PROCEEDING ISBN : 978-602-1037-00-3

S – 9

Forecasting Consumer Price Index of Education, Recreation, and Sport, using Feedforward Neural Network Model

Dhoriva Urwatul Wutsqa[1)], Rosita Kusumawati[2)], Retno Subekti[3)]
Department of Mathematics, Yogyakarta State University, Indonesia
dhoriva@yahoo.com[1)], rosita.kusumawati@gmail.com[2)], safina.rere@gmail.com[3)]

Abstract

The aim of this research is to forecast the consumer price index (CPI) of education, recreation, and sport in Indonesia using feedforward neural network (FFNN) model. We consider two FFNN models which are differed from the inputs. The inputs of the first model are generated by considering the inputs such as in a time series model, those are the lags of the CPI. Regarding that the pattern of the CPI data follow the segmented linear function, we generate the second model with the inputs such as in truncated polynomial spline regression model, by taking into account the location of the knots. The results demonstrate that the first model has better performance both in training and testing data. In addition, the first model is adequate model, means that the model delivers no autocorrelation error. Otherwise, the other model is not adequate model.

Keywords: feedforward neural network, CPI of education, recreation, and sport, truncated polynomial spline regression

Introduction

Consumer Price Index (CPI) is an index that calculates the average of pricechange of goods and services consumed by the citizens or families within a certain time. One type CPI in Indonesiais education, recreation and sport.The CPI is one of the economic indicators in Indonesia. Forecasting their values in the future is important, since it can be used by government as a basis to make decisions.

The CPI is a kind of time series data that has complex pattern. It tends to have linear trend in some period, but in some points it extremely increases or decreases.The typical linear model is not appropriate to this kind data. Wutsqa and Yudhistirangga (2013) demonstrate that truncated polynomialspline regression (TPSR) is an appropriate model of the CPI of education, recreation and sport in Yogyakarta. Sarle (1994) suggest an alternative approaches, that is neural network (NN). Feedforward Neural Network (FFNN) and Recurrent Neural Network (RNN) are examples of NN models. The main difference of those models is that the RNN has feedback connection in one or more layers, while the FFNN does not (Haykin, 1999). Wutsqa, Subekti, and Kusumawati (2014) combine the TPSR and RNN for modelling CPI of education, recreation and sport in Yogyakarta

In this paper, we apply FFNN for forecasting CPI of education, recreation and sport in Indonesia considering that FFNN has successfully provided good performance in forecasting many kind of data. Chen, Racine, and Swanson (2001) on inflation data in the United States,Suhartono, Subanar, and Rezeki (2005) on air passenger data. Wutsqa (2005) apply it on inflation data and Wutsqa and Abadi (2011) apply it on the number of Prambanan temple data.

The mathematical formula of the FFNN model is characterized from the inputs, activation function, and the number of hidden layer. In this research we examine two kinds of input design. Those are inputs generated from the histories data or usually

This paper has been presented at ***International Seminar on Innovation in Mathematics and Mathematics Education 1st ISIM-MED 2014*** *"Innovation and Technology for Mathematics and Mathematics Education"* ***Department of Mathematics Education,Yogyakarta State University***
Yogyakarta, November 26-30, 2014

called lags of variable and inputs generated such as in TPSR model. We prove that in this case FFNN with input histories data gives better performance.

Feedforward Neural Network Model

Feedforward neural network (FFNN) is a very popular model and widely used in various fields, especially in forecasting time series data. This model is commonly referred to as multilayer perceptron (MLP). The MLP model architecture consists of an input layer, one or more hidden layers, and output layer. In this model, the calculation of the response or output is performed by processing (propagating) inputfrom one layer forward to the next layer orderly. The complexity of the FFNN architecture depends on the number of hidden layers and number of neurons in each layer. The FFNN with one hidden layer is the most frequently usedmodel, because it has simplearchitecture, but it has been able to approach any continuous function at any degree of accuracy. This fact is supported by several theorems of Cybenko (1989), Funahashi (1989), and Hornik (1989).

The characteristic of neural network can be viewed from the activation function which is used to determine the output. The inputs of the FFNN model for time series data usually deal with thehistorical data. So, the FFNN modelwith a single hidden layer, a bipolar sigmoid activation function in the hidden layer and a linear function in the output layer can be written as

$$y = \sum_{h=1}^{q} v_h \frac{1-\exp\sum_{k=1}^{p} y_{t-k} w_{hk} + b_h}{1+\exp\sum_{k=1}^{p} y_{t-k} w_{hk} + b_h} + b_0 + \varepsilon \qquad (1)$$

where y is the output, y_{t-k} are independent variables (inputs), b_h are bias and w_{hk} are weights on hidden layer from input layer, while b_0 is bias and v_h are weights on output layer, k= 0, 1, 2, …, p, h = 1, 2, 3, …, q,and ε is the model error.

We can change the inputs structure as done by Wutsqa, Subekti, and Kusumawati (2014). The inputs structure is based on TPSR model developed by Wand(2000), The model is formulated by the following equation

$$y = \alpha_0 + \sum_{g=1}^{q} \alpha_g t^g + \sum_{j=1}^{m} \beta_j (t - K_j)_+^q + \varepsilon \qquad (2)$$

where α_0 is the polynomial intercept, α_k, g =1, 2,…, q are the polynomial parameters, β_j, j =1, 2,…, m are the truncated polynomial parameters, m is the number of knots, y is the response variable (output), and t and its polynomials are the predictorvariables (inputs). The knot points K_j, j=1, 2,…, $m$are defined as

$$(t - K_j)_+^q = \begin{cases} (t-K_j)^q; t \geq K_j \\ 0 \qquad\quad ; t < K_j. \end{cases} \qquad (3)$$

In general, the steps of the FFNN modeling consist of the inputs identification, the data split into two training and testing set, the determination of the number of neurons in the hidden layer, the determination of the optimal inputs, and the test of the model fit. The identification of inputs can be done such as in the time series analysis, i.e. by observing the significant autocorrelation on autocorrelation function (ACF) plot (Wei, 2006). The determination of the number of hidden neurons is performed by considering the smallest MSE and MAPE values both in the training and the testing

data. The next step is the input elimination to obtain of the optimal inputs by considering the least MSE and MAPE values. The test of the model fit deals with the evaluation of the random (white noise) properties of error by perceiving the residual ACF plot. The model is fit, if the autocorrelation in each lags isnot significant.

Empirical Result

This study uses CPI of education, recreation, and sportdata in Indonesia. They are the monthly indexof 118 period from January 2004 to October 2013. To model the CPI data, the first step is the identification of the model inputsby observing time series and ACF plots. Figure 1. presents the time series plot of CPI data and Figure 2. presents the ACF plot of CPI data.

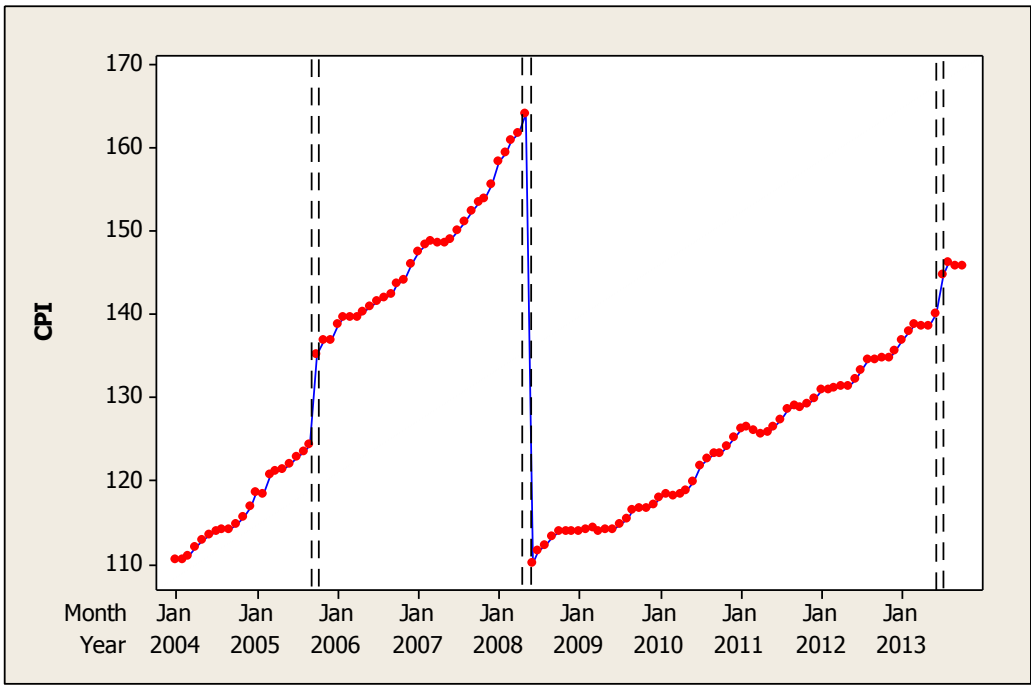

Figure 1.The time series and ACF plots of CPI data

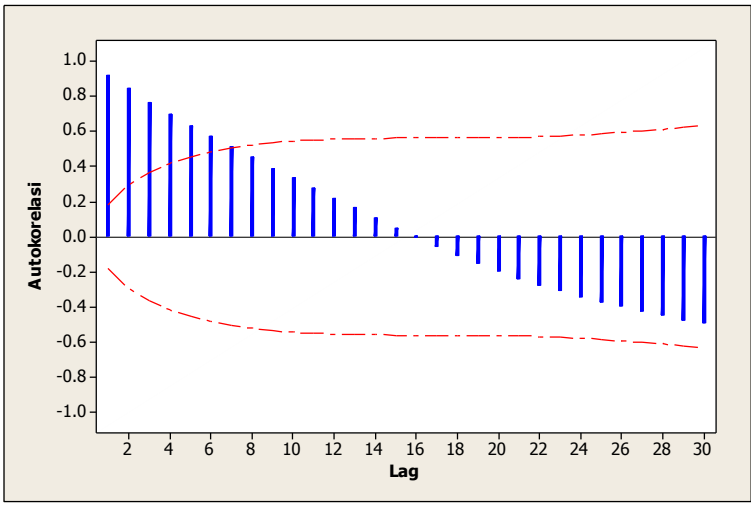

Figure 2.The time series andACFplots of CPI data

Here, we put two models which are differed from the inputs design. The inputs of the first FFNN modelareestablished based onthe ACFplotin Figure 2. The autocorrelationsare significant at lag 1, 2, 3, 4, 5, and6, so the inputs of FFNN are y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, and y_{t-1}.This model is denoted as the FFNN1 model. The inputs of second model FFNN2 are formed by regarding the points where the data change drastically. From Figure 1.we can see that the data increase dramatically in the periods 21, 22, 53, 54, 113, and 114. Those points are referred to as knot points in truncated polynomial spline regression (TPSR) term. The pattern of data between knot points tend to rise. Thus, we generate 7 inputs of FFNN2 based on model (2) and (3), those are t_0, t_1,t_2, t_3,t_4,t_5, and t_6,

Then the step is continued to cross validation process by splitting the data into training and testing set. The training data start from period 1 to period 84, and rest are the testing data. Learning is done to determine the number of hidden neurons by considering the least MSE and MAPE values of training and testing data. Those values are presented in Table 1.

Table 1. The MSE and MAPE values of FFNN learning

The number of Neuron	FFNN1Model				FFNN2Model			
	Training		Testing		Training		Testing	
	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
1	31.48	1.51	3.05	3.46	177.2	8.11	314.0	20.04
2	29.46	1.37	6.72	3.13	6.45	1.25	50.82	3.09
3	5.96	0.92	2.57	2.11	2.71	0.91	96.39	2.25
4	3.82	0.83	12.0	1.92	1.36	0.65	64.35	1.60
5	2.03	0.61	4.03	1.41	6.85	1.17	4.38	2.90
6	6.04	0.90	9.19	2.07	1.81	0.75	20.83	1.85
7	2.09	0.67	6.44	1.53	1.67	0.72	15.49	1.78
8	3.35	0.81	8.84	1.86	1.12	0.63	126.64	1.55

The results in Table1. shows that the models with high degree of accuracy are FFNN1 models with 5 neurons and FFNN2 with 7 neurons in the hidden layer.

The next step is doing learning again to determine the optimal inputs. The combinations of inputs having the least MSE and MAPE values lead to the optimal model.The result of this step are given in Table 2**.**

Tabel2. The MSE and MAPE values of FFNN learning

FFNN1 Model						Model FFNN2			
	Training		Testing		Input	Training		Testing	
Input (lag)	MSE	MAPE	MSE	MAPE		MSE	MAPE	MSE	MAPE
2, 3, 4, 5, 6	3.85	1.04	29.19	2.51	$t_1, t_2, t_3, t_4, t_5, t_6$	8.39	1.61	4620.0	3.98
1, 3, 4, 5, 6	3.97	0.80	8.62	1.83	$t_0, , t_2, t_3, t_4, t_5, t_6$	1.78	0.73	67.30	1.81
1, 2, 4, 5, 6	2.03	0.66	4.34	1.51	$t_0, t_1, , t_3, t_4, t_5, t_6$	1.73	0.69	87.74	1.71
1, 2, 3, 5, 6	21.06	1.61	46.45	3.69	$t_0, t_1, t_2, t_4, t_5, t_6$	3.55	0.89	37.66	2.21
1, 2, 3, 4, 6	16.78	1.41	4.09	3.24	$t_0, t_1, t_2, t_3, t_5, t_6$	3.12	0.89	41.56	2.21
1, 2, 3, 4, 5	3.02	0.83	23.70	1.91	$t_0, t_1, t_2, t_3, t_4, t_5$	1.55	0.66	90.10	1.63
1, 2, 3, 4,5 ,6	2.03	0.61	4.03	1.41	$t_0, t_1, t_2, t_3, t_4, t_5$	3.38	0.93	18.17	2.29
					$t_0, t_1, t_2, t_3, t_4, t_5, t_6$	1.67	0.72	15.49	1.78

Table 2. demonstrate that the best models of FFNN1 is model with inputs lag1,2,4,5, and 6, and with 5 hidden neurons, and best models of FFNN2 is model with inputs t_0, t_1, t_2, t_3,t_4,t_5, t_6, and with 7 hidden neurons.
Finally, the model fit is checked from residual ACF plot. Those of the models are delivered in Figure 3.

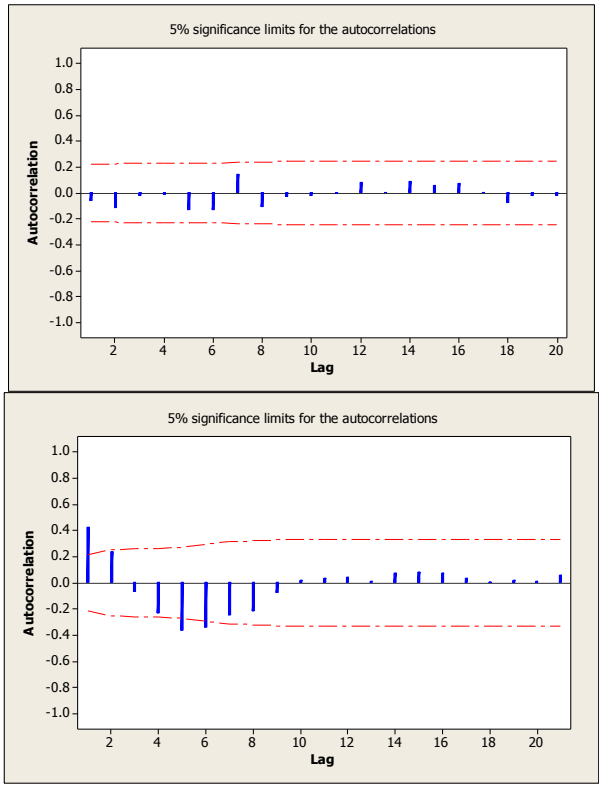

Figure3. The ACF plots of FFNN1and FFNN2residual

Figure 3. shows that only the FFNN1 model is appropriate, means that the model produce random error.The forecast of the FFNN1 model can be calculated using the formula

$$\hat{y} = \sum_{k=1}^{5} v_k \frac{1-\exp\left(-\left(y_{t-1}w_{1k(b)} + y_{t-2}w_{2k(b)} + y_{t-4}w_{4k(b)} + y_{t-5}w_{5k(b)} + y_{t-6}w_{6k(a)} + b_k\right)\right)}{1+\exp\left(-\left(y_{t-1}w_{1k(b)} + y_{t-2}w_{2k(b)} + y_{t-4}w_{4k(b)} + y_{t-5}w_{5k(b)} + y_{t-6}w_{6k(a)} + b_k\right)\right)} + b_0$$

where

$$\mathbf{b}=[b_h] = \begin{bmatrix} 6.3316 \\ 3.1862 \\ -0.1717 \\ -0.4356 \\ 1.8985 \end{bmatrix}, \mathbf{w}= [w_{hk}] =$$

$$\begin{bmatrix} -2.7237 & 1.7741 & 0.3835 & -3.4142 & 0.7912 \\ -1.0773 & -0.5011 & -0.7621 & 0.6211 & 0.0788 \\ -0.8117 & 0.2034 & 0.2875 & 0.6236 & 0.0292 \\ -1.4041 & 0.2366 & 0.3757 & -0.1741 & -0.0019 \\ 1.0282 & -0.3743 & -0.2558 & -0.3832 & 0.2607 \end{bmatrix}$$

$\mathbf{v}= [v_k] = [\ 5.1334 \quad -2.5262 \quad 0.2761 \quad -1.2021 \quad -0.0189]$, and b_0=–2.8708.

Furthermore, we also can explore the performance of the model from the plot of forecasts and actual data in training as well as in testing data.Figure 3. gives the plots for the training data and Figure 4. gives the plots for testing data.

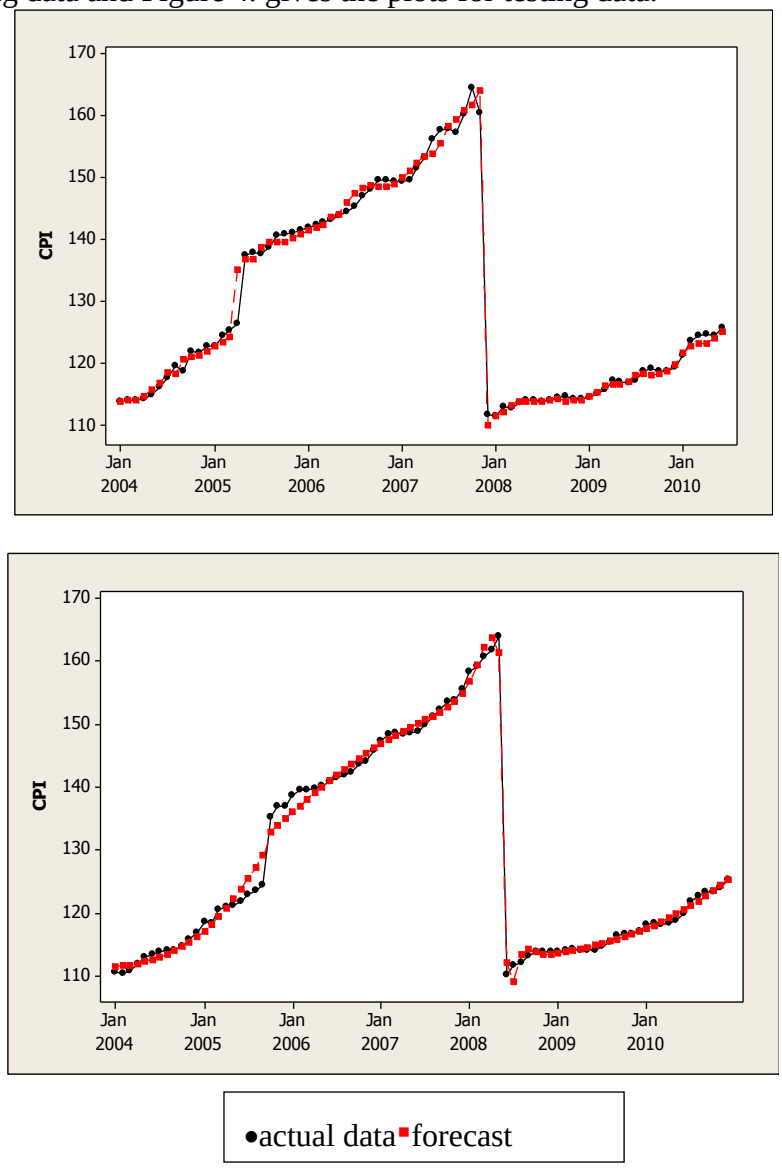

Figure 3.The time series plot of actual data and forecasts of FFNN1 and FFNN2 in training data.

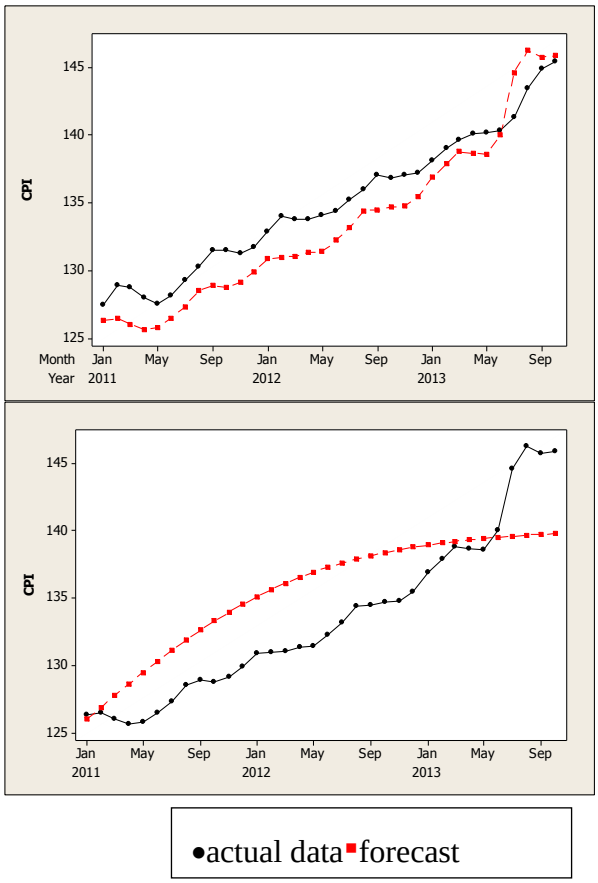

Figure 4 .The time series plot of actual data and Forecasts of FFNN1 and FFNN2 in testing data.

We can see that the plots of the forecastsand actual data in training data coincide. This fact demonstrate that all models yield accurate forecasts. Meanwhile, only the plot of FFNN1 show good performance in testing data. The plot of FFNN2hasexponential trend, which is very different from the pattern of the actual data. So, we recommend the FFNN1 to be the optimal model for CPI of education, recreation, and sport datain Indonesia.

Conclusion

The CPI of education, recreation, and sport in Indonesia has an upward trend and has many jump points. So, here we examines the modelswith two possible input design, those are inputs generated from the lags of variableand the inputs such as in the TPSR model.The result shows that the best model is FFNN models with inputsgenerated fromthe lags of variables. This model deliver high accuracy both in training and testing data.

Refference

Chen X., Racine J., and Swanson N. R. (2001). Semiparametric ARX Neural-Network Models with an Application to Forecasting Inflation. *IEEE Transaction on Neural Networks*, Vol. 12, No. 4, pp. 674-683.

Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems*, Vol. 2, pp. 304–314.

Funahashi, K. (1989). On the approximate realization of continuous mappings by *neural network*s. *Neural Networks*, 2, 183–192.

Hornik K. (1989). Multilayer Feedforward Networks are Universal Approximation. *Neural Networks*, 2, 359 – 366.

HaykinS. (1999). Neural Networks, A comprehensive foundation, Ontario:Pearson Education.

Suhartono, Subanar, Rejeki S. (2005). Feedforward Neural Networks Model for Forecasting Trend and Seasonal Time series. *IRCMSA Proceedings*. Sumatra Utara Indonesia.

Wei, W.W.S. (2006). *Time Series Analysis, Univariate and Multivariate Methods*, 2nd ed. New York: Pearson.

Wand M. P. (2000). "Comparison of regression spline smoothing procedures," Computational Statistics, 15, 443- 462.

Wutsqa D.U. (2005). Comparison between the Neural Network (NN) and ARIMA Models for Forecasting the inflation in Yogyakarta. *Proceeding ICAM*, ITB, Bandung.

Wutsqa D.U. and Abadi A. M. (2011). Modeling Islamic Lunar Calendar Effect in Tourism Data of Prambanan Templeby Using Neural Networks And Fuzzy Models. Presented in ICCT, Nidge, Turkey.

Wutsqa D.U. and Yudhistirangga, "Modeling consumer prices index data in Yogyakarta using truncated polynomial spline regression," Proceeding SEACMA, ITS, Surabaya, 2013.

Wutsqa D.U., Subekti R., and Kusumawati R. (2014). The Application of Elman Recurrent Neural Network Model for Forecasting Consumer Price Index of Education, Recreation and Sports in Yogyakarta, Proceeding 10th International Conference on Natural Computation (ICNC), Xiamen University, China, 2014, pp. 192-196