

An Optimal Stock Market Portfolio Proportion Model Using Genetic Algorithm

Wahyono^{*1}, Chasandra Puspitasari², M. Dzulfikar Fauzi³, Kasliono⁴, Wahyu Sri Mulyani⁵, Laksono Kurnianggoro⁶

¹Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

^{2,3,4,5}Master Program in Computer Science, FMIPA UGM, Yogyakarta, Indonesia

⁶Department of Electrical Engineering, University of Ulsan, Ulsan, Korea

e-mail: ^{*1}wahyo@ugm.ac.id, ²chasandra.p@mail.ugm.ac.id,

³muhammad.dzulfikar.f@mail.ugm.ac.id, ⁴kasliono@mail.ugm.ac.id,

⁵wahyu.sri.m@mail.ugm.ac.id, ⁶kurnianggoro@mail.ulsan.ac.kr

Abstrak

Untuk menekan jumlah kerugian karena resiko investasi, seorang investor atau pialang saham biasanya membentuk portofolio saham optimal. Teknik tersebut dilakukan untuk mendapat return of investasi yang maksimal atas saham yang akan dibeli. Akan tetapi dalam membentuk suatu portofolio saham diperlukan perhitungan yang cukup kompleks serta keahlian tertentu. Penelitian ini bertujuan memberikan alternatif solusi dalam permasalahan pembentukan komposisi portofolio saham yang efisien dan optimal dengan cara merancang sebuah sistem yang mampu membantu pengambilan keputusan bagi investor ataupun pialang saham dalam menyusun portofolio saham yang sesuai dengan kebijakan dan resiko investasi. Pada penelitian ini penentuan komposisi portofolio saham optimal dilakukan dengan menggunakan Algoritma Genetika. Data yang digunakan dalam proyek ini adalah 4 saham terpilih yang terdapat pada indeks LQ45 pada tahun 2017. Sedangkan, perhitungan tingkat keuntungan dan tingkat kerugian menggunakan teori single index model. Efektifitas algoritma telah dievaluasi berdasarkan jumlah populasi serta probability proses crossover dan mutasi pada algoritma genetika. Dari hasil eksperimen didapatkan bahwa framework yang diusulkan bisa dijadikan alternatif solusi dalam penentuan portofolio saham yang optimal.

Kata kunci—algoritma genetika, indeks LQ45, portofolio saham, single index model

Abstract

To reduce the amount of loss due to investment risk, an investor or stockbroker usually forms an optimal stock portfolio. This technique is done to get the maximum return on investment on shares to be purchased. However, in forming a stock portfolio required a fairly complex calculations and certain skills. This work aims to provide an alternative solution in the problem of forming the optimal and efficient stock portfolio composition by designing a system that can help decision making of investors or stockbrokers in preparing stock portfolio in accordance with the policy and risk investment. In this work, determination of optimal stock portfolio composition is constructed by using Genetic Algorithm. The data used in this work are the 4 selected stocks listed on the LQ45 index in 2017. Meanwhile, the calculation of profit and loss rate utilizes a single index model theory. The efficiency of the algorithm has been examined against the population size and crossover and mutation probabilities. The experimental results show that the proposed algorithm can be used as one of the solutions to select the optimal stock portfolio.

Keywords—genetic algorithm, LQ45 index, stock market portfolio, single index model

1. INTRODUCTION

Investment is a commitment of investors in the form of some funds or other resources currently undertaken to obtain some profits in the future. The purpose of investors in investing is to get the maximum return on the stock they bought. However, various problems arise due to a large number of stock investment instruments circulating in the capital market. These instruments have risks to be considered by each investor, while the analytical capability of investors is still relatively limited. Thus, this limitation is very influential on stock investment decisions.

To reduce risk investment, investors should have more stock options to invest. One of the strategies for avoiding risk on an investment is through diversification of stocks by forming a portfolio, called stock portfolio. An optimal stock portfolio will produce optimal returns with a moderate risk that can be accounted for. However, investors are faced with uncertainty selection when it comes to choosing the stock to be formed into their preferred portfolio. With thousands of stocks in the market, deciding which portfolio should be chosen is quite difficult. The combination of multiple stocks with different qualities arises difficulty for investors to determine the proportion of funds to be invested. On the other hand, creating the mathematical model is more difficult and very complex. Therefore, the artificial intelligence approach is needed to solve the problem.

For solving major problems in forming of an optimal stock portfolio, a recommendation system [1] is required by utilizing the appropriate algorithm. This algorithm should be able to be used in selecting the investment portfolio composition in accordance with the investment policy of an investor [2]. There are several methods can be used for this problem, such as linear programming, mathematical modeling, data mining, and optimization using a genetic algorithm (GA). Among these algorithms, the GA is preferred to be chosen as it has a capability to find the optimal and high-quality solution in solving both constrained and unconstrained optimization problems within large solution space [3]. There have been previous studies that have implemented genetic algorithms [2, 4, 5, 6] as well as data mining [7] on a stock portfolio recommendation system. Unfortunately, the proposed method is still a decision whether to take or not a particular stock. In contrast to previous research, the proposed method provides a recommendation system in the form of the proportion of funds to be invested in certain shares. In this work, the Genetic Algorithm is applied to get the best chromosome in the form of the proportion of each stock to get the portfolio composition with the optimum level of profit and the level of loss that can be accounted for. The data used in this project are 4 selected stocks listed on the LQ45 index in 2017. LQ45 is a stock market index in Indonesia Stock Exchange (BEI) consisting of 45 companies that meet certain criteria. Meanwhile, the calculation of profit and loss rate using single index model theory.

The rest of this paper is organized as follows. In Section 2, the detail explanation of stock portfolio and how to compute the risk and return rate are provided. In Section 3, the proposed framework of the genetic algorithm is explained. The the results of our proposed method are offered in Section 4. Lastly, Section 5 presents the conclusion of this paper.

2. METHODS

2.1 Stock Portfolio

The selection of stock data is based on the companies listed in the LQ45 index. In this case, PT Telekomunikasi Indonesia Tbk (TLKM), PT Jasa Marga (Persero) Tbk (JSMR), PT Perusahaan Gas Negara (Persero) Tbk (PGAS), and PT Semen Indonesia (Persero) Tbk (SMGR) were selected. The data were taken in the form of value of close price and stock

dividend in the range between January 1, 2017 and April 30, 2017. Therefore, there will be 82 data from each company.

A portfolio is a grouping of financial assets such as stocks, bonds, commodities, currencies and cash equivalents, as well as their fund counterparts, including mutual, exchange-traded and closed funds. The stock portfolio is an investment consisting of different stocks of different companies in the hope that if one stock decreases while the other increases, the investment will not incur losses [4]. In addition, the correlation of return between one stock and other stock will also reduce the variance of the portfolio return.

Investors should construct an investment portfolio in accordance with their risk tolerance and their investing objectives. The investment objective is to maximize return/profit, without neglecting the risk factor (risk) of investment. Return is one of the factors that motivates investors to invest and is a reward for the courage investors bear the investment risk.

To determine the return and risk of the stock portfolio, the following systematic steps are considered [8]:

1. *Determination of profit rate of actual return each year on each company.* To determine the actual return, the following formula is used:

$$R_{ij} = \frac{P_t - P_{t-1} + D}{P_{t-1}} \quad (1)$$

where

- R_{ij} = The profit level of individual stocks for an investment term
- P_t = The closing price of the stock period t
- P_{t-1} = Stock closing price of previous period
- D = Shared share dividends

2. *Calculating the expected profit value (expected return) of each company.*

$$E(R_i) = \frac{\sum R_{ij}}{N} \quad (2)$$

where

- $E(R_i)$ = The expected rate of return
- R_{ij} = The profit level of each individual stock
- N = The number of investment period

3. *Risk Calculation*

After calculating the actual and expected returns, the next step is to determine the value of risk by calculating the value of variance and standard deviation. The risk is expressed by a standard deviation which can be calculated by the following equation:

$$\sigma^2 = \frac{\sum [R_{ij} - E(R_i)]^2}{N} \quad (3)$$

4. *Stock portfolio combination from stock index LQ45.*

In this stage, the company's stocks are combined in forming portfolios. It then determines the proportion of funds to be invested in each company. The proportion of funds is determined randomly and should be equal to one. The calculation of the proportion of funds will be used to calculate expected returns (profit) from a combination of the stock portfolio.

5. *Determining expected return of portfolio risk.*

$$E(R_p) = \sum_{i=0}^N [X_i E(R_i)] \quad (4)$$

where

- $E(R_i)$ = The expected profit (return) from i-th stock.
 $E(R_p)$ = The expected profit from portfolio
 X_i = The proportion of funds invested in the i-th stock

6. *Calculating of correlation coefficient value of inter-company stock price*

$$\rho_{12} = \frac{N \sum XY \cdot \sum X \cdot \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (5)$$

where

- ρ = Correlation coefficient
 N = The number of stock periods
 X = The profit level of the first stock
 Y = The profit level of the second stock

7. *Determining variance and standart deviation of stocks*

$$\sigma_p^2 = X_1 \sigma_1^2 + X_2 \sigma_2^2 + 2(X_1 X_2 \rho_{12} \sigma_1 \sigma_2) \quad (6)$$

8. *Calculation of coefficient of variance (CV)*

$$CV = \frac{\sigma_p}{E(R_p)} \quad (7)$$

2.2 Genetic Algorithm

The genetic algorithm is one of computationally intelligence techniques, an optimization technique based on natural genetics. In a genetic algorithm, the search process is carried out between a number of alternative optimal points based on a probabilistic function to produce an optimal solution [3]. Compared to other optimization algorithms, genetic algorithms in many cases provide fairly high accuracy as in the case of scheduling [9], financing [10], and even in robot control systems [11]. So the use of genetic algorithm in the case of stock portfolio selection is expected to provide optimal results with high accuracy.

Algorithm 1. Stock portfolio selection using genetic algorithm

- 1: **Generate** n initial chromosome
 - 2: **Compute** fitness value of each chromosome
 - 3: **Sort** chromosome in descending order based on fitness value
 - 4: **repeat**
 - 5: **for** $i=1$ to k **do**
 - 6: **select** randomly two chromosomes (P_1, P_2)
 - 7: $child_i \leftarrow$ crossover(P_1, P_2)
 - 8: $child_i \leftarrow$ mutation($child_i$)
 - 9: **compute** fitness value of $child_i$
 - 10: **end for**
 - 11: **merge** k new children to the population
 - 12: **sort** $n+k$ chromosomes in descending order based on fitness value
 - 13: **remove** k worst chromosomes from the population
 - 14: **until** a fixed number of iteration
-

The framework of the GA used in this work for selecting optimal portfolio proportion is shown in Algorithm 1. The selection of an optimal portfolio relies on the investment policy of the investor. The first stage is to generate n initial chromosomes, where each chromosome consists of l genes. Each gene is representing the proportion of fund to be invested from i^{th}

stock. For each chromosome, the fitness value is calculated. Thus, the chromosomes can be sorted out based on their fitness value.

Fixed number of iteration is used as the termination criteria. In each iteration step, k -offspring (children) are generated based on the crossover and mutation processes. These process will be explained further in section 2.3.4 and 2.3.5. These new offspring are added to the population and then sorted out together with the previously n chromosomes in descending order. To maintain the quality of the population, k worst chromosomes are removed from the population. Hence, the number of population member is not changing in each iteration step (i.e. generation).

Reproduction process is done by using crossover and mutation operators according to their probability values. Furthermore, individuals from the initial and new offspring populations of crossover and mutation results are combined for the selection process. Selection is done by calculating the fitness of each individual. The best individuals are individuals who have the best fitness after a fixed number of iteration. Factors affecting fitness are the expected return (profit) and risk portfolios.

2. 2.1 Parameters Initialization

At the initialization stage, there are five parameters which should be considered such as population size (n), generation size, probabilities of crossover and mutation (pc and pm), and the length of chromosomes (l). The population size is the number of chromosomes in a population. Generation sizes (iteration) are used to determine how many generations to apply. The probabilities of crossover and mutation are used to determine the number of offspring (chromosomes) produced. In addition, the length of the chromosome is equal to the number of stock to be included in the portfolio. In this paper, the length of chromosomes is set to be four.

2. 2.2 Chromosome Representation

In this work, the chromosome is represented by real-values, which represents the proportion of funds for certain stock. The length of the chromosome is the number of stocks in the portfolio. A chromosome is formed by generating a random value within interval $[0,1]$. In this case, the total gene values should be equal to 1. Otherwise, the chromosome should be classified as invalid chromosome. Table 1 is an example of a chromosome representation used in this work.

Table 1 Chromosome Representation Sample

| TLKM | JSMR | PGAS | SMGR |
|------|------|------|------|
| 0,35 | 0,15 | 0,27 | 0,23 |

As can be seen in Table 1, there are 4 stocks to be formed into a stock portfolio, the chromosome has a length of 4 genes. The contents of each gene shown by index 1st to the 4th index indicate the proportion of stock.

2. 2.3 Fitness Function Calculation

In this case, the calculation of the profit and loss rate is performed based on the single index model theory [8]. Thus, the profit rate of the stock portfolio can be calculated by the following equation:

$$R_p = \alpha_p + \beta_p R_m \quad (8)$$

While the risk of stock portfolio can be calculated by the following formula:

$$\sigma_p^2 = \beta_p^2 \sigma_m^2 + \sum_{i=1}^N W_i^2 \sigma_{ei}^2 \quad (9)$$

The purpose of forming a stock portfolio is to obtain an optimal level of profit with low-level risk. Therefore, the formula in equation (10) is used as fitness function, where the higher value of fitness is better.

$$Fitness = \frac{R_p}{\sigma_p^2} \quad (10)$$

2. 2.4 Crossover

Crossovers are used to generate new individuals with different genes from the previous individual i.e. parents. In this work, the crossover is done with extended intermediate crossover. Extended intermediate crossover produces offspring from the combination of two parent chromosome [3]. The number of offspring generated in the crossover process is $pc \times n$. Suppose that P_1 and P_2 are two chromosome parents, then offspring C_1 and C_2 can be generated as follows:

$$\begin{aligned} C_1 &= P_1 + \alpha(P_2 - P_1) \\ C_2 &= P_2 + \alpha(P_1 - P_2) \end{aligned} \quad (11)$$

Table 2 Illustration of Crossover ($\alpha = 0.5$)

| | TLKM | JSMR | PGAS | SMGR |
|--------------------|------|-------|-------|-------|
| Parent 1 (P1) | 0.35 | 0.15 | 0.27 | 0.23 |
| Parent 2 (P2) | 0.35 | 0.10 | 0.30 | 0.25 |
| Child 1 (C1) | 0.35 | 0.125 | 0.285 | 0.27 |
| Normalized Child 1 | 0.34 | 0.121 | 0.277 | 0.262 |

The value of α is randomly set at predetermined intervals [0, 1]. In some cases, the total gene values can be more than one. To overcome this problem, the obtained offspring must be normalized using equation (12). Table 2 illustrates how the crossover process is performed with value $\alpha = 0.5$

$$newGene = \frac{oldGene}{\sum_{i=1}^4 oldGene} \quad (12)$$

2. 2.5 Mutation

The mutation operator on the genetic algorithm modifies an offspring (a child's chromosome) after crossover operation. The mutation is performed based on probability. To perform this process, reciprocal exchange mutation is utilized. It selects two positions (exchange point / XP) randomly on the parent chromosomes and then exchanges the value at each position [3]. For example, let the chromosome (0.34, 0.121, 0.277, 0.262) be selected for mutation, and genes 0.34 and 0.121 are swapped in cluster 1, and genes 0.277 and 0.262 are swapped in cluster 2; then the mutated chromosome becomes (0.121, 0.34, 0.262, 0.277).

2. 2.6 Selection

The selection process is done to obtain the best chromosome in the population for next generation. The selection process is performed by elitism, which sorts out the chromosomes based on the highest to the lowest fitness values and then keeps n best chromosome as a new population in the next generation, while the remaining chromosomes will be removed. After the selection process, the best chromosome with the highest fitness value is selected as optimal solution of stock portfolio proportion in each generation.

3. RESULTS AND DISCUSSION

3.1 Experiment Setting

The proposed method has been implemented using Java programming language using PC under Windows operating system. The experiments were conducted by evaluating the effect of population size as well as the effect of crossover probability and mutation probability.

3.2 Stock Market Data Acquisition

The selection of stock data is based on the companies listed in the LQ45 index. In this case, PT Telekomunikasi Indonesia Tbk (TLKM), PT Jasa Marga (Persero) Tbk (JSMR), PT Perusahaan Gas Negara (Persero) Tbk (PGAS), and PT Semen Indonesia (Persero) Tbk (SMGR) were selected. The values of close price and stock dividend in the range between 1 January 2017 and 30 April 2017 are selected as the data. Therefore, there are 82 data from each company.

3.3 Population Size Effect

The first experiment was conducted by using probability crossover 0.5 and probability mutation 0.2 to obtain optimal population size. The population size is set to be between 20 until 140 chromosomes, within increment 20. The genetic algorithm is a stochastic method which will produce different results each time it is run. So that, each population size will be performed 5 times and compute the average of the best fitness value on each running. The experimental results using several population sizes are shown in Table 3 and Figure 1.

Table 3 Results in Different Population Size

| Population Size | Fitness Value |
|-----------------|---------------|
| 20 | 0.99 |
| 40 | 3.48 |
| 60 | 0.4 |
| 80 | 1.61 |
| 100 | 3.81 |
| 120 | 1.93 |
| 140 | 0.74 |

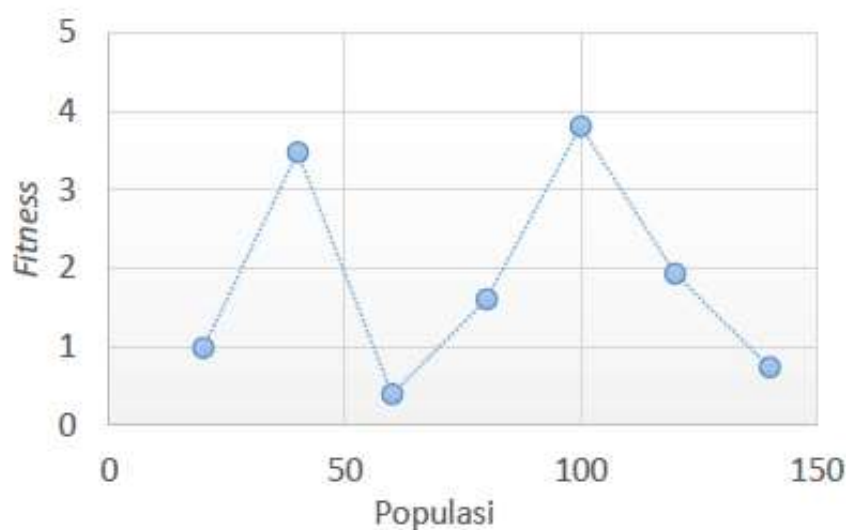


Figure 1 Results in different population size

The best average fitness is 3.81 when the population size is equal to 100. The obtained population sizes in this experiment is the optimal population that can be used for the determination of the optimal stock portfolio proportion. This population size will be used for the next experiment.

3.4 Crossover and Mutation Probability Effect

In the second experiment, the crossover probability (pc) and the mutation probability (pm) were examined by combining the probability value of both operators. Both probabilities were set to the combination of 0.1 until 1 with increment 0.1, respectively. So there will be 10 experiments.

To evaluate the effect of crossover and mutation probability, the population size is set to be 100 population, the best population size previously obtained. Each combination pc and pm sizes is tested 5 times and we select the average of the highest fitness value.

Table 4 Results in Combination of Crossover and Mutation Probability

| i^{th} trial | PC | PM | Fitness Value |
|-----------------------|-----|-----|---------------|
| 1 | 1 | 0 | 1.43 |
| 2 | 0.1 | 0.9 | 1.35 |
| 3 | 0.2 | 0.8 | 1.38 |
| 4 | 0.3 | 0.7 | 2.47 |
| 5 | 0.4 | 0.6 | 1.69 |
| 6 | 0.5 | 0.5 | 3.9 |
| 7 | 0.6 | 0.4 | 2.21 |
| 8 | 0.7 | 0.3 | 3.9 |
| 9 | 0.8 | 0.2 | 2.7 |
| 10 | 0.9 | 0.1 | 1.97 |

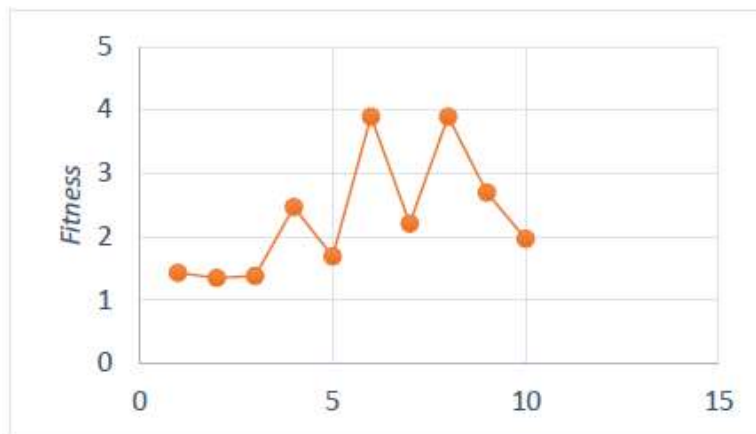


Figure 2 Results on combination of crossover and mutation probabilities

From Table 4 and Figure 2, it can be seen that the highest fitness value is 3.9 with a crossover probability value of 0.3 and a mutation probability of 0.7 or 0.5: 0.5.

It can be seen on the experimental results in Figure 3, an individual with the highest fitness values has chromosomes $\{0; 1; 0; 0\}$, so the portfolio formed from the chromosome is 100% in JSRM stock, while the proportion of the other three stocks are 0 (zero). It can be concluded that it is not recommended to take the stocks of TLKM, PGAS, and SMGR.

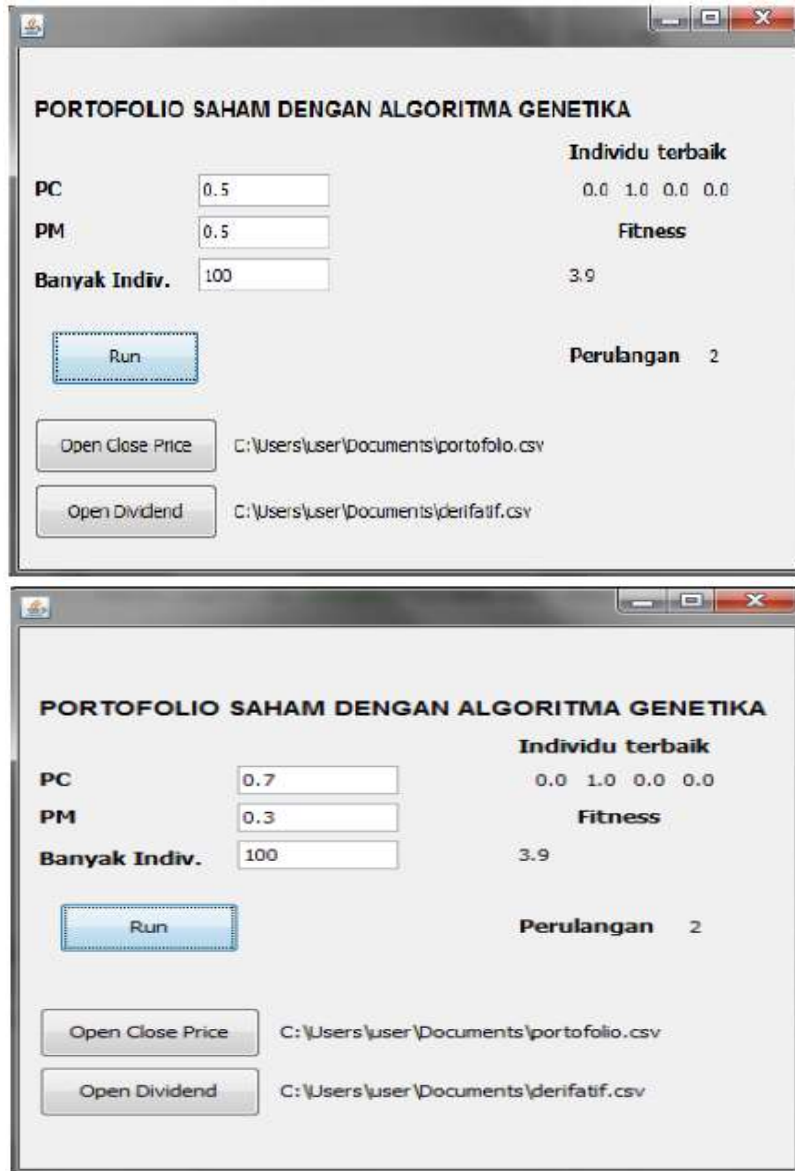


Figure 3 The best result on crossover and mutation probability combination of 0.5:0.5 (top) and 0.7:0.3 (bottom)

4. CONCLUSIONS

Genetic algorithm with real-valued chromosomes representation has been implemented. The GA utilizes extended intermediate crossover, reciprocal exchange mutation, and elitism selection. It was applied to determine the proportion of funds on stock portfolio. Based on the experiments, it was found that the best population size is 100 chromosomes with the ratio of crossover and mutation probabilities either 0.5:0.5 or 0.7: 0.3. The proportion of stock acquired is 100% of JSMR's stock. According to this results, it can be concluded that the genetic algorithm can determine the proportion of stocks with greater profit levels and smaller losses.

However, the number of stock involved in this work is still limited to four. Therefore, in future work, more stocks can be considered to determine the proportion of its investment.

ACKNOWLEDGMENTS

The authors would like to thank the Department of Computer Science and Electronics, FMIPA, Universitas Gadjah Mada that have given financial support to this research and publication.

REFERENCES

- [1] L. McGinty, and B. Smyth, "Adaptive selection: analysis of critiquing and preference based feed back in conversation on recommender system." *International J Electron Commerce* vol. 11, no. 2, p. 35-57. 2006
- [2] R. Wahyuni, W. F. Mahmudi, and B. D. Setiawan, Determination of Stock Portfolio Using Genetic Algorithm," (in Indonesian) Penentuan Portofolio Saham Optimal Menggunakan Algoritma Genetika, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Universitas Brawijaya, Malang, 2017.
- [3] W. F. Mahmudi, "The introduction of genetic algorithm," (in Indonesia) Dasar-dasar Algoritma Evolusi, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya, Malang, 2015.
- [4] C. Fiarni and Bastiyan, "Recommender System of Stock Portfolio Based on Genetic Algorithm," (in Indonesian) Sistem Rekomendasi Portofolio Investasi Berbasis Algoritma Genetika, *Seminar Nasional Sistem Informasi Indonesia.*, 2013.
- [5] A. Shritashava and A. Singh, "An Optimal Stock Portfolio Construction Model Using Genetic Algorithm," *Proceedings of 2013 International Conference on Machine Intelligence and Research Advancement (ICMIRA)*, 21-22 December 2013.
- [6] Y.-H. Chou, S.-Y. Kuo, and Y.-T. Lo, "Portfolio Optimization Based on Funds Standardization and Genetic Algorithm," *IEEE Access*, vol. 5, p. 21885-21900, 2017.
- [7] P. Paranjape-Voditela and U. Deshpande, "A stock market portfolio recommender system based on association rule mining," *Applied Soft Computing*, vol. 13, no. 2, February 2013, p. 1055-1063, [Online] Available: <https://www.sciencedirect.com/science/article/pii/S1568494612004322> [Accessed: 31-May-2018]
- [8] S. Husnan, "The Basic Theory of Stock Portfolio and Analysis", (in Indonesian) Dasar-Dasar Teori Portofolio dan Analisis Sekuritas, Unit Penerbit dan Percetakan AMP YKPN, Indonesia. 2005.
- [9] A. K. Wardana and S. Hartati, "Scheduling System of Pencak Silat Based on Genetic Algorithm," (in Indonesia) Sistem Penjadwalan Pertandingan Pencak Silat Berbasis Algoritma Genetika, *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 11, no. 2, p. 177, July. 2017 [Online]. Available: <https://jurnal.ugm.ac.id/ijccs/article/view/24214>. [Accessed: 24-May-2018]
- [10] Z. Indra and Subanar, "Optimasi Biaya Distribusi Rantai Pasok Tiga Tingkat dengan Menggunakan Algoritma Genetika Adaptif dan Terdistribusi," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 8, no. 2, July. 2014 [Online]. Available: <https://jurnal.ugm.ac.id/ijccs/article/view/1062>. [Accessed: 24-May-2018]
- [11] K. Hakiim, A. Darmawan, and Faizah, "Optimasi Kendali PID menggunakan Algoritma Genetika untuk Penerbangan Quadrotor," *IJEIS (Indonesian J. Electron. Instrum. Syst.)*, vol. 7, no. 2, October. 2017 [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/2321>. [Accessed: 29-May-2018]