Prediction of Regional Economic Growth in East Kalimantan using Genetic Algorithm

Emmilya Umma Aziza Gaffar

Faculty of Economic and Business, Universitas Mulawarman, East Kalimantan, Indonesia
Email: emmilya.gaffars@gmail.com

ARTICLE INFO

Article history:
April 29, 2016
May 5, 2016
May 30, 2016

Keywords:
Agriculture Share (SA)
Industry Share (SI)
Regional Economic Growth (REG)
Genetic Algorithm (GA)

ABSTRACT

This paper outlines and presents the development of genetic algorithms (GA) that are used for analyzing and predicting Regional Economic Growth (REG) with an agriculture share (SA) and an industry share (SI) as independent variables covers 13 districts/cities in East Kalimantan Province of 2002-2012 datasets. The genetic algorithm (GA) was used for modeling of REG datasets. The results of experiment shows that GA was produced prediction value of 92.389. This results indicate that the average price fluctuation is decreased in 2012. Meanwhile, the southern region is significantly increased in 2013. The results indicated that GA was good algorithm for prediction of REG. This paper is concluded by recommending some future works that can be applied in order to improve the prediction accuracy.

© 2016 International Journal of Computing and Informatics (IJCANDI).
All rights reserved.

I. Introduction

Regional economic growth can be seen as a complex system. Regional economic growth is an increasing public revenue as a whole in the region which increase through the entire value-added. Regional revenue is to describe the remuneration for the factors of production operations, such as land, capital, labor and technology [1, 2]. In other words, describe the prosperity of the area. Through regional high economic growth, it is expected that the public welfare will be gradually improved. Furthermore, the implementation of the concept of Growth Pole may have an impact on improving regional economic growth. The most important factor in the growth pole is the leading industry (i.e., the agriculture, industry and services shares), polarization, and spread effect. Growth pole is pointed out that at the center of industrial growth there must be major or driving industry that dominates other economic units. If economic activities are interrelated concentrated in a particular place, then the economic growth of the region concerned will increase faster. Instead, slow, if economic activities are dispersed and scattered to all corners of the region [3].

East Kalimantan province consists of 13 districts or cities, divided in three groups of constructions area: (1) North Development Region with development center in Tarakan, includes Bulungan, Nunukan, Malinau and Berau; (2) Central Development Region with development center in Samarinda, includes Samarinda City, Kutai Kartanegara, Kutai Timur, Kutai Barat, and Bontang districts; and (3) Development Region South with development center in Balikpapan City, includes Pasir and Penajam Paser Utara districts. Then, Special Development Region covers areas on the border does not have the growth center but refers to the northern and central part of the growth [4].

In order to help the problem of development, one of solution is forecasting economic growth patterns of each region [5]. Their economic growth pattern of each region in the future is expected to be a reference for the government in determining the appropriate policy in promoting economic growth of each region. In a literature review, several studies related to forecasting in determining the pattern of economic growth has implemented. In the case of power industry, forecasting for the transmission and distribution operator was required for power system planning and construction in Taiwan Regional (i.e., northern, central, southern, and eastern regions). Meanwhile, the datasets were from 1981 to 2000 of historical annual. For forecasting, artificial neural network (ANN) method has been used. The study indicated that the ANN one of model in the forecasting power industry and produces good forecasting [6]. In the case of local transport, statistical methods such as Multiple Linear Regression (MLR), Non-Linear Regression (NLR), and Simple Linear Regression (SLR) have been used to for predicting the regional freight transportation demand (RFTD) in China. The results showed that RFTDP models based NLR has produced a good predictor than the method MLR and SLR [7].

In some cases forecasting statistical methods have given good results. However, the researchers sought to improve forecasting results. For forecasting, machine learning methods (i.e., neural network, fuzzy, genetic algorithm, etc.) are alternatives methods because in some cases the machine learning has shown better forecasting results than statistical methods [2, 5, 8]. Machine learning is used for forecasting the cargo port as an economic indicator of Guangdong province in China. The results showed that machine learning (i.e., genetic algorithm-GA, and backpropagation neural network-BPNN) has been successfully implemented in order to overcome the shortage of statistical methods [9, 10].

In this paper, the economic growth in East Kalimantan would be predicted using machine learning methods. This prediction will utilize the leading factor (i.e., agriculture and industry) as variables that directly influences in order to the growth of regional economies. The remainder of this paper is structured as follows. Section 2 methodology related to forecast model. In Section 3, genetic algorithm is applied to predict regional economic datasets with the technique performance discussed. Finally, conclusions and recommendations for future work are provided in Section 4.

II. Methodology

A. The Principle of Genetic Algorithm

The basic concept of GA is found at the University of Michigan, United States of America by John Holland in 1975. Then after it was popularized by David Goldberg in the 1980s. The approach taken by this algorithm is to randomly combine a wide selection of the best solutions in a set to get the next generation of the best solution based on a condition that maximizes compatibility called fitness. Then, this generation will represent improvements on the initial population [11-13].

In general, using GA in the early stages is to represent the problem-solutions into the chromosome. This chromosome is a solution that shaped symbol. Initial population, are randomly generated. Meanwhile, the next population is the result of the evolution of chromosomes through iterations called generations. In each generation, the chromosomes will go through an evaluation process using a measure called the fitness function. Where, the fitness value of a chromosome will show the quality of the chromosomes in the population. The next stages is offspring, the next generation is known as the children called an offspring. Where, formed from the combination of two chromosomes present generation that acts as the parent by using the crossing operator called a crossover. In addition to crossing operator's crossover, a chromosome can also be modified by using mutation operators. Furthermore, in principle, GA is how to change the values of the genes in each individual in each generation to produce the most optimal solution [14-17]. In this study, types of reproduction is used includes crossover, mutation of genes (i.e., parents and children), and elitism with self-divide.

B. The Genetic Algorithm Cycle

In general, the implementation of the GA will go through a simple cycle consisting of four stages that include (1) initialization of population, (2) evaluation of chromosomes using fitness function value, (3) selection process, and (4) reproduction operations. The GA cycle briefly describe.

Regional economic growth data can be categorized as a data time series with seasonal patterns and non-stationary. Meanwhile, the independent variable in the form of agricultural and industrial shares. The equation time series model for regional economic growth can be expressed in Eq. 1.

Table 1. Agriculture Share (2002 – 2012)

Year	Tarakan	Berau	Malinau	Bulungan	Nunukan	Samarinda	Bontang	Kutai Barat	Kutai Kartanegara	Kutai Timur	Balikpapan	Paser	Penajam Paser Utara	East Kal.
2002	10.044	23.031	61.770	32.349	37.351	2.322	0.107	24.999	6.976	6.661	2.631	22.958	11.322	242.521
2003	11.319	23.329	57.935	31.667	36.867	2.208	0.113	24.611	7.510	6.386	2.852	22.365	11.617	238.779
2004	11.459	23.073	53.686	31.712	33.803	2.303	0.116	22.126	7.199	5.253	2.829	18.748	11.916	224.223
2005	10.878	22.821	50.066	25.649	30.987	2.271	0.110	21.000	7.348	4.473	2.609	17.574	12.247	208.033
2006	10.309	22.507	47.168	31.933	29.930	2.242	0.107	20.574	7.306	3.916	2.599	16.851	13.773	209.215
2007	10.032	21.875	44.682	28.044	28.957	2.172	0.114	19.790	7.325	3.695	2.621	15.974	14.365	199.646
2008	9.773	16.503	38.093	28.147	27.928	2.209	0.115	17.010	7.241	3.723	3.146	15.085	14.741	183.714
2009	9.873	16.198	28.035	28.052	28.405	2.211	0.117	16.096	7.145	3.804	3.049	15.060	14.104	172.149
2010	9.501	15.441	24.806	28.895	27.304	2.209	0.122	27.404	7.465	3.558	2.941	13.330	14.663	177.639
2011	9.449	14.603	20.731	2.917	26.302	1.831	0.136	14.707	7.702	3.382	2.765	12.650	13.709	130.884
2012	9.312	13.922	17.237	27.272	25.249	1.622	0.149	14.166	7.526	3.203	2.728	12.335	13.993	148.714

Table 2. Industry Share (2002 – 2012)

Year	Tarakan	Berau	Malinau	Bulungan	Nunukan	Samarinda	Bontang	Kutai Barat	Kutai Kartanegara	Kutai Timur	Balikpapan	Paser	Penajam Paser Utara	East Kal.
2002	10.673	15.218	0.062	25.753	0.035	28.591	95.666	2.086	1.851	0.444	50.033	2.344	19.732	252.488
2003	10.932	14.709	0.060	26.131	0.037	27.228	95.175	2.100	1.944	0.499	46.233	2.336	19.121	246.505
2004	11.769	14.666	0.061	27.182	0.036	25.891	94.646	2.231	2.015	0.451	46.707	1.936	17.318	244.909
2005	11.628	14.298	0.073	28.421	0.035	24.680	94.207	2.141	2.049	0.382	44.483	1.853	17.974	242.224
2006	11.393	13.937	0.092	14.504	0.038	23.692	93.826	2.191	2.116	0.346	41.318	1.766	19.121	224.340
2007	11.373	13.669	0.095	11.427	0.140	23.139	93.279	2.308	2.235	0.293	37.772	1.667	20.187	217.584
2008	11.280	10.136	0.097	12.264	0.209	21.859	93.022	2.424	2.265	0.286	38.846	1.653	20.672	215.013
2009	11.066	9.573	0.096	8.935	0.341	21.093	92.492	2.354	2.301	0.285	34.465	1.640	21.109	205.750
2010	10.792	8.878	0.090	4.422	0.452	21.559	91.884	2.915	2.403	0.261	32.384	1.478	21.118	198.636
2011	10.344	8.243	0.090	0.149	0.468	19.606	90.069	2.992	2.576	0.237	31.534	1.420	20.098	187.826
2012	10.011	7.709	0.086	0.142	0.517	19.339	89.493	2.225	2.498	0.212	27.309	1.403	19.969	180.913

Table 3. Regional Economic Growth (2002 – 2012)

Year	Tarakan	Berau	Malinau	Bulungan	Nunukan	Samarinda	Bontang	Kutai Barat	Kutai Kartanegara	Kutai Timur	Balikpapan	Paser	Penajam Paser Utara	East Kal.
2002	19.189	5.000	12.663	3.588	9.389	0.400	4.061	13.495	5.682	18.304	4.498	4.834	5.339	106.442
2003	4.412	4.158	6.031	0.010	5.972	9.521	3.961	10.493	0.213	3.110	2.132	5.249	4.912	60.174
2004	2.646	2.634	1.236	0.011	9.181	9.001	0.007	5.436	1.962	23.809	6.072	23.569	2.784	88.348
2005	7.631	5.108	3.630	0.012	9.559	8.046	6.500	8.235	2.672	20.585	3.214	8.927	4.163	88.282
2006	7.506	5.079	3.073	0.012	1.298	5.495	4.681	6.108	1.041	22.391	4.619	11.936	1.632	74.871
2007	6.912	5.701	6.314	5.885	3.766	3.107	4.810	6.453	0.117	8.075	2.079	12.918	3.787	69.924
2008	6.832	36.581	8.033	5.637	4.400	4.819	10.361	6.834	0.333	4.419	12.374	7.101	4.992	112.716
2009	4.633	5.976	8.953	4.655	3.931	4.495	2.610	6.889	2.003	0.262	1.703	7.738	3.507	57.355
2010	7.930	8.036	14.274	5.632	5.628	6.163	6.761	6.877	0.594	9.325	5.193	17.313	7.275	101.001
2011	7.626	7.934	11.894	6.056	6.738	13.772	9.128	6.891	0.882	11.433	7.439	10.847	11.681	112.321
2012	6.819	7.933	12.052	9.381	7.130	2.607	7.777	7.214	3.403	12.682	2.502	8.614	3.614	91.728

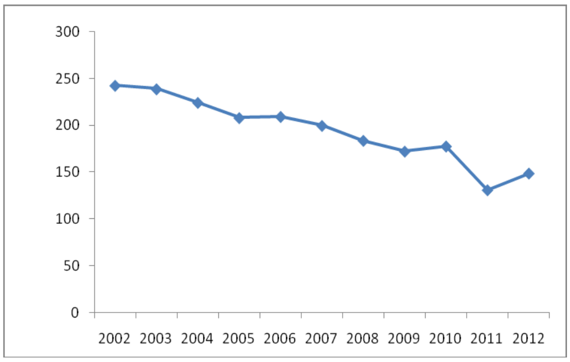

Fig. 1. Plot of Agriculture Share

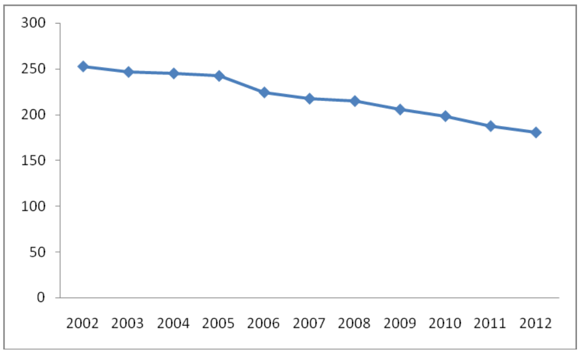

Fig. 2. Plot of Industry Share

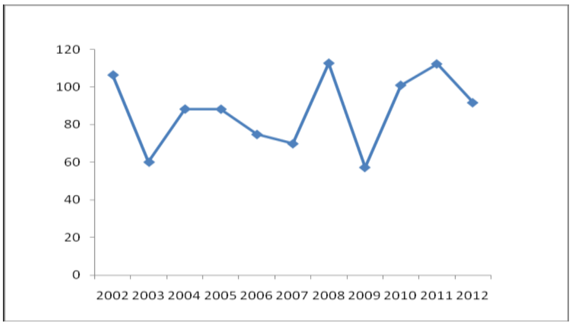

Fig. 3. Plot of Regional Economic Growth (REG)

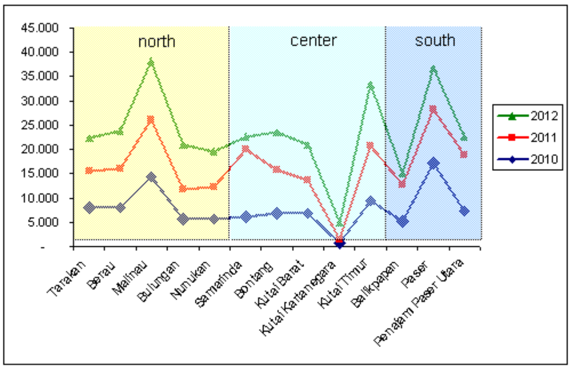

Fig. 4. Plot of Regional Economic Growth (REG) in Three Regions

$$y(t) = a.x_1(t) + b.x_2(t) \tag{1}$$

Where, y is regional economic growth; x_1 is agriculture share; x_2 is industry share; a, b are coefficients of the equation. Then, MISO-AR (Multiple Input Single Output-Autoregressive with Exogenous Input) is expressed in Eq. 2.

$$y(t) = -\sum_{i=1}^{n} a_i y(t-i) + \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij} x_i(t-j) + e(t) \tag{2}$$

Where, m is total input variable; n is data observation. Meanwhile, GA is used to find $a_1 \ldots a_{3n}$ values with $e(t) \approx 0$, Eq. 3.

$$y(t) = \sum_{i=1}^{n} \left(a_i y(t-i) + a_{n+i} x_1(t-i) + a_{2n+i} x_2(t-i) \right) + e(t) \tag{3}$$

B.1. Initialization of Chromosomes

Initialization population is required to select a number n of individuals that will be designated as first generation by randomly. In this study, Regional economic growth (REG) data are inclusive of optimization problem especially in order to determination coefficient $e(t) \approx 0$ of time series data. In this study, initialization chromosomes of REG is representation of the solution in REG with the initial population more than 2n in order to choose the most excellent fitness values., as shown in Fig. 5.

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + a_3 x_1(t-1) + a_4 x_1(t-2) + a_5 x_2(t-1) + a_6 x_2(t-2)$$

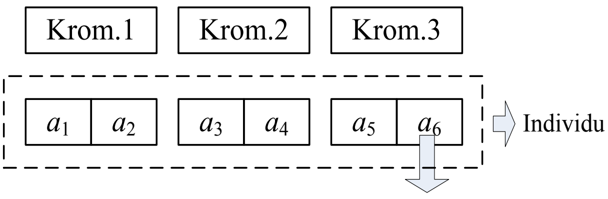

Fig. 5.Initialization of Chromosomes

Decoding

In this study, each gene in an individual is a model coefficients that need decoding in order to obtain approximate solutions. Each coefficient multiplied by the corresponding time series data in order to obtain a predicted value by 13.

B.2. Predefine Fitness Function

Fitness value is a measure of good or not a solution described in one individual. In this study, the value of fitness, Eq. 4.

$$Fitness = 1/(1+Fr) \Rightarrow Fr = e(t) \tag{4}$$

Where, Fr is objective function; if $Fr \approx 0$ then $Fitness \approx 1$. In this study, prediction error is calculated using, Eq. 5.

$$e(t) = MSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n} e_i(t)^2} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2} \tag{5}$$

Where, n is the number of districts/cities or actual and predictive values to i in the districts/cities. The fitness will be calculate by Eq. 6.

$$Fitness = 1/(1+MSE) \tag{6}$$

B.3. Selection

All individuals in each generation are the main candidates. In this study, parent selection is using Roulette Wheel. The purpose of the selection of the parent is to ensure the existence of a superior individual. The process is replacing the parent with fitness worst value will be replaced with elitism.

Elitism

Elitism is the selection of an individual that have the best Fitness value as superior individuals. Later, in every generation, the best Fitness value will be compared with elitism value. If the elitism value is greater, then it will divide.

B.4. Generate New Population

New population is the result of the whole process of reproduction coupled with elitism. Every individual in the new population would be individual candidates for new generation through the evaluation of the value of Fitness. Flowchart of GA, as shown in Fig. 7.

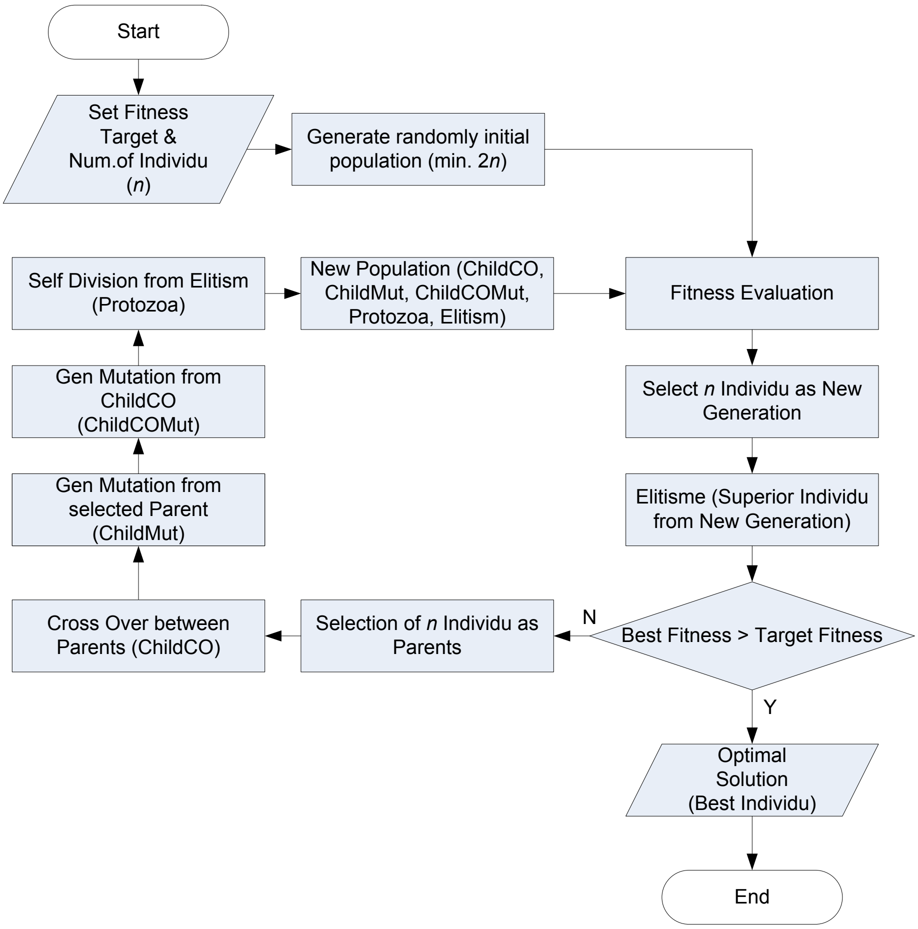

Fig. 6.Flowchart of GA

B.5. Training Data

In this study, economics regional includes regional economic growth, agriculture, and industrial shares for 13 districts/cities in East Kalimantan from 2002-2012 were obtained from Central Bureau of Statistics (BPS) East Kalimantan in 2013. The economics regional data can be seen in Table 1, 2, and 3. Each data variables are identified as: D1, D2, ..., D13. Then, the training data are 2002-2010, and target data is 2011, and test data is 2012. Furthermore, the total number of genes in an individual determined based on data from 2002-2010. That is, each variable or chromosome will have a number of data $((2012-2002)+1)\times 13 = 117$. The total number of genes in an individual is $117\times 3 = 351$. The data scheme illustrated in Fig 7.

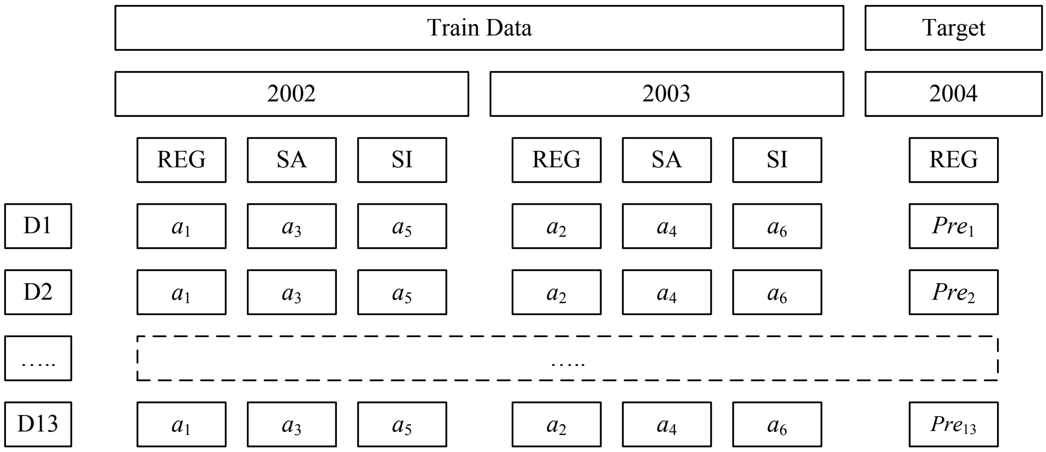

Fig. 7. Training Data Scheme

B.5. Performance Model

In this experiment, mean absolute error (MAE) was engaged the predicted output with the desired output. The MAE can be expressed, Eq. 7 and 8.

$$BestFit = (1 - MAE) \times 100\% \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} abs\left(e_i(t)\right) \tag{8}$$

III. Results and Discussions

This section describes the test results of REG data using the GA. The parameters of GA; total population of 20, fitness target of 0.95, and populations of 50 were implemented. The results of training and testing as shown in Fig. 8. Furthermore, MSE was used to measure the degree of accuracy of prediction.

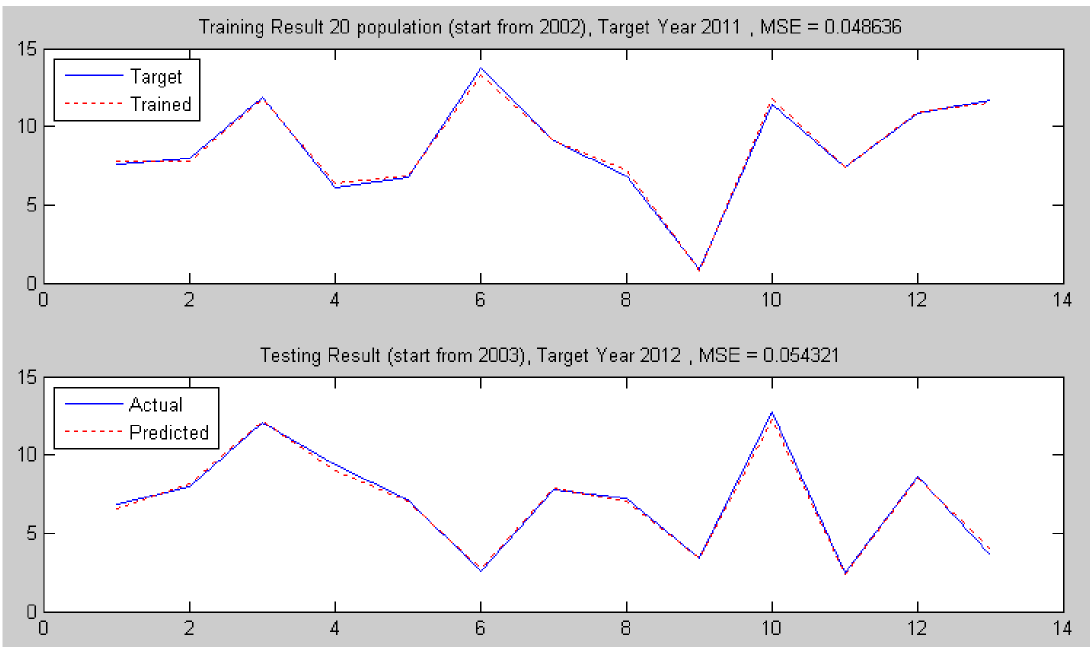

Fig. 8.Plot Training and Testing Results for REG

Based on experiment, the training MSE of 0.0486 and testing MSE 0.0543 has been resulted. The results of prediction in 2002-2012 produced MSE of 0.0616 has been as shown in Fig. 9.

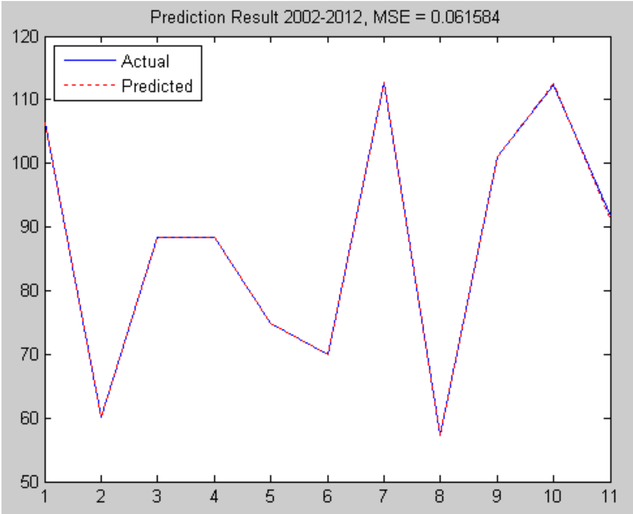

Fig. 9. Plot Prediction Result of REG 2002-2012

The best fitness of REG has been predicted in 2002 – 2012 by using Eq. 7, as shown.

$$
\begin{aligned}
BestFit &= (1-MAE)\times 100\% \\
&= (1-0.0939)\times 100\% \\
&= 90.61\%
\end{aligned}
$$

Based on best fitness calculation shows that the GA has been produced a good model for time series prediction of REG with 92.389 %. The prediction results on 2013 as indicated in Fig. 10.

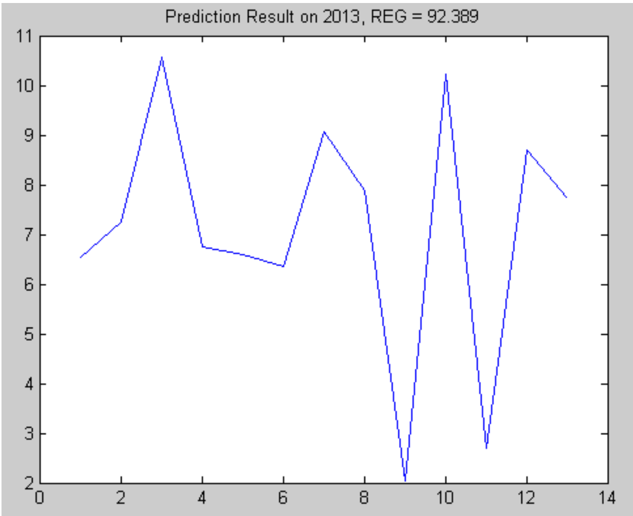

Fig. 10. Plot of REG Prediction Results 2013

The prediction results in each regions in 2013, Table 4, as follows: for northern regions (Tarakan is 6.527, Berau is 7.255, Malinau is 10.564, Bulungan is 6.749, and Nunukan is 6.597); for central regions (Samarinda is 6.353, Bontang is 9.057, Kutai Barat is 7.889, Kutai Kartanegara is 2.046, and Kutai Timur is 10.233); and for southern regions (Balikpapan is 2.700, Paser is 8.701, and Penajam Paser Utara is 7.718). In this experiment, the average fluctuation of REG was also investigated as shown in Fig. 11.

Table 4. Regional Economic Growth (REG) 2002 – 2013

Year	Tarakan	Berau	Malinau	Bulungan	Nunukan	Samarinda	Bontang	Kutai Barat	Kutai Kartanegara	Kutai Timur	Balikpapan	Paser	Penajam Paser Utara	East Kal.
2002	19.189	5.000	12.663	3.588	9.389	0.400	4.061	13.495	5.682	18.304	4.498	4.834	5.339	106.442
2003	4.412	4.158	6.031	0.010	5.972	9.521	3.961	10.493	0.213	3.110	2.132	5.249	4.912	60.174
2004	2.646	2.634	1.236	0.011	9.181	9.001	0.007	5.436	1.962	23.809	6.072	23.569	2.784	88.348
2005	7.631	5.108	3.630	0.012	9.559	8.046	6.500	8.235	2.672	20.585	3.214	8.927	4.163	88.282
2006	7.506	5.079	3.073	0.012	1.298	5.495	4.681	6.108	1.041	22.391	4.619	11.936	1.632	74.871
2007	6.912	5.701	6.314	5.885	3.766	3.107	4.810	6.453	0.117	8.075	2.079	12.918	3.787	69.924
2008	6.832	36.581	8.033	5.637	4.400	4.819	10.361	6.834	0.333	4.419	12.374	7.101	4.992	112.716
2009	4.633	5.976	8.953	4.655	3.931	4.495	2.610	6.889	2.003	0.262	1.703	7.738	3.507	57.355
2010	7.930	8.036	14.274	5.632	5.628	6.163	6.761	6.877	0.594	9.325	5.193	17.313	7.275	101.001
2011	7.626	7.934	11.894	6.056	6.738	13.772	9.128	6.891	0.882	11.433	7.439	10.847	11.681	112.321
2012	6.819	7.933	12.052	9.381	7.130	2.607	7.777	7.214	3.403	12.682	2.502	8.614	3.614	91.728
2013	6.527	7.255	10.564	6.749	6.597	6.353	9.057	7.889	2.046	10.233	2.700	8.701	7.718	**92.389**

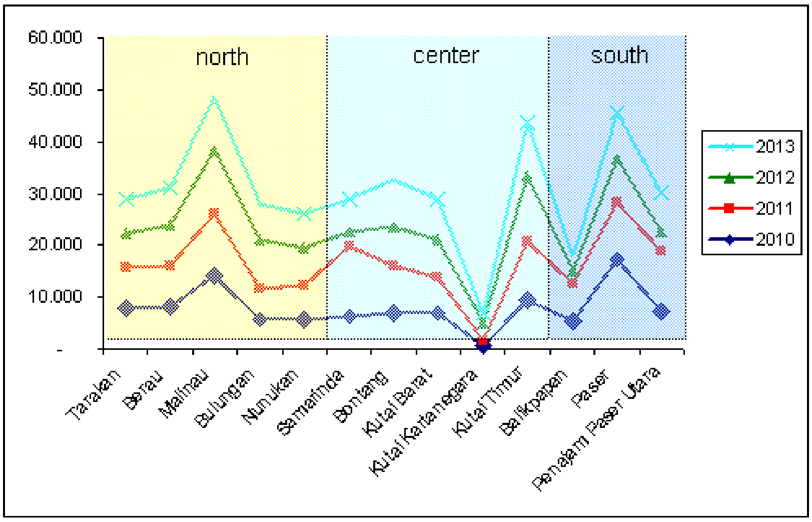

Fig. 11. Plot of Prediction Fluctuations of REG 2010-2013

Based on Table 4, the average fluctuation REG based group of construction area, have been obtained for the north is 0252, the center is 36 377, and south is 3,661. Then, Central and Northern regions will decline, while the South will rise quite dramatically in 2013, Fig. 12.

Table 5. Average fluctuation of REG based on development area in East Kalimantan

Year	North	Center	South
2011	1.09	45.95	22.16
2012	10.29	41.11	-52.00
2013	-12.14	22.06	40.83
mean	-0.252	36.377	3.661
2011	1.09	45.95	22.16
2012	-0.25	36.38	3.68
2013	-12.14	22.06	40.83

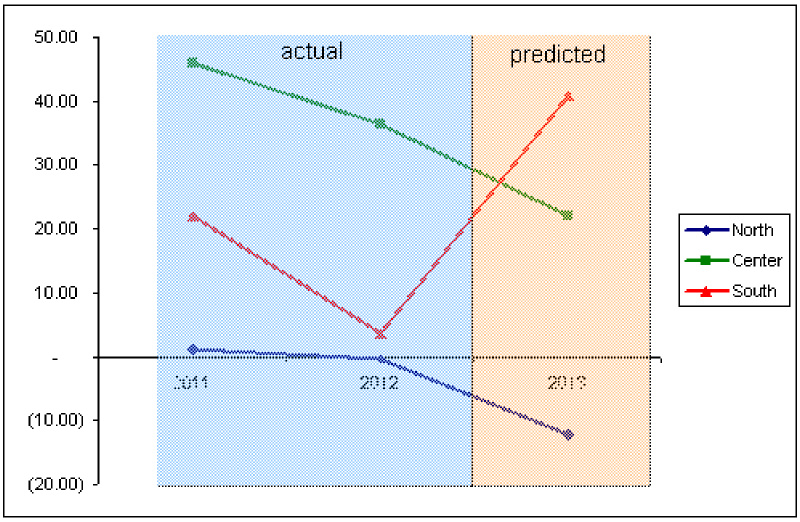

Fig. 12. Plot of Average Fluctuation of REG

IV. Conclussions

Genetic algorithm (GA) has been able as a forecasting model with time series datasets on Regional economic growth (REG) in 13 districts / cities, East Kalimantan Province. Test performance of GA model that have been trained and tested by using the Best-Fit criteria. The test results show that the performance of GA model of 90.61% can be categorized very well. Therefore, future research is expected to combine with neural network method in order to generate a higher accuracy of MSE and more efficient in the forecasting of Regional economic growth (REG).

References

[1] J. Adams-Kane and J. J. Lim, "Growth Poles and Multipolarity," 2011.

[2] L. Deng and B. Wang, "Regional capital flows and economic regimes: Evidence from China," *Economics Letters,* vol. 141 (2016), pp. 80–83, 2016.

[3] U. G. Emmilya, "Study for Effect of Growth and Regional Development Against Regional Economic Growth in East Kalimantan Province," Ph.D., Economic, Hasanuddin University – Makassar – South Sulawesi, Makassar, 2014.

[4] BPS. (2013). *Kaltim dalam Angka*.

[5] L. Feng and J. Zhang, "Application of artificial neural networks in tendency forecasting of economic growth," *Economic Modelling,* vol. 40 (2014), pp. 76–80, 2014.

[6] C.-C. Hsu and C.-Y. Chen, "Regional load forecasting in Taiwan—applications of artificial neural networks," *Energy Conversion and Management,* vol. 44 (2003), pp. 1941–1949, 2003.

[7] Y. Yang, "Development of the regional freight transportation demand prediction models based on the regression analysis methods," *Neurocomputing,* vol. 158 (2015), pp. 42–47, 2015.

[8] Haviluddin, R. Alfred, J. H. Obit, M. H. A. Hijazi, and A. A. A. Ibrahim, "A Performance Comparison of Statistical and Machine Learning Techniques in Learning Time Series Data," *Advanced Science Letters,* pp. 3037-3041, 2015.

[9] F. F. Ping and F. X. Fei, "Multivariant forecasting mode of Guangdong province port throughput with genetic algorithms and Back Propagation neural network," in *13th COTA International Conference of Transportation Professionals (CICTP 2013)*, 2013, pp. 1165 – 1174.

[10] Haviluddin and R. Alfred, "A Genetic-Based Backpropagation Neural Network for Forecasting in Time-Series Data," in *The 2015 International Conference on Science in Information Technology (ICSITech 2015)*, Yogyakarta, Indonesia, 2015, pp. xxx-xxx.

[11] Y. Perwej and A. Perwej, "Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm," *Journal of Intelligent Learning Systems and Applications,* vol. 4, (2012), pp. 108-119, 2012.

[12] E. J. Gill, E. B. Singh, and E. S. Singh, "Training Back Propagation Neural Networks with Genetic Algorithm for Weather Forecasting," in *8th International Symposium on Intelligent Systems and Informatics*, September 10-11, 2010, Subotica, Serbia, 2010, pp. 465-469.

[13] F. Song and H. Wang, "Hybrid Algorithm Based On Levenberg-Marquardt Bayesian Regularization Algorithm and Genetic Algorithm," in *The 2013 International Conference on Advanced Mechatronic Systems*, Luoyang, China, 2013, pp. 51-56.

[14] C.-X. Yang and Y.-F. Zhu, "Using Genetic Algorithms for Time Series Prediction," in *2010 Sixth International Conference on Natural Computation (ICNC 2010)*, 2010, pp. 4405-4409.

[15] R. Alfred, "Summarizing Relational Data Using Semi-Supervised Genetic Algorithm-Based Clustering Techniques," *Journal of Computer Science,* vol. 6, pp. 775-784, 2010.

[16] A. Sedki, D. Ouazar, and E. El Mazoudi, "Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting," *Expert Systems with Applications,* vol. 36, (2009), pp. 4523–4527, 2009.

[17] R. Alfred and D. Kazakov, "Aggregating Multiple Instances in Relational Database Using Semi-Supervised Genetic Algorithm-based Clustering Technique," in *ADBIS 2007*, Technical University of Varna, 2007, pp. 136-147.