A Heuristic Network for Predicting the Percentage of Gross Domestic Product Distribution

Emmilya Umma Aziza Gaffar, Irwan Gani
Department of Economic Study
Mulawarman University
East Kalimantan, Indonesia
emmilya.gaffars@gmail.com, ig.irwangani@gmail.com

Haviluddin
Faculty of Computer Science and Information Technology
Mulawarman University
East Kalimantan, Indonesia
haviluddin@gmail.com

Achmad Fanany Onnilita Gaffar
Department of Information Technology
State Polytechnic of Samarinda
East Kalimantan, Indonesia
onnygaffar212@gmail.com

Rayner Alfred
Associate Professor of Computer Science
Faculty of Computing and Informatics
University Malaysia Sabah
ralfred@ums.edu.my

***Abstract*—In a country, economic growth indicators are measured by GDP growth rates. In order to predict GDP growth then, economic situation and economic development strategy have always been used. In this study, GDP contribution at the current price by the industrial sector has been predicted by using heuristic network method. There were nine GDP growth variables including (1) Agriculture, Livestock, Forestry, Fishery, (2) Mining & Quarrying, (3) Manufacturing Industry, (4) Electricity, Gas, Water supply, (5) Construction, (6) Trade, Hotel, Restaurant, (7) Transport & Communication, (8) Finance, Real Estate & Business Services, and (9) Services in the period 2001-2016 have been analyzed. Experimental results show that the error rate forecasting in 2017 is less than 10%. The results show that intelligent computing method (heuristic network) can be an alternative method for predicting the contribution of GDP. This method predicts fairly quickly, significantly and produces an acceptable error prediction.**

Keywords*—*GDP; prediction; heuristic network

I. Introduction

Gross Domestic Product (GDP) is a basic measure of added value arising from economic activities at the national level. At the regional level (province / district / city) is called Gross Regional Domestic Product (GRDP). There are three ways to compile a country's GDP statistics; although in principle it gives the same result. The easiest way is the production approach, which measures the added value generated by different types of economic activity. The expenditure approach works based on the principle that all products must be purchased by a person. In this case the total value of the product must equal the total expenditure. The income approach works based on the principle that the income factor must equal production value, where GDP is determined by calculating the total revenue of all producers [1]. In other words, GDP / GRDP is the total value added generated by all economic activities and how it is used. GDP and its aggregate are presented in two forms, namely based on current prices and on a constant basis. All aggregates are valued at current prices, while constant prices in the base year are shown by assessing all aggregates at the base year price [2].

The recording of national statistics has made a change in the base year of GDP in Indonesia from 2000 to 2010. This basic change was made in conjunction with the adoption of United Nations (UN) recommendations as contained in the 2008 National Accounts System (SNA 2008) [3]. GDP by industrial sector is detailed according to the total added value of all existing economic sectors. By 2015, GDP by industrial sector has undergone a change of classification from 9 to 17 industrial sectors [4]. To meet the needs of the sample data, the study still uses 9 industry sector classifications with adjustment of each sector's detail. The classification in question is: (1). agriculture, livestock, forestry, & fishery, (2). mining & quarrying, (3). manufacturing industry, (4). electricity, gas, & water supply, (5). construction, (6). trade, hotel, & restaurant, (7). transportation & communication, (8). finance, real estate, & business services, (9). other services.

The economic growth of a country is measured by the GDP growth rate. The economic growth set as an increase in GDP strongly helps the government to predict the economic situation and development of economic development strategies. This measurement can be done by combining the computational mathematical concepts and computer technology to produce qualitative and quantitative predictions of economic growth scientifically and appropriately. The use of the scientific method proved very powerful to predict the development of GDP of a particular economy in the future [5].

Statistical methods are conventional methods that prove to be good enough to be used in predictive and forecasting activities. This method is based on the assumption that time series data has stationary properties and satisfies linearity rules when transformed into other forms. That is, when a prediction value is obtained, it should be returned to its original value using the reverse transformation principle. The time series data used needs to be modeled using two approaches, namely parametric and non-parametric approaches [6]. Time series data modeling with parametric approach is usually based on the ARIMA (Auto Regressive Integrated Moving Average) - Bob Jenkins [7]. To obtain all the regression coefficients can be done by using OLS (Ordinary Least Square) tested using GARCH (Generalized Auto Regressive Conditional Heterokedasticity). T-Test table is also commonly used to compare the calculation results of all regression coefficients with hypothetical values [8]. The fundamental weakness of statistical methods is when the time series data used does not meet the stationary and linearity requirements. In this case is usually used a non-parametric approach, but requires a fairly complicated computation. Machine learning method is a modern method that can overcome the weaknesses of statistical methods, as well as able to improve prediction or forecasting results. Sometimes, some researchers also combine machine learning methods with statistics. Some of the research that has been done related to this is stated in [1, 9-16].

Both statistical and machine learning methods involve various computational algorithms. There are two main roots in logic and reasoning within the scope of the philosophy of science and mathematics which is the basis of all computational activity, i.e., ampliative and non-ampliative reasoning. The term "ampliative" used in logic philosophy which means "extending" or "adding to the already known". Heuristics are classified as ampliative reasoning. Unlike algorithms, heuristics are problem-solving, learning, or discovery approaches that use practical methods that are not guaranteed to be optimal or perfect, but significant enough for goal attainment [17]. The heuristic approach has been widely applied in various studies, both in the areas of predictive problems, selection, and search problems, some of which have been published in [18-23].

In general, the predicted percentage of GDP distributions at current prices is usually by applying statistical concepts, i.e., sequential. In this study, the predicted percentage of GDP distribution at current prices by the industrial sector has been done by applying the concept of heuristic network. Meanwhile, time series data of GDP distributions are modeled using a weighted network, where the values of each network weighting are obtained using the heuristic approach. Furthermore, the purpose of this study is to simultaneously predict the nine variables of industrial sector GDP by applying a heuristic network in order to provide an overview of economic conditions for investors in East Kalimantan Province.

II. METHOD

A. Time Series Data Modeling

Time-series data can be represented by various model, one of them is AR (Auto Regressive) model. The AR model is one type of ARIMA - Bob Jenkins model that an observation at time t expressed by a linear function to a certain amount of n time preceding plus a residual random e_t (white noise). White noise is independent and has a normal distribution with mean = 0 and covariance σ. White noise is also called a prediction error. AR model expressed by (1).

$$y(t) = a_1.y(t-1) + a_2.y(t-2) + \cdots + a_n.y(t-n) + e(t) \tag{1}$$

Where, $y(t-1) \dots y(t-n)$ is the result of observation of a number of n previous data, $a_1 \dots a_n$ is the weighting of each data, $y(t)$ is the current data, and $e(t)$ is the residual/prediction error of the AR model. By searching for the value of $a_1 \dots a_n$ such that $e(t) \to 0$ then the model obtained can be used to predict the next data $\big(y(t+1)\big)$.

If (1) is expressed in discrete form where $t = k+1$ and $n = 3$ then using (2).

$$y(k+1) = a_1.y(k) + a_2.y(k-1) + a_3.y(k-2) + e(k+1) \tag{2}$$

Equation (2) can be rearranged in (3).

$$\begin{aligned} y(k+1)_1 &= a_1.y(k) + f\begin{pmatrix} a_2.y(k-1) + \\ a_3.y(k-2) \end{pmatrix} + e(k+1)_1 \\ y(k+1)_2 &= a_2.y(k-1) + f\begin{pmatrix} a_1.y(k) + \\ a_3.y(k-2) \end{pmatrix} + e(k+1)_2 \\ y(k+1)_3 &= a_3.y(k-2) + f\begin{pmatrix} a_1.y(k) + \\ a_2.y(k-1) \end{pmatrix} + e(k+1)_3 \end{aligned} \tag{3}$$

Where, $y(k+1)_1$ and $e(k+1)_1$ are the dominant contribution and the dominant residual error of $y(k)$ for $y(k+1)$, respectively, and so on. By summing all the forms in (3) then obtained can be (4).

$$y(k+1) = \begin{aligned} &F_1\left(u_1.y(k) + u_2.f_1\begin{pmatrix} a_2.y(k-1) + \\ a_3.y(k-2) \end{pmatrix}\right) \\ &+F_2\left(u_3.y(k-1) + u_4.f_2\begin{pmatrix} a_1.y(k) + \\ a_3.y(k-2) \end{pmatrix}\right) \\ &+F_3\left(u_5.y(k-2) + u_6.f_3\begin{pmatrix} a_1.y(k) + \\ a_2.y(k-1) \end{pmatrix}\right) \\ &+e(k+1) \end{aligned} \tag{4}$$

Where the variables $u_1 \dots u_6, a_1 \dots a_3$ are the weighted constants to be searched and $f_1 \dots f_3, F_1 \dots F_3$ are logistics function to ensure the sum of all data multiplication with each weighting within an acceptable range of data. Equation (4) can be represented in a MISO Heuristic Network (Multi Input Single Output) as shown in Fig. 1.

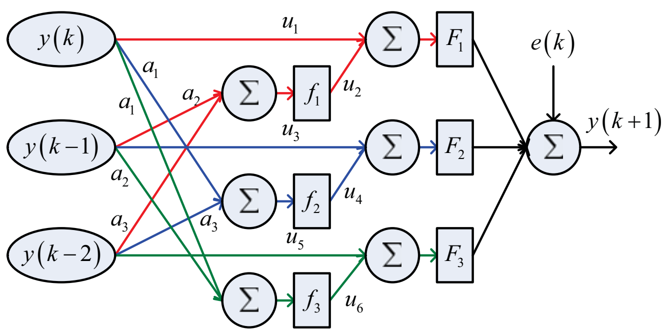

Fig. 1. Heuristic Network of MISO-AR Model

B. *Heuristic Network*

For predicting time series data involving multiple outputs, the heuristic network can be described as shown in Fig. 2.

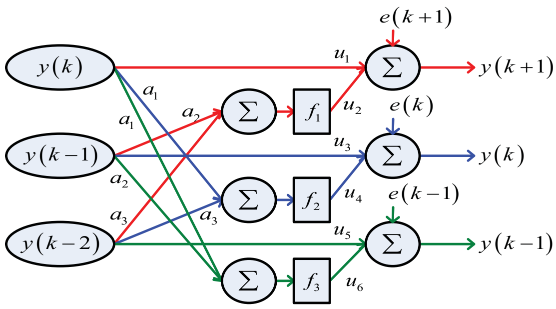

Fig. 2. Heuristic Network of MIMO-AR Model

In general, the heuristic networks used for predictive MIMO models (Multi Input Multi Output) are shown in Fig. 3.

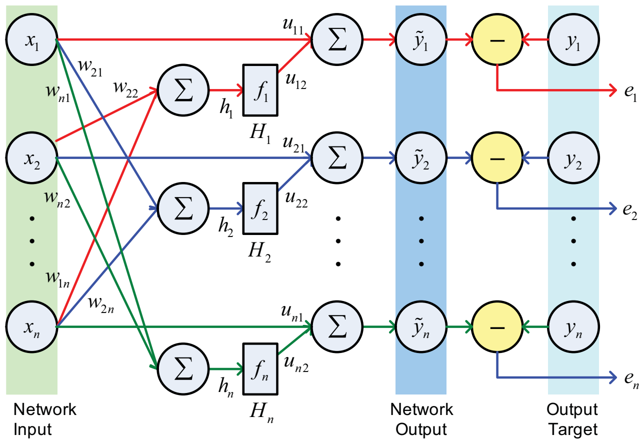

Fig. 3. Heuristic Network of MIMO Model

Mathematically expressed by (5).

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ \dots \\ h_n \end{bmatrix} = \begin{bmatrix} 0 & w_{12} & w_{13} & \dots & w_{1n} \\ w_{21} & 0 & w_{23} & \dots & w_{2n} \\ w_{31} & w_{32} & 0 & \dots & w_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & w_{n3} & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} \tag{5}$$

$$\boldsymbol{h} = \boldsymbol{W}_{(diag=0)} * \boldsymbol{x} \quad \boldsymbol{H} = \boldsymbol{f}(\boldsymbol{h})$$

Network output expressed by (6).

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \\ \dots \\ \tilde{y}_n \end{bmatrix} = \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \\ \dots \\ u_{n1} \end{bmatrix} .* \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} u_{12} \\ u_{22} \\ u_{32} \\ \dots \\ u_{n2} \end{bmatrix} .* \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ \dots \\ H_n \end{bmatrix} \tag{6}$$

$$\tilde{\boldsymbol{y}}_{(n,1)} = \boldsymbol{u}_{(n1,1)} .* \boldsymbol{x} + \boldsymbol{u}_{(n2,1)} .* \boldsymbol{H}_{(n,1)}$$

Network error expressed by (7).

$$e = y - (\tilde{y}) \tag{7}$$

Error function used is *SSE* (Sum Squared Error) expressed by (8).

$$SSE = E = \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{N}\left(y_j^{(k)} - \tilde{y}_j^{(k)}\right)^2 = \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{N}\left(e_j^{(k)}\right)^2 \tag{8}$$

Where, K is the number of training data, N is the number of heuristic network output.

The heuristic network is built to optimize all network weights in such a way as to obtain $E \to 0$ through a training process. Weighting adjustment is done by dividing network error based on the weighting proportion. From Fig. 3, then the division of network error expressed by (9) and (10).

$$du_{i1} = e_i .* \frac{x_i}{x_i + H_i} \quad du_{i2} = e_i .* \frac{H_i}{x_i + H_i} \quad i = 1 \dots n \tag{9}$$

$$dw_{ij} = du_{i2} .* \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \tag{10}$$

The heuristic network weighted adjustment is expressed by (11).

$$\begin{aligned} u_{i1(new)} &= u_{i1(old)} + du_{i1} \\ u_{i2(new)} &= u_{i2(old)} + du_{i2} \\ u_{ij(new)} &= w_{ij} + dw_{ij} \\ i,j &= 1 \dots n \end{aligned} \tag{11}$$

The iteration process stops, if the target is close to zero. The training process is used (12).

$$E = E_{target} \tag{12}$$

The performance of predicted results is calculated using *MAPE* (Mean Absolute Percentage Error) expressed by (13).

$$\begin{aligned} APE(i) &= \frac{|y_{actual}(i) - y_{pred}(i)|}{y_{actual}(i)} \times 100\% \\ MAPE &= \frac{1}{N}\sum_{i=1}^{N} APE(i) \end{aligned} \tag{13}$$

In order to ensure the convergence of training results, the average input should be close to zero. Then, the best data range were {0 ... 1}. Meanwhile, the log-sigmoid and tangent function that represents the interval have been applied using (14).

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{14}$$

In this study, both training input data and training targets were normalized using the following (15).

$$x_n(i) = \frac{x(i)}{max(X)} \tag{15}$$

Where, $x(i) \in X$.

C. Implementation strategy

In this study, the GDP distribution at current price by the industrial sector in period 2001-2016 have been captured from the Indonesia Statistics catalog [2-4, 24] as shown in Fig. 4. Based on machine learning concept, the dataset divided into three parts, includes training, testing and validation stages. In this experimental, data from 2001-2014 were used as training input data, while data from 2002-2015 were used as training targets. Data from 2016 were applied as training results validation data.

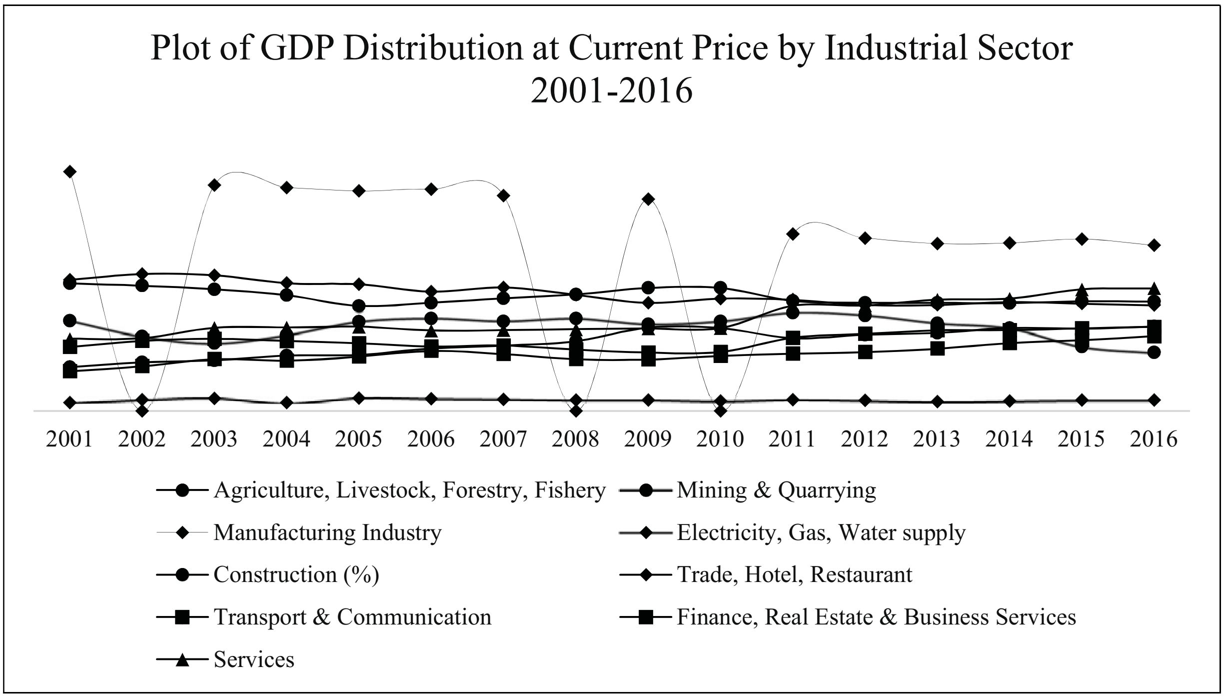

Fig. 4. Plot of GDP Distribution at Current Price by Industrial Sector 2001-2016

Since the predicted percentage is the percentage of GDP per sector, there will be 9 inputs and outputs. Before use, training data needs to be normalized using (15). Since the heuristic approach used in this study is the adjustment of network weights by dividing network errors by weighted proportions, the initialization weights $\left(\boldsymbol{W}_{(diag=\ )}, \boldsymbol{u}_{(n1,1)}, and\ \boldsymbol{u}_{(n2,1)}\right)$ are given the same value. Experiments were also conducted using the initial value variations to compare the performance of the results. The heuristic network used in this study is shown in Fig. 5.

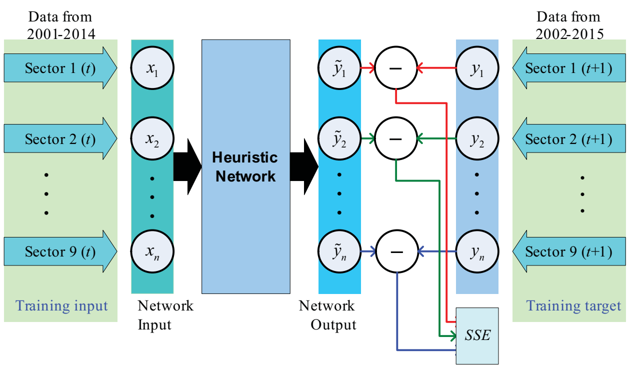

Fig. 5. The planned heuristic network of MIMO model used

III. RESULTS AND DISCUSSION

MATLAB software was used to implement the planned heuristic network method. By using $E_{target} = 10^{-4}$ and all weights were given initial value = 1, then after training the training results were obtained as shown in Fig. 6. The performance of training error was shown in Fig. 6. It was found that the final performance of the training error of 9.91×10^{-5} which reached at 386 iterations.

The trained heuristic network was used to predict all sectors in 2016 with data from 2015 as network input. The predicted results were then validated using data in 2016, with validation results as shown in Fig. 7. Table II shows the total percentage of GDP distribution was 100%, while the predicted result is 100.0024%. By using (13), the error of the total percentage of GDP distribution was obtained:

$$APE(i) = \frac{|100 - 100.0024|}{100} \times 100\% = 0.0024\%$$

Because this error was very small that it can be ignored. The validation of predicted results in 2016 has a value of MAPE = 8.33%. If it is determined that the maximum acceptable prediction error is 10% then this MAPE value was satisfactory. A trained and validated heuristic network is then used to predict the percentage of GDP distributions in 2017. The results were shown in Table I. It means that the GDP distribution at current price by the industrial sector has grown of 0.21% at 2017, Table II.

TABLE I. VALIDATION OF PREDICTION RESULTS IN 2016

Sector	Actual (%)	Predicted (%)	APE (%)
1	13.4500	13.3404	0.8148
2	7.2100	9.8371	36.4365
3	20.5100	21.0800	2.7791
4	1.2200	1.1600	4.9211
5	10.3800	9.8599	5.0105
6	13.1900	13.4269	1.7963
7	8.8400	7.9190	10.4181
8	9.9300	9.6891	2.4257
9	15.2700	13.6899	10.3475
Total		100.0024	
MAPE (%)			8.3277

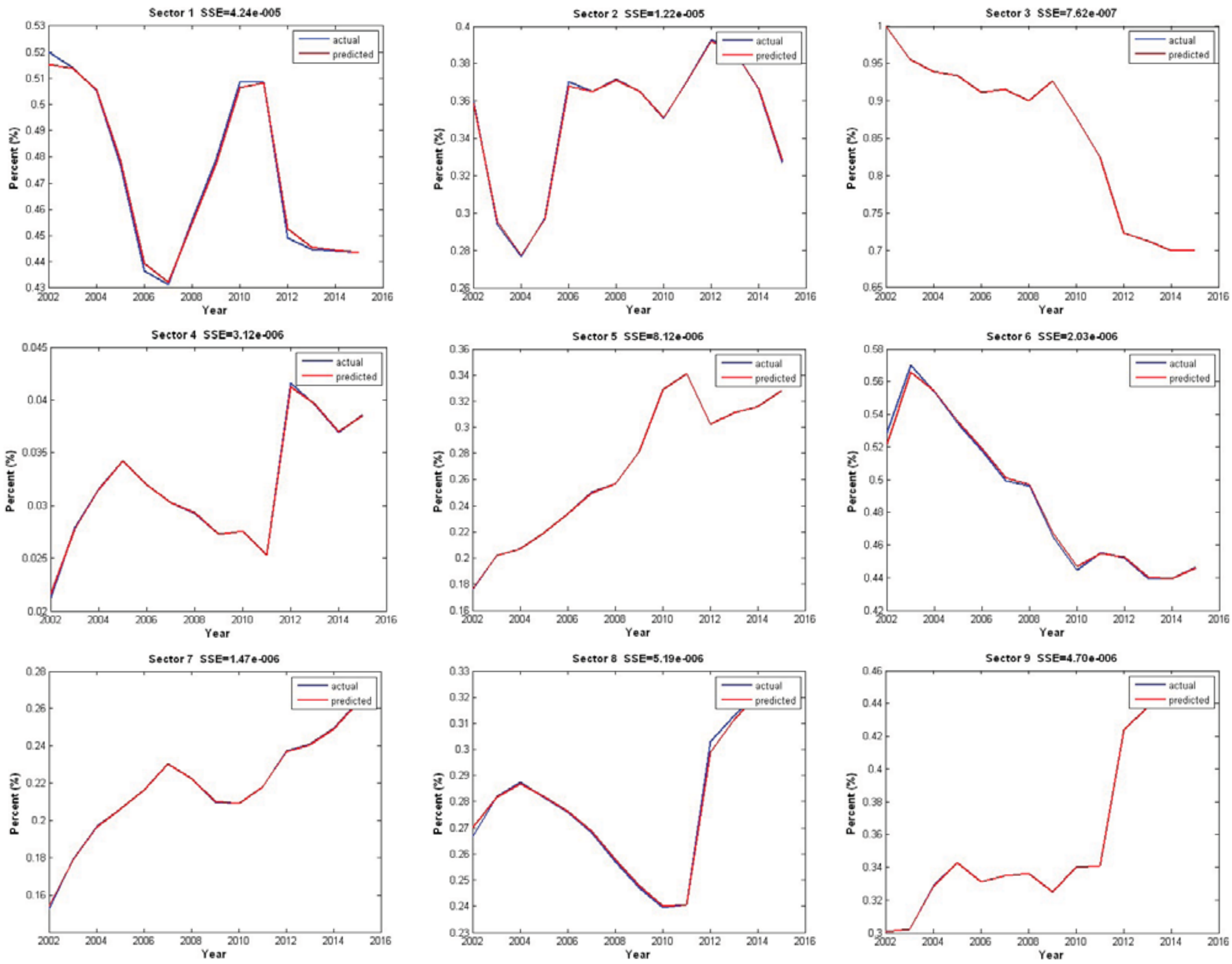

Fig. 6. The planned heuristic network of MIMO model

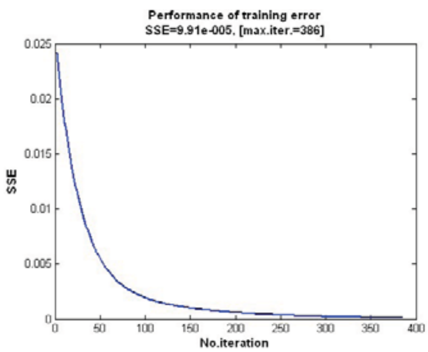

Fig. 7. Performance of training error heuristic network

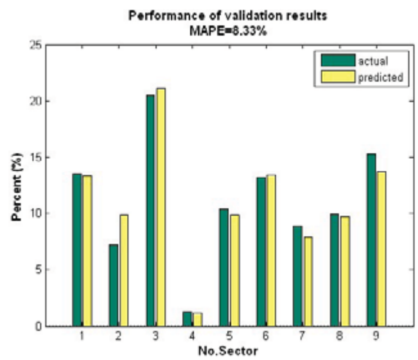

Fig. 8. Prediction results in 2016

TABLE II. PREDICTION RESULTS IN 2017

Sector	2016	Predicted 2017	Growth (%)
1	13.4500	13.4437	0.0471
2	7.2100	7.3001	1.2493
3	20.5100	20.5101	0.0006
4	1.2200	1.2188	0.0961
5	10.3800	10.3792	0.0076
6	13.1900	13.2066	0.1255
7	8.8400	8.8274	0.1423
8	9.9300	9.9134	0.1673
9	15.2700	15.2670	0.0195
Total	100.0000	100.0663	
Average growth (%)			0.2061

Table III shows that various weighted initialization values have been used to obtain predicted results. The experimental results show that the average APE has an error prediction below 10%.

TABLE III. PERFORMANCE OF HEURISTIC NETWORK WITH VARIOUS INITIAL WEIGHTING

No	Initial weighting	SSE	Maks. iteration	MAPE of validation (%)	APE of the total GDP distribution (%)
1	2.0000	9.9972E-05	1682	8.3086	0.0009
2	1.0000	9.9115E-05	386	8.3277	0.0024
3	0.5000	9.9662E-05	477	8.3518	0.0023
4	0.1000	9.9994E-05	352	8.7259	0.0420
5	0.0500	9.9518E-05	315	8.9103	0.0599
6	0.0100	9.9388E-05	161	9.1448	0.1256
7	0.0050	9.9362E-05	110	9.1735	0.1262
8	0.0010	9.9384E-05	42	9.1741	0.1962

IV. CONCLUSION

The heuristic network method as an optimal prediction is presented. This method has a fast computation process and good predictive error accuracy. Thus, the GDP growth in 2017 variables includes (1) agriculture, livestock, forestry, fishery of 0.0471, (2) Mining & Quarrying of 1.2493, (3) Manufacturing Industry of 0.0006, (4) Electricity, Gas, Water supply of 0.0961, (5) Construction of 0.0076, (6) Trade, Hotel, Restaurant of 0.1255, (7) Transport & Communication of 0.1423, (8) Finance, Real Estate & Business Services of 0.1673, and (9) Services of 0.0195 have been achieved. In other words, the percentage of GDP distribution in 2017 shows 0.21% growth has been proven. However, these results will be validated with the Indonesian Statistics Catalog 2018. Then, an experimental heuristic network using multivariate time series data will be the next research.

REFERENCES

[1] J. Vrbka, "Predicting Future GDP Development by Means of Artificial Intelligence," *Littera Scripta [online],* vol. 9, pp. 154-167, 2016.

[2] BPS-Jakarta, "Statistical Yearbook of Indonesia 2005/2006," BPS-Jakarta, Ed., ed. Jakarta-Indonesia: BPS - Statistics Indonesia, 2006.

[3] BPS-Jakarta, "Statistical Yearbook of Indonesia 2009," BPS-Jakarta, Ed., ed. Jakarta-Indonesia: BPS-Statistics Indonesia, 2009.

[4] BPS-Jakarta, "Statistical Yearbook of Indonesia 2015," BPS-Jakarta, Ed., ed. Jakarta-Indonesia: BPS-Statistics Indonesia, 2015.

[5] T. Wang, "Forecast of Economic Growth by Time Series and Scenario Planning Method—A Case Study of Shenzhen," *Modern Economy,* vol. 07, pp. 212-222, 2016.

[6] S. Saeed, L. Hussain1, I. A. Awan, and A. Idris, "Comparative Analysis of different Statistical Methods for Prediction of $PM_{2.5}$ and PM_{10} Concentrations in Advance for Several Hours," *IJCSNS International Journal of Computer Science and Network Security,* vol. 17, 2017.

[7] K. Kwon, W.-S. Cho, and J. Na, "ARIMAX and ARX Models with Social Media Information to Predict Unemployment Rate," *Journal of Advanced Management Science,* pp. 401-404, 2016.

[8] A. A. Abdurehman and S. Hacilar, "The Relationship between Exchange Rate and Inflation: An Empirical Study of Turkey," *International Journal of Economics and Financial Issues (IJEFI),* vol. 6, pp. 1454-459, 2016.

[9] B. Al-hnaity and M. Abbod. (2016, Predicting Financial Time Series Data Using Hybrid Model. *Intelligent Systems and Applications 650,* 19-41.

[10] N. Buduma, *Fundamentals of Deep Learning - Designing Next Generation Artificial Intelligence Algorithms*, First Edition ed. United States of America: O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2015.

[11] A. Camara, W. Feixing, and L. Xiuqin, "Energy Consumption Forecasting Using Seasonal ARIMA with Artificial Neural Networks Models," *International Journal of Business and Management,* vol. 11, p. 231, 2016.

[12] A. Dingli and K. S. Fournier, "Financial Time Series Forecasting - A Machine Learning Approach," *Machine Learning and Applications: An International Journal,* vol. 4, pp. 11-27, 2017.

[13] E. U. A. Gaffar, "Prediction of Regional Economic Growth in East Kalimantan using Genetic Algorithm," *International Journal of Computing and Informatics (IJCANDI)* vol. 1, pp. 58-67 May, 2016 2016.

[14] S. R. Y. Mayankkumar B Patel, "Stock Price Prediction Using Artificial Neural Network," *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), (An ISO 3297: 2007 Certified Organization),* vol. 3, June, 2014 2014.

[15] X.-N. Mergani Khairalla, Nashat T. AL-Jallad, "Hybrid Forecasting Scheme for Financial Time-Series Data using Neural Network and Statistical Methods," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 8, 2017.

[16] A. R. A. Murillo, "Short-Term Forecasting of Financial Time Series with Deep Neural Networks," Master, Faculty of Engineering, Department of Systems and Industrial Engineering, Universidad Nacional de Colombia, 2016.

[17] L. Magnani, *Heuristic Reasoning* vol. 16: Springer International Publishing Switzerland, 2015.

[18] A. Mittal, R. Aggarwal, and S. Kaur, "Analysis of energy efficient connected target coverage algorithm for static and dynamic nodes in IWSNs," *International Journal of Advances in Intelligent Informatics,* vol. 2, pp. 94-102, 2016.

[19] C. Mira, P. Feijao, M. A. Souza, A. Moura, J. Meidanis, G. L. §, R. Schmitz, R. P. Bossolan, and I. T. Freitas, "A GRASP-based heuristic for the Project Portfolio Selection Problem," in *IEEE 15th International Conference on Computational Science and Engineering*, 2012, pp. 36-41.

[20] M. B. Ridha, "The Role of Heuristic Methods as a Decision-Making Tool in Aggregate Production Planning," *International Journal of Business Administration,* vol. 6, 2015.

[21] I. Sendiña-Nadal, P. Wang, L.-J. Zhang, X.-J. Xu, and G. Xiao, "Heuristic Strategies for Persuader Selection in Contagions on Complex Networks," *Plos One,* vol. 12, p. e0169771, 2017.

[22] C. Thirumalai, K. S. Harsha, M. L. Deepak, and K. C. Krishna, "Heuristic prediction of rainfall using machine learning techniques," pp. 1114-1117, 2017.

[23] G. d. Tollo, "Hybrid Metaheuristic for Portfolio Selection: Comparison with an exact solver and search space analysis," vol. 5, pp. 579-588, 2015.

[24] BPS-Jakarta, "Statistical Yearbook of Indonesia 2017," BPS-Jakarta, Ed., ed. Jakarta - Indonesia: BPS - Statistics Indonesia, 2017.