International Journal of Advances in Intelligent Informatics
Vol. 8, No. 1, March 2022, pp. 33-44

Portfolio optimization based on self-organizing maps clustering and genetics algorithm

Fajri Farid [a,1] , Dedi Rosadi [a,2,*]

[a] Department of Mathematics, Faculty of Mathematics and Natural Science, Gadjah Mada University, Yogyakarta, Indonesia
[1] fajrifarid94@gmail.com; [2] dedirosadi@gadjahmada.edu
* corresponding author

ARTICLE INFO

Article history
Received November 9, 2020
Revised September 27, 2021
Accepted January 8, 2022
Available online March 31, 2022

Keywords
Portfolio
Optimization
Clustering
Genetic Algorithm
Self-organizing Maps

ABSTRACT

In this modern era, gaining additional income is necessary to fulfill daily needs since inflation is unavoidable. Investing in stocks can give passive income to help people deal with the increasing prices of necessities. However, selecting stocks and constructing a portfolio is the major problem in investing. This research will illustrate the stock selection method and the optimization method for optimizing the portfolio. Stock selection is carried out by clustering using Self-organizing Maps (SOM). Clustering will show the best stocks formed for a portfolio to be optimized. The best stocks that have the best performance are selected from each cluster for the portfolio. The best performance of the stock can be determined using the Sharpe Ratio. Optimization will be carried out using a Genetic Algorithm. The optimization is carried out using software R i386 3.6.1. The optimization results are then compared to the Markowitz Theory to show which method is better. The expected return on the portfolio generated using Genetic Algorithm and Markowitz Theory are 3.348458 and 3.347559975, respectively. While, the value of the Sharpe Ratio is 0.1393076 and 0.13929785, respectively. Based on the results, the best performance of the portfolio is the portfolio produced using Genetic Algorithm with the greater value of the Sharpe Ratio. Furthermore, the Genetics Algorithm optimization is more optimal than the Markowitz Theory.

1. Introduction

The fairly fast economic development in Indonesia affects the prices of necessities, from the agricultural sector, the industrial sector, the development sector and so on. The increasing need for clothing, food and shelter due to the influence of economic development has made people look for other alternatives to get additional income. Moreover, most Indonesians have a very consumptive attitude that requires a large amount of money to meet these needs. One of the things that can increase income is investing. Broadly speaking, investment is divided into two, namely investment in real assets and investment in financial assets. Investing can be done by everyone (Investor). However, in investing there are things that need to be considered, one of which is risk. Such as the slogan which is very popular among investors, namely "High Risk High Return". The greater the risk faced by an investor, the greater the rate of return that investor will get. There are several types of investors in facing risk, namely, risk-seeking investors, risk-neutral investors, and risk-averse investors [1]. In addition, the investment strategy also needs to be considered, for example, in the portfolio construction. Developments in research to determine the optimal portfolio have been carried out a long time ago [2], [3]. Harry Max Markowitz, a person who wrote theory about portfolios [4] entitled Portfolio Selection (1952) in The Journal of Finance and a book [5] entitled Portfolio Selection: Efficient Diversification of Investment (1959). Since then, the theory of portfolios has grown and the determination of several assets to be invested in order to get a profit has been increasingly varied [6]. However, the Markowitz Theory only illustrates how to optimize a portfolio and did not illustrate the stock selection itself.

Stock selection in portfolio construction is necessary in order to get the desired rate of return. The stocks chosen are stocks that have the best performance as measured by the rate of return and the risk that will be borne by investors. In this research, clustering will be carried out based on the categories of stocks to be selected [7]. Clustering in this research will be carried out using the Self-organizing Maps method. After the stocks are clustered, several stocks that have the best performance from each cluster will be selected to become a portfolio which will later be weighted using the Genetic Algorithms and the results will be compared to the Markowitz Theory. Several previous studies used as a reference in this research, such as research conducted in Cheong, et al. [8]. They clustered stocks in the Korean Stock Market by investor information using k-means clustering and then formed four portfolios based on different weighting methods, namely Equal Weights, Market Capital Weights, Minimum Variance Weights and Sharpe Ratio Weights [9]. Then they compared the performance of the four portfolios and found that the portfolio using the Sharpe Ratio Weights weighting method as a fitness function is the portfolio that has the best results [10]. However, they did not compare the results to the Markowitz Theory. Based on this, this research will use the Sharpe Ratio Weights weighting method which already proved that have the best results between other weighting methods and will compared it to Markowitz Theory [11], [12]. Research conducted in Nair, et al. [13]. They used the Self-organizing Maps method for clustering high-dimensional data to a lower dimension and obtained 16 recommendation systems that can determine which stocks are profitable or not based on time series data information on stock prices [14]. Their research only used SOM to determine which stocks are profitable. This research uses SOM to determine the best stock in each cluster that will be formed for a portfolio to be optimized. The objectives of this research are: (1) Obtain the results of clustering and knowing the best stock from each cluster; (2) Determine the proportion of stock weight in the portfolio using a Genetic Algorithm; (3) Comparing the resulting portfolio performance.

2. Method

The portfolio construction in this research uses the SOM method for clustering and the Genetic Algorithm for optimization.

2.1. Markowitz Theory

A portfolio is a collection of investments owned by institutions or individuals. The combination of these assets can be in the form of real assets, financial assets or both. Portfolios aim to assist investors in diversifying stocks to reduce the risk borne by the hope that if one stock value falls while the value of other stocks rises, the loss can be offset by the rate of return obtained [15]. The Markowitz portfolio theory or commonly known as the Markowitz Model was first put forward by Harry Max Markowitz [4] in his article entitled Portfolio Selection (1952) in The Journal of Finance. Markowitz stated that if an investor wants a maximum expected return on his portfolio, then that investor must use the funds on assets that have a high expected return. However, this does not guarantee that the portfolio will have low risk [16]–[18]. Portfolio selection is very important to increase expected return or reduce the risk that will be borne by investors. If an asset is added to a portfolio [19], [20], the total portfolio risk will be reduced but the expected return remains the weighted average of the expected return of each asset in the portfolio. In other words, diversification will reduce the total risk without sacrificing return [21]. An investor always wants a high rate of return with low risk. A portfolio that provides a greater expectation of returns with the same risk, or provides a smaller risk with the same return expectation is called an efficient portfolio and an investor will always choose an efficient portfolio. Meanwhile, a portfolio chosen by investors from many efficient portfolios is called the optimal portfolio. The selection of the portfolio is of course in accordance with the preferences of investors on the level of profit and risk they are willing to bear.

2.2. Self-organizing Maps

Self-organizing Maps (SOM) is a neural network type that is used for data classification. In 1982, Teuvo Kohonen introduced SOM for the first time. Unlike other types of neural networks, the structure of Kohonen SOM consists of two layers, they are the output and input layer (Fig. 1).

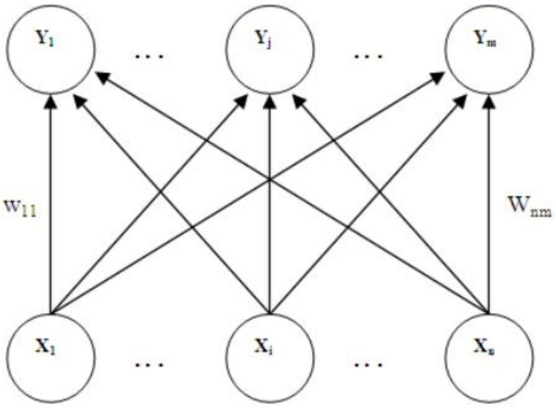

Fig. 1. Self-organizing Maps (SOM)

A hidden layer is not required in SOM because every neuron within the input layer is solely mapped by SOM onto every neuron in the output layer. So, every neuron represents a cluster in the output layer. Every input neuron is received onto the output neuron through a weight that is connected directly to the input. So, the weight vector has identical dimensions with the input vector [15]. SOM is unsupervised learning and a technique that is efficient for clustering high-dimensional data into low-dimensional data [13]. The SOM algorithm is performed by initializing the randomly obtained weight (w_{ij}) for each neuron. After being given a weight (w_{ij}), the network is then given input (x_i). After the input is received by the network, the weight vector is calculated with the input vector. Distance calculations can be calculated using the formula:

$$D(j) = \sum_{i=1}^{n}(w_{ij} - x_i)^2 \quad (1)$$

After the distance between neurons is obtained, the neuron that has the minimum value of the vector distance is selected $D(j)$ and make weight changes $w_{ij}\ (t + 1)$ by using the formula:

$$W_{ij}(t+1) = W_{ij}(t) + h(t) * [x_i(t) - W_{ij}(t)] \quad (2)$$

where $h(t)$ is the neighbor function and $h(t) \to 0$ if $t \to 0$. Average width and shape of $h(t)$ determine the "stiffness" of the "elastic surface" to be attached to the data points. Fig. 2 is an example of the form of the neighbor function:

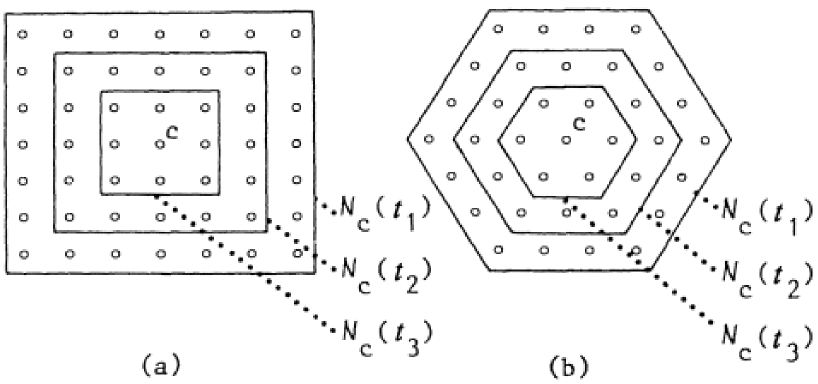

Fig. 2. Neighbor Function Topology

Let $N(t)$ be the set of neighbors at time t. There are tow types of neighboring functions $h(t)$, namely: (1) a simple function expressed by $h(t) = \alpha(t)$ if i in natural numbers and the value of $h(t) = 0$ if i is not in natural numbers. $\alpha(t)$ is the learning rate which is the amount $0 \leq \alpha \leq 1$. The value of $\alpha(t)$ and the radius of N is monotonically decreasing with time during the process. (2) other functions that are widely applied and "smoother" can be written in the form of a gaussian function as follows:

$$h(t) = \alpha(t).\exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \tag{3}$$

where r_c are the winning neurons and r_i is *neuron* to-i. $\alpha(t)$ and $\sigma^2(t)$ as a function whose value gets smaller over time t. In general value $\alpha(t)$ and $\sigma^2(t)$ can be calculated using a formula:

$$\alpha(t) = \alpha_0 e^{\frac{-t}{\tau}} \tag{4}$$

$$\sigma(t) = \sigma_0 e^{\frac{-t}{\tau}} \tag{5}$$

where t is iteration on time to-t, τ is the time constant used to reduce the radius and rate of learning. The use of the Gaussian function in SOM is because the SOM process is a process that follows natural phenomena so that the distribution of the data will follow a normal distribution.

2.3. Genetic Algorithm

The Genetic Algorithm which is part of the Evolution Algorithm is a metaheuristic method used to solve an optimization problem. Basic principles of Genetic Algorithm use a natural selection process and the principles of genetic science. Individuals in the population compete for survival and reproduction. Individuals who survive are “fit” individuals so that they have a higher chance of living and reproducing. On the other hand, individuals who do not survive die and become extinct. This principle is called "Survival of the fittest". The genetic algorithm was first introduced by John Holland in the 1960s, inspired by the concept of Darwin's theory of evolution. By imitating this theory of evolution, genetic algorithm is an effective technique to solve optimization problems [22]. In the theory of evolution, each individual in the population has different characteristics. Individuals who survive in the population will carry out a transformation based on the principles of genetic science to get new individuals. There are two types of transformation: mutation, the formation of a new individual by changing the chromosome structure in one individual, and crossover, the formation of a new individual by combining the chromosomes of two individuals. The new individual is called the offspring who will be evaluated. Furthermore, the population is formed from individuals who are more "fit" than the parent population and offspring populations [23]. After several generations, the algorithm will get the best individual results which will represent the optimal solution to the problem at hand [24].

2.3.1. Encoding

The first step that must be taken to apply the Genetic Algorithm is to create an encoding. How to encode a solution of a problem into chromosome form is the key to using Genetic Algorithm [25]. Holland was first coded into chromosomes using the binary strings representation. This representation symbolizes the number zero (0) and one (1) of the elements that are on a chromosome. Each sequence of elements has a special meaning that indicates the fitness value of the chromosome in question [21].

2.3.2. Fitness Function

Fitness function is the objective function of a problem at hand. In a genetics algorithm, the chromosomes in the population will be optimized to find a solution that can solve a problem. Each chromosome will have a fitness value that will determine how good or not the solution is. The higher the value of the fitness function, the better the solution is [26].

2.3.3. Selection

Chromosomes in the population whose fitness value has been evaluated will be selected to be the parent. Chromosomes that have a large fitness value will become the parent which will produce offspring with a greater fitness value. Furthermore, the chromosomes that have the worse fitness value will be replaced by new, better chromosomes. In genetic algorithm, the selection process uses the general principle of selection, namely fitness proportional selection, which means that the chances of each chromosome being selected are proportional to the fitness value. So, if a chromosome has twice the value of another chromosome, then the chromosome has twice the chance to reproduce [15]. In 1989, Goldberg [26] introduced a selection technique in genetic algorithm. This selection technique is called roulette wheel selection. This technique is illustrated as a technique for playing roulette discs. Each chromosome will be placed in a slot with a roulette disc with the scale of the slot same to the ratio of the fitness value of a chromosome to the total fitness value of all chromosomes.

2.3.4. Genetic Algorithm Operators

The use of genetic algorithm aims to get the best offspring which will produce the best solution for the problems at hand. In achieving this goal, the thing to pay attention to is to avoid premature convergence, where the optimum solution will be reached before the time occurs. That is, the solution obtained is the local optimum result. In the formation of child chromosomes, the parent chromosomes will carry out a transformation, namely crossing (crossover) or mutation (mutation). Within a generation, these processes can occur sequentially or parallel [27]. The meaning of sequentially is that the crossing process will occur first and then the mutation process will continue. This process is called mutation embedded within crossover. Meanwhile, in parallel it means that the process of crossing and mutation will occur separately [22].

2.3.5. Crossover

Moving crossovers is the main operator in genetic algorithm. This operator produces two daughter chromosomes by exchanging some elements (genes) on a pair of parent chromosomes. This operation is not always performed on all existing individuals. Individuals will be randomly selected for crossing over. If crossing over is not done, then the value of the parent will be passed down to the offspring. The process of crossing over is carried out with a certain probability, namely P_c. Crossing can only be done if a number is random [0, 1) which is raised is less than the value P_c which is determined. Generally, value P_c used ranges from 0.6 to close to 1. Determination of value Pc the right one really depends on the problem at hand [28].

2.3.6. Mutation

Mutation is the support operator in genetic algorithm that play a role in changing the chromosome structure directly. This direct change forms a mutant, which is a new chromosome that is genetically different from the previous chromosome. This operator works only on one chromosome. To get the optimal solution, mutations are needed to restore the lost genes in the previous generation, and give rise to new genes that have never appeared in the previous generation. The mutation probability denoted as P_m is the ratio between the number of genes expected to undergo the mutation and the total number of genes in the population. Since mutation is a supporting operator, the P_m value used is quite low, ranging from 0.001 to 0.2. If the chance of mutation is low, it is less likely that new genes will emerge. Although in fact new genes are needed to obtain optimal solutions. Conversely, if the chance of mutation is too high, then a lot of mutants will emerge. This will result in many characteristics of the parent chromosome which may disappear in the next generation so that the genetic algorithm will lose the ability to remember or learn from previous processes.

2.4. SOM Algorithm

In this research, the clustering was carried out using R i386 3.6.1 software. After the input data has been entered, clustering can be done using the SOM method. The steps for the SOM algorithm are as follows; (1) The initial process is to determine the data that will be used as input (x_i); (2) Initialize the weights of w_{ij}, α_0, σ_0 randomly; (3) For each x_i perform steps 4 through 6; (4) For every j count all the values of $D(j)$; (5) Determine the value of j such that $D(j)$ is minimum; (6) Update the weights of the winning and neighboring neurons; (7) Update learning rate and neighbor function. Test the stop conditions, perform steps 2 through 7 if the stop conditions are not met. The condition for stopping the test is done by calculating the difference between the t-weight and the weight to-($t+1$), if the value changes only slightly, it means that the test has reached convergence so it can be stopped [3], [29], [30]. Fig. 3 is the flowchart of algorithm in SOM.

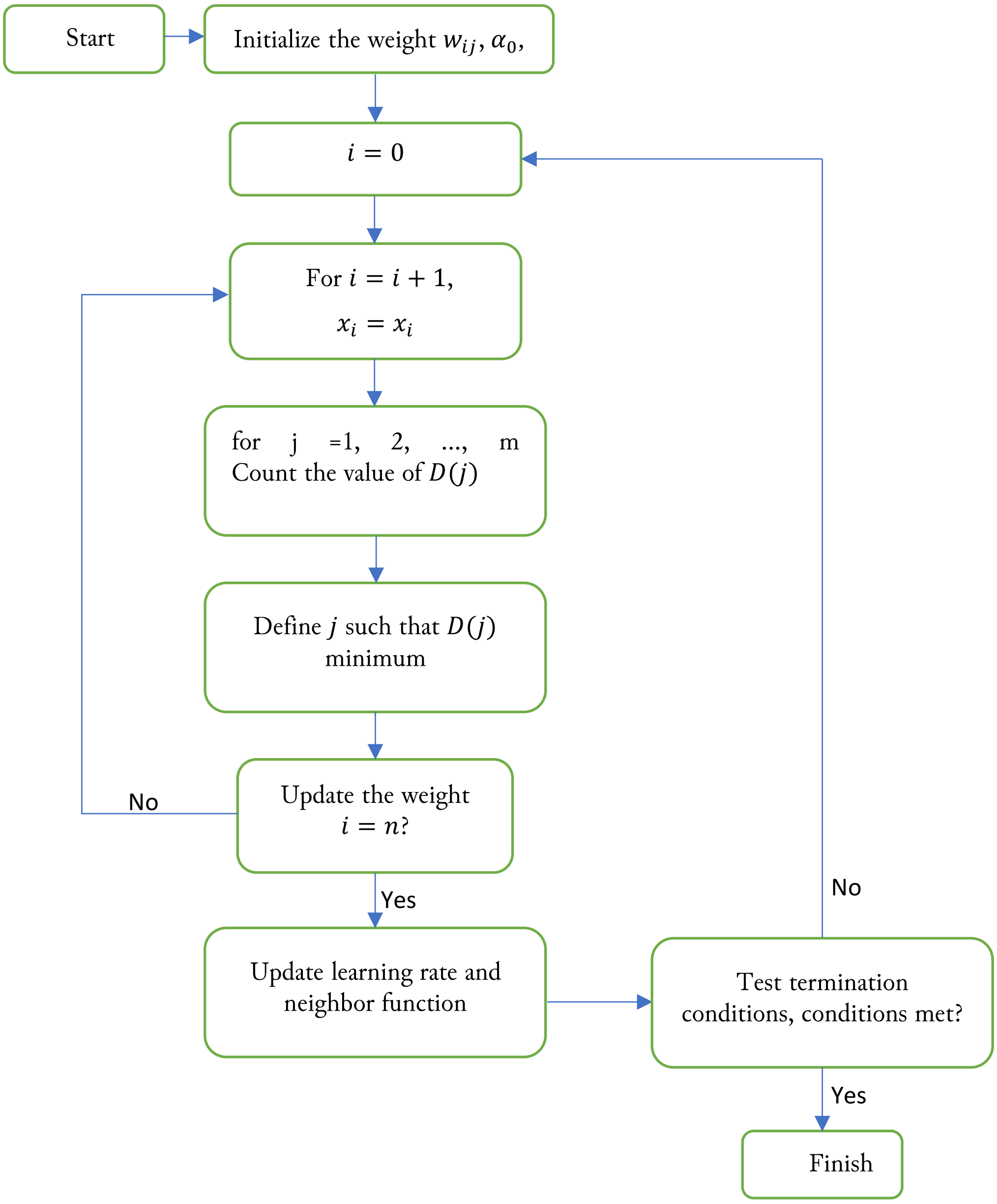

Fig. 3. SOM Algorithm Flowchart

2.5. Portfolio Optimization

The portfolio weights are carried out using the Genetic Algorithm and Markowitz Theory. Portfolio optimization using Genetic Algorithm was carried out using R i386 3.6.1 software. Meanwhile, portfolio optimization using Markowitz Theory was carried out using the Solver command in the Ms. Excel software. The assumptions used when optimizing the portfolio are as follows; (1) There is no short selling; (2) No stock split; (3) There is no influence of inflation, exchange rates and economic crises.

2.5.1. Portfolio Optimization using Genetic Algorithm

This method will look for the optimal stock portfolio weight. The steps of the Genetic Algorithm are as follows; (1) The first step is encoding. In this step, the weight of the portfolio will be represented by chromosomes in the form of binary strings. The chromosome consists of several genes which indicate the weight of each stock; (2) Initialize population and maximum iteration; (3) Determine the fitness function. The fitness function in optimizing the stock portfolio is to maximize the Sharpe Ratio. The fitness function can be written as follows:

$$maxS = \frac{R_P - Rf}{\sigma_p} \tag{6}$$

$$= \frac{W^T R - R_f}{\sqrt{(W^T CW)}} \tag{7}$$

with constraint

$$\sum_{i=1}^{N} w_i = 1, w_i \geq 0, i = 1,2, \dots, N \tag{8}$$

where R_p is the portfolio return, R_f is the risk-free rate, σ_p is the standard deviation or portfolio risk, W is matrix $N \times 1$ of the portfolio weight, C is the covariance matrix, $C = |\sigma_{ij}|; i = 1,2,\dots,N; j = 1,2,\dots,N$.

- After the fitness function is determined, then perform chromosome selection. The chromosomes that have the best fitness value will be selected as the parent and will carry out the next process, namely producing offspring
- The process of producing offspring using genetic algorithm operators is the main process of genetic algorithm. As previously explained, there are two types of genetic algorithm operators, namely crossover and mutation. In this research, the probability of a crossover is denoted by P_c equal to 0.8. Meanwhile, the chance of a mutation which is denoted by P_m equal to 0.1
- Calculate the value of the fitness function of the offspring
- Repeat steps 4 through 6 until maximum iteration.

2.5.2. Portfolio Optimization using Markowitz Theory

Optimizing the stock portfolio in Markowitz Theory will minimize the risk that will be borne by investors and maximize the return [31]. The objective function for portfolio optimization using Markowitz Theory can be written as follows:

$$Min\ \sigma_p = \sqrt{W^T CW} \tag{9}$$

with constraint

$$\sum_{i=1}^{N} w_i = 1, w_i \geq 0,\ i = 1,2, \dots, N \tag{10}$$

where σ_p is the standard deviation or portfolio risk, W is matrix $N \times 1$ of the portfolio weight, C is the covariance matrix; $C = |\sigma_{ij}|; i = 1,2,\dots,N; j = 1,2,\dots,N$.

3. Results and Discussion

3.1. Data Description

This research uses secondary data in the form of stock data in LQ45 shares. The data used are historical data on daily stock prices (Closed Price) for a year from November 26, 2018 to November 25, 2019. The data is taken from the websites *www.yahoo.finance.com* and *www.idx.co.id*. BI Rate data is obtained from the Bank Indonesia *website* *www.bi.go.id*. The data that have been obtained from the website will then be processed using Ms. Excel to find the value of return, expected return, risk (variance and standard deviation), coefficient of variance and Sharpe Ratio.

3.2. Clustering based Self-organizing Maps

Before constructing portfolio, the LQ45 shares that have been obtained will be clustered first using the SOM method. Before the SOM process begins, a data plot will be created (data plot can be seen in Fig. 4).

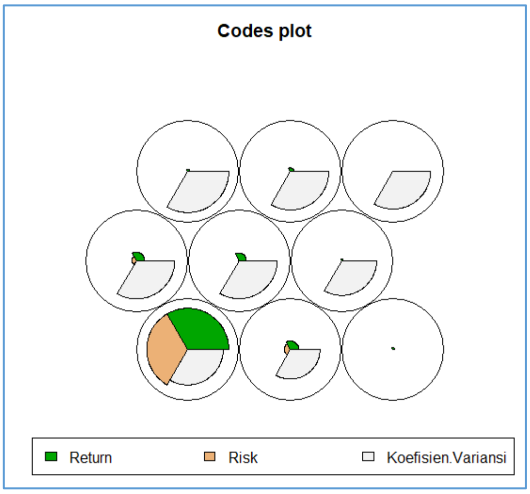

Fig. 4. Data Plot

Based on Fig. 4 the green one is the return variable, the cream one is the risk variable, and the white one is the variable of the coefficient of variance [32], [33]. Furthermore, authors choose to make three clusters from the data so that portfolio that will be produced will have three stocks and the numbering clusters can be done freely based on researchers' preference. The plot of the clustering results can be seen in Fig. 5.

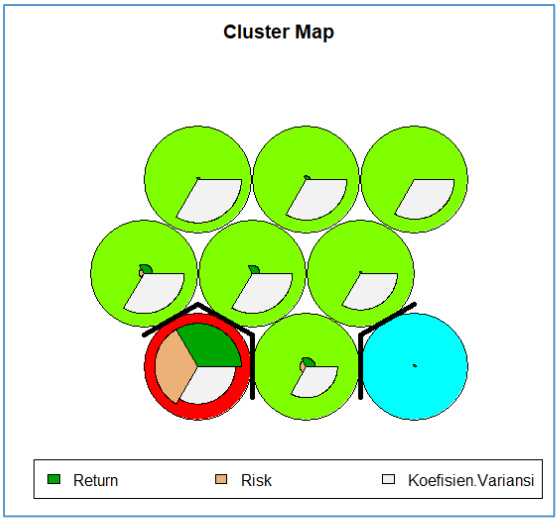

Fig. 5. Data Plot after Clustering

Based on the results, there are 42 stocks in the first cluster, 1 stock in the second cluster, and 2 stocks in the third cluster.

3.3. Portfolio Construction

The stocks from each cluster that have the best performance will be selected, that is the highest Sharpe Ratio of each cluster. In the first cluster, BRPT.JK stock is the stock that have the highest Sharpe Ratio of 0.141508151, while for the second cluster which only has 1 stock, AKRA.JK has a Sharpe Ratio of -0.00514189, then the highest Sharpe Ratio of the third cluster is UNVR.JK stock amounting to -0.00028894. Thus, the resulting portfolio is a portfolio of 3 stocks, those are BRPT.JK, AKRA.JK, and UNVR.JK. The resulting portfolio from the clustering results is then optimized using the Genetic Algorithm method with the Sharpe Ratio as a fitness function [34], [35]. Optimization using the Genetic Algorithm method will find the weight of each stock in the portfolio. The results obtained from this optimization can be seen in Table 1.

Table 1. Optimization results using the Genetic Algorithm method

Stock	BRPT.JK	AKRA.JK	UNVR.JK
Weight	0.999270409	7.12425E-05	0.000319853

Based on Table 1, the weights of the portfolio for BRPT.JK, AKRA.JK, and UNVR.JK stocks are 0.999270409, 0.0000712425, and 0.000319853, respectively. The plot of Fitness Value or prospective results from portfolio optimization using Genetic Algorithms can be seen in Fig. 6.

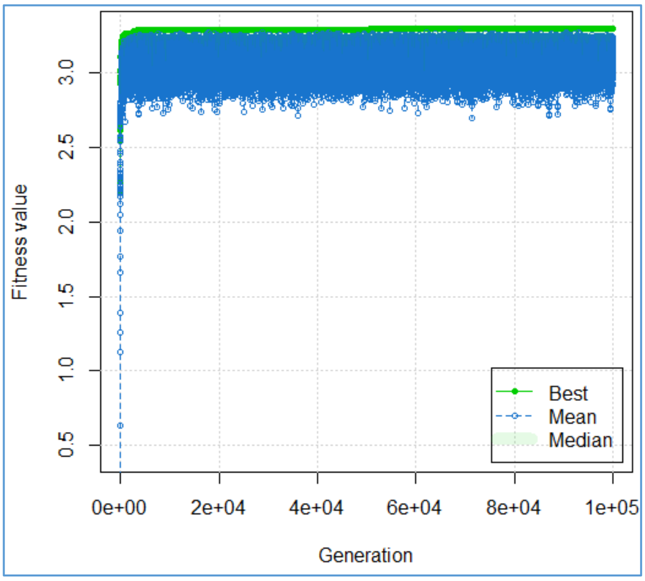

Fig. 6. Fitness Value (portfolio optimization using Genetic Algorithm)

The green line is the best fitness value for a certain generation. Meanwhile, the blue line is the average of the fitness value. The plot of the return comparison between the results obtained with each stock can be seen in Fig. 7. The black line on the plot is the return value resulting from the resulting portfolio weights. Meanwhile, the blue, red, and green lines are the stock return lines of BRPT.JK, AKRA.JK, and UNVR.JK. On Fig. 7 shows that the black line is almost the same as the blue line. This is because the weight generated in the portfolio is mostly allocated to BRPT.JK stock.

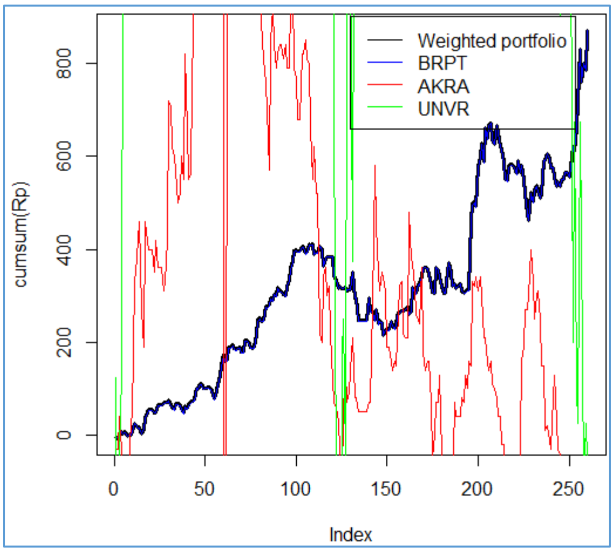

Fig. 7. Comparison of returns between the results obtained with each stock

The weight comparison between portfolio optimization using the Genetic Algorithm method and Markowitz Theory can be seen in Table 2.

Table 2. Portfolio Weight with Different Method

Method	Genetic Algorithm	Markowitz Theory
BRPT.JK	0.999270409	0.999574509
AKRA.JK	7.12425E-05	0.000425491
UNVR.JK	0.000319853	0

Based on Table 2, the comparison of the resulting portfolio weights for BRPT.JK stock is 0.999270409 for the Genetic Algorithm and 0.999574509 for the Markowitz Theory. AKRA.JK stock amounting to 0.0000712425 for the Genetic Algorithm and 0.000425491 for the Markowitz Theory. UNVR.JK stock of 0.000319853 for the Genetic Algorithm and 0 for the Markowitz Theory. This shows that in the Markowitz Theory the resulting portfolio only consists of 2 stocks, namely BRPT.JK and UNVR.JK. Furthermore, the comparison of expected return, risk and Sharpe Ratio of the portfolio can be seen in Table 3.

Table 3. Expected Return, Risk and Sharpe Ratio

Method	Genetic Algorithm	Markowitz Theory
E(Rp)	3.348458	3.347559975
Variance	560.6248	560.3980032
Standard Deviation	23.67752	23.67272699
Sharpe Ratio	0.1393076	0.13929785

Based on Table 3, the amount of expected return on the portfolio generated using Genetic Algorithm and Markowitz Theory are 3.348458 and 3.347559975, respectively. The variance of the portfolio is 560.6248 and 560.3980032, respectively. The standard deviation of portfolio is 23.67752 and 23.67272699. Meanwhile, the value of Sharpe Ratio is 0.1393076 and 0.13929785, respectively. Based on these results, the Genetic Algorithm method produces higher expected return and Sharpe Ratio than Markowitz Theory. However, with higher expected return and Sharpe Ratio, the risk of the portfolio generated using Genetic Algorithm is higher than Markowitz Theory [36]. This means that the portfolio produced by Genetic Algorithm has a better performance and better return than Markowitz Theory. Furthermore, optimization using Genetic Algorithm is more optimal than using Markowitz Theory.

4. Conclusion

Three clusters are made using SOM to get three stocks for portfolio. The first cluster consists of 42 stocks, the second cluster consists of 1 stock and the third cluster consists of 2 stocks. The stocks of each cluster which have the largest Sharpe Ratio, namely BRPT.JK, AKRA.JK, and UNVR.JK with Sharpe Ratio of 0.1411508151, -0.00514189, and -0.00028894, respectively. These stocks will be used as a portfolio which will then be optimized. Portfolio optimization is carried out using the Genetic Algorithm method and produces stock weighting of BRPT.JK, AKRA.JK, and UNVR.JK of 0.999270409, 0.0000712425, and 0.000319853 respectively. The expected return and the risk are 3.348458 and 23.67752, respectively. The value of Sharpe Ratio is 0.1393076. Meanwhile, there are only two stocks produced using the Markowitz theory, namely BRPT.JK and AKRA.JK stocks with weights of 0.999574509 and 0.000425491 respectively with the expected return, risk and Sharpe Ratio respectively 3.347559975, 23.67272699, and 0.13929785. Based on the Sharpe Ratio measuring, the portfolio which has the best performance is the portfolio that is optimized using the Genetic Algorithm method, which is 0.1393076. This means that optimization using Genetic Algorithm is more optimal than using Markowitz Theory.

Acknowledgment

The authors are thankful to the Department of Mathematics at Gadjah Mada University, Indonesia for their research lab and other crucial technical support along with the Department of Mathematics at Gadjah Mada University.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] W. Zhou, W. Zhu, Y. Chen, and J. Chen, "Dynamic changes and multi-dimensional evolution of portfolio optimization," *Econ. Res. Istraz.* , vol. 0, no. 0, pp. 1–26, 2021, doi: 10.1080/1331677X.2021.1968308.

[2] A. U. Khan, B. Gour, and K. K. Dwivedi, "Self Organizing Map Neural Network and Fuzzy based Method to Identify Profit Making Stocks in Stock Market," *Adv. Comput. Sci. Technol.*, vol. 10, no. 5, pp. 863–873, 2017. Available at: Google Scholar.

[3] S. Goudarzi, M. J. Jafari, and A. Afsar, "A Hybrid Model for Portfolio Optimization Based on Stock Clustering and Different Investment Strategies," vol. 7, no. 3, pp. 602–608, 2017. Available at: Google Scholar.

[4] H. M. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 60, 1952. doi: 10.2307/2975974

[5] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.Inc, 1959. Available at: Google Scholar.

[6] M. Kritzman and D. Turkington, "History, Shocks, and Drifts: A New Approach to Portfolio Formation," *J. Portf. Manag.*, vol. 49, no. 1, 2022, doi: 10.3905/jpm.2021.1.321.

[7] W. Bakry, A. Rashid, S. Al-Mohamad, and N. El-Kanj, "Bitcoin and Portfolio Diversification: A Portfolio Optimization Approach," *J. Risk Financ. Manag.*, vol. 14, no. 7, p. 282, 2021, doi: 10.3390/jrfm14070282.

[8] D. Cheong, Y. M. Kim, H. W. Byun, K. J. Oh, and T. Y. Kim, *Using genetic algorithm to support clustering-based portfolio optimization by investor information*, vol. 61. Elsevier B.V., 2017. doi: 10.1016/j.asoc.2017.08.042

[9] D. Allen, C. Lizieri, and S. Satchell, "In Defense of Portfolio Optimization: What If We Can Forecast?," *Financ. Anal. J.*, vol. 75, no. 3, pp. 20–38, 2019, doi: 10.1080/0015198X.2019.1600958.

[10] O. Ledoit and M. Wolf, "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," *Rev. Financ. Stud.*, vol. 30, no. 12, pp. 4349–4388, 2017, doi: 10.1093/rfs/hhy016.

[11] S. O. Adebiyi, O. O. Ogunbiyi, and B. B. Amole, "Artificial intelligence model for building investment portfolio optimization mix using historical stock prices data," *Rajagiri Manag. J.*, vol. ahead-of-p, no. ahead-of-print, 2021, doi: 10.1108/ramj-07-2020-0036.

[12] V. Brătian, "Portfolio Optimization. Application of the Markowitz Model Using Lagrange and Profitability Forecast," *Expert J. Econ.*, vol. 6, no. 1, pp. 26–34, 2018. Available at: Google Scholar

[13] B. B. Nair, P. K. S. Kumar, N. R. Sakthivel, and U. Vipin, "Clustering stock price time series data to generate stock trading recommendations: An empirical study," *Expert Syst. Appl.*, vol. 70, pp. 20–36, 2017, doi: 10.1016/j.eswa.2016.11.002.

[14] L. Chin, E. Chendra, and A. Sukmana, "Analysis of portfolio optimization with lot of stocks amount constraint: Case study index LQ45," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 300, no. 1, 2018, doi: 10.1088/1757-899X/300/1/012004.

[15] I. . Krislianto, "Analisis dan penilaian kinerja portofolio optimal saham lq 45 pernyataan mengenai tesis dan sumber informasi serta pelimpahan hak cipta," Bogor Agricultural University, 2015. Available at: Google Scholar

[16] D. A. Milhomem and M. J. P. Dantas, "Analysis of New Approaches Used in Portfolio Optimization: a Systematic Literature Review," *Production*, vol. 30, no. 1, pp. 1–16, 2020, doi: 10.1590/0103-6513.20190144.

[17] M. Ozyesil, "Markowitz Portfolio Optimization Model: An Application On Listed Firm On Borsa Istanbul-30 National Stock Index (Bist-30)," no. February, 2021. Available at: Google Scholar

[18] M. S. Paolella, "The univariate collapsing method for portfolio optimization," *Econometrics*, vol. 5, no. 2, pp. 1–33, 2017, doi: 10.3390/econometrics5020018.

[19] L. H. Pedersen, A. Babu, and A. Levine, "Enhanced Portfolio Optimization," *Financ. Anal. J.*, vol. 77, no. 2, pp. 124–151, 2021, doi: 10.1080/0015198X.2020.1854543.

[20] P. Xidonas, R. Steuer, and C. Hassapis, "Robust portfolio optimization: a categorized bibliographic review," *Ann. Oper. Res.*, vol. 292, no. 1, pp. 533–552, 2020, doi: 10.1007/s10479-020-03630-8.

[21] K. Dowd, *Measuring Market Risk*. England: John Wiley & Sons Ltd, 2002. Available at: Google Books

[22] M. Gen and R. Cheng, *Genetics Algorithms and Engineering Optimization*. New York: John Wiley & Sons.Inc, 2000. Available at: Google Scholar

[23] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic algorithm: Reviews, implementations and applications," *International Journal of Engineering Pedagogy*, vol. 10, no. 6. 2021, doi: 10.3991/IJEP.V10I6.14567.

[24] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimed. Tools Appl.*, vol. 80, no. 5, 2021, doi: 10.1007/s11042-020-10139-6.

[25] M. Kordos, J. Boryczko, M. Blachnik, and S. Golak, "Optimization of warehouse operations with genetic algorithms," *Appl. Sci.*, vol. 10, no. 14, 2020, doi: 10.3390/app10144817.

[26] D. . Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989. Available at: Google Scholar

[27] M. Zanjirdar, "Overview of Portfolio Optimization Models," *Adv. Math. Financ. Appl.*, vol. 5, no. 4, pp. 419–435, 2020. Available at: Google Scholar

[28] W. . Sharpe, "Mutual Fund Performance," *J. Bus.*, vol. 39, no. 1, pp. 119–138, 1966, doi: 10.1086/294846.

[29] S. K. Kumari, P. Kumar, J. Priya, S. Surya, and A. K. Bhurjee, "Mean-value at risk portfolio selection problem using clustering technique: A case study," in *AIP Conference Proceedings*, 2019, vol. 2112, doi: 10.1063/1.5112363.

[30] M. A. Jishan, K. R. Mahmud, A. K. Al Azad, M. S. Alam, and A. M. Khan, "Hybrid deep neural network for bangla automated image descriptor," *Int. J. Adv. Intell. Informatics*, vol. 6, no. 2, 2020, doi: 10.26555/ijain.v6i2.499.

[31] N. Bnouachir and A. Mkhadri, "Efficient cluster-based portfolio optimization," *Commun. Stat. Simul. Comput.*, vol. 50, no. 11, 2021, doi: 10.1080/03610918.2019.1621341.

[32] Y. E. Putra, D. Saepudin, and A. Aditsania, "Portfolio Selection of KOMPAS-100 Stocks Index Using B-Spline Based Clustering," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 375–382, 2021, doi: 10.1016/j.procs.2021.01.019.

[33] A. Pickens and S. Sengupta, "Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means, Fuzzy C-Means and Evolutionary Neural Networks," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 3, pp. 695–719, 2021, doi: 10.3390/make3030035.

[34] L. Gubu, D. Rosadi, and Abdurakhman, "Classical portfolio selection with cluster analysis: Comparison between hierarchical complete linkage and Ward algorithm," in *AIP Conference Proceedings*, 2019, vol. 2192, doi: 10.1063/1.5139174.

[35] Sukono, S. Supian, H. Napitupulu, Y. Hidayat, and A. S. Putra, "The application of genetic algorithm optimization on quadratic investment portfolio without a risk-free asset under Value-at-Risk," in *Journal of Physics: Conference Series*, 2018, vol. 1090, no. 1, doi: 10.1088/1742-6596/1090/1/012026.

[36] B. Y. Qu, Q. Zhou, J. M. Xiao, J. J. Liang, and P. N. Suganthan, "Large-Scale Portfolio Optimization Using Multiobjective Evolutionary Algorithms and Preselection Methods," *Math. Probl. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/4197914.