficient is [16]:

$$r_{AB} = \rho_{AB} = \frac{\sigma_{R_A,R_B}}{\sigma_{R_A}\cdot\sigma_{R_B}} \tag{2.5}$$

Where:

r_{AB} : Correlation coefficient of returns between stocks A and B
σ_{R_A,R_B} : Covariance between stock A and stock B

σ_{R_A} : Standard deviation of stock A
σ_{R_B} : Standard deviation of stock B

Markowitz Portfolio Theory

The Markowitz method is used to determine the optimal allocation of funds among various stocks within a portfolio. The optimal proportions are determined through an optimization process. In this study, Python programming is employed to solve the optimization problem of identifying the optimal stock weights. The goal is to find the combination that yields the portfolio with the lowest risk among the five selected stocks. Thus, the objective function used is the portfolio risk function based on the Markowitz model, which is formulated in the optimization problem as follows [16]:

$$\min \sum_{i=1}^{n} w_i {\sigma_i}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{2.6}$$

Subject to the constraints:
(1) $\sum_{i=1}^{n} w_i = 1$
(2) $w_i \geq 0$ for each $i = 1,2,..,n$
(3) $\sum_{i=1}^{n} w_i R_i = R_p$

K-Means Clustering and Portfolio Diversification

K-Means clustering is one of the simplest and most widely used unsupervised machine learning algorithms. K-Means has the aim of grouping similar or identical data points and discover underlying patterns. In grouping similar data points, K-Means uses the right number of target clusters in a data set, where refers to the number of centroids required in the data set. The algorithm works by partitioning observations into clusters based on closest means, creating Voronoi cells in the data space. By minimizing internal variance cluster, K-Means clustering converges convergently to optimal local usage heuristic algorithm [9].

3. RESEARCH METHOD

The data utilized in this study consists of secondary data in the form of daily closing stock prices from 17 companies listed in the IDX-MES BUMN17 index, covering the period from January 2023 and December 2023, obtained from the website https://finance.yahoo.com/. The calculations and modeling in this research were conducted using the Python programming language. The steps undertaken in this study are as follows:

1. Gather daily closing stock price data from 17 companies.
2. Transform daily closing price data into return data.
3. Calculate the return and risk for each data.
4. Apply K-Means Clustering.
5. Calculate the correlation coefficient for each cluster and examine the Markowitz Mean-Variance model.
6. Combine stocks into a portfolio by selecting potential stocks from each cluster.
7. Generate the output of portfolio optimization.
8. Determine the Minimum Variance Portfolio (MVP) and Tangency Portfolio.

4. RESULT AND DISCUSSION

In this study, we utilized closing price data from companies from IDX-MES BUMN17 as previously described. By employing Visual Studio Code software and the Python programming language, the stock price movement charts for each company's closing prices can be visualized in Figure 1 below.

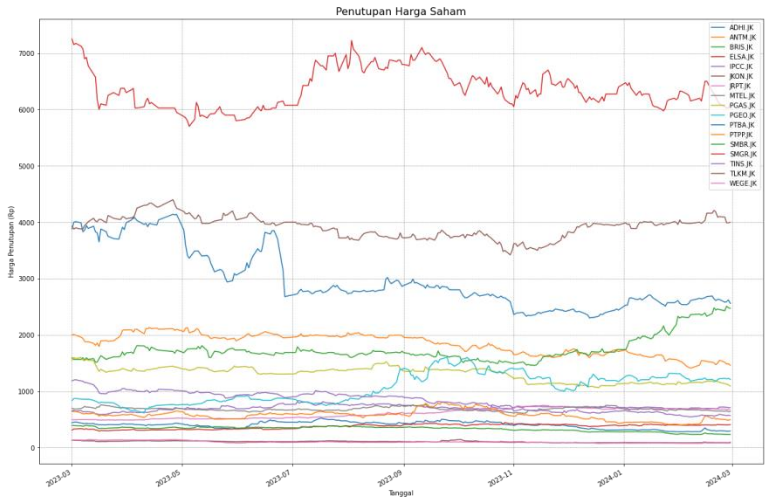


Figure 1. Stock Price Data

This study applies a clustering method to assist investors in analysing time series data of stock closing prices, aiming to group stocks with similar price movement patterns. This clustering serves as a basis for diversifying the selected stock portfolio. Based on the clustering results using the K-Means method, six clusters were identified in accordance with the Elbow method, because the decrease in inertia began to slow down starting at cluster = 6 as shown in Figure 2. Figure 3 shows the visualization of the clustering results consisting of six distinct clusters. It is evident that these six clusters represent different characteristics based on the grouping of their risk levels and returns. In Figure 4, we also presented the clustering result by price for each group.

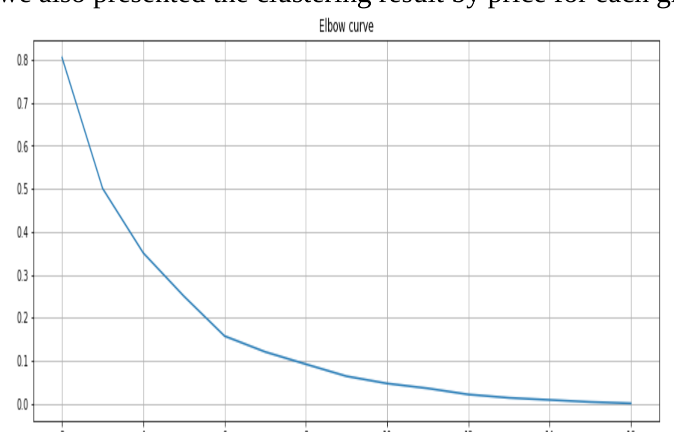


Figure 2. Elbow Method

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Muhammad Rifki Nisardi, Restu Ananda Putra, Sri Muslihah Bakhtiar, Hartina Husain, Wahyuni Ekasasmita, Kusnaeni

Figure 3. Clustering Result

Figure 4. Clustering by the Prices

Based on Figures 3 and 4, cluster 1 and cluster 3 are groups of stocks that exhibit upward price trends during the specified time interval. Both clusters are among those with higher return levels compared to the other clusters. However, the difference between the stocks in cluster 1 and cluster 3 lies in their volatility levels, where stocks in cluster 3 tend to have higher volatility compared to those in cluster 1. Meanwhile, the stocks in cluster 2 and cluster 6 have medium return levels compared to the stocks in the other clusters. The difference between these clusters is that cluster 2 tends to have higher volatility than stocks in cluster 6. The last two clusters, cluster 4 and 5, consist of stocks with lower return levels compared to the other four clusters.

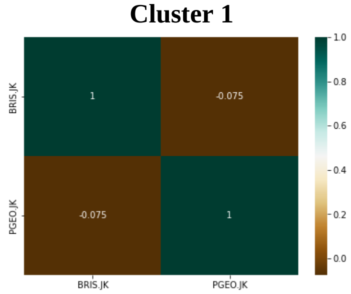
Table 1. Stock List for Each Cluster

Cluster	Stock List
1	BRIS, PGEO
2	PTPP, ADHI, JKON
3	ELSA, JRPT
4	PGAS, ANTM, SMBR, PTBA, WEGE
5	TINS
6	IPCC, SMGR, MTEL, TLKM

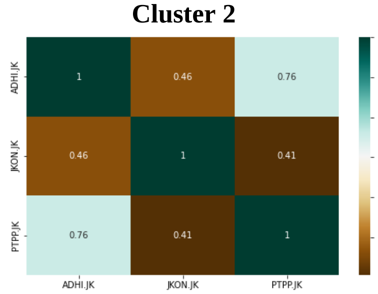
Subsequently, a correlation analysis was performed within each cluster to evaluate the strength of relationships among the constituent stocks. The results, illustrated in Figure 5, reveal that most clusters do not exhibit strong internal correlations, supporting the effectiveness of the clustering process in grouping stocks with distinct price movement patterns. In Cluster 1, the stocks BRIS and PGEO demonstrated a weak negative correlation of -7.5%, indicating a negligible relationship between their price movements. Cluster 2 includes ADHI, JKON, and PTPP, which showed positive correlations, with the highest correlation—76%—observed between ADHI and PTPP. This high correlation suggests that these two stocks exhibit similar return patterns and, therefore, may not be suitable for inclusion in the same investment portfolio due to limited diversification benefits.

Cluster 3 comprises ELSA and JRPT, with a correlation coefficient of just 4%, indicating a very weak association and suggesting potential suitability for portfolio diversification. In Cluster 4, which includes ANTM, PGAS, PTBA, SMBR, and WEGE, the correlation coefficients range from 12% to 36%. Although these are moderate, they indicate a tendency toward similar price trends, which may reduce the effectiveness of diversification if these stocks are held within the same portfolio. Lastly, Cluster 6 consists of IPCC, MTEL, SMGR, and TLKM. While the correlations among these stocks vary, all coefficients remain relatively low, suggesting limited shared movement and potential for diversification.

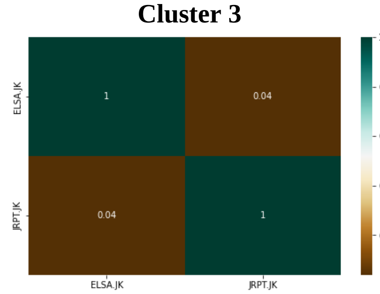
Cluster 1



Cluster 2



Cluster 3



Cluster 4

Cluster 5

Cluster 6

Figure 5. Stock Correlation by Cluster

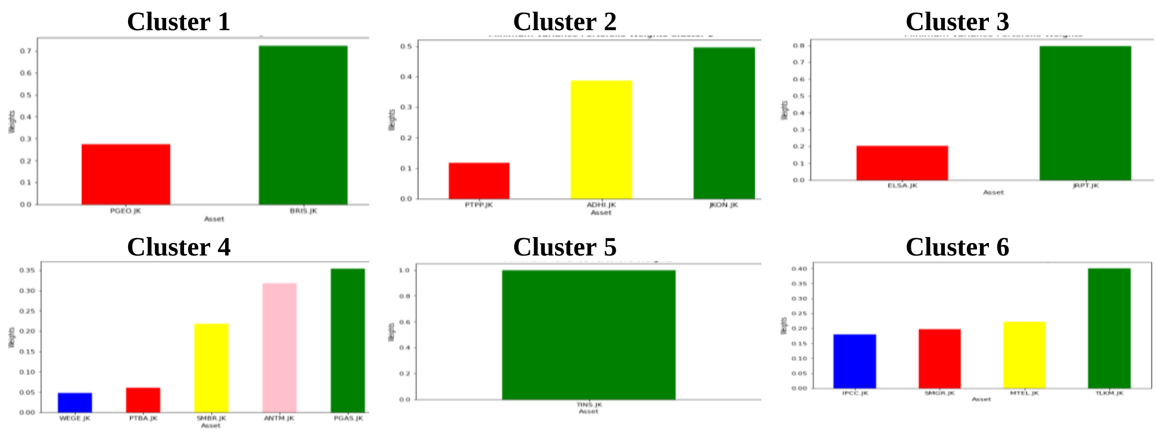


Figure 6. Stock Proportion for each Cluster

As a comparison, portfolios were constructed based on the clustering results obtained through the K-Means method, with one portfolio representing each cluster. This approach aimed to evaluate the respective levels of return and risk under the assumption that each portfolio is optimized to minimize risk, in accordance with the Mean-Variance Portfolio framework. The asset allocation proportions, as well as the resulting return and risk values for each portfolio, are presented in Figure 6 and Table 1.

Table 2. Return and Risk by the Cluster

Cluster	Stock List	Return MVP	Risk MVP	Stock With Highest Proportion
1	BRIS, PGEO	0.5731	0.275	BRIS
2	PTPP, ADHI, JKON	-0.2914	0.4028	JKON
3	ELSA, JRPT	0.4353	0.135	JRPT
4	PGAS, ANTM, SMBR, PTBA, WEGE	-0.2842	0.1808	PGAS
5	TINS	-0.5187	0.3018	TINS
6	IPCC, SMGR, MTEL, TLKM	0.0104	0.128	TLKM

Based on the results presented in Table 1, the portfolio derived from Cluster 1 yielded the highest return, amounting to 57.31%, accompanied by a risk level of 27.5%. Within this portfolio, BRIS had the highest allocation proportion, comprising approximately 70% of the total investment. In contrast, Cluster 5 exhibited the lowest return, recorded at -51.87%**,** indicating poor performance relative to the other clusters. In terms of risk, Cluster 6 demonstrated the lowest level of portfolio risk, measured at 12.8%, making it the most stable portfolio among the six. Conversely, Cluster 2 recorded the highest level of risk, reaching 40.28%. Based on these findings, the stocks with the highest allocation proportions from each cluster will be combined into a composite portfolio, for which the optimal allocation of each constituent stock will be determined in order to construct a more efficient and well-diversified portfolio.

Minimum Variance Portfolio

In this section, the stock with the highest proportion in each cluster is selected and compiled into a portfolio. Subsequently, using the Markowitz model, two types of portfolios are constructed based on different investor preferences: the Minimum Variance Portfolio (MVP) and the Tangency Portfolio. Figure 7 and Table 3 present the simulation results for the MVP. This portfolio consists of JRPT (41.50%), TLKM (33.72%), TINS (12.37%), PGAS (4.86%), BRIS (4.37%), and JKON (3.18%). The resulting combination of stocks carries a risk level of 10.8% with an expected return of 8.14%. This implies that if an investor allocates their capital according to the stock selection and weightings mentioned above, they can anticipate an expected return of approximately 8.14% of the total investment, with a corresponding risk exposure of 10.8%. Compared to the individual cluster portfolios, the MVP exhibits the lowest level of risk, making it particularly suitable for investors with a conservative or risk-averse profile. The resulting Sharpe Ratio for this portfolio is 0.1531.

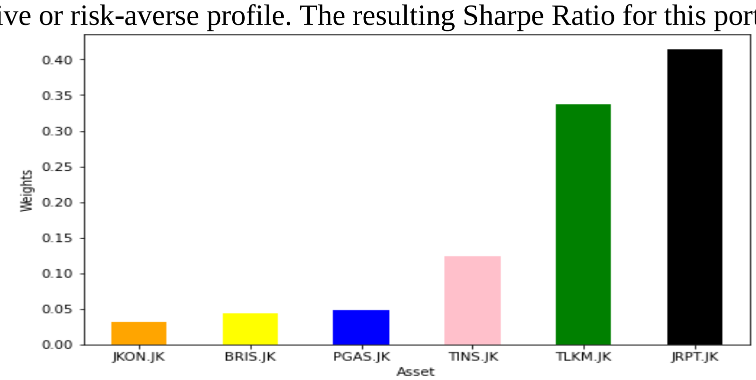


Figure 7. Proportion Minimum Variance Portfolio (MVP) After Clustering

Table 3. Proportion Each Stock

JKON	BRIS	PGAS	TINS	TLKM	JRPT
0.0318	0.0437	0.0486	0.1237	0.3372	0.4150

Subsequently, the Tangency Portfolio is constructed based on the selected set of stocks. The Tangency Portfolio represents the optimal portfolio positioned at the point of tangency between the Capital Market Line (CML) and the Efficient Frontier (Figure 9). This portfolio yields the maximum Sharpe Ratio, making it the most efficient combination of risky assets and the risk-free asset. Using an assumed risk-free rate of 6.5% (as represented by World Government Bonds), the resulting portfolio comprises JRPT (46.17%), BRIS (38.79%), TLKM (5.92%), TINS (5.47%), PGAS

(2.55%), and JKON (1.1%), as shown in Figure 8 and Table 4. This portfolio achieves an expected return of 35.27% with a corresponding risk level of 15.07%. The resulting Sharpe Ratio is 1.91, indicating that the portfolio delivers strong returns relative to the level of risk undertaken. This suggests that the portfolio is well-constructed and exhibits high efficiency in risk-adjusted performance.

The comparative analysis of two investment portfolios reveals significant differences in risk-adjusted performance as measured by the Sharpe Ratio. Tangency Portfolio exhibits a Sharpe Ratio of 1.91, indicating a highly efficient conversion of risk into returns, as it generates nearly twice the excess return per unit of risk taken. This suggests that Tangency Portfolio effectively balances risk and reward, providing strong performance relative to the volatility endured. In contrast, MVP's Sharpe Ratio of 0.15 reflects poor risk-adjusted returns, implying minimal compensation for the risk assumed by the investor. Such a low ratio suggests that MVP may suffer from inadequate risk management or a return profile that scarcely exceeds the risk-free benchmark. These findings underscore the importance of considering risk-adjusted metrics over absolute returns when evaluating portfolio performance, as superior Sharpe Ratios translate to more desirable investment outcomes for risk-conscious investors

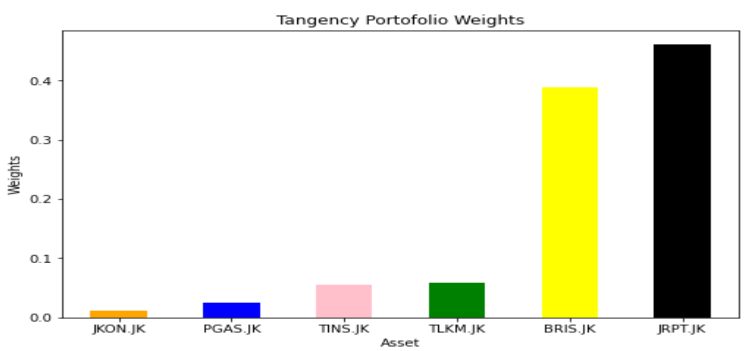


Figure 8. Tangency Portfolio Proportion

Table 4. Stock Proportion

JKON	PGAS	TINS	TLKM	BRIS	JRPT
0.0110	0.0255	0.0547	0.0592	0.3879	0.4617

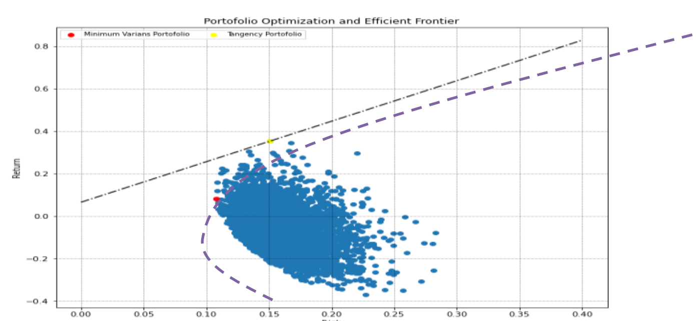


Figure 9. Efficient Frontier

Figure 9 illustrates the Efficient Frontier curve. The Efficient Frontier is a curve plotted on a Cartesian diagram that represents the combination of portfolio risk on the X-axis and portfolio return on the Y-axis. The results shown in Figure 9 demonstrate a simulation involving 5,000 portfolio combinations, generating an orange curve that represents the set of portfolios available to investors. These portfolios are considered efficient because they offer higher returns for the same level of risk. The red dot on the curve indicates the portfolio with the minimum level of risk. The red point corresponds to the previously obtained Minimum Variance Portfolio (MVP), with an expected return of approximately 8.14% and a portfolio risk of 10.8% and the yellow point is the tangency portfolio that has 35.27% for expected return with a corresponding risk level of 15.07%.

5. CONCLUSION

Based on the results and discussion of this study, it can be concluded that the application of K-Means clustering in portfolio diversification is effective in reducing potential risk. The clustering process resulted in six clusters, from which the selected stocks were JKON, BRIS, PGAS, TINS, TLKM, and JRPT. Two types of portfolios were constructed: the Minimum Variance Portfolio (MVP) and the Tangency Portfolio. Using the Markowitz method, the MVP—representing the portfolio with the lowest risk—was formed with the following asset allocations: JRPT (41.50%), TLKM (33.72%), TINS (12.37%), PGAS (4.86%), BRIS (4.37%), and JKON (3.18%). With this composition, the expected return and risk of the portfolio were 8.14% and 10.8%, respectively. Meanwhile, the Tangency Portfolio was composed with the following proportions: JRPT (46.17%), BRIS (38.79%), TLKM (5.92%), TINS (5.47%), PGAS (2.55%), and JKON (1.1%), yielding an expected return of 35.27% and a risk level of 15.07%. Future research could explore alternative clustering techniques and optimization methods to further enhance portfolio performance.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

REFERENCES

[1] Balqis, V.P., Subiyanto, S. & Supian, S., 2021. Optimizing stock portfolio with Markowitz method as a reference for investment community decisions. *International Journal of Research in Community Services*, Vol.2, No.2, 71–76. doi:10.46336/ijrcs.v2i2.213

[2] Birungi, C. & Muthoni, L., 2021. Analysis of risk measures in portfolio optimization for the Uganda Securities Exchange. *Journal of Financial Risk Management*, Vol.10, No.2 , 135–152. doi:10.4236/jfrm.2021.102008.

[3] Chaweewanchon, A. & Chaysiri, R., 2022. Markowitz mean-variance portfolio optimization with predictive stock selection using machine learning. *International Journal of Financial Studies*, Vol.10, No.3, 64. doi:10.3390/ijfs10030064.

[4] Hasan, N., Pelleng, F.A.O. & Mangindaan, J.V., 2019. Analisis Capital Asset Pricing Model (CAPM) sebagai dasar pengambilan keputusan berinvestasi saham (Studi pada Indeks Bisnis-27 di Bursa Efek Indonesia). *Jurnal Administrasi Bisnis*, Vol.8, No.1, 36–43. doi:10.35797/jab.8.1.2019.23498.

[5] Hasbiah, S., Anwar, A. & Bado, B., 2022. Model Markowitz dalam keputusan investasi saham pada Index LQ45 di Bursa Efek Indonesia. *Jurnal Ekonomi Pembangunan dan Pendidikan*, Vol.5, No.1, 69. doi:10.26858/jekpend.v5i1.24709

[6] Kulali, D.I., 2016. Portfolio optimization analysis with Markowitz quadratic mean-variance model. *European Journal of Business and Management*.

[7] Melta, F. & Murni, D., 2021. Pembentukan portofolio optimal model Markowitz menggunakan metode Sharpe (Studi kasus pada saham Jakarta Islamic Index), Vol. 6, No.3.
[8] Muthohiroh, U., Rahmawati, R. & Ispriyanti, D., 2021. Pendekatan metode Markowitz untuk optimalisasi portofolio dengan risiko expected shortfall (ES) pada saham syariah dilengkapi GUI MATLAB, Vol.10, No.3.
[9] Murphy, K.P., 2012. *Machine learning: A probabilistic perspective*. London: The MIT Press.
[10] Nainggolan, Y.T., Juliana, A. & Alantina, C.A., 2020. Analisis Value At Risk dalam pembentukan portofolio optimal (Studi kasus perusahaan perbankan di Indonesia). *Jurnal Manajemen*, Vol.10, No.2, 124–136. doi:10.30656/jm.v10i2.2433.
[11] Nanda, S.R., Mahanty, B. & Tiwari, M.K., 2010. Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*, Vol.37, No.12, 8793–8798. doi:10.1016/j.eswa.2010.06.026.
[12] Nisardi, M.R., Husain, H., Kusnaeni, K. and Resky, A., 2024. Penentuan portofolio saham optimal menggunakan metode Markowitz sebagai dasar keputusan investasi. *Square*, Vol.6, No.1, 33–40. doi.org/10.21580/square.2024.6.1.20441
[13] Nourahmadi, M. & Sadeqi, H., 2023. Portfolio diversification based on clustering analysis, Vol.7, No.3.
[14] Putri, V.Y., 2018. Effectiveness of portfolio selection using Markowitz model and broker recommendation in Indonesia Stock Exchange (Mining, agriculture, and infrastructure industry). *KnE Social Sciences*, Vol.3, No.11, 1603. doi:10.18502/kss.v3i11.2872.
15] Siregar, B. & Pangruruk, F.A., 2021. Portfolio optimization based on clustering of Indonesia Stock Exchange: A case study of index LQ45, Vol.1, No.1.
[16] Yunita, I., 2018. Markowitz model dalam pembentukan portofolio optimal (Studi kasus pada Jakarta Islamic Index). *Jurnal Manajemen Indonesia*, Vol.18, No.1. doi:10.25124/jmi.v18i1.1262.