

Prediction of Indonesian Inflation Rate Using Regression Model Based on Genetic Algorithms

Faisal Dharma Adhinata¹, Shabrina², Astrid Noviana³, M. Tahir⁴, Nirwana Hendrastuty⁵, Wahyono^{6*}

^{1, 2, 3, 4, 5}Master Program in Computer Science, Universitas Gadjah Mada, Yogyakarta, Indonesia

⁶Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Article history:

Received Sep 3, 2019

Revised May 17, 2020

Accepted June 28, 2020

Keywords:

Inflation Level

Prediction

Regression

Genetic Algorithm

Mean Squared Error

ABSTRACT

Inflation occurs when there is a general and continuous increase in the price of goods or services in a country. This study aims to predict future inflation levels to aid the government in formulating future economic policies. The method used involves studying historical Consumer Price Index (CPI) data and applying a genetic algorithm-based regression model. The research has found that regression methods are frequently used to solve prediction problems, and the problem of finding the optimal prediction model can be addressed using genetic algorithms. The model is trained and evaluated using real CPI data obtained from the Indonesian Central Bank. Based on the experiment, it was proved that the proposed model is effective in predicting the inflation rate, as it achieved a Mean Squared Error (MSE) value of 0.1099. Understanding these predictions is crucial for maintaining economic stability and planning.

Corresponding Author:

Wahyono,
 Department of Computer Science and Electronics,
 Universitas Gadjah Mada,
 Sekip Utara Bulaksumur, Yogyakarta, Indonesia
 Email: wahyo@ugm.ac.id

1. INTRODUCTION

Developed countries have a strong economy, stability and equitable economic growth. The Indonesian state, through Bank Indonesia, focuses on the goal of maintaining rupiah stability. One of the main indicators of currency stability is through goods and services reflected through the development of the inflation rate [1].

Inflation can be an indicator in describing the general trends of price developments. One indicator that is often used to measure inflation is the Consumer Price Index (CPI). Changes in the CPI over time represent the rate of increase or decrease in inflation of goods and services [2], [3]. Inflation has a positive and negative impact on the Indonesian economy. A decline in inflation will have a positive effect on driving the economy better. Conversely, when an increase in inflation happens, it will cause the economy to deteriorate resulting in higher prices. Uncontrolled inflation will cause the decline of the Indonesian economy. Therefore, predictions of future inflation rates are very necessary for the government to formulate economic policies in the future.

Prediction of inflation rates can be done by studying historical CPI data. This method is often known as the time-series data prediction method. Prediction is defined as the forecasting process of a variable (event) in the future based on previous variable data. One prediction method uses the regression method. The regression method is a measurement method based on the relationship between one variable and many other variables. Regression methods have often been used, such as research using historical data regression methods, to predict stock prices [4], [5], goods stock [6], [7], and gold prices (using multiple linear regression) [8], [9].

The problem of finding the optimal prediction model can be seen as an optimization problem. Genetic algorithms are generally used to deal with the optimization problems. "Genetic Algorithm" is a solution-search algorithm with a system that adopts a natural selection process, in which there is a group of individuals in a population representing a set of solutions that will later experience natural selection. Some studies that use genetic algorithms include producing the best coefficients that can produce stock prediction prices that are close to the original price [4]. Then research genetic algorithms successfully predict currency exchange rates [10], [11].

The structure of this paper in Section 2 discusses inflation theory and research methods. Section 3 contains a discussion of the results of the implementation of the method while Section 4 concludes the paper.

2. METHOD

2.1 Inflation

Inflation is a tendency for the average price of goods and services to increase steadily over time. If the price of goods and services in the country increases, so will the inflation. The rising price of goods and services causes a decrease in the value of money. Thus, inflation can also be interpreted as a decrease in the value of money towards the value of goods and services in general [2]. The indicator that is often used to measure inflation is the Consumer Price Index (CPI). Equation (1) is an inflation calculation based on the CPI.

$$\text{Inflation} = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} \times 100\% \quad (1)$$

where

CPI_t : Current month CPI.

CPI_{t-1} : Previous month CPI.

For example, the CPI in March 2019 was 135.87, then the CPI in April 2019 was 136.47. The calculation process of inflation can be written as follows:

$$\text{Inflation} = \frac{136.47 - 135.87}{135.87} \times 100\% = 0.44$$

2.2 Regression-Based Approach

The regression method is a measurement method based on the relationship between one variable or many other variables. For example, if we know something that can cause changes in value in the past, we will try to identify the relationship between these variables and the change in value that occurs. This regression method seeks to connect between the existing dependent variables and the given independent variable.

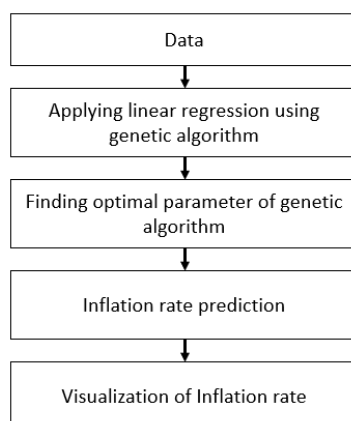


Figure 1. The flowchart of inflation prediction stages.

Multiple regression analysis was used to predict the value of one dependent variable based on two or more independent variables. Regression analysis produced an equation / regression model. This study used Genetic algorithms for optimization in MLR to predict inflation rates in Indonesia. In Figure 1 the stages of the study are shown.

2.3 Genetic Algorithm

Genetic algorithms are computational algorithms that are inspired by the evolutionary process of finding optimal solutions naturally [12], [13]. This algorithm has been widely used, because it can handle various problems that have complex search space, using mathematical models with high accuracy values [14]. So that genetic algorithms are suitable to be applied in various fields, especially cases such as prediction or forecasting for the future, such as predictions on currencies [11], stock prices [15], marketing [10], image [16], document clustering [17], and transportation [18]. For this reason, genetic algorithms were considered

appropriate and chosen to find the optimal solution in solving the inflation rate problem based on the Consumer Price Index data in Indonesia. Figure 2 shows the stages of the Genetic algorithm.

The first stage in genetic algorithms is the initialization of individuals, namely creating individuals with a random array of genes (chromosomes). This chromosome represents the solution to the problem. The

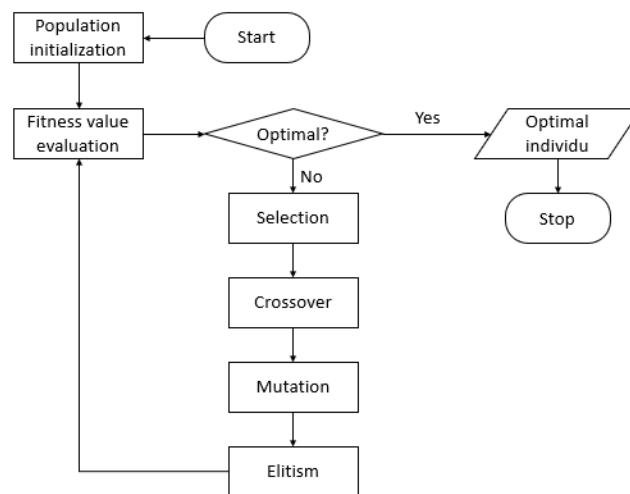


Figure 2. The flowchart of genetic algorithm.

next stage is reproduction in the form of a crossover and mutation process to produce new individuals in the population. Each chromosome has fitness, the greater the fitness value the better for the chromosome to be a solution. Evaluation is a step to calculate this fitness value. The final stage is selection, which involves choosing individuals from the population set and offspring. The best individuals as a result of selection are kept alive in the next generation [4], [18], [19].

2.3.1 Population Initialization

In general, for the case of prediction using genetic algorithms, there is the main process, namely the search for optimal historical data patterns using the regression method [20]. This process aimed to find a pattern that most closely characterizes the historical data of the inflation rate in Indonesia. Multiple linear regression (MLR) was used in this study, aiming to find a pattern that most closely characterize the historical inflation rates in Indonesia, which was represented in the form of linear functions as in equation (2). The initialization process using this function was used to build a prediction model using training data.

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n \quad (2)$$

where:

- Y : Prediction of the Consumer Price Index on month B
- $x_1 \dots x_n$: Consumer Price Index in months B₁ to B_n
- $\theta_0 \dots \theta_n$: Representation of each gene on a chromosome with random values

2.3.2 Chromosome Representation

Chromosome representations use real numbers that are set randomly from 0 to 1. Representation of real numbers is suitable for use because they can represent predictive functions built to predict the inflation rate in this study.

Table 1. Chromosome Representation

θ_0	θ_1	θ_2	θ_3	θ_4
0.8531	0.4113	0.2690	0.9638	0.2079

Table 1 shows that the number of genes on the chromosome can be flexible according to the user's input. This allows many patterns to be compared with each other to get the best chromosome pattern from the entire set of experiments to predict the inflation rate. For example, if we wish to predict the fifth month, the number of genes will be 5.

2.3.3 Fitness Value Calculation

The fitness value (f) used in this study employed MSE (Mean Square Error) with the actual value. The optimal value was obtained when getting the smallest of MSE value, so the fitness value would be greater and vice versa. It obtained a prediction error with equation (3),

$$f = \frac{1}{MSE + \varepsilon} \quad (3)$$

where ε was a very small number to avoid division with zero.

The purpose of predicting the inflation rate was to minimize the error of the prediction with the MSE (Mean Square Error) value, then we squared all the existing error data and divided them with the number of data errors. The MSE value used equation (4).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2 \quad (4)$$

where:

- n : the number of data
- Y : data prediction
- Y' : data ground truth

Prediction calculations were using equation (2).

$$\begin{aligned} Y &= \theta_0 + \theta_1(0.32) + \theta_2(0.24) + \theta_3(0.14) + \theta_4(0.11) \\ &= 0.8531 + 0.4113(0.32) + 0.2690(0.24) \\ &\quad + 0.9638(0.14) + 0.2079(0.11) \\ &= 1.2070 \end{aligned}$$

In this study, the prediction data was 9 years or 108 months. It predicted in the fifth month, so the value of n became 104.

$$\begin{aligned} MSE &= \frac{1}{104} \sum_{i=1}^{104} ((1.2070 - 0.86)^2 + \dots + (1.4396 - 2.46)^2) \\ &= 0.3750 \end{aligned}$$

Then, the value of fitness was written as follows:

$$f = \frac{1}{0.3750 + \varepsilon} = 2.6661$$

2.3.4 Parent Selection

The parent selection was done using the Roulette Wheel Selection. In the Roulette Wheel method, individuals were mapped in a sequential line segment according to their fitness value. Then random numbers were generated as many as N rounds as the parent of the crossover processed. The crossover process could not have the same parent. Then checking was done to remove the redundant parent.

2.3.5 Crossover

The formation of new individuals in a generation was carried out by the crossover method, which was by taking part or pieces of the allele line from the genome DNA to be combined into a hereditary genome or offspring. The crossover process was carried out using the whole arithmetic crossover method and predetermined alpha values. The number of individuals was adjusted by the specified crossover probability (P_c). The selection of genes to be used for the crossover process was chosen randomly, with the following equation:

$$child\ 1 = \alpha \cdot x_i + (1 - \alpha) \cdot y_i, \quad 1 \leq i \leq n \quad (5)$$

$$child\ 2 = \alpha \cdot y_i + (1 - \alpha) \cdot x_i, \quad 1 \leq i \leq n \quad (6)$$

Suppose 2 parents are ready to do the crossover process, namely parent 1 and parent 2. The chromosomes are shown in Table 2.

Parent 1	0.8531	0.4113	0.2690	0.9638	0.2079
Parent 2	0.1988	0.2243	0.1578	0.1399	0.3479

We specified the value of α , which was a random value between 0 - 1. We gave an α value of 0.4. Then, the calculation process for each chromosome can be written as follows:

$$\text{Child 1} = 0.4 \cdot 0.8531 + 0.6 \cdot 0.1988 = 0.4605$$

$$\text{Child 2} = 0.4 \cdot 0.1988 + 0.6 \cdot 0.8531 = 0.5913$$

Crossover was done by selecting random gene values so that the results for all chromosomes of child 1 and child 2 are as shown in Table 3. Crossover calculations are also performed on all genes.

Table 3. Child chromosome from crossover operation ($\alpha=0.4$)

Child 1	0.4605	0.2991	0.2022	0.4694	0.2919
Child 2	0.5913	0.3365	0.2245	0.6342	0.2639

2.3.6 Mutation

The mutation process was carried out on individuals after the results of parent crosses or crossover. This process changed the value of one or several genes in a population that aimed to avoid premature convergence, which was the achievement of value or outcome that had not been or was not maximal [14]. In the mutation process, the number of individuals depended on the probability of the mutation (P_m) that had been determined. This study used the uniform mutation method.

This mutation worked by replacing the selected gene value for mutations with random numbers that already had a limit. The limit was determined so that the resulting solution was a legal solution. The random numbers used in this study were from 0 to 1. For example, the fourth-gen was changed with 0.241. Mutation examples used the uniform method in Table 4.

Table 4 Example of mutation operation

Individual	0,4605	0,2991	0,2022	0,4694	0,2919
Mutated Individual	0,4605	0,2991	0,2022	0,2454	0,2919

2.3.7 Elitism

The elitism process is an individual selection process that will survive and will be brought to the next generation (next iteration). In this study, the selection process was done by sorting the best fitness values of all individuals in the population. Then it was taken to the next generation, only the first population, so it did not increase the number of individuals in the population.

For example, individual reproductions were 20, while the first generation of individuals was 100. The total number of individuals was 120. Therefore, selection was made by sorting the best fitness values to select individuals brought to the next generation with a fixed amount of pop size of 100.

3. RESULTS AND DISCUSSION

In this part, we discussed how the data was collected, the evaluation strategy, and the effect of several parameters in the proposal genetic algorithm schema.

3.1 Data Acquisition

The data used in this study were the historical inflation data in Indonesia from 2006-2018 based on the Consumer Price Index (CPI) taken from the website of the Central Statistics Agency. Data acquisition for this study were grouped into two parts, training data had been taken between 2006 - 2014 or for 108 months and testing had been taken between 2015 - 2018 or for 48 months.

3.2 Evaluation Strategy

The purpose of testing was to get individuals with the best chromosome values from the learning process. The better the chromosome was obtained, the better the results of the prediction. Tests were carried out by taking into account the parameters of the genetic algorithm that were entered to obtain the best fitness value. The parameters included three things, namely the number of genes, the probability of crossover, and the probability of mutation.

The results of the predictions were compared to the existing data so that accuracy could be obtained for the relevant inflation data. The accuracy value of predictions was measured using MSE (Mean Squared Error). Predictions are said to be more accurate and better if the smaller the MSE value is obtained.

3.3 The Effect of The Number of Gene

In the first experiment, variations in the number of genes were carried out to see the best fitness value. The size of the population was 100 individuals and the number of iterations was 50. The value in this population

was chosen so that the number of individuals in the population was not too large and not too small. Variations in the number of genes were carried out with the number of genes from 2 to 10. In this study, a crossover probability of 50% and a mutation probability of 2% was used to obtain the optimal gene. Table 5 shows the results of gene variation experiments.

Table 5. Results of experiment variations in the number of genes

# Trial	# Gene	MSE	<i>Fitness Value</i>
1	2	0.2009	4.9842
2	3	0.2702	3.6999
3	4	0.3357	2.9788
4	5	0.3750	2.6661
5	6	0.3822	2.6159
6	7	0.4713	2.1215
7	8	0.7512	1.3310
8	9	0.8104	1.2338
9	10	1.0583	0.9449

The experimental results obtained that the most optimal number of genes was 2. In the number of genes 2 was the smallest MSE value, which was equal to 0.2009, and the best fitness value was 4.9842. These values indicate that the number of genes 2 has the least error rate. The use of genes with more than 10 resulted in a worse fitness value because it was proven in this experiment that increasing the number of genes decreased the fitness value continuously.

The population size generated and many iterations in this experiment were optimal results used to determine the inflation value for the following months. The population size and many of these iterations were used in subsequent experimental variations.

3.4 The Effect of Mutation Probability

The second experiment was carried out with several combinations of input parameters on mutations namely 0.1 until 0.9 and 0.01 until 0.09 as many as ten tests. The aim was to find a combination of parameters that produced the most optimal prediction function. The number of genes was 2. This was the best result of the previous experiment. The crossover probability used in this experiment was 50%. Table 6 shows the results of variations in mutation experiments.

Table 6. The Effect of Mutation Probability

#Trial	Pm	MSE	<i>Fitness Value</i>
1	0.9	0.5753	1.7381
2	0.09	0.4266	2.3436
3	0.7	0.5753	1.7381
4	0.07	0.3636	2.7497
5	0.5	0.5753	1.7381
6	0.05	0.2976	3.3596
7	0.3	0.5745	1.7405
8	0.03	0.2095	4.7619
9	0.1	0.4809	2.0794
10	0.01	0.1099	8.9285

For each test result, the mutation probability value of 0.01 obtained the optimum fitness value compared to all the mutation probability values. The mutation probability greater than 0.1 caused poor fitness value. This was because the value of the mutation was greater, making the individuals in each iteration vary (the fitness value changed rapidly). So, it could also affect the number of fewer iterations when using a greater probability of mutation.

3.5 The Effect of Crossover Probability

The crossover probability value was 50% in the first and second experiments. So the obtained maximum fitness value tent to be greater with the combination of other parameters tested, namely gene parameters, and mutations interrelated with each other. This certainly happened to these two parameters. The crossover value parameter certainly affected the variation in the results of the child.

Table 7. The effect of crossover probability

#Trial	Pc	MSE	<i>Fitness Value</i>
1	0.2	0.2015	4.9756
2	0.4	0.2005	4.9906
3	0.6	0.2044	4.9302

4	0.8	0.2025	4.9601
5	1.0	0.2007	4.9882

For cases with a large number of iterations, using the appropriate crossover probability can be seen as a decrease in the iteration, which has decreased slightly. However, in the third experiment conducted in this study, the experiment used the number of genes 2 with an iteration of 50 times, a population size of 100, and a mutation probability of 0.03. The experimental results showed no significant change in the probability of a crossover. Table 7 shows the results of variations in the probability of a crossover.

3.6 Inflation Prediction Evaluation

The prediction was made using the best individuals from previous experiments. The best genes had the smallest MSE value and the greatest fitness value because these values showed that the best genes had the lowest error rate with the most optimal level of prediction. The program saved the number of genes, the best fitness value, and chromosome as data for predictions. The results of several experiments showed the number of genes 2 with a 50% crossover probability and a 1% mutation probability obtained by MSE 0.1099. The closer the MSE value was to 0, the better the prediction was. Prediction results in the first month of 2019 were 0.49, the second month was 0.52, the third month was 0.54, and the fourth month was 0.55. Figure 3 shows the visualization of Indonesia's inflation predictions for the next four months (February – May 2019).

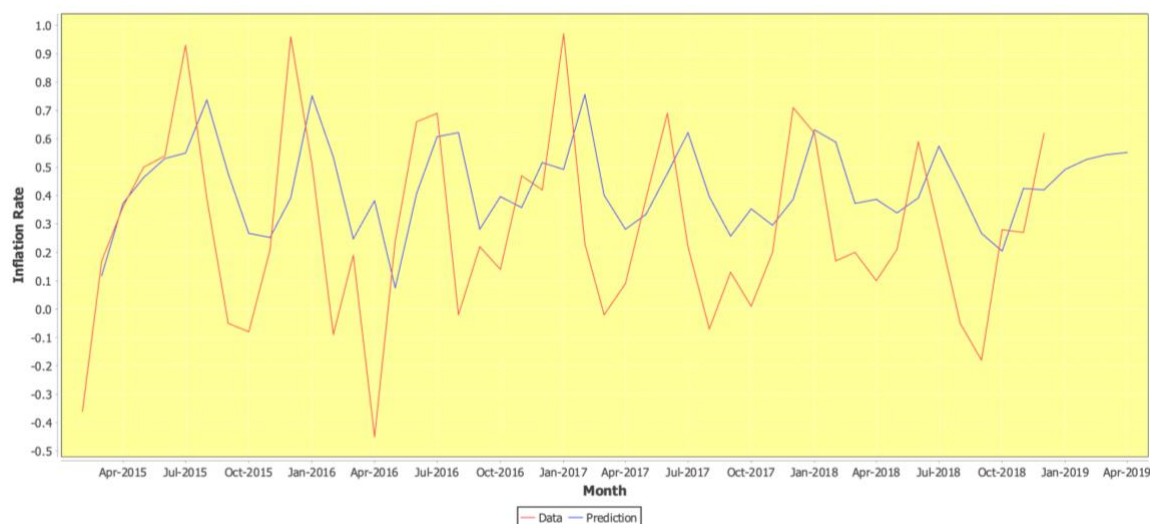


Figure 3. Indonesia's inflation prediction for the next 4 months (February – May 2019)

4. CONCLUSION

Genetic algorithm-based regression models were proven to be used to predict the inflation rate with MSE of 0.1099. In the training stage, the number of genes that had the best fitness value was 2. The probability of mutation was quite influential on the speed of getting the optimum fitness value. When the probability value of the mutation entered was greater, the number of mutations that occurred increased. This larger mutation value made the individual in each iteration more varied. However, the change in crossover probability did not significantly influence the fitness value.

5. REFERENCES

- [1] Bank Indonesia, "Data Inflasi," bi.go.id. Accessed: May 06, 2019. [Online]. Available: <https://www.bi.go.id/id/statistik/indikator/data-inflasi.aspx>
- [2] Badan Pusat Statistik, "Harga Indeks Konsumen (Umum)," bps.go.id. Accessed: May 06, 2019. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MiMy/indeks-harga-konsumen--umum-.html>
- [3] H. S. Noor and C. Komala, "Analisis Indeks Harga Konsumen (IHK) Menurut Kelompok Pengeluaran Nasional Tahun 2018," *Jurnal Perspektif*, vol. 3, no. 2, pp. 110–119, Dec. 2019.
- [4] A. Rahmi and W. F. Mahmudy, "Pembentukan model regresi harga saham menggunakan algoritma genetika," in *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 2016, pp. 18–19.
- [5] S. Sable, A. Porwal, and U. Singh, "Stock price prediction using genetic algorithms and evolution strategies," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Apr. 2017, pp. 549–553. doi: 10.1109/ICECA.2017.8212724.

-
- [6] M. R. Islam, I. F. Al-Shaikhli, R. B. M. Nor, and V. Varadarajan, "Technical approach in text mining for stock market prediction: A systematic review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 770–777, 2018, doi: 10.11591/ijeecs.v10.i2.pp770-777.
 - [7] A. Shah and C. Bhavsar, "Predicting Stock Market using Regression Technique," *Res. J. Financ. Account*, vol. 6, no. 3, pp. 27–34, 2015.
 - [8] Z. Ismail, A. Yahya, and A. Shabri, "Forecasting gold prices using multiple linear regression method," *Am J Appl Sci*, vol. 6, no. 8, p. 1509, 2009.
 - [9] J. Manoj and K. K. Suresh, "Forecast model for price of gold: Multiple linear regression with principal component analysis," *Thailand Statistician*, vol. 17, no. 1, pp. 125–131, 2019.
 - [10] R. Ispandi and S. Wahono, "Application of genetic algorithms to optimize parameters in support vector machine to increase direct marketing predictions," *J. Intell. Syst*, vol. 1, no. 2, pp. 115–119, 2015.
 - [11] F. Y. Septiawan, A. Hayati, and H. Kusuma, "Forecasting of currency exchange rate in forex trading system using genetic algorithm," in *International interdisciplinary conference on science technology machineering management pharmacy and humanities held*, 2017.
 - [12] A. Lambora, K. Gupta, and K. Chopra, "Genetic Algorithm- A Literature Review," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, IEEE, Feb. 2019, pp. 380–384. doi: 10.1109/COMITCon.2019.8862255.
 - [13] S. N. Sivanandam and S. N. Deepa, "Genetic Algorithms," in *Introduction to Genetic Algorithms*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 15–37. doi: 10.1007/978-3-540-73190-0_2.
 - [14] A. K. Wardana and S. Hartati, "Sistem Penjadwalan Pertandingan Pencak Silat Berbasis Algoritma Genetika," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11, no. 2, p. 177, Jul. 2017, doi: 10.22146/ijccs.24214.
 - [15] W. Wahyono, C. Puspitasari, M. D. Fauzi, K. Kasliono, W. S. Mulyani, and L. Kurnianggoro, "An Optimal Stock Market Portfolio Proportion Model Using Genetic Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 2, p. 171, Jul. 2018, doi: 10.22146/ijccs.36154.
 - [16] N. Mohananthini and G. Yamuna, "Comparison of multiple watermarking techniques using genetic algorithms," *Journal of Electrical Systems and Information Technology*, vol. 3, no. 1, pp. 68–80, 2016.
 - [17] A. K. Santra and C. J. Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering," *IJCSI International Journal of Computer Science Issues*, 2012.
 - [18] A. K. Ariyani, W. F. Mahmudy, and Y. P. Anggodo, "Hybrid Genetic Algorithms and Simulated Annealing for Multi-trip Vehicle Routing Problem with Time Windows," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, p. 4713, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4713-4723.
 - [19] A. I. Permatasari and W. F. Mahmudy, "Pemodelan regresi linear dalam konsumsi Kwh listrik di Kota Batu menggunakan algoritma genetika," *DORO Repos. J. Mhs. PTIIK Univ. Brawijaya*, vol. 5, no. 14, pp. 1–9, 2015.
 - [20] R. Rismala, "Prediksi Time Series Tingkat Inflasi Indonesia Menggunakan Evolution Strategies," *Jurnal Ilmiah Teknologi Infomasi Terapan (JITTER)*, vol. 1, no. 2, 2015.