Clustering Performance Analysis in Portfolio Optimization based on the Swarm Intelligence Algorithm

Abdurakhman*, *Member, IAENG,* Mufrih Nur Huda Tri Putra, and Evita Purnaningrum

Abstract—**Portfolio optimization often faces computational challenges due to the complexity of risk minimization and the increasing number of constraints. Swarm intelligence (SI) algorithms, inspired by nature, have emerged as a promising approach to address this NP-hard problem. This study investigates the integration of cluster analysis (k-means++ and k-medoids) in stock selection for portfolio optimization, aiming to enhance the performance of SI methods. We employ three diverse SI algorithms—Firefly Algorithm (FA), Grey Wolf Optimization (GWO), and Whale Optimization Algorithm (WOA)—representing air, land, and sea domains. Our empirical analysis demonstrates that the performance of these SI models, when combined with clustering pre-selection, is comparable to the traditional mean variance (MV) model in terms of returns. Notably, the combination of k-means++ clustering with GWO achieves the highest return (1.55094) and Sharpe ratio (3.40573). Additionally, in short-term market scenarios (7 days), the k-medoids clustering strategy coupled with SI algorithms minimizes losses, with an average loss of -0.23%. These findings suggest that integrating clustering techniques with SI algorithms offers a promising avenue for optimizing portfolio performance, particularly in long-term investment horizons.**

Index Terms—**Portfolio Optimization, Swarm Intelligence, Clustering, Grey Wolf Optimization (GWO), k-medoids.**

I. Introduction

PORTFOLIO optimization, the intricate process of selecting assets to balance risk and return, remains a central challenge in financial decision-making [1], [2]. The dynamic nature of markets and the increasing complexity of investment instruments necessitate innovative approaches to address this NP-hard problem. While traditional models like Markowitz's mean-variance approach offer a foundational framework, their reliance on parameter estimation and asset pre-selection can limit their effectiveness in real-world scenarios [3], [4], [5]. The integration of machine learning techniques, particularly clustering, has shown promise in enhancing portfolio optimization by revealing underlying price dynamics and enabling more informed asset selection [6], [7], [8]. However, the selection of optimal assets and their allocation within a portfolio remains a complex optimization task.

Manuscript received May 8, 2024; revised November 5, 2024.

Abdurakhman is an associate professor of Mathematics Department, Universitas Gadjah Mada, Indonesia. (corresponding author; e-mail: rachmanstat@ugm.ac.id).

Mufrih Nur Huda Tri Putra is an undergraduate student of Mathematics Department, Universitas Gadjah Mada, Indonesia. (e-mail: mufrih.n.h@mail.ugm.ac.id).

Evita Purnaningrum is a PhD candidate of Mathematics Department, Universitas Gadjah Mada, Indonesia. (e-mail: evitapurnaningrum0890@mail.ugm.ac.id).

To address this challenge, Swarm Intelligence (SI) algorithms, inspired by the self-organizing behavior of natural systems, have emerged as powerful tools. Unlike traditional evolutionary approaches, SI algorithms mimic the collective intelligence observed in nature, such as bird flocks or ant colonies, to effectively navigate the complex search space of portfolio optimization [9]. While prior research has explored the application of various SI algorithms to portfolio optimization, the potential of strategically combining diverse SI models, representing different natural domains (air, land, and sea), with clustering techniques remains largely untapped. This integration holds significant promise for enhancing the adaptability and robustness of portfolio optimization strategies, particularly in dynamic market environments.

This study makes a significant contribution by systematically investigating the synergy between cluster analysis (k-means++ and k-medoids) and three distinct SI algorithms (Firefly Algorithm (FA), Grey Wolf Optimization (GWO), and Whale Optimization Algorithm (WOA)) in the context of portfolio optimization. The novelty of our approach lies in the combination of diverse SI models with clustering-based asset pre-selection, aiming to enhance the identification of optimal portfolios. By evaluating the performance of these combinations against the traditional mean-variance model, we provide valuable insights into the potential benefits and limitations of integrating clustering and SI for portfolio optimization.

The significance of this research extends beyond algorithmic comparisons. We aim to address the limitations of existing portfolio optimization methods by incorporating a data-driven asset pre-selection step, potentially improving the robustness and adaptability of SI-based approaches. Furthermore, our findings have practical implications for investors and portfolio managers, offering insights into the selection of appropriate SI models and clustering techniques for different market scenarios and investment horizons. By demonstrating the potential of this integrated approach, we contribute to the development of more effective and resilient portfolio optimization strategies in an increasingly complex financial landscape. In addition, clustering analysis will be utilized for selecting the appropriate stocks before optimization. This paper is organized as follows: Section 2 presents a literature review related to portfolio optimization using IS. Section 3 describes the applied methodology by showing the algorithm and workflow. Section 4 shows the results and discussion regarding portfolio optimization. Finally, Section 5 draws relevant conclusions.

II. Literature Review

A. Clustering Analysis

Cluster analysis allows searching for correlations between financial time series data. Tolun [6] developed a cluster algorithm based on Euclidean distance which has superior performance compared to other cluster algorithms based on different datasets, back testing cycles, and time. Bnouachir and Mkhadri [10] used hierarchical clustering to estimate errors in stability and reduce the associated risks. Bjerring et al. [11] obtained a fixed number of clusters before running the portfolio optimization model. It can be concluded that integrating clustering methods into dimensionality reduction tools or preprocessing of portfolio models can improve the performance of optimization models. This research uses two types of cluster methods at the portfolio pre-selection stage, namely K-means++ [12] and K-medoids [13].

1) K-means++: The K-Means++ method is an extension of the K-Means algorithm that spreads cluster centers while iteratively selecting them with the assumption that data points that are far apart tend to come from different clusters by defining $D(x)$ as the object point's closest distance to the nearest center point or centroid that has been determined. The centroid is calculated by selecting $x \in x_j$ with probability.

$$\frac{D(x)^2}{\Sigma_{x \in x_j} D(x)^2} \tag{1}$$

Then carry out the K-means algorithm which is generally defined as minimizing the objective function with the following equation:

$$SSE(C) = \sum_{k=1}^{K} \sum_{x_j \in C_k} \|x_j - C_k\|^2 \tag{2}$$

C_k is the number of clusters formed from the K-means method, x_j is the j-th variable.

2) K-medoids: The K-medoids algorithm itself is similar to K-means in that it minimizes a predetermined objective function. However, K-medoids minimizes the absolute error criterion compared to the sum of squared errors (SSE) used in K-means. This means that K-medoids focuses on minimizing the distance between data points and their respective medoids.

$$S = \sum_{L=1}^{L} \sum_{x_j \in C_l} |x_j - med_{lj}| \tag{3}$$

C_l is the number of clusters formed from the K-medoids method.

B. Portfolio Optimization

A portfolio is a collection of investment assets in the form of property, shares, gold, bonds, deposits, options and other investment instruments in order to minimize risk. Portfolio formation is intended to diversify investment capital so that it has as little risk as possible. Therefore, when an investor wants to form a portfolio, he must know the characteristics of the assets and be able to calculate the possibility of loss from these assets so that he can determine the optimum weight for each asset that can minimize the risk of loss. Selecting the optimum portfolio or portfolio optimization can be different for each investor depending on their individual preferences, whether risk seeker, risk averse or risk neutrality.

C. Swarm Intelligence (SI) on portfolio optimization

Swarm Intelligence (SI), which is based on population metaheuristics, is an interesting field of artificial intelligence [14]. This term was introduced and coined by Jing Wang and Gerardo Beni in 1989 while they were working on a mobile robotic system. SI-based algorithms typically mimic the behavior of population-based agents and are inspired by nature, especially biological systems. In their environment, these agents interact with each other. However, there is no central control system that controls how agents behave locally. The unpredictability of agents' behavior leads to the disclosure of intelligence that each agent is not aware of. Microbial intelligence and bacterial growth, flocks of birds, ant colonies, animal herding, eagle hunting, and schools of fish are some of the well-known inspirations. Several SI algorithms are used to complete various variants of portfolio optimization. Some of them [9] are particle swarm optimization (PSO), ant colony optimization (ACO), bacterial foraging optimization (BFO), artificial bee colony (ABC), cat swarm optimization (CSO), firefly algorithm (FA), invasive weed optimization (IWO), bat algorithm (BA) and fireworks algorithm (FA). This research uses a type of SI method that comes from air or can fly FA [15], land GWO [16], and sea WOA [17].

1) Firefly Algorithm (FA): One of the SI algorithms used for optimization is the Firefly Algorithm (FA). FA was first introduced in 2007 by Yang et al. This algorithm is derived from the natural behavior of fireflies which relies on the bioluminescence phenomenon. Moreover, FA is also a kind of stochastic, nature-inspired meta-heuristic algorithm, which can be applied to solve the most difficult optimization problems (also NP-hard problems) [18]. Heuristic means 'discover' or 'find a solution by trial and error' [19]. In fact, there is no guarantee that an optimal solution will be found within a reasonable timeframe. Finally, meta-heuristic means 'higher level', where the search process used in the algorithm is influenced by certain trade-offs between randomization and local search [20]. The most important thing in FA design is Light Intensity and Attractiveness Function. Attractiveness is influenced by the level of light intensity. In the simple case related to maximum optimization, if y is a firefly then $I(y) = f(y)$, where I is the level of light intensity proportional to the solution of the objective function $f(y)$ in the optimization problem. Meanwhile, the β firefly attractiveness function has the equation:

$$\beta = \beta_0 e^{-\gamma r^2} \tag{4}$$

Where r is the distance between two fireflies i and j at y_i and y_j respectively which is the Cartesian distance. $r_{i,j} = \|y_i - y_j\| = \sqrt{\Sigma_{k=1}^{n} (y_{ik} - y_{jk})^2}$. So the movement of fireflies that are attracted to the brighter ones is determined by equation [18]:

$$y_i = y_j + \beta_0 e^{-\gamma r_{i,j}^2} (y_i - y_j) + \alpha \epsilon_i \tag{5}$$

Where ϵ_i is a random number taken from a Gaussian distribution.

2) Grey Wolf Optimization (GWO): The Grey Wolf Optimization (GWO) algorithm was developed based on social hierarchy and cooperation in wolf packs and is inspired by grey wolves' social behavior when foraging for prey

[21]. Wolf groups are assumed to be omega, delta, beta, and alpha in a pyramid order. This hierarchy contains distinct roles and functions inside the group. The method of systematically pursuing organized groups of grey wolves serves as the foundation for developing a GWO algorithm, with alpha, beta, and delta indicating the best solution and omega representing the other solutions. The mathematical model for GWO identifies the best-fit solution as alpha (α), the second and third best solutions as beta (β) and delta (δ), and the other potential solutions as omega (ω) [22]. In other words, the optimal solution is determined based on adjusting the strategy and dynamics of the wolf pack. Social ranking and hunting tactics motivate mathematical model optimization techniques. Grey wolves will surround their prey in an ambush (prey-encircling), which can be mathematically expressed as follows [23]:

$$d = |\vec{c}\vec{w_p}(t) - \vec{w}(t)| \tag{6}$$

$$\vec{w}(t+1) = \vec{w_p}(t) - \vec{A}\vec{d} \tag{7}$$

where $\vec{A}$ and $\vec{c}$ are vector coefficients, t indicates the current iteration, $\vec{w_p}$ is the position of the prey in vector form, and $\vec{w}$ indicates the position of the grey wolf in the vector around the prey. Vectors $\vec{A}$ and $\vec{c}$ can be calculated using the following equation.

$$\vec{A} = 2\vec{a}\vec{r_1} - \vec{a} \tag{8}$$

$$\vec{c} = 2\vec{r_2} \tag{9}$$

$\vec{a}$ is decreased linearly from 2 to 0 during iteration, while $\vec{r_1}$ and $\vec{r_2}$ are random vectors between $[0, 1]$. While the hunting mode mathematical model is structured based on the assumption that the top three agents have good knowledge of potential prey locations, the other wolves (i.e. omega wolves) are required to update their locations based on the positions of the alpha wolf $\vec{w_\alpha}(t)$, beta wolf $\vec{w_\beta}(t)$, and delta wolf $\vec{w_\delta}(t)$ using the position update equation below.

$$\vec{w}(t+1) = \frac{\vec{w_1}(t) + \vec{w_2}(t) + \vec{w_3}(t)}{3} \tag{10}$$

where the calculation of $\vec{w_1}(t), \vec{w_2}(t), \vec{w_3}(t)$ follows the equation

$$\begin{cases} \vec{w_1}(t) = \vec{w_\alpha}(t) - \vec{a_1}(d_\alpha) \\ \vec{w_2}(t) = \vec{w_\beta}(t) - \vec{a_2}(d_\beta) \\ \vec{w_3}(t) = \vec{w_\delta}(t) - \vec{a_3}(d_\delta) \end{cases} \tag{11}$$

and

$$d_\alpha = |\vec{c_1}\vec{w_\alpha}(t) - \vec{w}(t)| \tag{12}$$

$$d_\beta = |\vec{c_1}\vec{w_\beta}(t) - \vec{w}(t)| \tag{13}$$

$$d_\delta = |\vec{c_1}\vec{w_\delta}(t) - \vec{w}(t)| \tag{14}$$

3) Whale Optimization Algorithms (WOA): The Whale Optimization (WOA) algorithm is an optimization algorithm inspired by the hunting behavior of humpback whales [24]. This algorithm was inspired by a hunting method used by humpback whales called bubble-net feeding. The bubble-net feeding method used by humpback whales involves two main movements, namely upward-spirals and double loops. The double loops process consists of coral loop, lobtail, and capture loop which are used as the basis for mathematical modeling as an optimization process. WOA assumes that the target prey location is the best solution. Identical to the GWO algorithm, the WOA phase circling the prey can be expressed as follows [23]:

$$d = |\vec{c}\vec{w_p}^*(t) - \vec{w}(t)| \tag{15}$$

$$\vec{w}(t+1) = \vec{w_p}^*(t) - \vec{A}\vec{d} \tag{16}$$

$\vec{w_p}^*$ is the temporary best solution vector, and $\vec{w}$ is the position vector. Vectors $\vec{A}$ and $\vec{c}$ can be calculated using the following equation.

$$\vec{A} = 2\vec{a}\vec{r} - \vec{a} \tag{17}$$

$$\vec{c} = 2\vec{r} \tag{18}$$

where $\vec{r}$ is random vectors between $[0, 1]$, while $\vec{a}$ is decreased linearly from 2 to 0 during iteration. Afterward, the prey attack acknowledged as the Bubble-Net Attack consists of shrinking encircling and spiral updating position. The probability that a humpback whale swims around its prey in decreasing circles and along a spiral path is 50%. The mathematical model is as follows:

$$\vec{w}(t+1) = \begin{cases} \vec{w}^*(t) - \vec{A}\vec{d} & \text{if } p < 0.5 \\ (\vec{w}^*(t) - \vec{w}(t))e^{bl}\cos 2\pi l + \vec{w}^*(t) & \text{otherwise} \end{cases} \tag{19}$$

where p is a random number between $[0, 1]$, b is a constant to describe logarithmic spiral, otherwise l is random number betwen $[-1, 1]$. The similar tactics based on vector A variations might be used to look for other potential prey throughout the exploration phase of a global search. The mathematical model is as follows.

$$d = |\vec{c}\vec{w_{rand}} - \vec{w}| \tag{20}$$

$$\vec{w}(t+1) = \vec{w_{rand}} - \vec{A}\vec{d} \tag{21}$$

where $\vec{w_{rand}}$is a vector of random positions or whales selected at random from the existing whale population.

III. Proposed methodology

A. Approach

The approach employed in this research consists of four main elements. The experimental data to be assessed using the SI models is identified in the first phase. Data pre-processing with standardization is also executed . In the second stage, mean and standard deviation are calculated to identify portfolio pre-selection by clustering analysis, such as K-means++ and K-medoids. The Silhouette score determines the number of groups in each approach. Following that, representative shares from each cluster are found using Sharpe Ratio profiling. The next phase is optimizing each SI method to produce representative share weights. Finally, in the fourth stage, the findings from the various models are compared and assessed. The main measures for comparison include return, standard deviation, and Sharpe ratio and simulation of portfolio performance on the Stock Market. These four steps are explained in more detail in the flow diagram Fig.1.

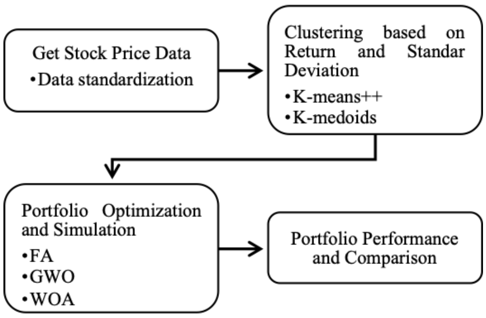


Fig. 1: An approach to comparing Swarm Intelligence models (FA, GWO, WOA)

B. Data Standardization

Data standardization is implemented while the pre-processing step to equalize the scale of variables, making comparisons easier and preventing big-scale variables from having more of an effect on the analysis or model. This method, also known as normalization, involves decreasing the variable by the mean and dividing it by the standard deviation to ensure that the variable values range from 0 to 1. The data standardization equation can be expressed as follows.

$$z = \frac{x - \hat{x}}{s} \tag{22}$$

where x is the data value, $\hat{x}$ is the average value, and s is the data's standard deviation. This standardization or normalization method is not affecting the distribution structure of the existing data. Standardization is unable to transform data which was not previously normally distributed [25].

C. Experiment Data

This experiment analyzes closing stock prices to evaluate the algorithm. The observation period was from January 1, 2023, to January 3, 2024. The 45 shares utilized in this study are from the LQ-45 index list for the period August 2023 to January 2024. Cluster analysis examines returns and standard deviations for each stock over a given time. Portfolio performance is also calculated using a reference interest rate, specifically the Bank Indonesia (BI) 7-Days Reverse Repo Rate, which is 6% per year as of January 2024.

IV. Experimental Results

A. Determination of Optimal Clusters and Cluster Profiling.

The optimal number of clusters for k-means++ and k-medoids was determined using the silhouette score as the evaluation metric. The silhouette analysis (Figure 2) revealed that the optimal number of clusters for k-means++ is two, with a silhouette value of 0.684, and for k-medoids, it's three, with a silhouette value of 0.538. The higher silhouette score for k-means++ suggests that it forms more distinct and well-separated clusters compared to k-medoids. However, the presence of an additional cluster in k-medoids indicates its potential to capture more subtle patterns or subgroups within the data, which could be advantageous in achieving a more diversified portfolio.

TABLE I: Cluster Profiling Based on Return and Standard Deviation

Methods	C	frequency	return	s
K-means	2	5	0.00076	0.04171
		40	-0.00029	0.02034
K-medoids	3	17	-0.00038	0.02425
		5	0.00325	0.04006
		23	-0.00041	0.01670

C : Number of Clusters
s : standard deviation

The optimal number of clusters identified in the silhouette analysis was used to cluster 45 stocks based on their return and standard deviation using k-means++ and k-medoids. The characteristics of the resulting clusters were then analyzed by examining their centroids, which represent the average return and standard deviation of the stocks within each cluster. The centroid values for each cluster, as presented in Table I, provide insights into the risk-return profiles of the clusters.

The cluster profiles, summarized in Table I, reveal distinct risk-return characteristics for both k-means++ and k-medoids clustering methods. In the case of k-means++, two clusters were identified. Cluster 1 exhibits a high-risk, high-reward profile, characterized by a higher average return (0.00076) but also accompanied by a higher standard deviation (0.04171). In contrast, Cluster 2 displays a low-risk, low-reward profile with a lower average return (-0.00029) and a lower standard deviation (0.02034).

The k-medoids approach, on the other hand, results in three clusters. Cluster 2 stands out with the highest average return (0.00325) and a relatively high standard deviation (0.04006), mirroring the high-risk, high-reward profile observed in k-means++'s Cluster 1. The remaining two clusters, 1 and 3, both exhibit lower average returns (-0.00038 and -0.00041, respectively) and lower average standard deviations (0.02425 and 0.01670, respectively), aligning with lower-risk profiles. Notably, Cluster 1 demonstrates a slightly higher average standard deviation compared to Cluster 3, suggesting a marginally elevated risk level for a similar return. The presence of this additional moderate-risk, moderate-reward cluster in k-medoids highlights its potential to uncover a broader spectrum of investment opportunities compared to k-means++.

B. Portfolio Optimization

1) Representative Shares: To select representative stocks for each cluster, we employed a data-driven approach based on the Sharpe ratio, a widely recognized measure of risk-adjusted return that considers both the potential profit and volatility of an investment. Our analysis identified two distinct clusters using both k-means++ and k-medoids clustering algorithms. Notably, the two algorithms consistently selected the same representative stocks for the high-risk, high-reward and low-risk, low-reward clusters, reinforcing the robustness of our findings.

As illustrated in Figure 3, the high-risk, high-reward cluster is represented by BRPT.JK (Sharpe ratio = 0.07540),

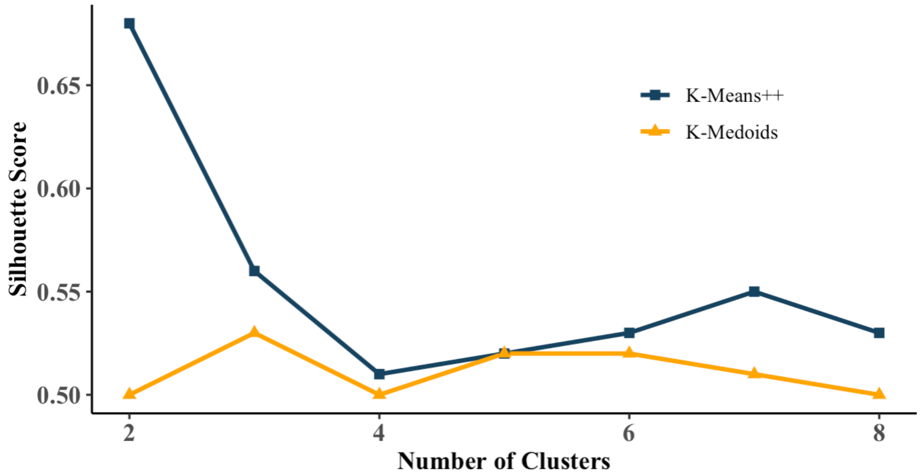


Fig. 2: Comparison of silhouette scores

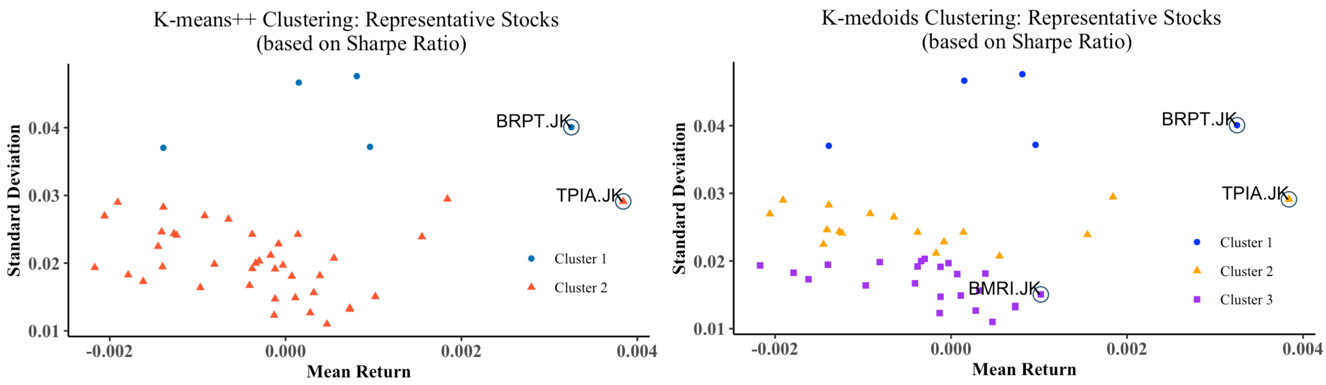
while the low-risk, low-reward cluster is represented by TPIA.JK (Sharpe ratio = 0.12386). The k-medoids clustering algorithm additionally identified a third cluster corresponding to a moderate-risk, moderate-reward profile, represented by BMRI.JK (Sharpe ratio = 0.05206). This highlights the potential of k-medoids to uncover a wider range of investment options, enabling the construction of more diversified portfolios. Subsequently, we employed portfolio optimization techniques to determine the optimal weights for each representative stock, aiming to minimize portfolio risk (as reflected in the objective function: $\min fitness = \sigma_p^2$, where σ_p^2 represents the portfolio variance) and maximize risk-adjusted returns.

2) Portfolio Weighting: The analysis of investment weights, as presented in Table II, provides insights into the allocation strategies of the three SI models (FA, GWO, and WOA) in comparison to the Mean Variance (MV) model. The average absolute difference in weight assignments was used to quantify the alignment between the SI models and MV. Within the k-means++ clusters, FA and GWO exhibited the closest alignment with MV, with an average absolute difference of only 0.003. This suggests that these two SI models, particularly when combined with k-means++ clustering, tend to produce portfolio allocations that are remarkably similar to those derived from the traditional MV model. In contrast, WOA displayed a larger average absolute difference from MV within the k-means++ clusters, indicating a more distinct allocation strategy.

Similarly, for the k-medoids clusters, GWO once again demonstrated the closest alignment with MV, showcasing an average absolute difference of 0.005. This further supports the notion that GWO, when coupled with k-medoids clustering, may serve as a viable alternative to the MV model in terms of weight allocation. Overall, these findings highlight that while the SI models employ distinct optimization mechanisms, their resulting weight allocations remain broadly comparable to those of the MV model. The observed variations in weight assignments across different SI models underscore their unique approaches to balancing risk and return, as well as their sensitivity to the underlying cluster structures generated by k-means++ and k-medoids. The choice between different SI models ultimately hinges on factors such as computational efficiency, specific investment goals, and individual risk preferences.

C. Portfolio Performance and Comparison

The evaluation of portfolio performance, encompassing key metrics such as return, standard deviation, and Sharpe ratio (Table III), offers a comprehensive assessment of the various clustering and optimization combinations. The KMP GWO (K-means++ - Grey Wolf Optimization) combination stands out, exhibiting the highest return (1.55094) and Sharpe



(a) Figure 3a (b) Figure 3b

Fig. 3: Representative Shares

TABLE II: Stock Weighting

C	Stock	MV	FA	GWO	WOA
K-means					
1	BRPT.JK	0.801	0.798	0.804	0.74
2	TPIA.JK	0.199	0.202	0.196	0.26
K-medoids					
1	TPIA.JK	0.775	0.766	0.775	0.742
2	BRPT.JK	0.054	0.07	0.046	0.076
3	BMRI.JK	0.171	0.164	0.178	0.182

TABLE III: Portfolio Performance Comparison

Portfolio (C - O)	Return	Std. Deviation	Sharpe Ratio
KMP MV	1.54949	0.43846	3.40279
KMP FA	1.54872	0.43848	3.40086
KMP GWO	1.55094	0.43851	3.40573
KMP WOA	1.52680	0.43974	3.34131
KMD MV	0.50322	0.20472	2.17717
KMD FA	0.50862	0.20498	2.20084
KMD GWO	0.50429	0.20477	2.18188
KMD WOA	0.53318	0.20557	2.31398

KMP is indicated as the K-Means++ clustering

KMD is indicated as the K-Medoids clustering.

ratio (3.40573). This suggests that the synergy between k-means++ clustering and the GWO algorithm effectively captures the market's risk-return dynamics, leading to superior portfolio performance in terms of both absolute gains and risk-adjusted returns. However, it's important to acknowledge the trade-off between risk and return. The KMP GWO portfolio also demonstrates the highest standard deviation, indicating greater volatility compared to other models. The choice of the optimal model, therefore, hinges on investors' risk tolerance and investment objectives.

In contrast, k-medoids-based portfolios generally exhibited lower performance compared to k-means++, potentially due to the inclusion of the moderate-risk cluster. While this cluster offers diversification benefits, it might also dampen overall returns, particularly in a bullish market. Notably, the KMD WOA (K-medoids - Whale Optimization Algorithm) combination emerged as the top performer within the k-medoids category, achieving a commendable Sharpe ratio of 2.31398. This suggests that WOA's optimization strategy might be particularly adept at handling the complexities of a three-cluster structure.

D. Portfolio Performance in the Stock Market

To further evaluate the real-world implications of our optimized portfolios, we simulated their performance in the stock market by calculating profit/loss based on the closing prices of the selected stocks. We assumed a hypothetical investment of Rp 1,000,000,000.00 on January 3, 2024, and observed the portfolio performance over a seven-day period from January 4 to January 12, 2024.

TABLE IV: Daily Portfolio Performance in the Stock Market (in Rupiah)

Date	MV	FA	GWO	WOA
K-Means++				
04/01/24	9049000	9105570	9143290	**9203370**
05/01/24	-17847400	-17756375	-17804875	**-16771050**
08/01/24	-83860000	-83642855	-84001435	**-79450680**
09/01/24	-263560480	-263344740	-263730580	**-258955400**
10/01/24	-288533300	-288168935	-288911195	**-280068685**
11/01/24	-336317040	-335933850	-336728850	**-327305980**
12/01/24	-356116540	-355665045	-356,625,265	**-345320975**
Average	-191026537	-190772319	-191236987	**-185524200**
K-Medoids				
04/01/24	33823940	**33893970**	33864560	32824720
05/01/24	**37435225**	37401900	37387375	35407200
08/01/24	19896215	**20133620**	19483085	16738895
09/01/24	**-23694260**	-24861150	-24035180	-32501145
10/01/24	**-22390820**	-22922585	-23175230	-31254610
11/01/24	**-29888090**	-30718480	-30709430	-40291940
12/01/24	**-18199780**	-18937675	-19211890	-29440210
Average	**-431081**	-858629	-913816	-6931013

To further evaluate the practical implications of our optimized portfolios, we simulated their performance in the stock market by calculating profit/loss based on the closing prices of selected stocks over a seven-day period from January 4 to January 12, 2024. We assumed a hypothetical investment of Rp 1,000,000,000.00 on January 3, 2024.

The daily performance of each portfolio combination is presented in Table IV. This detailed view allows us to assess the fluctuations and trends experienced by each strategy over time, providing a more comprehensive understanding of their strengths and weaknesses under various market conditions.

Interestingly, the k-means++ portfolio paired with the Whale Optimization Algorithm (WOA) consistently generated the highest returns on days with positive market movements and the lowest losses on days with negative market movements. This suggests that WOA, when combined with k-means++, may be particularly adept at capturing upside potential while mitigating downside risk. However, it's important to note that this strategy also exhibited higher volatility overall, as evidenced by its larger fluctuations in daily profit/loss.

On the other hand, the k-medoids-based portfolios demonstrated greater resilience, incurring a lower average loss (-0.23%) compared to the k-means++ portfolios (-18.96%). This superior performance can likely be attributed to the inclusion of the stable stock BMRI.JK in the k-medoids cluster, highlighting the critical role of diversification in mitigating downside risk, particularly in volatile market conditions.

In summary, the performance analysis underscores the trade-offs between risk and return inherent in different clustering and optimization combinations. The k-means++ portfolio with WOA excels in capturing market upside and mitigating downside, but with higher volatility. Conversely, k-medoids-based portfolios, particularly KMD WOA, prioritize stability and downside protection. The selection of the optimal approach ultimately hinges on the investor's specific investment goals, risk tolerance, and time horizon.

The integration of clustering techniques with SI algorithms presents a promising avenue for enhancing portfolio optimization. However, our findings emphasize that the choice of the most suitable approach necessitates a careful balance between risk and return, tailored to the individual investor's financial objectives and risk appetite.

V. Conclusion

This study investigated the potential of integrating clustering analysis with swarm intelligence (SI) algorithms for enhancing portfolio optimization. Our findings reveal that combining k-means++ clustering with the Grey Wolf Optimization (GWO) algorithm yields the highest return and Sharpe ratio, outperforming other SI and traditional Mean Variance (MV) models. However, the simulation of portfolio performance in the stock market highlighted the inherent volatility and potential for short-term losses, even with optimized portfolios.

Importantly, we observed that k-medoids-based strategies, particularly when combined with the Whale Optimization Algorithm (WOA), demonstrated greater resilience and minimized losses in a short-term market scenario. This underscores the importance of diversification and risk management in volatile market conditions.

Our research contributes to the field by providing a comprehensive analysis of the synergy between clustering techniques and diverse SI algorithms in the context of portfolio optimization. The findings have practical implications for investors and portfolio managers, offering insights into the selection of appropriate SI models and clustering techniques for balancing risk and return in different investment horizons.

Future research should explore the impact of different clustering methods and parameter settings on the performance of SI-based portfolio optimization strategies. Additionally, investigating the robustness of these strategies across various market conditions and timeframes would further enhance their practical applicability.

References

[1] K. Erwin and A. Engelbrecht, "Multi-guide set-based particle swarm optimization for multi-objective portfolio optimization," *Algorithms*, vol. 16, no. 2, p. 62, 2023.

[2] T. Bodnar, S. Mazur, and Y. Okhrin, "Bayesian estimation of the global minimum variance portfolio," *European Journal of Operational Research*, vol. 256, no. 1, pp. 292–307, 2017.

[3] F. Yang, Z. Chen, J. Li, and L. Tang, "A novel hybrid stock selection method with stock prediction," *Applied Soft Computing*, vol. 80, pp. 820–831, 2019.

[4] B. Chen, J. Zhong, and Y. Chen, "A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from shanghai stock exchange," *Expert Systems with Applications*, vol. 145, p. 113104, 2020.

[5] A. Baykasoğlu, M. G. Yunusoglu, and F. B. Özsoydan, "A grasp based solution approach to solve cardinality constrained portfolio optimization problems," *Computers & Industrial Engineering*, vol. 90, pp. 339–351, 2015.

[6] S. T. Tayalı, "A novel backtesting methodology for clustering in mean–variance portfolio optimization," *Knowledge-Based Systems*, vol. 209, p. 106454, 2020.

[7] C. Iorio, G. Frasso, A. D'Ambrosio, and R. Siciliano, "A p-spline based clustering approach for portfolio selection," *Expert Systems with Applications*, vol. 95, pp. 88–103, 2018.

[8] M. Ashrafzadeh, H. M. Taheri, M. Gharehgozlou, and S. H. Zolfani, "Clustering-based return prediction model for stock pre-selection in portfolio optimization using pso-cnn+ mvf," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 9, p. 101737, 2023.

[9] O. Ertenlice and C. B. Kalayci, "A survey of swarm intelligence for portfolio optimization: Algorithms and applications," *Swarm and evolutionary computation*, vol. 39, pp. 36–52, 2018.

[10] N. Bnouachir and A. Mkhadri, "Efficient cluster-based portfolio optimization," *Communications in Statistics-Simulation and Computation*, vol. 50, no. 11, pp. 3241–3255, 2021.

[11] T. T. Bjerring, O. Ross, and A. Weissensteiner, "Feature selection for portfolio optimization," *Annals of Operations Research*, vol. 256, pp. 21–40, 2017.

[12] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Systems*, vol. 252, p. 109358, 2022.

[13] F. G. Duarte and L. N. De Castro, "A framework to perform asset allocation based on partitional clustering," *IEEE access*, vol. 8, pp. 110 775–110 788, 2020.

[14] K. Mazumdar, D. Zhang, and Y. Guo, "Portfolio selection and unsystematic risk optimisation using swarm intelligence," *Journal of Banking and Financial Technology*, vol. 4, no. 1, pp. 1–14, 2020.

[15] J. Li, X. Wei, B. Li, and Z. Zeng, "A survey on firefly algorithms," *Neurocomputing*, vol. 500, pp. 662–678, 2022.

[16] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural computing and applications*, vol. 30, pp. 413–435, 2018.

[17] F. S. Gharehchopogh and H. Gholizadeh, "A comprehensive survey: Whale optimization algorithm and its applications," *Swarm and Evolutionary Computation*, vol. 48, pp. 1–24, 2019.

[18] I. Fister, I. Fister Jr, X.-S. Yang, and J. Brest, "A comprehensive review of firefly algorithms," *Swarm and evolutionary computation*, vol. 13, pp. 34–46, 2013.

[19] X.-S. Yang, *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.

[20] ——, "Firefly algorithm, stochastic test functions and design optimisation," *International journal of bio-inspired computation*, vol. 2, no. 2, pp. 78–84, 2010.

[21] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46–61, 2014.

[22] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, 2016.

[23] O. O. Obadina, M. A. Thaha, K. Althoefer, and M. H. Shaheed, "Dynamic characterization of a master–slave robotic manipulator using a hybrid grey wolf–whale optimization algorithm," *Journal of Vibration and Control*, vol. 28, no. 15-16, pp. 1992–2003, 2022.

[24] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51–67, 2016.

[25] P. Bruce, A. Bruce, and P. Gedeck, *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media, 2020.