IJCAT - International Journal of Computing and Technology, Volume 06, Issue 06, June 2019
ISSN (Online) : 2348-6090
www.IJCAT.org
Impact Factor: 0.835

Pattern Prediction Comparison of Time Series Data Using Artificial Neural Network (ANN) – Multilayer Perceptron (MLP) and Support Vector Regression (SVR)

[1] Ika Oktavianti; [2] Ermatita; [3] Dian Palupi Rini

[1, 2, 3] Department of Computer Engineering
Faculty of Computer Science, Sriwijaya University, Indonesia

Abstract - Support Vector Regression (SVR) and Artificial Neural Network (ANN) - Multilayer Perceptron method has been proven to be able to solve time series forecasting problems to achieve generalization performance. Both methods are also able to handle small and large sample data. In this study, these two methods will be compared to find the most accurate. The data used is time series data in the form of secondary data about licensing, the season before delegation of authority. The data will be trained and tested using the Python programming language using the SVR method with several kernel functions and MLP methods. The composition of the use of training data and testing data is 70% and 30%. Then with these data, a Simple Linear Regression calculation will be conducted to test the extent to which the causal relationship between the number of permits issued and the number of available officers, so that the effectiveness of licensing service performance can be known. The results showed that with small sample data, the accuracy of the MLP method proved to be able to do better forecasting than the SVR method with several kernel functions with an accuracy level of MLP score -0.17 and the values of MSE, MAE and RMSE are 251.09, 11.45, and 15.84. Based upon SLR calculations, it is evident that there is a significant relationship between the variables x and y, namely the number of service personnel influencing the effectiveness of service performance and the number of permits that have been processed.

Keywords - Pattern prediction; time series data; multilayer perceptron; support vector regression

1. Introduction

Public services in Indonesia still receive special attention from the government. One example is health services, education services, and licensing services. The many problems in the field of licensing services encourage the government to carry out bureaucratic reform. One of them is to establish that institutions that carry out all licenses in an integrated through single submission online are One-Stop Integrated Services. However, it turned out that the establishment of licensing implementing agencies had not been able to overcome the speed of the licensing process. The number of permits issued is less than optimal because of the proportion of service personnel.

Based on these problems, this research wants to know the prediction pattern by comparing two methods, namely Support Vector Regression (SVR) and Multilayer Perceptron (MLP) and processed using the Python programming language. One of the main characteristics of SVR to achieve generalized performance is attempts to minimize generalized error bound and to avoid overfitting problems. The generalized error is the difference between empirical risk on training data and the risk of output hypothesis. In other words, it measures of how accurate the model to predict outcome values for previously data [1].

Recently, Artificial Neural Network (ANN) also have been used in time series forecasting. ANN can solve a problem such as classification or pattern recognition. The major advantage of ANN is their capability for nonlinear modeling and can find more information from larger data compared to statistical models. The ANN model does not require larger data samples to work properly. In addition, ANN is suitable for various types of datasets because the model is adapted to form based on attributes presented in the data [2].

Research is expected to produce conclusions, which method is most accurate in forecasting and whether there is a significant relationship between the number of service officers and the effectiveness of service performance,

where the indicator is the number of permits issued and the Community Satisfaction Index.

The remainder of this paper is organized as follows: Section 2 discusses the literature review of data mining, prediction, forecasting, time series, regression analysis, support vector regression, multilayer perceptron, calculation of performance criteria and related works. Section 3 describes the method of this research. Section 4 posits the prediction result using SVR and MLP. The conclusion summarized in Section 5.

2. Literature Review

2.1 Data Mining

Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from various large databases [3]. Various other names from Data Mining, such as knowledge extraction, data analysis/patterns, business intelligence, and others. Data mining techniques are used to examine large databases to find new and useful patterns.

Data mining is a series of processes to explore added value in the form of information that has not been known manually from a database. Information is obtained by extracting and recognizing patterns of data contained in the database [4]. Data mining techniques are considered appropriate to be used in this study considering the data process will use the help of machine learning methods.

The stages of data mining can be seen in Figure 1 below, which starts from data selection, data processed, then modified (transformation), data mining processes, and evaluation. After evaluation, the prediction patterns found will be new information/knowledge that can be used by researchers [5].

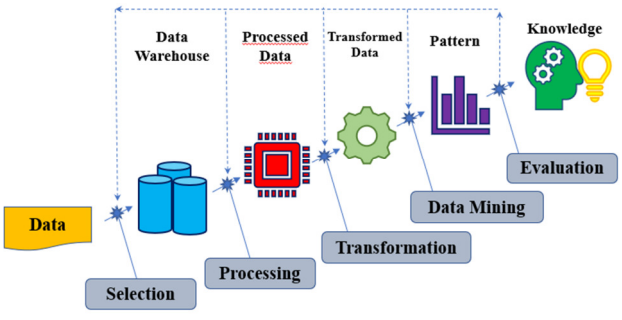

Figure 1. Flow Concept of Data Mining

Data mining can be processed using several methods including description, prediction, estimation, classification, clustering, and associations [6]. In this study the prediction process will be used to look for licensing prediction patterns using support vector regression and artificial neural network - multilayer perceptron. The process of data mining can be seen in Figure 2.

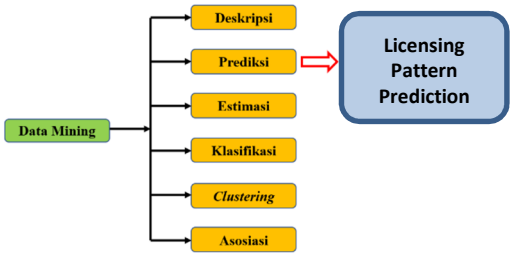

Figure 2. Data Mining Process

2.2 Prediction

Prediction is a systematic process of estimation of something that is most likely to occur in the future based on past data / information and current data so that errors in predictions can be minimized. Predictions do not have to provide definitive answers to events that will occur, but rather try to find answers as close as possible that will happen [7].

Predictions in data mining have similarities with the classification process. What distinguishes is data that is classified in the prediction process based on behavior or predicted value for the future. The process of estimating predictive values based on patterns that exist in a set of data. In this study, forecasting was determined based on the prediction pattern of the licenses produced.

2.3 Forecasting

Analysis of time series data is intended to make a forecast for the future. The time series data obtained which will then be analyzed is used as a reference in predicting what changes will occur in the next period that will occur based on patterns that occur in the present and past. Predicted changes will be expected to help related parties to determine a decision to face the possibility that will occur in the future.

Data forecasting for the future carried out follows systematic steps and follows the model that corresponds to the nature and pattern possessed by the original data. So that the forecasting results obtained from the model become appropriate and possible to do. Forecasting steps can be seen in Figure 3 below.

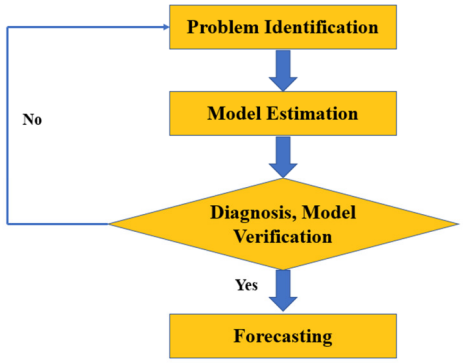

Figure 3. Forecasting Step

2.4 Time Series

Based upon time, data can be divided into three, namely cross section data, time series and pooled (panel). Time series data is data obtained from observing one object from several time periods. Time series form data can be recorded based on daily, weekly, monthly, annual, or other specified time periods in the same time period [8]. In time series data the value of observations of a time period is assumed to be influenced by the value of observations in the previous time period. Thus, time series data analysis makes it possible to forecast in the future.

Time series data can also be used for forecasting certain cases, where the results of these predictions have a large influence on the decisions or policies to be taken. In this case study, time series data from the licensing data sample is 7 (seven) years. Examples of time series data can be seen in Figure 4 below.

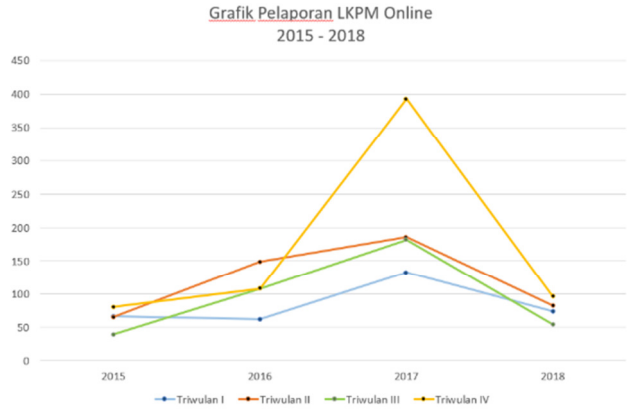

Figure 4. Example of Data Time Series

The example above is an analysis using time series data that has a trend pattern, meaning that the data plot spreads a constant up and down pattern. Even so, to get an accurate prediction pattern a large database is needed. The more data available, then the prediction pattern produced will also be more accurate. For this reason, in a study the availability of baseline data greatly influences the accuracy of the results of the research.

2.5 Regression Analysis

Regression analysis is used in almost all areas of life, both in agriculture, economics, and information technology. The usefulness of regression analysis is to find out the key variables that have an influence on a model, estimation or forecasting. The stages in regression analysis are problem formulation, selection of relevant potential variables, data collection, modeling, appropriate methodology, validation, and application of selected models to solve problems.

One type of regression analysis that is often used is Simple Regression Analysis. Simple Regression Analysis is an approach method for modeling relationships between one dependent variable and one independent variable. In the regression model, the independent variable explains the dependent variable. In a simple regression analysis, the relationship between variables is linear, where changes to the variable X will be followed by changes to the variable Y permanently. While the nonlinear relationship changes in variable X is not followed by changes in variable Y.

Simple Regression Analysis is based on functional or casual relationships between one independent variable with one dependent variable. The general simple linear regression equation is:

$$y = a + b\,x \quad (1)$$

Where:
y = Subject in the predicted dependent variable.
a = y value if $x = 0$ (constant value).
b = Regression coefficient, which shows the number of increases or decrease in the dependent variable based on the variable independent.
x = Subject on an independent variable that has a certain value.

The regression equation that has been found can be used to make predictions (forecasting) how individuals in the dependent variable will occur if the individual in the independent variable is set.

With this equation, a regression line can be made. Examples of regression lines can be seen in Figure 5 below.

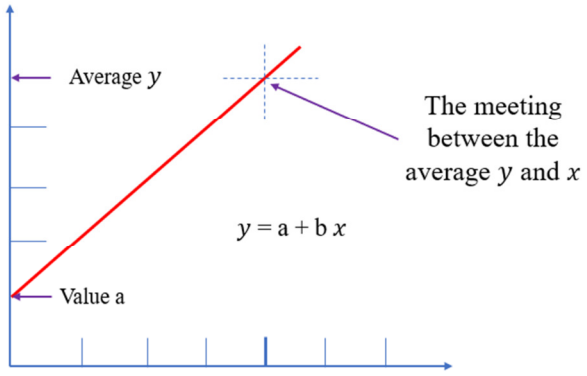

Figure 5. Regression Sample from Equation

2.6 Support Vector Regression

Support Vector Machine (SVM) is a good technique for forecasting or prediction, both for classification and regression cases [9]. SVM is a one of machine learning system using a high dimensional feature space [10]. SVM implementation in a regression case is called Support Vector Regression (SVR) and is a method that can produce maximum performance because it can overcome overfitting problems. Overfitting can occur if the data used is different so that it can reduce accuracy and the results of the research are not as expected. SVR has been proposed by Vapnik, Alex Smola and Steven Golowich in 1997 [11]. The basic concept of SVM method is a linear classifier and then developed so that can be used on non-linear problems using the kernel trick concept in high dimensional feature space.

According to Scholkopt and Smola (2002), SVR aims to find a function f (x) as a hyperplane from a regression function which corresponds to all input data in order to produce the smallest error (Ɛ). According to Santosa (2007), suppose that it has 1 training data set:

(x_i, y_i), I = 1, 2, ... where x_i is an input vector:

$$x = \{ x_1, x_2, \ldots, x_n \} \ \Re^n \quad (2)$$

and scalar output:

$$y = \{ y_i, \ldots, y_i \} \Re \quad (3)$$

2.7 Multilayer Perceptron (MLP)

ANN are computational or mathematical models that are inspired by the process of human's brain [12]. The system designed to solve more complex problems. Feed forward neural networks are the well-known in the various of applications among in the different types of connections for artificial neurons. It is also known as Multilayer Perceptron (MLP).

Multi-Layer Perceptron (MLP) is a feed-forward neural network consisting of several neurons that are connected by connecting weights. Each weight is connected to each weight in subsequent layers. Each MLP consists of a minimum of three layers of input layers. These neurons are arranged in layers consist of one input layer, one or more hidden layers, and one output layer. The input layer receives a signal from the outside, then passes it to the first hidden layer, which will continue until reaches the output layer. An architecture of MLP is shown in Figure 7.

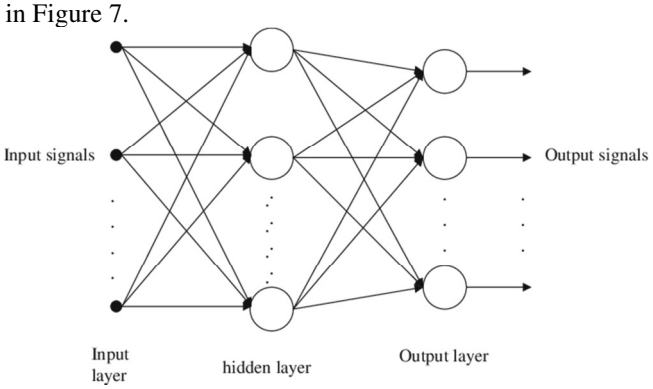

Figure 7. Multilayer Perceptron Architecture

Neuron in the input layer as shown in figure 2 only distributing the input signal x_i to neurons in the hidden layer. Each neuron$_j$ in the hidden layer has the original input multiplied by a weight w_{ji} with threshold added and calculate using activation function and computes its output y_j as a function f [13].

$$Y_j = f(\sum w_{ji} x_i) \quad (4)$$

There is transfer function f such as linear, log-sigmoid and tan-sigmoid. The activation function must be differentiable.

$$Linear: \quad f(x) = x \quad (5)$$

$$Log - Sigmoid: \quad f(x) = \frac{1}{1+e^{-x}} \quad (6)$$

$$Tan - Sigmoid: \quad f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (7)$$

2.8 Calculation of Performance Criteria

Based on the time series data used in this study and using the Python application program, calculation of the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of each dataset and using the SVR method

with several kernel functions and the MLP-ANN method. Mean Absolute Error (MAE) and Mean Squared Error (MSE) are calculated using the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (8)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|^2 \quad (9)$$

Where n is the amount of data, f_i is the value of the forecasting result and y_i is the actual value. MAE and MSE intuitively calculate error averages by giving equal weight to all data (i = 1....n). While RMSE uses the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \quad (10)$$

The smaller the value of RMSE, MAE and MSE, the closer the time series data is to the actual value.

2.9 Related Works

The application of the Support Vector Machine (SVR) method for forecasting has the advantage that SVR can make conclusions in general that are quite good. But in the case of financial time series forecasting, there is a lot of data noise and time series that often change. To overcome this problem SVR is combined with independent component analysis (ICA) because ICA can reduce data noise by producing independent components. Noisy Independent Component (IC) can be discarded and the remaining IC is used as input in the SVR for forecasting. The results of this study indicate that by combining SVR and ICA, forecasting can be done better [14].

The application of the Support Vector Regression (SVR) method was also conducted in research to predict the number of hotel guests in Demak Regency. SVR using Ɛ-insensitive and quadratic loss functions with linear kernel functions and polynomials are considered to produce good accuracy in forecasting training data and testing data. In this study, forecasting of training data produced MAPE values of 10.2806% and forecasting of data testing resulted in MAPE values of 11.622%. The research results obtained a forecasting model of the number of hotel guests in Demak Regency and obtained forecasting results for the next 12 months (September 2014 - August 2015) for data on the number of hotel guests in Demak Regency [15].

Application of Support Vector Machine (SVM) to predict energy consumption in the tropics. The results of this study are the salient features of SVM implemented to minimize the upper limit of generalization errors rather than training errors as applied in the Neural Network. Therefore, SVM is feasible to be used to predict monthly bills of energy users in the tropics [16].

Other research is about the prognosis of damage to rolling bearings using the Support Vector Regression (SVR) method. In this study, the SVR method will be used to carry out the prognosis process, namely the ability to assess the good and badness of a machine part and to predict the damage time of the engine part. The input data comes from the extraction of several statistical features from the trend data of the vibration bearing rolling signal. The resulting feature data is used in the learning process and testing process with the SVR method, after that it will produce a prognosis of rolling bearing damage that approaches the ideal value of the Root Mean Square Error (RMSE) and Coefficient of Correlation (R).

The results of this study are that the RMS feature is a good feature used to carry out prognosis. The RMSE and R values for both features approach the ideal value, namely the value of RMSE for the RMS feature is 0.0129 while the R value for the RMS feature is 0.9709.

3. Methodology

3.1. Data Collection

The method of data collection is a way to obtain information needed to examine a problem, where the data can be in the form of statements, circumstances, certain activities and the like. Some methods of data collection include interviews, observations, questionnaires, or document studies.

The data source used in this study is secondary data. Where secondary data comes from the dataset of permits and non-permits that have been issued on the Investment Services and One-Stop Integrated Services of South Sumatra Provincial Government for ± 7 (seven) years. It contains 78 data and 3 attributes (month, number of licenses and number of employees). Then the collection of data and information is also based on literature studies of reference books, journals, and articles that are in accordance with the object of research.

3.2. Model Framework

Various opinions are used as the basis for consideration in determining the method of approach used by the author. In this case the author uses the Support Vector Regression (SVR) and Artificial Neural Network

(ANN) - Multilayer Perceptron methods. Following the research conceptual framework for the use of methods in the search for licensing prediction patterns is described in Figure 8.

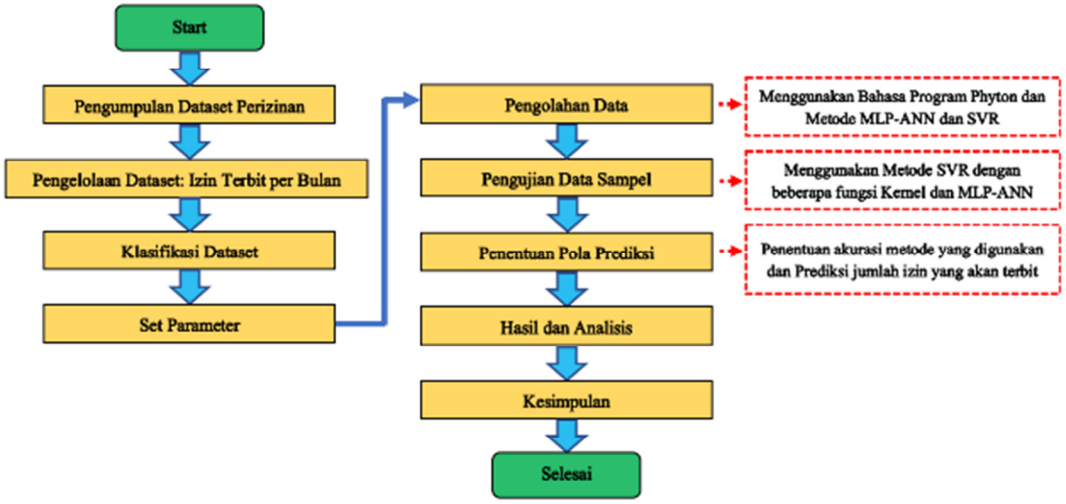

Figure 8. Model Framework

Proposed model consists of comparison analysis of ANN and SVR for pattern prediction of licenses. For ANN, we use Multilayer Perceptron (MLP) model and in case of SVR, were used different kernels (linear, RBF, Linear Model-Lasso), Linear Model-Elastic Net and Linear Model-Ridge) and compare their performance and accuracy.

This research used a licensing dataset that has been published every month from August 2009 to January 2016. Separating data into training and testing sets is by default. 70% of 78 data are used as the training set and 30% of 78 data are used as the test set. The training data will be used as input for Multilayer Perceptron (MLP) and Support Vector Regression (SVR), while the testing data will be used for prediction. And Performance of ANN MLP and SVR measure with Mean Square Error (MSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Then the accuracy of ANN and SVR will be compared to see models that are more suitable for time series data of licensing dataset.

4. Experiment and Result

4.1 Time Series Data Plot

An observed time series data plot can be shown in figure 4. In this study, data plots were used to determine licensing patterns every month for 7 (seven) years. The number of licenses every month is expected to be able to describe and produce a model. The time series data plot is also used to determine the right technique for forecasting.

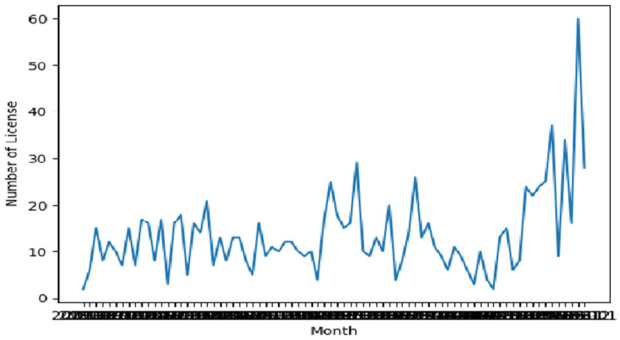

Figure 9. Time Series Data Plot

4.2 Forecasting Using MLP-ANN and SVR

Pattern Predictions of licensing process was using python programming language. Prediction time is for one month ahead. In order to make a comparison between the accuracy of MLP-ANN and SVR methods, MSE, MAE and RMSE are used. An MLP-ANN algorithm which includes 10 inputs, 5 neurons in hidden layer and 1 output layer was developed. For training, limit epoch is assigned at 200 and learning rate is set at 0.001 while momentum is set at 0.9. Sigmoid activation function was used in this study. The parameters are used because this expected gives the better result. Figure 9 showed prediction performance from MLP-ANN model.

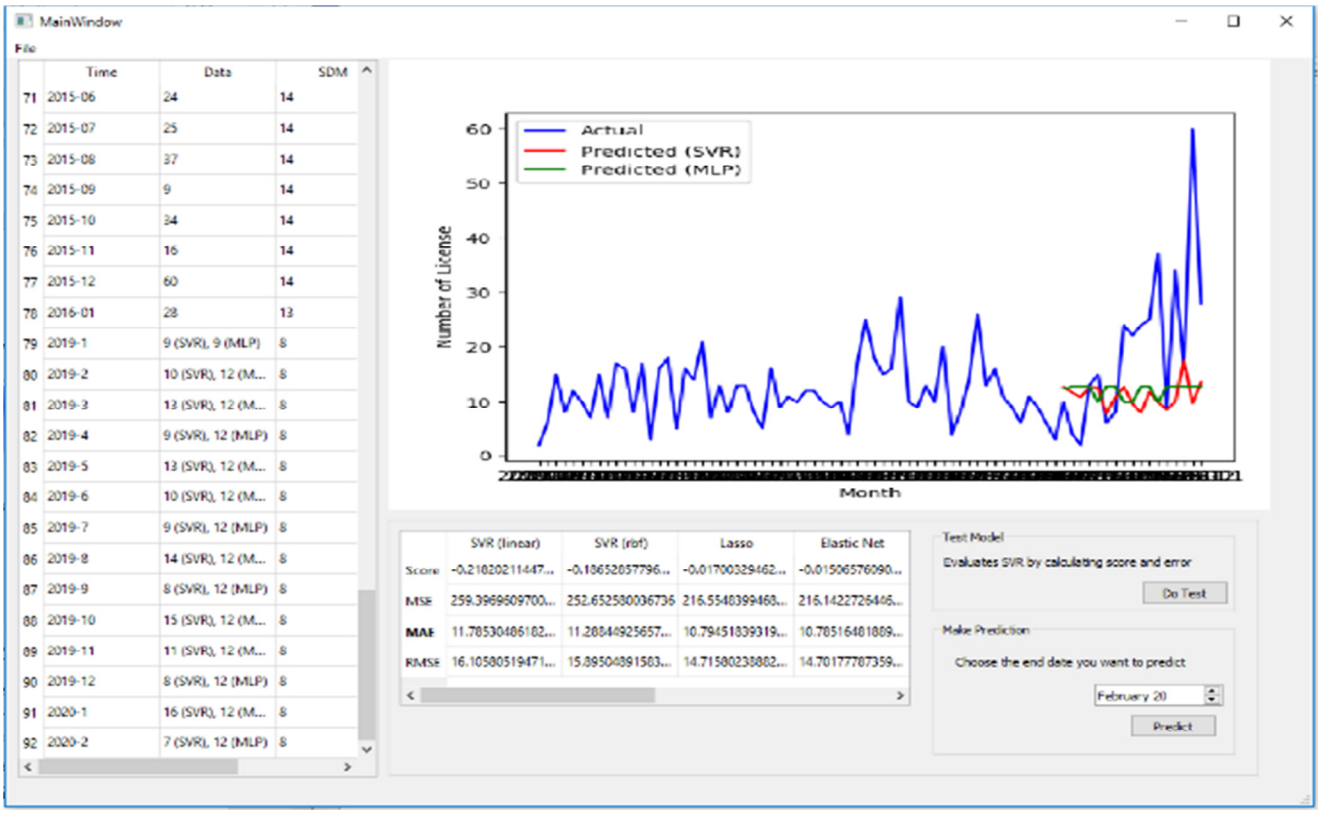

Figure 9. Output from SVR and MLP Model

Of the 78 actual types of permit data available, the data trained was 70%, or as 55 data and the data tested to calculate MAE, MSE, and RMSE were only 30%, which were 23 data. By using equations (8), (9), and (10), the calculation results for MAE, MSE, and RMSE are processed through Python programming language. The calculation results using SVR with several kernel and MLP functions, can be seen in table 1 below.

Table 1. Score, MSE, MAE, dan RMSE – *Dataset* 1

Metode	SVR (Linear)	SVR (RBF)	SVR (Lasso)	SVR (Elastic Net)	SVR (Ridge)	MLP
Score	-0.21	-0.18	-0.017	-0.015	-0.011	-0.17
MSE	259.39	252.65	216.55	216.14	215.33	251.09
MAE	11.78	11.28	10.79	10.78	10.76	11.45
RMSE	16.10	15.89	14.71	14.70	14.67	15.84

Based upon table 1, it can be seen that the accuracy of MLP is better than SVR-Linear which is a score of -0.17 and the values of MSE, MAE and RMSE are 251.09, 11.45, and 15.84. It was proven that MLP error value was lower than SVR-Linear which was 15.84%. But when compared with other kernel function SVRs, SVR-Ridge is more accurate than MLP with a score of -0.011 and the values of MSE, MAE and RMSE are 215.33, 10.76, 14.67. The SVR-Ridge error value is smaller than MLP which is 14.67%.

4.3 Simple Linear Regression Calculations

In the SLR statistical method, the causal factor is expressed as x or called the Predictor and the resulting variable is expressed as y or called Response. By using this method, it is expected to obtain a forecasting model that can predict relationships between variables.

The model of Simple Linear Regression equation has been explained in equation (1). To get the regression equation, the first thing to do is to determine the sample dataset to be used and the researcher will use the data on the number of permits in December every year from 2009 - 2018, which is 10 data. The sample dataset used to determine the regression equation can be seen in table 2 below.

Table 2. Sample Dataset Regression Equation

No.	Xi	Yi	X²i	Y²i	XiYi
1	12	10	144	100	120
2	5	10	25	100	50
3	9	12	81	144	108
4	15	11	225	121	165
5	13	13	169	169	169
6	13	14	169	196	182
7	60	14	3600	196	840
8	88	13	7744	169	1144
9	99	16	9801	256	1584
10	68	15	4624	225	1020
Total	382	128	26582	1676	48896
Rataan	38.2	12.8	2658.2	167.6	488.96

After calculated using data and equations, the values of a= -127, 409 and the value of b= 3.67, so the results of the equation become:

$$y = -127, 409 + 3.67x \quad (11)$$

After the data is entered into the above equation by using random data from all testing data as many as 20 data, the forecasting results obtained that based on the number of licenses issued as many as 76 permits required 152 people to service (HR) or 2 times the number of licenses issued. This is needed so that the licensing service process performance can run optimally. The results of calculations by means of simple linear regression prove that the licensing service officer who is currently serving in DPMPTSP South Sumatra Province is still lacking. Based on this, the government can take the best policy or decision in order to improve the quality of licensing services.

5. Conclusion

The main purpose of this study was to compare the performances of the Perceptron-Multilayer Neural Network (MLP-ANN) and Support Vector Regression (SVR) to predict licensing patterns and forecast the number of licenses to be issued in the following month and year. This study used data series based on monthly data from August 2009 to January 2016. In addition, we used various kind of kernels to improve performance of SVR model. That is Linear, Linear Model Lasso, Elastic, Ridge and RBF kernel. The results showed the ability of using time series data which proves that the MLP-ANN method is more accurate than the SVR method, where the accuracy value is -0.17 and the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values are 11.45 and 15, 84. The major difficulty of the work was to achieve minimize generalized error bound so that the difference of actual value and the predicted value is close to zero. Based upon SLR calculations, it is evident that there is a significant relationship between the variables x and y, namely the number of service personnel influencing the effectiveness of service performance and the number of permits that have been processed.

As a future work, we recommend using hybrid approaches, which are the combination of ANN and SVR. Otherwise it is expected to select the most appropriate parameters of the kernel function and ANN model. Another future improvement can be proposed such as using metaheuristic algorithm, deep learning and genetic algorithm techniques in the forecasting accuracy.

Acknowledgement

The authors would like to acknowledge Verlly and Azhary as a team for supporting this research and South Sumatra Investment and Licensing Department for providing licensing data.

References

[1] A. Xu and M. Raginsky, "Information-theoretic analysis of generalization capability of learning algorithm," *31st Conference on Neural Information Processing System (NIPS) CA, USA,* 2017.

[2] M. Khashei, S. R. Hejazi and M. Bijari, "A new hybrid artificial neural network and fuzzy regression model for time series forecasting," *Elsevier Fuzzy Set and System,* vol. 159, pp. 769-786, 2008.

[3] Turban, E. 2005. *Decision Support Systems and Intelligent Systems.* Edisi Bahasa Indonesia Jilid 1. Andi: Yogyakarta.

[4] **Tri Wulandari, Retno. 2017.** Data Mining Theory and Rapidminer Application. Gava Media: Yogyakarta.

[5] Fayyad, Usama. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press.

[6] Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.

[7] Kasmir, Jakfar. 2003. Business Feasibility Study. Jakarta: Prenada Media.

[8] Cryer, J.D., dan Chan, KS. 2008. *Time Series Analysis: With Application in R: Second Edition.* USA: *Spinger Science dan Business Media, LLC.*

[9] Santosa, Budi. 2007. Data Mining Techniques for Using Data for Business Purposes. Graha Ilmu: Yogyakarta.

[10] D. Basak, S. Pal and D. C. Patranabis, "Support vector regression," *Neural Informating Processing*. 11, no. 10, 2007.

[11] V. Vapnik, S. Golowich and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Neural Information Processing Systems*, MIT Press Cambridge, vol. 9, 1997.

[12] P. Kumar and P. Sharma, "Artificial neural network-a study,"*International Journal of Emerging Engineering Research and Technology,* vol. 2, issue 2, pp. 143-148, 2014.
[13] J. M. Nazzal, I. M. El-Emary and S. A. Najim, "Multilayer perceptron neural network (MLPs) for analyzing the properties of Jordan oil shale," *Wolrd Applied Sciences Journal,* vol. 5, (5) pp. 546-552, 2008.
[14] C.J. Lu., T.S. Lee., C.C. 2009. Chiu. *Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression.* Taiwan.
[15] Adiningtyas, Desy Trishardiyanti. 2015. Forecasting the Number of Hotel Guests in Demak Regency Using the Support Vector Regression Method. Diponegoro University, Semarang.
[16] Dong, B., Cao, C., Lee, S.E., 2005. *Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region*. *Energy and Buildings* 37, 545-553.