ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

Prediction of Rupiah Exchange Rates Using Genetic Algorithm Based Multiple Regression

Nikolaus Aldo Halim , Wahyono*

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

ABSTRACT: Genetic Algorithm is an evolutionary algorithm that could be used for solving many problems, especially problems of searching and optimization. It is often used in prediction with multiple regression to find the optimal weight for every variable used in the system equation. With many options of window size, multiple regression will give different results depending on the window size. This research calculates the accuracy of the foreign currency to rupiah prediction system using genetic algorithm and multiple regression. The variables used this system are the past currency value and the currency interest rate as an external factor besides the past currency value. This research uses Mean Absolute Error (MAE) and Minimum Squared Error (MSE) as the error values to calculate the accuracy of the system created.

KEY WORDS: Genetic Algorithm, Linear Regression, Currency, Prediction

I.INTRODUCTION

The foreign exchange rate is one of many aspects that determine the economic level of a country [1]. The foreign exchange rate also determines how the pricing of things work.Foreign exchange rates are so volatile that their value cannot be predicted in the future and tends to be fluctuate on every pair of currency. Datasets of the foreign exchange rates usually come in complex time-series. Moreover, in this era of globalization, the competition in the global market is very intense, such that the value of the exchange rates will become more difficult to predict. The impact of the fluctuation will cause prices of goods to change too. The impact of globalization will strengthen the bonds between countries, in fact, now Indonesia has a close relationship with China and has many exchange students and labors with America, which are two of the most influential countries in the world. Therefore, in this research, the datasets used are Indonesian Rupiah to Chinese Yuan, and Indonesian Rupiah to the US Dollar.

One of the factors that determine the foreign exchange rate is the interest rate by the central bank in the country. In general, if the interest rate of one currency increases, the number of investors who will invest in the currency will increase too, such the demand of the currency will increase as well, causing the currency rate will increase [2]. The interest rate of a country is now is more volatile compared to the past, because of globalization and the fast advancement of technology.

The fast growth of information technology provides people with easier access to trade foreign currencies. Usually people gain profit from trading currencies by investing in the gradual changes of the values. In this foreign currency investment, the risk of failure is high, so the investors need a way to predict the value with a better accuracy. By using a good approach, we can minimize the failure risk significantly [3].

The growth of information technology also provides various ways predict changes in exchange rates. Artificial intelligence especially become so reliable because it can learn by itself using past dataset and predict the future. The methods include fuzzy logic, machine learning, genetic algorithm, etc. The prediction of exchange rate is also a challenge because of its time-series nature [4].

This research uses multiple regression with genetic algorithm method because according to past researches, modelling of exchange rate prediction problem using genetic algorithm gives accurate results. Multiple regression method is used because in time-series dataset has strong connection between time and value, such that multiple regression method could be used to get a good result. The genetic algorithm is a computation algorithm invented by John Holland in 1960, where the concept is based on the evolution theory of biology using natural selection [5]. In other researches, the variables used are usually only the past time-series data without using other factors. This research, however, uses an additional variable which is the interest rate of the national bank, so the results are expected to be more accurate.

The paper mainly focuses on how genetic algorithm technique can be applied to predict the United States Dollar and Chinese Yuan exchange rates to Indonesian Rupiah. The study of literature survey is presented in section III, Methodology is explained in section IV, section V covers the experimental results of the study, and section VI discusses the future study and Conclusion.

II. RELATED WORKS

Exchange rate predictions using hybrid methods are usually done by researchers using different combination of techniques. The combination of techniques used by Chang[7] is different fromTorregoza [1] but with the same combined methods genetic algorithm and the artificial neural network. However, it uses the genetic algorithm as a method to select the variables that are used as input parameters in the ANN, thus this method is called the Genetic Algorithm Back Propagation Network (GABPN). GABPN uses 10 variables that have been selected by the genetic algorithm before. The reference [7] also compares the GABPN method to the pure genetic algorithm and the pure ANN. The result shows the GABPN method has the best accuracy. The difference with reference [1] is the prediction is done by using the genetic algorithm and the ANN where the genetic algorithm is used to declare the weight of the nodes. Based on their research, we know that the accuracy of the hybrid method is likely higher than the pure ANN.

Prediction using the ANN is also used byNanayakkara [4], but the method used to update the weight in the network is the Generalized Auto Regressive Conditional Heteroskedasticity (GARCH). The ANN structure used is also different with Chandar[3], with 4 input nodes and 2 hidden layers. The average MAE value is 0.0627. We can see that the MAE with the same method, but different technique gives a different result. Besides the GARCH method, Chandar[3] implements the exchange rate prediction using ANN with 3 hidden layers, and 3 input nodes, and the backpropagation method to update weights. The resulting average MAE value is 0.0085, which means this method has high accuracy. The research of Jauhari [6] also uses an advanced genetic algorithm called genetic programming to predict exchange rates. The method uses the chromosome representation as a tree that becomes an equation for prediction. The resulting MAPE value is 0.08, showing that this method also has high accuracy. Septiawan[5] does the prediction differently than all the method above. Their prediction uses multiple regression and basic genetic algorithm to declare the weight of the data. Based on the research, the MAPD (Mean Absolute Percentage Deviation) of 0.6% is achieved. Septiawan[5] only uses the past currency value as a variable of the prediction, but this research uses an additional variable, the interest rate of the country, because the interest rate is one of the main factors that influence the currency value.

III. METHODOLOGY

A.ALGORITHM DESIGN

The system design is an implementation of the genetic algorithm [8] to the multiple regression method for predicting exchange rates. Foreign exchange rates are predicted by the system using multiple regression with 2 to 7 sliding windows variation. Then the system calculates the fitness value for the genetic algorithm using MAE value [9]. If the value is lower, the fitness value will be higher, and vice versa [10]. The system is trained using the training, followed by the calculation of MAE, MSE, and MAPD. These values are used to calculate the accuracy of the system. The outline of the algorithm is described in the following 10 steps:

1. The data is be separated into 2 parts, one the training dataset, and the other one is the testing dataset.
2. The starting population is generated using a given number whose value is declared randomly.
3. The population's fitness value is also calculated.
4. The chromosomes are selected to do the crossover function using the Roulette Wheel Selection.
5. The selected chromosomes are crossovered using the Simple Arithmetic Crossover.
6. The offspring from the crossover function are mutated using the uniform mutation.
7. The fitness values of the offspring are calculated using the training dataset.
8. The population is updated by eliminating some chromosomes that have the lowest fitness values until the population number is the same as the initial population. This method is called elitism.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

9. The process is repeated until the generation number is reached.
10. The chromosome with the highest fitness value is selected as a solution and is tested using the testing dataset. And MAE, MSE, and MAPD value will be calculated.

B.ENCODING PROCEDURE

In the first step, the genetic algorithm starts with the defined number of populations which is the chromosome representation with the length of each chromosome varying based on the number of sliding windows used.

$\theta 1$	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7

Fig.1. Chromosome representation of 3 sliding window

$$Y_h = K_t \, x \, \theta 1 + \sum_{n=1}^{w} \theta_{n+1} \, x \, C_{h-n} + \sum_{n=1}^{w} K_i \, x \, \theta_{n+4} \, x \, I_{h-n} \quad (1)$$

Every chromosome using n sliding windows has ((n*2) +1) length of genomes. the formula is, in the first genome, the θ_1 which will be multiplied with its multiplying constant K_t, is always the constant variable for the multiple regression formula in equation (1). Y_h is the forecast value, and θ_2 until θ_{1+n} are the multiplying weights for the currency rates Ci. θ_{n+2} until $\theta_{(n*2)+1}$ are the multiplying weights for the interest rates Ii where the weights will also be multiplied with the interest multiplying constant K_i to normalize the value.

C.SELECTION CROSSOVER AND MUTATION

The second step after the population is generated is the selection function, where the function uses the roulette wheel selection to select the chromosomes for input to the crossover step. The crossover function uses the simple arithmetic crossover. Then the mutation function is then applied to every single chromosome in the population and the genomes mutated will be selected randomly based on the variable mutation probability Pm.

D.FITNESS FUNCTION

In this research, we define the minimal error value for the fitness function, where the error value used is the MAE.

$$MAE = \sum \frac{\left| \hat{S}_t - S_t \right|}{n} \quad (2)$$

Based on equation (2) where the MAE is the sum of the absolute value of the predicted value subtracted by the real value divided by the number of predictons made.

E.GENERATION UPDATE AND STOPPING CRITERIA

The system updates the population using the elitism method. This method ranks the chromosomes based on the fitness value and removes the chromosomes one by one from the lowest fitness value until the number of chromosomes reaches the same number as the first population initialized. The stopping criteria in this research is the number of generations, and if the number is reached, then the training process terminates and continues to the testing method.

F.TESTING

In the testing step, this research runs some tests to find the optimum value of the following parameters before the main testing:

1. Sliding window
2. Mutation Probability
3. Number of Generations
4. Crossover Point
5. Alpha
6. Theta 1 Multiplying Constant

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

7. Interest Rate Multiplying Constant

The testing uses the data ranged from January 2017 until January 2018. MAE, MSE, and the MAPD values for the main testing.

V. EXPERIMENTAL RESULTS

A.EXPERIMENTAL SETTINGS

This research and its implementation are executed using a computer Intel Core i5 2.50 GHz under Windows 64-bits operating system with RAM 8GB. The method was implemented using Java programming language in Intellij IDEA Ultimate Edition.

B.DATA DETAILS

The data used in this research are the foreign exchange rates of CNY and USD against IDR, made available by the Bank of Indonesia, and the interest rates of the two currencies (CNY and USD). A total of 4 years of data are used, where the first 3 year are used for training, and the remaining 1-year data used to test the model. The plots of exchange rates for the currencies are shown in Fig.2 and Fig.3.

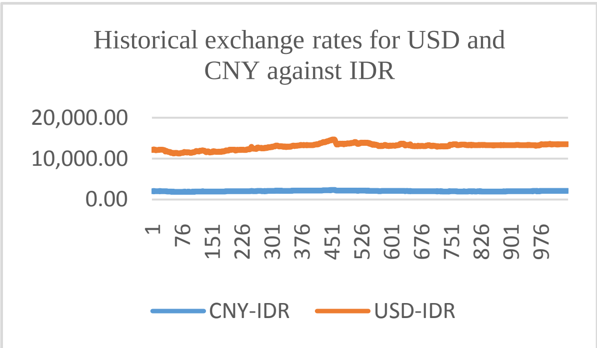


Fig.2. Exchange rates for USD & CNY to IDR.

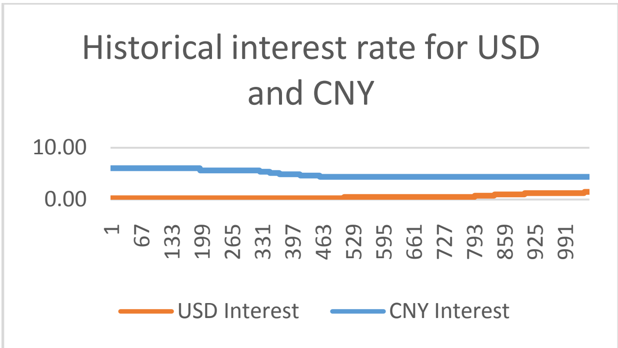


Fig.3. Interest rate for USD and CNY.

C.TRAINING RESULTS

The training result is based on the error value when the model is trained using the training data. Each training uses the optimal parameters obtained from the previous training.

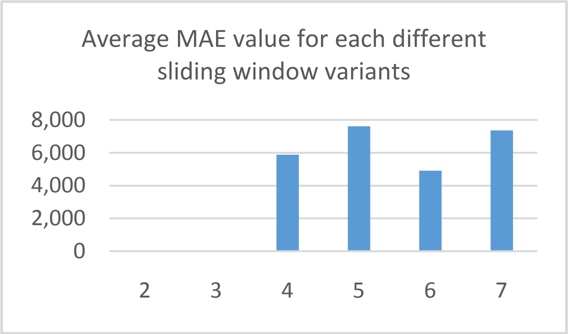
C.1)Sliding Window

The sliding window training is used to find the most suitable sliding window size in the prediction model by calculating the MAE value from each window size. The window sizes trained are 2, 3, 4, 5, 6, and 7. Based on Fig.4, it can be seen that the most accurate window size is when window size = 2 which gives us the average MAE value of 17,2.

C.2)Mutation Probability

The mutation probability training is used to find the most suitable number of mutation probability, the mutation probability is used on the mutation function which ranges from 0-1 in real number. The mutation probability determines the chances of a genome to be mutated. Based on Fig.5, we could see that the most accurate mutation probability to be used in this model is the 0,4 with the lowest value of MAE.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019



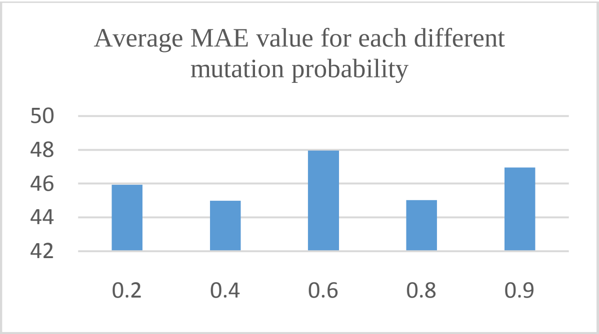


Fig.4. Average MAE for Each Different Sliding Window

Fig.5. Average MAE for Each Different Mutation Probability Value

C.3) Generation Number

The generation number is the number to determine how many loops the genetic algorithm runs. With more generations, the solution is expected to be more accurate, but the runtime should increase significantly. Based on Fig.6 we could see that the most fit generation number to be used in this model is 100,000, resulting the MAE value of 43,66. Increasing number of generation to 1,000,000, the MAE value increases 43.70, which costs higher runtime with no significant difference of the result, so the number will be used is 100,000.

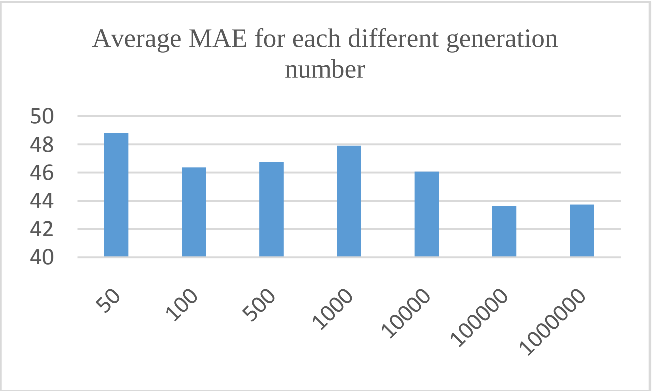


Fig.6. Average MAE for Each Different Generation Number

C.4) Crossover Point

The crossover point in the crossover function is used to cut the two parents selected by the selection function. Because the size of window used is 2, thus the variations of the crossover points available are 0, 1, 2, 3, and 4. Based on Fig.7 it can be seen that the most accurate crossover point for this model is the point 3.

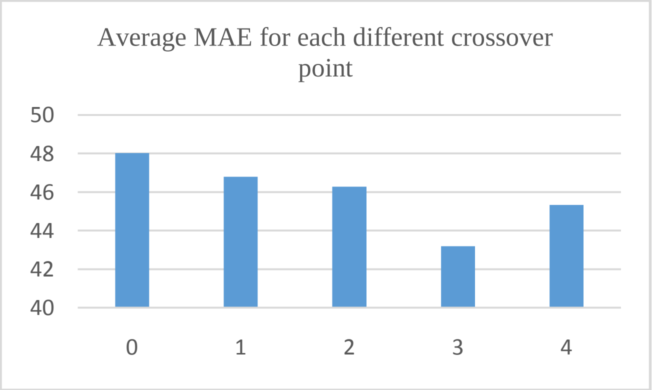


Fig 7.Average MAE for Each Different Crossover Point

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

C.5) Alpha

The alpha value training is used to find the most suitable alpha value. The alpha number is used in the simple arithmetic crossover function as the equation value to calculate the value of each genome of the offspring gained from the crossover. The Alpha number is ranged from 0 to 1, so the variants are 0.2, 0.4, 0.6, and 0.8. Based on Fig.8, we could see that the most accurate alpha number to be used is 0,2 with the lowest value of MAE.

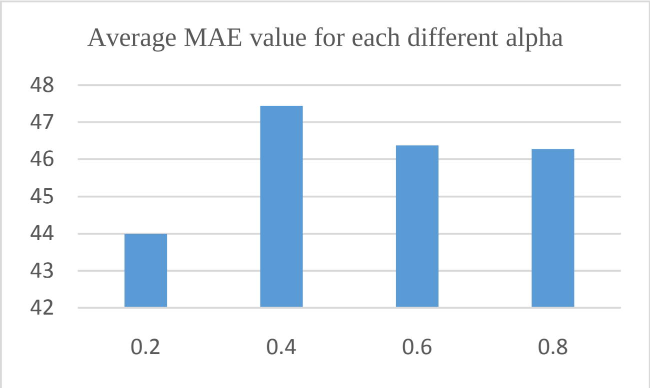


Fig.8. Average MAE for Each Different Alpha Number

C.6) Theta 1 Multiplying Constant

The constant is used as the multiplier in the regression equation. The constants used are 500, 1000, 2000, 3000, and 4000. In Fig.9, it could be seen that the most accurate value is 500 with the lowest value of MAE.

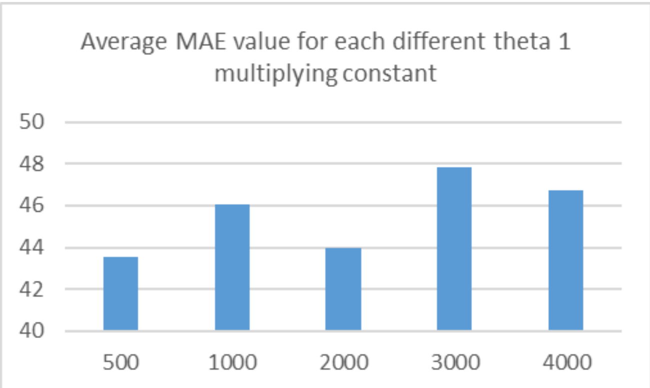


Fig.9. Average MAE for Each Different Theta 1 Multipying Constant

C.7) Interest Rate Multiplying Constant

The interest rate multiplying constant training is used to find the most suitable constant value. The constant is used in the regression function to normalize the interest rate to match the value of the currency. The constants are the same as the theta 1 training. Based on Fig.10, it can be seen that the constant that gives the best value is 2000.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

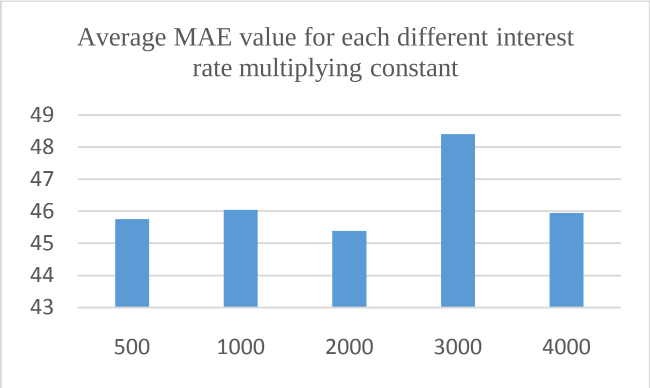


Fig.10. Average MAE for Each Different Interest Rate Multipying Constant

D.OPTIMAL PARAMETERS

Based on the parameters gained from the results above, the optimal parameters used for the testing are:

1. Sliding Window size: 2
2. Mutation Probability: 0.8
3. Generation Number: 100000
4. Crossover Point: 3
5. Alpha Value: 0.2
6. Theta 1 Multiplying Value: 500
7. Interest Rate Multiplying Value: 2000

E.TESTING RESULTS

The testing is done by using the optimal parameters. It is run 5 times and the error is calculated the average value. This research compares the error value of using the interest rate is used and when it is not used the results are shown on Table I.

TABLE I
Testing Results of each currency

Currency	MAE AVG	MSE AVG	MAPD AVG
CNY-IDR	6.81	97.19	0.3516%
CNY-IDR without Interest Rate	7.82	111.62	0.38%
USD-IDR	18.23	734.16	0.1361%
USD-IDR without Interest Rate	19.07	803.53	0.14%

F.DISCUSSION

Based on Table I, using interest rate on predicting exchange rates gives higher accuracy. The prediction of different currency pairs with the same model gives us a different result.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 8 , August 2019

VI.CONCLUSION AND FUTURE WORK

From this research, it can be concluded that:

1. Prediction of USD-IDR using this method with interest rate results in an average MAE value of 18.23, MSE value of 734.16, and MAPD value of 0.136%.
2. Prediction of USD-IDR using this method without the interest rate results in an average MAE value of 19.07, MSE value of 805.53, and MAPD value of 0.141%.
3. Prediction of CNY-IDR using this method with interest rate results in an average MAE value of 6.81, MSE value of 97.19, and MAPD value of 0.35%.
4. Prediction of CNY-IDR using this method without the interest rate results in average MAE value of 7.28, MSE value of 111.62, and MAPD value of 0.376%.
5. The genetic algorithm prediction model with interest rate provides a better result than the genetic algorithm model without using the interest rate.

This research has proven that interest rate has a positive impaction predicting exchange rates. In future works, the methodology could be improved by adding other economic factors such as inflation rate, economy rate, etc.

REFERENCES

[1] Torregoza, Mark Lorenze R., dan Dadios, Elmer P. 2014. Comparison of Neural Network and Hybrid Genetic Algorithm-Neural Network in Forecasting of Philippine PesoUS Dollar Exchange Rate. 7th IEEE International Conference Humanoid, Nanotechnology, Information Technologi Communication and Control, Environment and Management (HNICEM), Palawan, Filipina. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Twowey, Brian. 2012. Inside The Currency Market. John Wiley & Sons, Inc : New Jersey. K. Elissa, “Title of paper if known,” unpublished.

[3] Chandar, Kumar S., Mahadevan, Sumathi, dan Sivanandam S. N. 2015. Forecasting of foreign currency exchange rate using neural network. International Journal of Engineering and Technology (IJET).

[4] Nanayakkara, K. A. D. S. A., Chandrasekara, N. V., dan Jayasundara, D. D. M. 2014. Forecasting Exchange Rates using Time Series and Neural Network Approaches. European International Journal of Science and Technology.

[5] Septiawan, Fauzi Yudhi, Dewa, Chandra Kusuma, dan Afiahayati. 2017. Prediction Of Currency Exchange Rate In Forex Trading System Using Genetic Algorithm. International Interdisciplinary Conference on Science Technology Engineering Management Pharmacy and Humanities.

[6] Jauhari, Daneswara, Hanafi, Anang, Alam, M. Fahrul, Satria, Arrofi Reza, Hakim, Lukman, and Cholissodin, Imam. 2016. Exchange Rate Prediction Rupiah to USD using Genetic Programming (in Bahasa Indonesia). Journal of Information Technology and Computer Science. Vol. 3(4):285-291..

[7] Chang, Jui-Fang, Kuan, Chi-Ming, dan Lin, Yu-Wen. 2009. Forecasting Exchange Rates by Genetic Algorithms Based Back Propagation Network Model. 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

[8] Chambers, Lance. 2001. The Practical Handbook of Genetic Genetic Algorithms Applications Second Edition. CHAPMAN & HALL/CRC : London.

[9] Elamir, Elsayed A. H. 2012. Mean Absolute Deviation about Median as a Tool of Explanatory Data Analysis. Proceedings of the World Congress on Engineering 2012 Vol I.

[10] Gen, Mitsuo, dan Cheng Runwei. 2000. Genetic Algorithms & Engineering Optimization. John Wiley & Sons : New York.