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. 3, May 2026, pp. 671-685
DOI: 10.20956/j.v22i3.50039

Optimal Portfolio Formation Using a Combination of Genetic Algorithms and Particle Swarm Optimization Based on Cluster Analysis

Irfan N. Amiruddin*[1], Edy S. Rusdi[2]

*[1] *Actuarial Science Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Hasanuddin, Makassar, Indonesia*
[2] *Information Systems Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Hasanuddin, Makassar, Indonesia*
Email: *amiruddinin22h@student.unhas.ac.id[1], edy_saputra@sci.unhas.ac.id[2]*
**Corresponding author*
Received: 22 February 2026, revised: 10 April 2026, accepted: 15 April 2026

Abstract

The formation of an optimal portfolio is one of the strategies of investors in allocating their funds so as to minimize risk while maximizing profits. The optimal portfolio formation method has evolved over the years, ranging from simple calculation methods, to using complex optimization algorithms. This study aims to group LQ-45 stocks through a clustering algorithm, then determine the right weighting of representative shares of each cluster through the combination of GA-PSO so that an optimal portfolio is produced. The research begins with data pre-processing which includes transformation and reduction of data dimensions. The data from the dimension reduction is used to group stocks into clusters based on the best clustering algorithm. The stocks with the highest Sharpe ratio in each cluster are used to form the portfolio. The performance of the weighted portfolio using GA-PSO will be compared with the weighted portfolio using PSO. The results of the study showed that the K-Means algorithm became the clustering method with the highest silhouette score, which was 0.3614, and the optimal number of clusters produced was 6. Based on the results of the K-Means algorithm, the cluster representative stocks used for portfolio formation are ANTM, BRPT, EXCL, MDKA, MEDC, and PTBA. Furthermore, the results showed that the Sharpe ratio of stock portfolios using GA-PSO combined was greater than that of portfolios that used PSO alone for weighting. The K-Means algorithm is more suitable for grouping stocks than the DBSCAN and Agglomerative algorithms with Average Linkage. Furthermore, the combination of genetic algorithms and particle swarm optimization complements each other in weighting stocks so that it produces a portfolio with more optimal performance when compared to only using particle swarm optimization.
Keywords: GA-PSO Hybrid; Agglomerative Clustering; DBSCAN; K-Means; Stock Portfolio.

1. INTRODUCTION

An investment portfolio is defined as a series of combinations of several assets that are invested and held by investors, both individuals and institutions. Each asset in the capital market has its own characteristics. Therefore, every investor needs to create a portfolio that suits his preferences for return and risk levels. This condition makes it necessary for a systematic and quantitative framework to effectively manage the fundamental relationship between risk levels and returns [33].

Portfolio creation continues to evolve over time. The condition of asset returns that are not normally distributed generally results in the portfolio model proposed by Markowitz in 1952 being inappropriate for use. This is evidenced by the research of Athayde and Flôres as well as Jondeau and Rockinger, which explains that asset returns with highly skewed distributions will result in unsatisfactory portfolio weighting results [6,15]. Therefore, a number of methods using machine learning have been developed, which are grouped based on their function into supervised learning, unsupervised learning, and reinforcement learning [21]. The majority of raw data is large in size and unlabeled, so unsupervised learning methods are needed to analyze it [22].

Unsupervised learning is a machine learning method that analyzes unlabeled data and serves to identify patterns or hidden relationships in the data. One method of unsupervised learning is clustering, which is the grouping of data based on similarities so that a number of clusters are obtained in the data [8]. Rodriguez and his colleagues explain that clustering is expected to place each piece of data information into a single cluster [27]. In the case of stock data clustering, various clustering algorithms have been used. For example, grouping based on company performance ratios using HDBSCAN [12], or based on capital market valuations using K-Means [25]. This clustering method is also very helpful in grouping investment assets other than stocks, such as cryptocurrencies [5].

Clustering methods essentially only assist in grouping assets, but not in weighting them. Therefore, an optimization algorithm is also needed to do this. One group of optimization algorithms is meta-heuristic algorithms inspired by the way nature works (nature-based) [37]. Some metaheuristic algorithms inspired by nature are the grey wolf algorithm, bat algorithm, whale algorithm, ant lion colony algorithm, genetic algorithm, and particle swarm algorithm [37,30]. The advantage of this algorithm is its ability to combine one algorithm with another, creating a hybrid algorithm with better optimization performance [37].

Fajri Farid and Dedi Rosadi's research combine a clustering algorithm in the form of self-organizing maps (SOM) for stock selection and a genetic algorithm for stock weighting. The results of their research show that the portfolio created performs slightly better than the Markowitz portfolio resulting from clustering analysis [9]. Several previous studies have shown that metaheuristic optimization algorithms can be modified to produce more optimal asset portfolios than those created by conventional metaheuristic methods [5,1]. In addition, research by Kuo and Hong shows that combined metaheuristic algorithms produce better mutual fund portfolios than conventional metaheuristic methods [18].

This study adopts a two-layer approach: first, grouping stocks using the best clustering algorithm, and second, weighting representative stocks in each cluster using a combination of genetic algorithms and particle swarm optimization, then measuring the performance of the resulting portfolio.

2. MATERIALS AND METHODS

The type of research used is descriptive quantitative, which describes the performance of the optimal portfolio using the integration of the best clustering algorithm with a combination of genetic algorithms and particle swarm optimization to provide an overview of asset selection strategies and effective portfolio weighting. Therefore, the researcher used quantitative data obtained from the calculation of daily stock returns for the period from January 1, 2022, to December 31, 2024, which were consistently listed on the LQ-45 index during the period from

January 2022 to December 2024. In addition, the quantitative data used are the results of calculating the ROE, Debt-to-Equity, and Cash Ratio ratios based on the 2024 annual financial reports of companies consistently listed on the LQ-45 index during the period from January 2022 to December 2024. Finally, the risk-free rate used in calculating the Sharpe ratio is the BI 7-Days Repo Rate (BI7DRR) at the end of the research period, namely December 2024, which is 6%. For ease of data processing and analysis, the researcher used Microsoft Excel 2019 and R version 4.4.2 software.

2.1 Pre-Processing

2.1.1. Return Calculation

The calculation of return rates in this study focuses on the use of realized returns, which are calculated using the following formula [40]:

$$(R_i)_t = \frac{(P_i)_t - (P_i)_{t-1}}{(P_i)_{t-1}} \tag{2.1}$$

Expected return can be defined as a projection of future return rates. The expected return value can be used as a reference for investors in predicting the rate of return on a stock, but this prediction is not entirely accurate due to the potential for unexpected events or changes in market sentiment [13]. The value of the expected return on a stock over a period of T can be formulated as follows:

$$E[R]_i = \frac{1}{T}\sum_{t=1}^{T}(R_i)_t \tag{2.2}$$

One measure that can be used to measure stock risk is the standard deviation of stock returns. The standard deviation of stock returns is the square root of the variance of stock returns, which is formulated as follows:

$$\sigma_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}((R_i)_t - E[R]_i)^2} \tag{2.3}$$

2.1.2. Financial Ratio

a. Return-On-Equity

One of the most commonly used measures of a company's profitability is the ratio of its net income to its capital, or Return-On-Equity. This is because if the net income is less than the capital invested, investors will be hesitant to continue investing because it is considered unprofitable. The Return-On-Equity calculation can be formulated as follows [28]:

$$ROE = \frac{Net\ Income}{Shareholder's\ Equity} \tag{2.4}$$

b. Debt-to-Equity

The measure used to describe a company's ability to repay its debts in the future is known as the leverage ratio [26]. One of the most commonly used leverage ratios is the debt-to-equity ratio, which is the ratio of a company's debt to its total capital. The formula for the debt-to-equity ratio can be expressed as follows [4]:

$$Debt\ to\ Equity = \frac{Total\ Debt}{Shareholder's\ Equity} \tag{2.5}$$

c. Cash Ratio

The cash ratio is used to describe the ability to pay off short-term liabilities, especially during a crisis [26]. The higher a company's cash ratio, the easier it is for the company to liquidate its funds. This ratio is also often used by creditors who want to lend their funds to companies in the short term [28]. The value of the cash ratio can be obtained using the following formula [4]:

$$Cash\ Ratio = \frac{Cash}{Current\ Liabilities} \tag{2.6}$$

2.1.3. Data Transformation

a. Winsorizing

Winsorization is a method of handling outliers in data by changing extreme values in the data to reduce the effect of outliers [41]. Generally, this approach can be done by modifying extreme data into certain percentile values, specifically the 5th percentile and the 95th percentile. Thus, data classified as outliers are converted into more representative values, and the analysis of winsorized data can be more precise and interpretive [2].

b. Log and Yeo-Johnson Transformation

Log transformation is a technique for converting values in data by replacing them with the logarithm of that data. The base used for the logarithm can be Euler's number (e) or 10. Although Euler's number has the most enthusiasts, the base-10 logarithm is useful for converting data with values in the hundreds or thousands into new data with values in units [39]. In this study, Euler's number (e) is used as the base for the logarithmic transformation.

Data transformation techniques such as using logarithmic functions have several disadvantages. One of the main disadvantages is the inability to transform data with negative values. One form of modification of logarithmic transformation is the Yeo-Johnson transformation, which is defined as follows [43]:

$$G_\theta(x) = \begin{cases} \frac{(1+x)^\theta - 1}{\theta} & , if\ \theta \neq 0\ and\ x \geq 0 \\ \ln(x+1) & , if\ \theta = 0\ and\ x \geq 0 \\ \frac{-(-x+1)^{2-\theta} - 1}{2-\theta} & , if\ \theta \neq 2\ and\ x < 0 \\ -\ln(-x+1) & , if\ \theta = 2\ and\ x < 0 \end{cases} \tag{2.7}$$

c. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method that aims to reduce the dimensions of large data sets. This method transforms the original variables that are correlated with each other into a smaller set of variables, but still capable of covering most of the information contained in the original data. In general, PCA is used to analyze complex data structures and relationships between variables so that they are easier to understand [20].

2.2 Clustering Algorithm

2.2.1. K-Means

The K-Means algorithm is the most commonly used clustering method due to its high efficiency and applicability to almost all types of data [23]. Xu and Wunsch explain the steps of the K-means algorithm as follows [42]:

1. Initiate K cluster centroids randomly or based on the information provided.
2. Assign data points to the cluster with the nearest centroid. Measure the distance between data points and centroids using the specified distance metric, namely Euclidean.
3. Determine new centroids based on the previous clustering results using the following formula:

$$c_i = \frac{1}{N_{C_i}} \sum_{p_i \in C_i} p_i \tag{2.8}$$

4. Repeat steps two and three until there are no changes in the clusters formed. In other words, all new centroids are equal to the centroids from the previous iteration.

2.2.2. Agglomerative Clustering with Average Linkage

Agglomerative Hierarchical Clustering is a data clustering method that uses a bottom-up approach in forming clusters. The result of this algorithm is a dendrogram with each data point at the lowest level [44]. The distance between clusters can be measured using the average linkage approach, which is calculated using the following formula:

$$d_{(C_iC_j)C_k} = \frac{\sum_{p_i \in (C_iC_j)} \sum_{p_k \in C_k} d_{ik}}{N_{(C_iC_j)} N_{C_k}} \tag{2.9}$$

Johnson and Wichern describe the stages of agglomerative hierarchical clustering for grouping n data points as follows [14]:

1. Construct n clusters, each containing one data point, and a symmetric $n \times n$ matrix of distances (similarities) $\boldsymbol{D} = (d_{ij})$
2. Find the pair of clusters with the closest distance. Suppose the clusters in question are C_i and C_j, with their distance denoted by $d_{(C_iC_j)}$
3. Merge clusters C_i and C_j, then label the new cluster (C_iC_j). Update each entry in the distance matrix by: (a) deleting the rows and columns corresponding to clusters C_i and C_j; and (b) adding a row and column containing the distances between cluster (C_iC_j) and the remaining clusters.
4. Repeat steps 2 and 3 $n - 1$ times. Thus, all data points will be in one large cluster after the algorithm is complete. Assign an identity to each merged cluster and record the level at which the cluster merging occurred.

2.2.3. DBSCAN

The DBSCAN algorithm assumes clusters as the maximum set of densely connected points. Data points that are not included in clusters are considered noise [34]. The following are the general steps of the DBSCAN algorithm:

1. Determine the MinPts value, the minimum number of points required to form a cluster, and Eps, the distance used to define the cluster area

2. Calculate the distance between data points
3. Include data points with a distance less than or equal to Eps into the neighborhood being observed
4. Label each point as a core point if the number of neighborhoods is greater than or equal to MinPts. If there are data points that do not meet the core point criteria, then those points are categorized as noise
5. Include all data points that have a neighborhood relationship and are not noise into one cluster

2.2.4. Silhouette Score

The silhouette score is a metric that measures the quality of clusters in clustering analysis [29]. The silhouette score is the average of the silhouette coefficients of each data point. Kauffman and Rousseeuw define the silhouette coefficient at data point i, namely p_i, as follows [19]:

$$Silhouette\ Coefficient_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{2.10}$$

where a_i is the average distance of data point i to all other points within the cluster, and b_i is the smallest average distance of data point i to other clusters. The values of a_i and b_i can be calculated using the following formulas [19]:

$$a_i = \frac{\sum_{p_i, p_j \in C_i} d_{ij}}{N_{C_i}}; b_i = \min_{C_k \neq C_i} \frac{\sum_{p_i \in C_i, p_k \in C_k} d_{ik}}{N_{C_k}} \tag{2.11}$$

The silhouette score has a range of values from -1 to 1. A positive silhouette score that is closer to 1 indicates that the clusters formed are well defined. a score closer to 0 indicates overlapping clusters, and a silhouette score closer to -1 indicates that data points may have been grouped into the wrong clusters [12].

2.3 Portfolio Optimization using Metaheuristic Algorithm

2.3.1. Sharpe Ratio

The expected excess return and standard deviation used in the Sharpe ratio are actually based on the Markowitz portfolio model paradigm. This model assumes that the mean and standard deviation of historical return data are sufficient statistics for evaluating the performance of an investment portfolio. The Sharpe ratio can be used to rank the performance of investment portfolios. A higher ratio value indicates better portfolio performance, and vice versa. This is because the excess return that will be obtained in the future is much greater than the risk that must be borne [35].

$$Sharpe\ Ratio = \frac{E[R_a] - r_f}{\sigma_a} \tag{2.12}$$

where ER_a and σ_a are the annualized expected return and annualized standard deviation of return, respectively [3].

The Sharpe ratio will be used to determine representative stocks for each cluster. In addition, this ratio will be used as a fitness function for the metaheuristic algorithm. Furthermore, the Sharpe ratio is also used as a measure of the performance of the portfolio generated by the metaheuristic algorithm.

2.3.2. Genetic Algorithm (GA)

Genetic algorithms were first introduced by Holland in 1975 as a method for solving optimization problems, utilizing the concepts of genetics, natural selection, and the evolutionary process of living things [11]. Genetic algorithms are a global optimization method formed by simulating the evolutionary process of natural selection and the "survival of the fittest" population. Each possible solution is expressed as a "chromosome" so that a "group" consisting of these chromosomes is obtained. This group is limited to an environment with certain conditions. An objective function is used to evaluate the quality of each individual in that environment. Individuals who are more adaptive to the environment have a higher chance of survival. A number of individuals are randomly generated, then combined through crossbreeding based on the principle of "survival of the fittest" to produce offspring that are better than the previous generation. In this way, the "chromosome" population will continue to evolve gradually into better solutions. The incorporation of genetic operations such as gene mutation in the evolutionary process will improve the quality of offspring to be more adaptive to the given environmental conditions [36].

2.3.3. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) was first introduced by Kennedy and Eberhart in 1995 and is an optimization algorithm whose working method is inspired by the social behavior of animals that live in groups, such as birds and fish [17]. The behavior of these animals is translated into algorithmic operations to solve optimization problems by interpreting animal groups as a collection of particles, with each particle representing a candidate solution. The particle swarm explores the space in the given dimension and finds the best solution to the problem at hand [31].

The PSO algorithm begins by generating a set of random particles. Each particle has its own position and velocity and will remember its best position, which can be denoted as $(\boldsymbol{p}_{best}^{t})_i$, while the best position of other particles in the entire swarm is denoted as $\boldsymbol{g}_{best}^{t}$. Each particle will move in the search space towards the best position based on the fitness function value and towards the location of the best particle [17].

2.3.4. Hybrid GA-PSO

Genetic Algorithm-Particle Swarm Optimization (GA-PSO) is a hybrid metaheuristic algorithm that combines the workflows of genetic algorithms and particle swarm optimization. The core of this algorithm is to find optimal solutions using a set of particles that are viewed as a population, with each individual having its own chromosome. Some of the existing particles will be crossed and/or mutated at each iteration, resulting in a new population that will provide the best solution. The steps of this algorithm are as follows [18]:

1. Generate the initial population randomly. The number of populations or particles generated in this study is 40 individuals [24]. In the case of portfolio optimization, each particle has the following vector form:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}; w_i \in 0,1$$

2. Evaluate the performance of each particle using a fitness function, then sort them from largest to smallest based on their value as a form of selection. The fitness function used is adjusted to the optimal portfolio conditions—maximum return and minimum risk. The fitness function to be used for stock portfolio optimization is defined as follows:

$$Max(S_P) = Max\left(\frac{E[R_a]_P - r_f}{(\sigma_a)_P}\right) \tag{2.13}$$

3. Substitute half of the particles with the smallest fitness values with half of the particles with the largest fitness values, so that the top ½M is equal to the bottom ½M.

4. Save the top ½M particles and update the bottom ½M particles. Particle updates are performed using the following formula:

$$\boldsymbol{v}_i^{t+1} = \omega.\boldsymbol{v}_i^t + c_1.uniform[0,1].\left((\boldsymbol{p}_{best}^t)_i - \mathbf{x}_i^t\right) + c_2.uniform[0,1].\left(\boldsymbol{g}_{best}^t - \mathbf{x}_i^t\right) \tag{2.14}$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \boldsymbol{v}_i^{t+1} \tag{2.15}$$

The inertia weight (ω) used in this study was 0.8 [32], and the acceleration coefficients (c_1, c_2) were both 2 [38].

5. The update was performed using the PSO formula, followed by crossover, and ended with mutation. In this study, the crossover rate used was 60%, with the crossover formula defined as follows:

$$\mathbf{x}_i^{new} = \mathbf{x}_i^{old}.uniform0{,}1 + \mathbf{x}_{i+1}^{old}.(1 - uniform0{,}1), i = 1,2,\dots,\frac{1}{2}M - 1$$

$$\mathbf{x}_i^{new} = \mathbf{x}_i^{old}.uniform[0,1] + \mathbf{x}_1^{old}.(1 - uniform[0,1]), i = \frac{1}{2}M \tag{2.16}$$

Meanwhile, the mutation rate used is 20% with the formula defined as follows:

$$\mathbf{x}_i^{mut} = \mathbf{x}_i^{new} + uniform[0,1].\boldsymbol{N}(\mathbf{0},\mathbf{1}) \tag{2.17}$$

Where $\boldsymbol{N}(\mathbf{0},\mathbf{1})$ is a vector with random entries that are from $N(0,1)$.

6. Re-evaluate all particles using the fitness function

7. If it meets the criteria, the algorithm stops; if not, return to step 3.

3. MAIN RESULTS

The list of stocks used in this study can be found in Appendix A.

3.1. Transforming the Data

3.1.1. Handling the Outliers

The data to be used for the clustering process consists of five variables: average return, standard deviation of return, return-on-equity ratio, debt-to-equity ratio, and cash ratio. Each variable will be checked for outliers using boxplot visualization.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Irfan N. Amiruddin, Edy S. Rusdi

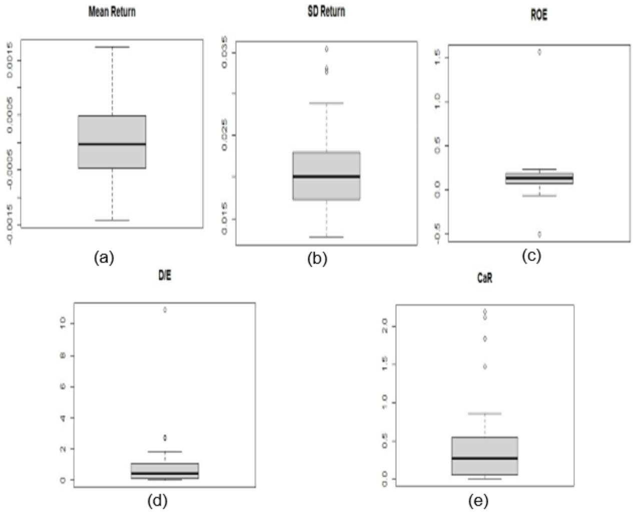

Figure 3.1. Boxplot on features: (a) Average return; (b) Standard deviation of return; (c) ROE; (d) Debt-to-equity ratio; (e) Cash ratio.

Figure 3.1 shows that only the average return variable does not have outliers. This means that the other four variables need to be transformed first. The standard deviation of return, debt-to-equity ratio, and cash ratio variables are each given winsorizing treatment, followed by logarithmic transformation with Euler's number as the base. Meanwhile, the return-on-equity ratio variable was winsorized, followed by a Yeo-Johnson transformation because it contained negative values.

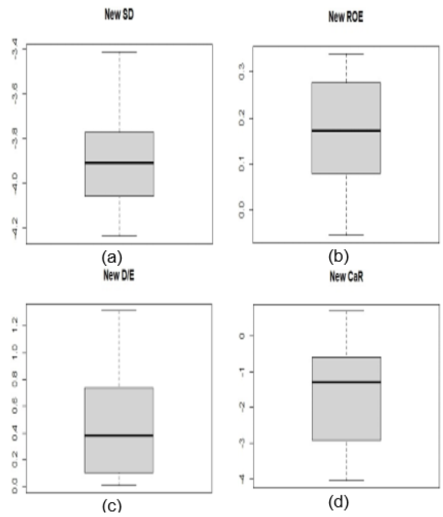

Figure 3.2. Boxplot of transformation results on features: (a) Standard deviation of return; (b) ROE; (c) Debt-to-equity ratio; (d) Cash ratio.

Figure 3.2 shows the results of the transformation of the four variables, each of which now has no outliers. This indicates that the transformed data, together with the average return variable, is ready to be used for the next step.

3.1.2. Dimensionality Reduction

The data consisting of average returns and the four other variables that have been transformed are then reduced in dimension using PCA. This is done to simplify data visualization and clustering processes [7].

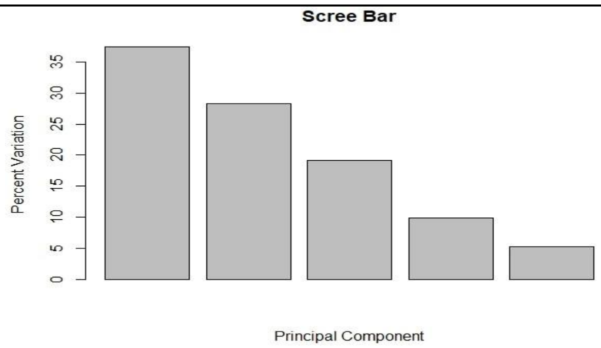

Figure 3.3. Scree bar for the five main components.

The results shown in Figure 3.3 indicate that the elbow method is not suitable for determining the sufficient number of principal components. Therefore, the researcher used the Kaiser-Guttman criterion or Kaiser's rule [10,16], which showed that only two principal components were sufficient for the next step.

3.2. Clustering Comparison

The data obtained from the principal component analysis then used for the clustering process. The clustering quality of the algorithms used is presented in Table 3.1.

Table 3.1. Clustering quality of several algorithms.

Clustering Algorithm	Optimal Number of Clusters	Silhouette Score
K-Means	6	0.3614
Agglomerative (with Average Linkage)	9	0.3260
DBSCAN	9	0.226

The results of grouping the data into six clusters obtained using K-Means are visualized in Figure 3.4.

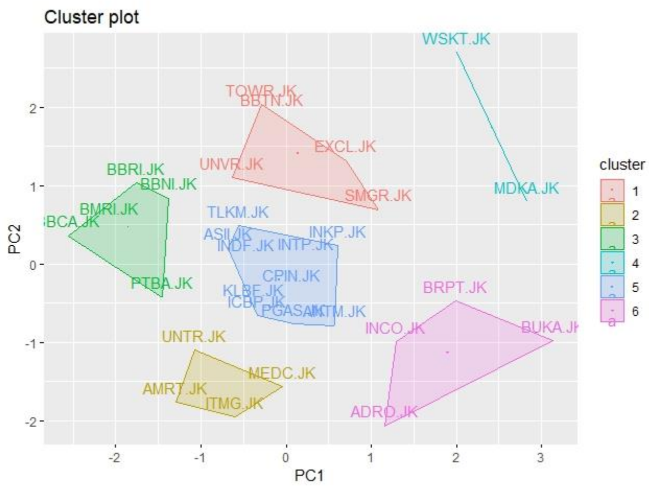

Figure 3.4. K-Means algorithm results' cluster plot.

By applying the Sharpe ratio to each stock in each cluster, ANTM.JK, BRPT.JK, EXCL.JK, MDKA.JK, MEDC.JK, and PTBA.JK were obtained as representatives of each cluster they belong to and will be used to form an optimal portfolio.

3.3. Comparing GA-PSO with PSO

The results of measuring the performance of the optimal portfolio formed using the six stocks obtained from the clustering process can be seen in Table 3.2.

Table 3.2. Performance of PSO and GA-PSO algorithm portfolios.

Optimization Algorithm	Annualized Expected Return	Annualized Standard Deviation of Return	Sharpe Ratio
PSO	0.359905	0.3823321	0.7844097
GA-PSO	0.5223849	0.5001662	0.9244624

The results presented in Table 3.2 show that the GA-PSO optimized portfolio yields an annual return of 52.23%, which is significantly higher than the PSO optimized portfolio. Indeed, the standard deviation of the annual return of the PSO-generated optimal portfolio is significantly smaller, but the ratio of annual return to annual risk—the Sharpe ratio—is still smaller than that of the GA-PSO portfolio. This shows that the GA-PSO algorithm works better in generating optimal portfolios than regular PSO.

4. DISCUSSION AND CONCLUSIONS

Each cluster formed using the three clustering algorithms has different members. However, banking stocks such as BBCA, BBNI, BBRI, and BMRI are always placed in one cluster. AMRT shares, mining-focused stocks such as ITMG and UNTR, and energy sector stocks such as MEDC are also always in the same cluster in all three clustering algorithms.

The K-Means method, agglomerative with average linkage, and DBSCAN all produce stock groups with fewer than 10 clusters. The results of the clustering quality comparison show that the K-Means method works better in grouping stock data. The silhouette scores of the agglomerative method with average linkage are not much different from those of the K-Means method because some of the clusters formed have the same or nearly the same members as the clusters produced by K-Means. For example, the cluster containing BBCA, BBNI, BBRI, BMRI, and PTBA; the cluster containing AMRT, ITMG, MEDC, and UNTR; and the cluster containing ADRO, ANTM, BRPT, and INCO. The DBSCAN method has the lowest silhouette score because the stock data is quite scattered, so it is not grouped well.

The selected stocks have different weights in the optimal portfolio. Based on the PSO and GA-PSO methods, MEDC shares have the largest weight compared to the other five shares, with a weight value above 0.5 or 50%. Next, BRPT, PTBA, and EXCL shares are ranked second, third, and fourth, respectively, in terms of weight. Finally, MDKA and ANTM shares have the smallest weight values. The optimal portfolios resulting from the GA-PSO and conventional PSO methods both produce positive expected returns and Sharpe ratios, but there are significant differences between them. A comparison of the performance of the two portfolios shows that the combined metaheuristic method works better in handling optimization problems than the conventional metaheuristic method.

Appendix A. Subset of Assets Used

Table A1. List of subset of assets considered.

30 LQ-45 Stocks	30 LQ-45 Stocks	30 LQ-45 Stocks
Adaro Energi Tbk. (ADRO.JK)	Bukalapak.com Tbk. (BUKA.JK)	Merdeka Copper Gold Tbk. (MDKA.JK)
Sumber Alfaria Trijaya Tbk. (AMRT.JK)	Charoen Pokphand Indonesia Tbk. (CPIN.JK)	Medco Energi Internasional Tbk. (MEDC.JK)
Aneka Tambang Tbk. (ANTM.JK)	XL Axiata Tbk. (EXCL.JK)	Perusahaan Gas Negara Tbk. (PGAS.JK)
Astra Internasional Tbk. (ASII.JK)	Indofood CBP Sukses Makmur Tbk. (ICBP.JK)	Bukit Asam Tbk. (PTBA.JK)
Bank Central Asia Tbk. (BBCA.JK)	Vale Indonesia Tbk. (INCO.JK)	Semen Indonesia Tbk. (SMGR.JK)
Bank Negara Indonesia Tbk. (BBNI.JK)	Indofood Sukses Makmur Tbk. (INDF.JK)	Telkom Indonesia Tbk. (TLKM.JK)
Bank Rakyat Indonesia Tbk. (BBRI.JK)	Indah Kiat Pulp & Paper Tbk. (INKP.JK)	Sarana Menara Nusantara Tbk. (TOWR.JK)
Bank Tabungan Negara Tbk. (BBTN.JK)	Indocement Tunggal Prakarsa Tbk. (INTP.JK)	United Tractors Tbk. (UNTR.JK)
Bank Mandiri Tbk. (BMRI.JK)	Indo Tambangraya Megah Tbk. (ITMG.JK)	Unilever Indonesia Tbk. (UNVR.JK)
Barito Pacific Tbk. (BRPT.JK)	Kalbe Farma Tbk. (KLBF.JK)	Waskita Karya Tbk. (WSKT.JK)

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

REFERENCES

[1] Anam, S., Bukhori, H.A., Maulana, A., Maulana, M.I., Rasikhun, H., 2025. An Enhanced Particle Swarm Optimization with Mutation for Mean-Value-at-Risk Portfolio Optimization in the Indonesian Banking Sector. *Jurnal Teknik Informatika (Jutif)*, Vol. 6, No. 4, 2648–2666.

[2] Ardhani, N.T., Notodiputro, K.A., Oktarina, S.D., 2025. Winsorization for Outliers in Clustering Non-Cyclical Stocks with K-Means and K-Medoids: Winsorization untuk Penanganan Pencilan dalam Penggerombolan Saham Sektor Consumer Non-Cyclical dengan K-Means dan K-Medoids. *Indonesian Journal of Statistics and Its Applications*, Vol. 9, No. 1, 46–60.

[3] Bacon, C.R., 2022. *Practical risk-adjusted performance measurement*, Second Edition. Wiley, Hoboken, NJ.

[4] Berk, J., DeMarzo, P., 2023. *Corporate finance*, Sixth Edition. Pearson, New York.

[5] Bulani, V., Bezbradica, M., Crane, M., 2025. Improving Portfolio Management Using Clustering and Particle Swarm Optimisation. *Mathematics*, Vol. 13, No. 10, 1623.

[6] De Athayde, G.M., Flôres, R.G., 2004. Finding a maximum skewness portfolio—a general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control*, Vol. 28, No. 7, 1335–1352.

[7] Ding, C., He, X., 2004. *K*-means clustering via principal component analysis. *Twenty-first international conference on Machine learning - ICML '04*, 29.

[8] Ebrahimi, P., Asjodi, A.H., Dolatshahi, K.M., 2025. Quantifying hybrid failure mode in cyclic-loaded reinforced concrete shear walls: Integrating unsupervised and supervised learning techniques. *Engineering Applications of Artificial Intelligence*, Vol. 162, 112346.

[9] Farid, F., Rosadi, D., 2022. Portfolio optimization based on self-organizing maps clustering and genetics algorithm. *International Journal of Advances in Intelligent Informatics*, Vol. 8, No. 1, 33.

[10] Guttman, L., 1954. Some Necessary Conditions for Common-Factor Analysis. *Psychometrika*, Vol. 19, No. 2, 149–161.

[11] Holland, J.H., 1975. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, University of Michigan Press, Ann Arbor.

[12] Jearanaitanakij, K., Poonpon, N., Wongtep, C., Buriyameathakul, K., Pimsupaporn, A., 2025. Stock Clustering Framework using Financial Ratios: A Case Study in the Stock Exchange of Thailand. *Engineering and Technology Horizons*, Vol. 42, No. 4,

[13] Jogiyanto, H., 2017. *Teori Portofolio dan Analisis Investasi*, Eleventh Edition. BPFE, Yogyakarta, Indonesia.

[14] Johnson, R.A., Wichern, D.W., 2019. *Applied multivariate statistical analysis*, Sixth Edition [Pearson modern classic edition]. Pearson, Upper Saddle River, New Jersey.

[15] Jondeau, E., Rockinger, M., 2006. Optimal Portfolio Allocation under Higher Moments. *European Financial Management*, Vol. 12, No. 1, 29–55.

[16] Kaiser, H.F., 1961. A Note On Guttman's Lower Bound For The Number Of Common Factors. *British Journal of Statistical Psychology*, Vol. 14, No. 1, 1–2.

[17] Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, 1942–1948.

[18] Kuo, R.J., Hong, C.W., 2013. Integration of Genetic Algorithm and Particle Swarm Optimization for Investment Portfolio Optimization. *Applied Mathematics & Information Sciences*, Vol. 7, No. 6, 2397–2408.

[19] Leonard K., Peter J. R., 2005. *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., New Jersey, USA.

[20] Lesnussa, T.P., Utubira, E.E.M., Kaseside, M., 2025. Penerapan Metode Principal Component Analysis (Studi Kasus: Tingkat Kemiskinan di Kepulauan Maluku). *Bilangan : Jurnal Ilmiah Matematika, Kebumian dan Angkasa*, Vol. 3, No. 2, 172–181.

[21] Morales, E.F., Escalante, H.J., 2022. A brief introduction to supervised, unsupervised, and reinforcement learning.

[22] Naeem, S., Ali, A., Anam, S., Ahmed, M.M., 2023. An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, Vol. 13, No. 1, 911–921.

[23] Palupi, I., Wahyudi, B.A., Indwiarti, I., 2019. The Clustering Algorithms Approach for Decision Efficiency in Investment Portfolio Diversification. *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 1–6.

[24] Piotrowski, A.P., Napiorkowski, J.J., Piotrowska, A.E., 2020. Population size in Particle Swarm Optimization. *Swarm and Evolutionary Computation*, Vol. 58, 100718.

[25] Pratama, Y., Sulistianingsih, E., Debataraja, N.N., Imro'ah, N., 2024. K-Means Clustering dan Mean Variance Efficient Portfolio dalam Portofolio Saham. *Jambura Journal of Probability and Statistics*, Vol. 5, No. 1, 24–30.

[26] Robinson, T.R., 2015. *International financial statement analysis*, Third Edition. John Wiley & Sons, Hoboken, N.J.

[27] Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.D.F., Rodrigues, F.A., 2019. Clustering algorithms: A comparative approach. *PLOS ONE*, Vol. 14, No. 1, e0210236.

[28] Ross, S.A., Westerfield, R., Jaffe, J.F., Jordan, B.D., 2022. *Corporate finance*, Thirteenth Edition, international student edition Edition. MCGraw Hill, New-York NY.

[29] Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, Vol. 20, 53–65.

[30] Setiawan, E.P., Rosadi, D., 2019. Model Pengoptimuman Portofolio Mean-Variance dan Perkembangan Praktisnya. *Jurnal Optimasi Sistem Industri*, Vol. 18, No. 1, 8–16.

[31] Shami, T.M., El-Saleh, A.A., Alswaitti, M., Al-Tashi, Q., Summakieh, M.A., Mirjalili, S., 2022. Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, Vol. 10, 10031–10061.

[32] Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, 69–73.

[33] Sunariyah, 2011. *Pengantar Pengetahuan Pasar Modal*, Sixth Edition. UPP STIM YKPN, Yogyakarta, Indonesia.

[34] Sutramiani, N.P., Arthana, I.M.T., Lampung, P.F., Aurelia, S., Fauzi, M., Darma, I.W.A.S., 2024. The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover. *Journal of Information Systems Engineering and Business Intelligence*, Vol. 10, No. 1, 13–24.

[35] Tandelilin, E., 2010. *Portofolio dan Investasi: Teori dan Aplikasi*, PT Kanisius, Sleman, Indonesia.

[36] Tang, C., Wu, H., 2022. Joint optimization of task caching and computation offloading in vehicular edge computing. *Peer-to-Peer Networking and Applications*, Vol. 15, No. 2, 854–869.

[37] Tomar, V., Bansal, M., Singh, P., 2024. Metaheuristic Algorithms for Optimization: A Brief Review. *RAiSE-2023*, 238.

[38] Wang, D., Tan, D., Liu, L., 2018. Particle swarm optimization algorithm: an overview. *Soft Computing*, Vol. 22, No. 2, 387–408.

[39] West, R.M., 2022. Best practice in statistics: The use of log transformation. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, Vol. 59, No. 3, 162–165.

[40] Widyaningrum, E.N., Hariyanto, H., Wahyudi, S., 2022. Model Optimasi Portofolio Saham dengan Aset Bebas Risiko dan Estimasi Risiko dengan Menggunakan Expected Shortfall. *Jurnal Sains dan Seni ITS*, Vol. 11, No. 2, A70–A76.

[41] Wilcox, R.R., 2021. *Introduction to robust estimation and hypothesis testing*, Fifth Edition. Elsevier, Inc, Philadelphia.

[42] Xu, R., Wunsch, D.C., 2009. *Clustering*, Wiley ; IEEE Press, Hoboken, N.J. : Piscataway, NJ.

[43] Yeo, I.-K., 2000. A new family of power transformations to improve normality or symmetry. *Biometrika*, Vol. 87, No. 4, 954–959.

[44] Zainudin, N.S., Ting, C.-Y., Khor, K.-C., Ng, K.-H., Tong, G.-K., Kalid, S.N., 2024. Clustering Defensive Shariah-compliant Stocks Using Financial Performance as the Indicator. *JOIV : International Journal on Informatics Visualization*, Vol. 8, No. 1, 115.