Modeling Crude Oil Prices (CPO) using General Regression Neural Network (GRNN)

Rezzy Eko Caraka[1,2]

[1]*School of Mathematics, Faculty of Science and Technology, the National University of Malaysia, Malaysia*
[2]*Bioinformatics and Data Science Research Center, Bina Nusantara University, Indonesia*

Abstract— *Modeling time series is often associated with the process forecasts certain characteristics in the next period. One of the methods forecasts that developed nowadays is using artificial neural network or more popularly known as aneural network. Use neural network in forecasts time series can be agood solution, but the problem is network architecture and the training method in the right direction. General Regression Neural Network (GRNN) is one of the network model radial basis that used to approach a function. GRNN including model neural network model with a solution that quickly, because it is not needed each iteration in the estimation weight. This model has a network architecture that wasa number of units in pattern layer in accordance with the number of input data. One of the application GRNN is to predict the crude oil by using a model GRNN.From the training and testing on the data obtained by the RMSE testing 1.9355 and RMSE training 1.1048.Model is good to be used to give aprediction that is quite accurate information that is shown by the close target with the output*

Keywords— Artificial Intelligence; CrudeOil; General Regression Neural Network; Prediction.

I. INTRODUCTION

Final energy consumption in Indonesia for the period 2000 – 2012 increased by an average of 2.9% per year. The mostdominant type of energy is petroleum products which include aviation fuel, avgas, gasoline, kerosene, diesel oil, and fuel oil. These types of fuel consumed mostly by thetransport sector. Today, most of the fuel prices are still subsidized. Fuel subsidies in 2013 have reached 199 trillion rupiahs. The government is also still subsidizing electricity for aparticular type of users. Total electricity subsidies in 2013 reached 100 trillion rupiahs. The energy subsidy (fuel and electricity) has been increasing steadily. Energy subsidies in 2011 amounted to 195.3 trillion rupiahs and increased to 268 trillion rupiahs in 2013. Total spending on energy subsidies is always greater than the allocated budget and it often causes problems by the end of each fiscal year.The government has issued a number of policies to reduce petroleum fuel usage. Such policies include the kerosene to gas conversion program for thehousehold sector, compressed natural gas (CNG) usage for thetransport sector, and the mandatory use of biofuels which applies to industrial, transportation and power generation sectors. However, there are still many difficulties that must be faced and the petroleum fuel consumption is still increasing as strong as ever. It is essential to find solutions to any energy problems that arise. Energy planning is vital in order to ensure energy availability at affordable prices for a long term period. In line with this objective, BPPT contributes through Indonesia Energy Outlook Book that published annually. IndonesiaEnergy Outlook 2014 contains a long-term projection for the period 2012-2035 on energy balance, energy demand, energy supply, and energy infrastructure based on potential energy reserves and resources as well as current conditions.

Crude oil price is based on 2010 data with 105 \$/barrel (current price) and it assumed to be rising linearly to 126 \$/barrel in 2035. Oil production continues to decline while the demand for energy continues to grow which led to the increase in import of crude oil and petroleum products. This was shown by the deficit 3, 5 billion Dollar at oil account in thesecond quarter which increased from 2, 1 billion Dollar deficit in thefirst quarter of 2014 financial year. On the other hand, fuel subsidy is relatively high, due to increased domestic consumption, the increase in international oil prices and the decline in the exchange rate against thedollar and other foreign currencies. It is estimated that fuel subsidies until the end of 2014 will exceed the budget allocation in 2014.

Scientists have two different perspectives about AI (Artificial Intelligence). The first believes that AI as part of that only a focus on the process to think. While the second believes that AI as knowledge that focuses on their way in demand. This point to two see AI wider because of an appearance must be preceded by a process to think. That is the most suitable AI Definition for the moment is acting rationally with the approach rational agent. This was based on a thought that computer can make a logical reasoning and can also do the action in a rational reasoning based on the result was. A major impediment to scientific progress in many fields is the inability to make sense of the huge amounts of data that have been collected via experiment or computer simulation. In the fields of statistics and machine learning there have been major efforts to develop automatic methods for finding significant and interesting patterns in complex data, and for forecasting the future from such data. In general, however, the success of such efforts has been limited, and the automatic analysis of complex data analysis and prediction can often be formulated as search problems.

II. GENERAL REGRESSION NEURAL NETWORK

General Regression Neural Network (GRNN) is one of the network model radial basis that is often used to approach a function. The basis of the operation GRNN essentially based on multiple regression nonlinear (kernel) where the estimation of the hope that output is determined by input association [8] Although GRNN produces output in vector multivariate Equality (1) contract logic GRNN in regression formula nonlinear:

$$E[y|x] = \frac{\int_{-\infty}^{\infty} y f(X,y)dy}{\int_{-\infty}^{\infty} f(X,y)dy} \quad (1)$$

In this case, y is output that is predicted by GRNN while X is vector input (x1,x2, ... , Xp) which consists of p variables particular E[y|x]Was the expectation valueof the output y given vector input X and f(X,y) is a joint pdf from X and y.General Regression Neural Network from the initial estimate density kernel multivariate [9]. The purpose of the estimation non-parametric are estimated function probability density $F(z_1^*, \dots, z_m^*)$ of m variables at random $z = (z_1, \dots, Z_m)^T$ by using n size of each variables. Estimator density kernel multivariate in that case m dimension is defined as

$$F(z^*) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_1 \dots h_m} k\left(\frac{z_{i1}-z_1^*}{h_1}, \dots, \frac{z_{im}-z_m^*}{h_m}\right) \quad (2)$$

Where K is a function multivariate kernel and bandwidth vector $\mathbf{h} = (h_1, \dots h_m)^T$. The original data Z ($X_i$,$Y_i$); i=1,..n will be divided into data training is used for development of the model, while data association training came from a process sampling which measures the output with additives random noise:

$$\boldsymbol{z_i} = E[Z|x,y] + \varepsilon_i \quad (3)$$

Where $\varepsilon_i \sim NID\ (0, \sigma^2)$

If f(x,y,Z) is a function probability density then mean conditional are:

$$E[Z|x,y] = \frac{\int_{-\infty}^{\infty} Z.f(x,y,Z)dZ}{\int_{-\infty}^{\infty} f(x,y,Z)dZ} \quad (4)$$

Density function f(x,y,Z) can be estimated from data by using estimator continuously nonparametric as follows:

$$f(x,y,Z) = \frac{1}{[(2\pi)^{(\frac{p+1}{2})}\sigma^p]} \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\frac{(X-X_i)^T}{2\sigma^2}\right) \exp\left[\frac{-(Y-Y_i)}{2\sigma^2}\right] \quad (5)$$

With maximum p lag input and n is the number measurements in data association Training is a long-term, as well as the distance metric (Di^2) are:

$$D_i^2 = (X - X_i)^T (X - X_i) \quad (6)$$

With the substitution estimation probability (5) into the mean (4) was conditional, estimator kernel Nadaraya-Watson as follows [5] :

$$Z_m(X,Y) = \frac{\sum_{i=1}^{n} z_i \exp\left(-\frac{D_i^2}{2h^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{D_i^2}{2h^2}\right)} \quad (7)$$

III. THE STRUCTURE AND THE ARCHITECTURE GRNN

Construction GRNN consists of four layers is processing neurons input, pattern, summation and output. Input layer receives vector input X and distributes data to pattern layer. Every neuron in his patterns layer and build output θ and send the results to summation layer. Neurons numerator and denominator summation layer count final results were entered into a simple and weighted which is based on values and θ wij that, based on learning through training .Neurons in the output layer and the distribution of final results were entered into a which had been calculated by neurons in summation layer.

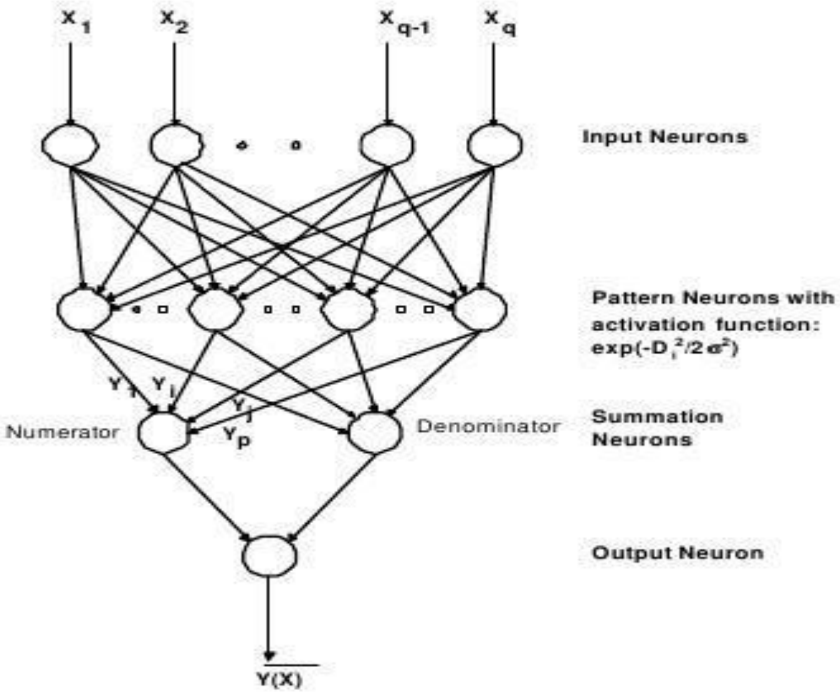

Fig. 1: Architecture GRNN

Each layer processing unit is marked with a specific function. The first Layer called neurons, responsible for receiving information. There is a neuron input for every single variable , in particular, is vector input X, there is no data processing is done in neurons input. Neurons input and then send data to thesecond layer of theprocessing unit, called neurons pattern. In this case, the number of neurons patterns together with the number of cases in training. Neurons pattern to i get data from neurons input and count output θ_i use transfer function:

$$\theta_i = e^{-(X-U_i)'(X-U_i)/2\sigma^2)} \quad (8)$$

In this case X is vector input from the particular is said to GRNN, Ui is vector training which is represented by neurons patterns, i and σ is parameter smoothing. The output of neurons patterns then continued to thethird layer of theprocessing unit, called neurons final results were entered into a (summation dialogs on) where theoutput from all neurons pattern is to be added. Technically there is two type weighted addition. In the topology GRNN processing unit there is a separate became clever arithmetic simple and weighted addition. Equality (9.a) and (9.b) each of them said mathematical operations

$$S_s = \sum_i \theta_i \quad (9.a)$$

$$S_w = \sum_i w_i \theta_i \quad (9.b)$$

$S_s$is simple arithmetic summation and S_w is weighted summation. Final results were entered into a produced by summation in a row dialogs on were sent to alayer of processing unit that neurons output. Neurons output to form the distribution here to get output regression GRNN y:

$$y = \frac{S_w}{S_s} \quad (9.c)$$

IV. PERFORMANCE INDICATORS

The mean square error (MSE), root mean square error (RMSE), between the modeled output and measures of the training and testing data set, are the most common indicators to provide a numerical description of the goodness of the model estimates. They are calculated and defined according to equations 10, 11

$$MSE = \frac{1}{Q}\sum_{k=1}^{Q}(t_k - y_k)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{Q}\sum_{k=1}^{Q}(t_k - y_k)^2} \quad (11)$$

Where:
Q = Number of observation
t_k= predicted value
y_k = observed value

V. RESULTS AND SIMULATION

Data daily basket price period January 2, 2003, to March 30, 2015, where there are 3158 data that noted. Some data preprocessing steps on raw set as shown below:

1. Firstly, 80% data were used to training
2. Secondly, data were normalized by min-max normalization into a specified range 0.0 to 1.0

The GRNN learning method simply stores the training patterns and processes them through a nonlinear smoothing function to determine the component output probability density functions. GRNN has been reported to deliver better performance than other neural networks and classical modeling methods, especially in noisy environment.Before the training of GRNN often necessary scaling on the input and the target such that the data input and the target falls within a certain range. It is intended that the data are processed in accordance with the activation functions are used. Therefore, the activation function used is radbas (Radial Basis), the radial-based function has a maximum value of 1, which occurs when the input received is worth 0. These instructions will produce a pn as the input data is normalized and tn as target data for training superbly normalized while the spread is the speed of learning. After performing preprocessing GRNN network to be formed.Newgrnn command will produce a network with two layers. The first layer contains neurons with radbas activation function, which will calculate the value weighted with dist and input commands with command netprod network

```
    [pn,meanp,stdp,tn,meant,stdt]=
prestd(P,T)
     net=newgrnn(pn,tn,spread)
     net_input_weighted
=net.IW{1,1}
     net_bias= net.b{1,1}
     side_effect = net.LW{2,1}
     net_bias_2=net.b{2,1}
     an=sim(net,pn);
     a=poststd(an,meant,stdt);
     e=T-a;
     msee=mse(e);
     ifmsee<min_mse
     min_mse=msee;
          nil_i=i;
     end
     end
```

The second layer is the activation function of neurons with purelin that will calculate the weighted with instructions normprod input and input with the command network netsum. Only the first layer will have a bias. With the spread is the spread of radial basis function with the default one. The next step is to do an instruction to seek input weights, layers, and bias. Then the result of the formation GRNN network will be simulated. Because the output network results GRNN has a mean = 0 and standard deviation = 1, then the next step to bring output Training Network GRNN according to its original condition (normalize) Next to the instruction to get the value of the mean square error (MSE) and the root mean square error (RMSE) and R2 process can be done in Matlab as follows: on the instruction would happen repetition (looping) to get the smallest error value. After the training, the model obtained using the new data (data testing), the new data should not be the same as the training data. Input test data stored in the matrix R target test data stored in the matrix TR. After the input and the target test data is normalized first. The next step brings the output of test results in accordance with its original condition (normalize)

```
     mse_training=min_mse
     rmse_training=sqrt(mse_train
ing)
     sse_training=norm(e.^2);
     sst_training=norm(T-
mean(a));
     R_Square_Training=(1-
(sse_training/sst_training))
     an=sim(net,pn);
     a=poststd(an,meant,stdt);
     R=[xt(s-k-1:s-1) xt(s-k-2:s-
     2) xt(s-k-12:s-12)]';
     TR=xt(s-k:s)';
     Rn=trastd(R,meanp,stdp);
     cn=sim(net,Rn);
     c=poststd(cn,meant,stdt);
     e2=TR-c;
     mse_testing=mse(e2)
     rmse_testing=sqrt(mse_testin
g)
     sse_testing=norm(e2.^2);
     sst_testing=norm(TR-
mean(c));
     R_Square_testing=(1-
(sse_testing/sst_testing))
```

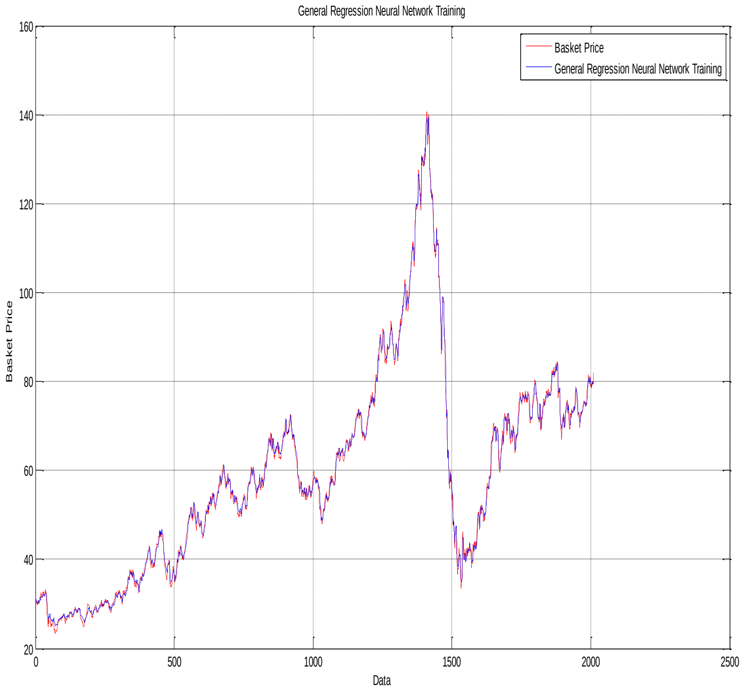

Fig.2: Testing Phase

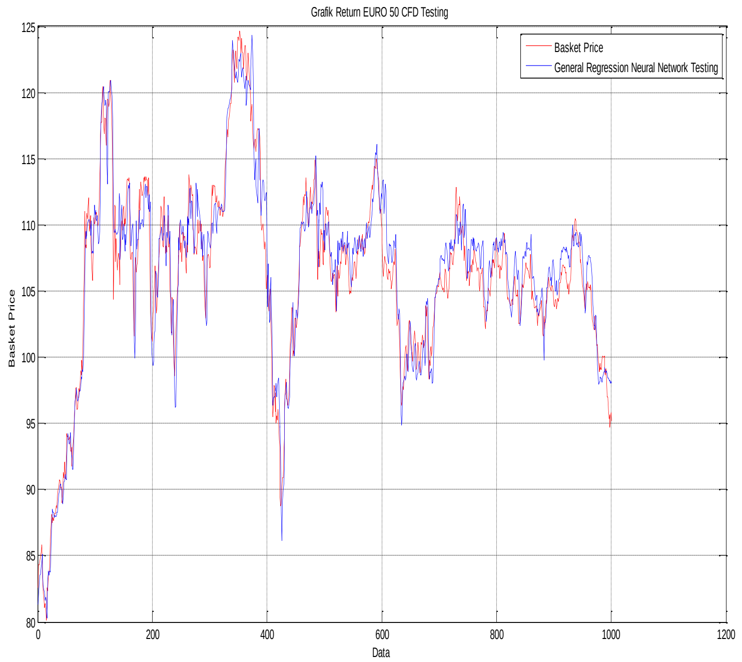

Fig..3: Training Phase

Based on Fig.2 comparison target and output training GRNN seen that the training network has prediction that is quite accurate information that is shown by the close target line (blue) with the output (red). But, even though it is still needed performance evaluation network in general to see the result of a network testing GRNN. By using GRNN, the MSE training value is 1.2205, RMSE training value is 1.1048. MSE testing value is 3.7463 and RMSE testing value is 1.9355. The main advantages of GRNN very simple and fast training procedure, prediction is unique not dependent on training procedure and initial conditions

VI. CONCLUSION

A number of weight bias input period with the number of input data that is 1010. The spread is function radial basis with the standard=1. Layers pattern consists of 1010 neurons that were established in the training network GRNN. Every neuron layer of learning methods to find out the distance between data integrity input

(D_i), after that, to seek the θ_i ; wherein θ_i modeling GRNN using Matlab is a function activating radial basis to get input from the distance between data to get input from a distance between the data input and breach weight bias input $\theta_i = e^{-(X-U_i)'(X-U_i)/2\sigma^2)}$. After that neurons pattern spins off signals that will be forwarded to the final results were entered into a (summation layer). This separation signals there is the one that is a heavy layer and there are also those who are not being given the weight layers. The weight layer (W_i) is equivalent to the target in training network GRNN. In the upper layers final results were entered into a (summation layer) there are 2 units of neurons are the simple arithmetic summation and weighted summation. On the neurons summation arithmetic summation (S_S), will be done addition of each value θ_iWhile in summation $S_S = \sum_{i=1}^{1010} \theta_i$in the upper layers output $S_W = \sum_{i=1}^{1010} \theta_i W_i$ consisting of 1 unit neurons that make the distribution weighted summation shared with simple arithmetic. With equality (9.c) it is possible to predict from the weight or parameters that unites the GRNN for time series can be written in the form similarities as follows equality (12).

$$y = \frac{s_W}{s_S} = \frac{\sum_{i=1}^{1010} W_i \times \theta_i}{\sum_{i=1}^{1010} \theta_i} =$$

$$\frac{\sum_{i=1}^{1010} W_i \times e^{-(b\sqrt{(X_{t-1}-X_{t-i}i)^2+(X_{t-2}-X_{t-2}i)^2+(X_{t-12}-X_{t-12}i)^2}}}{\sum_{i=1}^{1010} e^{-(b\sqrt{(X_{t-1}-X_{t-i}i)^2+(X_{t-2}-X_{t-2}i)^2+(X_{t-12}-X_{t-12}i)^2}}} \quad (12)$$

where

S_s=simple arithmetic summation

S_w =weighted summation.

Wi= weight training results in network layer GRNN

θ_i = values neurons in each layer pattern

The oil price is partly determined by actual supply and demand, and partly by expectation. Demand for energy is closely related to economic activity. If producers think the price is staying high, they invest, which after a lag boosts supply. Similarly, low prices lead to an investment drought. Crude oil price is based on 2010 data with 105 $/barrel and it assumed to be rising linearly to 126 $/barrel in 2035. Improvement in theeconomy will encourage petroleum fuel utilization, especially in the transport sector as its main user.This has to be supported by an adequate increase in crude oil supply.

ACKNOWLEDGEMENT

I want to send you the most wonderful message with good health and happiness. Happy birthday to the most amazing person and best friend KadiMey Ismail. Master of Aquaculture and Marine Resource Management, Wageningen University and Research Center. Who has taught me to live and enjoy life. Let's make it memorable!

REFERENCES

[1] Caraka,R.E.,Yasin,H "PrediksiProduksi Gas Bumidengan General Regression Neural Network (GRNN)", in Proc.National Statistics Conference (SNS IV).2014, ISSN : 2087-2590 pp 270-277.

[2] Caraka,R.E.,Yasin,H., dan Prahutama, A. "Pemodelan General Regression Neural Network (GRNN) pada Data Return IndeksHargaSaham Euro 50, JurnalGaussian,vol4,no2.2015, pp.89-94,2015.

[3] Caraka,R.E.,Yasin, H, &Prahutama,A. "Pemodelan*General Regression Neural Network* (GRNN) dengan Peubah Input Data *Return* Untuk PeramalanIndeksHangseng, Seminar Nasional Ilmu Komputer Universitas Negeri Semarang. ISBN:978-602-71550-0-9, 2015, pp.283-288, 2014.

[4] D. E. Rumelhart, G. E. Hinton, & R. J. Williams, "Learning representation by back-propagating errors", *Nature*, vol. 323, 1991, pp. 533–536.

[5] D. Tomandl, and A. Schober, "A Modified General Regression Neural Network (MGRNN) with new, efficient training algorithms as a robust 'block box'-tool for data analysis", *Neural Networks*, vol. 14, no. 6, 2001,pp. 1023–1034.

[6] E. Parzen, "On Estimation of a Probability Density Function and Mode", *Annals of Mathematical Statistics*, vol. 33, 1962, pp. 1065–1076

[7] E.A. Nadaraya, "On estimating regression", Theory Probab. Appl.,vol. 10, 1964, pp. 186-190.

[8] G. S. Watson, "Smooth regression analysis", *Sankhya*, Series A, vol.26, 1964, pp. 359-372

[9] L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates ofmultivariate densities", *Technometrics*, vol. 19, pp. 135–144, 1977.

[10] Leung, M.T.,Chen,A.N., And H. Daouk. *Forecasting Exchanges rates using general regression neural networks, Computers and Operation Research Vol. 27, 2000, pp:1093-1110*

[11] Sprecht, D.F.*A Generalized Regression Neural Network*, IEEE Transactions on Neural Networks, vol.2,No.6,1991, pp.568-576

[12] T. Cacoullos, "Estimation of a Multivariable Density", *Annals of.Institute of Statistical Mathematics*, vol. 18, no. 2, 1996, pp. 179–189.

[13] Yasin,H.,Caraka,R.E., Tarno, and Hoyyi,A. Prediction Of Crude Oil Prices Using Support Vector Regression (SVR) With Grid Search–Cross Validation Algortyhm. Vol.12 No.4; August. Global Journal of Pure and Applied Mathematics. Print ISSN : 0973-1768 Online ISSN: 0973-9750,2016,pp. 3009–3020