JOIN (Jurnal Online Informatika)
Volume 5 No.1 | June 2020: 45-52
DOI: 10.15575/join. join.v5i1.532

p-ISSN: 2528-1682
e-ISSN: 2527-9165

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

**Faisal Dharma[1], Shabrina[2], Astrid Noviana[3], M. Tahir[4], Nirwana Hendrastuty[5], Wahyono[6*]**
[1, 2, 3, 4, 5]Master Program in Computer Science, Universitas Gadjah Mada, Yogyakarta, Indonesia
[6]Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia

**Article Info**

***Article history:***

Received Mar 17, 2020
Revised June 20, 2020
Accepted June 28, 2020
Published July 15, 2020

***Keywords:***

Genetic Algorithm
Inflation Level
Mean Squared Error
Prediction
Regression

**ABSTRACT**

Inflation occurs where there is an increase in the price of goods or services in general and continuously in a country. Uncontrolled inflation will have an impact on the decline of the Indonesian economy. Therefore, the prediction of future inflation levels is necessary for the government to develop economic policies in the future. Prediction of inflation levels can be done by studying historical past Consumer Price Index (CPI) data. Regression methods are often used to solve prediction problems. The problem of finding the optimal prediction model can be seen as an optimization problem. Genetic algorithms are often used to deal with optimization problems. Thus, this work proposed to use a genetic algorithm-based regression model for predicting inflation levels. The model was trained and evaluated using real CPI data which obtained from the Indonesian Central Bank. Based on the experiment, it is proved that the proposed model is effective in predicting the inflation level as it gains MSE of 0.1099.

***Corresponding Author:***

Wahyono,
Department of Computer Science and Electronics,
Faculty of Mathematic and Natural Sciences,
Universitas Gadjah Mada, Yogyakarta, Indonesia
Email: wahyo@ugm.ac.id

## 1. INTRODUCTION

Developed countries are countries that have a strong economy and stability and equitable economic growth. The Indonesian state, through Bank Indonesia, focused 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 general trends regarding price developments. One indicator that is often used to measure inflation is the CPI (Consumer Price Index). Changes in CPI over time represent the rate of increase in inflation or the rate of decline in inflation of goods and services [2]. 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 will cause the economy to deteriorate, resulting in the price of expensive goods. Uncontrolled inflation will have an impact on the decline of the Indonesian economy. Therefore, predictions of future inflation rates are very necessary for the government to develop 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 or forecasting 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 or many other variables. Regression methods have often been used, such as research [3] using historical data regression methods to predict stock prices, [4] using regression methods to predict goods stock, and [5] predicting gold prices using multiple linear regression.

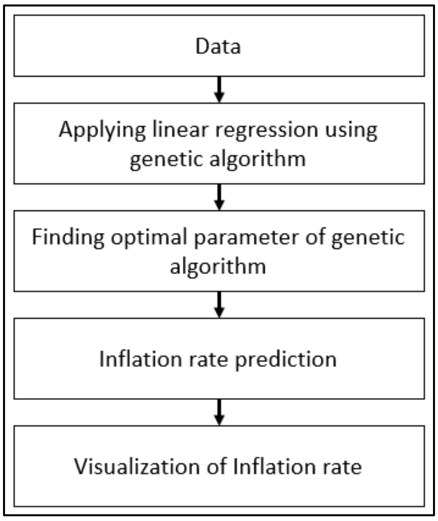


Figure 1. The flowchart of inflation prediction stages.

The problem of finding the optimal prediction model can be seen as an optimization problem. Genetic algorithms are often used to deal with optimization problems. Genetic Algorithm is a solution search algorithm with a system that adopts a natural selection process, which is where there is a group of individuals in a population that represents a set of solutions that will later experience natural selection. Some studies using genetic algorithms, including [3] produce the best coefficients that can produce stock prediction prices that are close to the original price. Then research [6] genetic algorithms successfully predict currency exchange rates.

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. Section 4 concludes the paper.

## 2. METHOD

### 2.1 Inflation

Inflation is a tendency to increase the price of goods and services in general, which continues continuously. If the price of goods and services in the country increases, then inflation will increase. 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 CPI.

$$Inflation = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} x100\% \qquad (1)$$

where

$CPI_t$ : Current month CPI.
$CPI_{t-1}$ : Previous month CPI.

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

$$\text{Inflation}=\frac{136.47-135.87}{135.87}\text{ x100\%=0.44}$$

### 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 dependent variables that exist with the given independent variable.

Multiple regression analysis is used to predict the value of one dependent variable based on two or more independent variables. Regression analysis will produce an equation/regression model. This study uses 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. This algorithm has been widely used, because it can handle various problems that have complex search space, using mathematical models with high accuracy values [7]. 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 [6], stock prices [8], marketing [9], and transportation [10]. For this reason, genetic algorithms are 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.

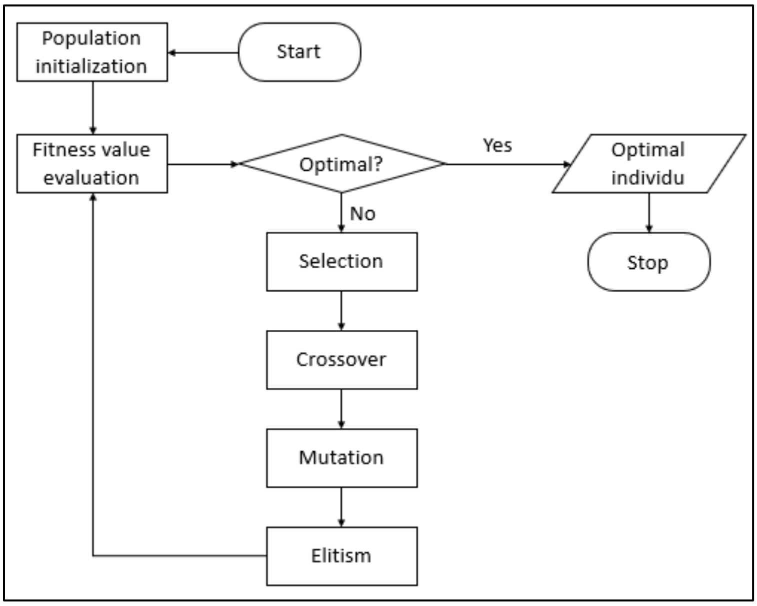


Figure 2. The flowchart of a genetic algorithm.

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 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. Then the final stage is selection, namely choosing individuals from the population set and offspring. The best individuals as a result of selection are kept alive in the next generation [11].

**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 [12]. This process aims to find a pattern that is closest to the characteristics of historical data on the inflation rate in Indonesia. Multiple linear regression (MLR) is used in this study [13], aiming to find a pattern that is closest to the characteristics of historical inflation rates in Indonesia, where the pattern is represented in the form of linear functions as in equation (2). The initialization process using this function will be used to build a prediction model using training data.

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

where:

Y : Prediction of the Consumer Price Index on month B
$x_1$ ... $x_n$ : Consumer Price Index in months $B_1$ to $B_n$
$\theta_0$ ... $\theta_n$ : Representation of each gene on a chromosome with random values

**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

| $\theta_0$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
|---|---|---|---|---|
| 0.8531 | 0.4113 | 0.2690 | 0.9638 | 0.2079 |

p-ISSN: 2528-1682
e-ISSN: 2527-9165

It can be seen in Table 1, that the number of genes on the chromosome can be flexible according to the input desired by the user. To enable 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. In Table 1, suppose we want to predict the fifth month, so the number of genes is 5.

**3) Fitness Value Calculation**

The fitness value (f) used in this study uses MSE (Mean Square Error) with the actual value [14]. The optimal value is obtained when getting the smallest of MSE value, so the fitness value will be greater and vice versa. It will get a prediction error with equation (3),

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

where $\varepsilon$ is a very small number to avoid division with zero.

The purpose of predicting the inflation rate is to minimize the error of the prediction with the MSE (Mean Square Error) value, then we will square all the error data that is there then divided by the number of data errors. The MSE value uses 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 are using equation (2).

$Y = \theta_0 + \theta_1 (0.32) + \theta_2 (0.24) + \theta_3 (0.14) + \theta_4 (0.11)$
$= 0.8531 + 0.4113(0.32) + 0.2690(0.24) + 0.9638(0.14) + 0.2079(0.11)$
$= 1.2070$

In this study, the prediction data is 9 years or 108 months. If it will be predicted in the fifth month, the value of n becomes 104.

$$MSE = \frac{1}{104}\sum_{i=1}^{104}((1.2070 - 0.86)^2 + \cdots + (1.4396 - 2.46)^2)$$
$$= 0.3750$$

Then, the value of fitness can be written as follows:

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

**4) Parent Selection**

The parent selection is done using the Roulette Wheel Selection. In the Roulette Wheel method, individuals are mapped in a sequential line segment according to their fitness value. Then random numbers are generated as many as N rounds as the parent of the crossover process. The crossover process cannot have the same parent. Then do a check to remove the redundant parent.

**5) Crossover**

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

$$child\ 1 = \alpha . x_i + (1 - \alpha) . y_i , \quad 1 \le i \le n \quad (4)$$
$$child\ 2 = \alpha . y_i + (1 - \alpha) . x_i , \quad 1 \le i \le n \quad (5)$$

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

Table 2. Parent chromosome in crossover operation

| | | | | | |
|---|---|---|---|---|---|
| *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 specify the value of α, which is a random value between 0 - 1. We give an α value of 0.4. Then the calculation process for each chromosome can be written as follows:

Child 1 = (0.4)(0.8531) + (0.6)(0.1988) = 0.4605
Child 2 = (0.4)(0.1988) + (0.6)(0.8531) = 0.5913

Crossover is 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 (α=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 |

#### 6) Mutation

The process of mutation is carried out on individuals after the process of the results of parent crosses or crossover. This process changes the value of one or several genes in a population that aims to avoid premature convergence, which is the achievement of value or outcome that has not or is not maximal [7]. In the process of mutation, the number of individuals depends on the probability of the mutation (Pm) that has been determined. This study uses the uniform mutation method.

This mutation works by replacing the selected gene value for mutations with random numbers that already have a limit. The limit is determined so that the resulting solution is a legal solution. The random numbers used in this study are from 0 to 1. For example, the fourth-gen is changed with 0.241. Mutation examples use 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 |

#### 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 is done by sorting the best fitness values of all individuals in the population. Then it will be taken to the next generation, only the first population, so it will not increase the number of individuals in the population.

For example, individual reproductions are 20, while the first generation of individuals is 100. The total number of individuals is 120. Therefore, the selection is 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 are 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 is grouped into two parts, and training data has taken between 2006 - 2014 or for 108 months, and testing has taken between 2015 - 2018 or for 48 months.

### 3.2 Evaluation Strategy

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

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

### 3.3 The Effect of The Number of Gene

In the first experiment to see the best fitness value, variations in the number of genes were carried out. The size of the population is 100 individuals and the number of iterations is 50. The value in this population is chosen so that the number of individuals in the population is not too large but also not 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 | *Fitness Value* |
|---|---|---|---|
| 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 is 2. In the number of genes 2 has the smallest MSE value that is equal to 0.2009, and the best fitness value is 4.9842. These values indicate that the number of genes 2 has the least error rate. The use of genes that more than 10 is resulting in a worse the fitness value because it is proven in this experiment that increasing the number of genes makes the fitness value decrease continuously.

The population size generated and many iterations in this experiment are optimal results that can be used to determine the inflation value for the following months. The population size and many of these iterations will be 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 [15]. The aim is to find a combination of parameters that produce the most optimal prediction function. The number of genes is 2. This is the best result of the previous experiment. The crossover probability used in this experiment is 50%. Table 6 shows the results of variations in mutation experiments.

Table 6. The Effect of Mutation Probability

| #Trial | Pm | MSE | *Fitness Value* |
|---|---|---|---|
| 1 | 0.90 | 0.5753 | 1.7381 |
| 2 | 0.09 | 0.4266 | 2.3436 |
| 3 | 0.70 | 0.5753 | 1.7381 |
| 4 | 0.07 | 0.3636 | 2.7497 |
| 5 | 0.50 | 0.5753 | 1.7381 |
| 6 | 0.05 | 0.2976 | 3.3596 |
| 7 | 0.30 | 0.5745 | 1.7405 |
| 8 | 0.03 | 0.2095 | 4.7619 |
| 9 | 0.10 | 0.4809 | 2.0794 |
| 10 | 0.01 | 0.1099 | 8.9285 |

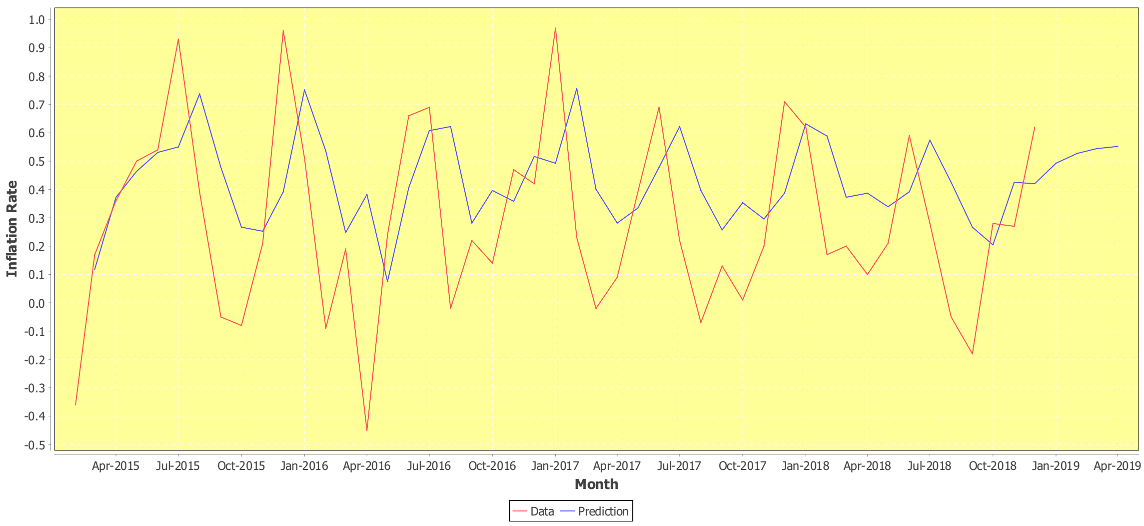


Figure 3. Indonesia's inflation prediction for the next 4 months (February – May 2019)

It can be seen for each test result that the mutation probability value of 0.01 is getting the optimum fitness value compared to all the mutation probability value. The mutation probability that greater than 0.1 causes poor fitness value. This is because the value of the mutation is greater, making the individuals in each iteration vary (the value of fitness changes rapidly). So that it can also affect the number of fewer iterations when using a greater probability of mutation.

### 3.5 The Effect of Crossover Probability

In the first and second experiments, the crossover probability value is 50%. So that the maximum fitness value obtained tends to be greater with the combination of other parameters tested, namely gene parameters, and mutations interrelated with each other. This certainly happens to these two parameters. The crossover value parameter will certainly affect the variation in the results of the child.

Table 7. The effect of crossover probability

| #Trial | Pc | MSE | Fitness Value |
|---|---|---|---|
| 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. But 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

Prediction is done using the best individuals from previous experiments. The best genes have the smallest MSE value and the greatest fitness value because these values show the best genes have the lowest error rate with the most optimal level of prediction. The program will save the number of genes, the best fitness value, and the chromosome as data for predictions. The results of several experiments show the number of genes 2 with a 50% crossover probability and a 1% mutation probability obtained by MSE 0.1099. The MSE value is getting closer to 0, the prediction is getting better. 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 4 months (February – May 2019).

## 4. CONCLUSION

Genetic algorithm-based regression models have been proven to be used to predict the inflation rate with MSE of 0.1099. In the training stage, the number of genes that have the best fitness value is 2. The probability of mutation is quite influential on the speed of getting the optimum fitness value. When the probability value of the mutation entered is greater, the number of mutations that occur becomes more. This larger mutation value makes the individual in each iteration more varied. Whereas the change in crossover probability does not significantly influence the fitness value.

## 5. REFERENCES


[1] Indonesian Bank, “Indonesia Bank and Inflation – Central Bank of the Republic of Indonesia,”. Available: https://www.bi.go.id/id/moneter/inflasi/bi-dan-inflasi/Contents/Single.aspx. [Accessed: 12-May-2019].

[2] Central Statistic Bureau (Badan Pusat Statistik), “Consumer Price Index (Indeks Harga Konsumen),” 2019. [Online]. Available: https://www.bps.go.id/subject/3/inflasi.html#subjekViewTab1. [Accessed: 12-May-2019].

[3] A. Rahmi, W. F. Mahmudy, and B. D. Setiawan, “Stock Price Prediction Based on Historical Data Using a Regression Model Built with Genetic Algorithms,” *Repos. J. Mhs. PTIIK Univ. Brawijaya*, vol. 5, no. 12, pp. 1–9, 2015.

[4] P. M. A. Shah, “Predicting Stock Market using Regression Technique,” *Res. J. Financ. Account.*, vol. 6, no. 3, pp. 27–34, 2015.

[5] Z. Ismail, A. Yahya, and A. Shabri, “Forecasting Gold Prices Using Multiple Linear Regression Method,” *Am. J. Appl. Sci.*, vol. 6, no. 8, pp. 1509–1514, 2009.

[6] F. Y. Septiawan, U. G. Mada, C. K. Dewa, U. G. Mada, U. I. Indonesia, and I. Introduction, “Prediction of Currency Exchange Rate in Forex Trading System Using Genetic,” in *International Interdisciplinary Conference on Science, Technology, Engineering, Management, Pharmacy and Humanities*, 2017, no. April, pp. 111–120.

[7] A. K. Wardana and S. Hartati, “Pencak Silat Match Scheduling System Based on Genetic Algorithms,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 177, 2017.

[8] 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 J. Comput. Cybern. Syst.*, vol. 12, no. 2, p. 171, 2018.

[9] R. Ispandi; Satrio 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.
[10] 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," *IJECE (International J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 4713, 2018.
[11] W. F. Mahmudy, *The Introduction of Genetic Algorithm*. Malang: Fakultas Ilmu Komputer Universitas Brawijaya. Malang, 2015.
[12] R. Rismala, "Time Series Prediction Indonesian Inflation Levels Using Evolution Strategies," *J. Ilm. Teknol. Infomasi Terap.*, vol. I, no. 2, 2015.
[13] K. Uyanık, G. Güler, Neşe. "A Study on Multiple Linear Regression Analysis," *Procedia - Social and Behavioral Sciences*. vol. 106. pp.234–240. 2013. doi:10.1016/j.sbspro.2013.12.027.
[14] A. Botchkarev. "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology". *arXiv:1809.03006 [stat.ME]*, 2018.
[15] A. Hassanat, et al., "Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach", *Information* 2019, 10, 390; doi:10.3390/info10120390