SVM Approach for Forecasting International Tourism Arrival In East Java

Evita Purnaningrum$^1$, Muhammad Athoillah$^2$

$^1$ Management Department, Universitas PGRI Adi Buana, Surabaya, City, 60234, Indonesia
$^2$ Statistics Department, Universitas PGRI Adi Buana, Surabaya, City, 60234, Indonesia

*E-mail: purnaningrum@unipasby.ac.id

Abstract. The maneuver of international tourists visiting Indonesia, especially the East Java area, in positive correlation with time. In accordance with the government’s work program to increase GDP from the tourism sector. However, the condition necessary to be supported by the existence of feasibility and adequate tourism services. Facility improvements certainly require a projection of international tourist arrivals visiting East Java. SVM is a machine learning method that has shown excellent performance in predicting time series. In addition, analysis results have shown that SVM had able to predict foreign tourist visits with a high degree of accuracy. Therefore, this research can be used as a support for tourism facility procurement policies, especially in East Java.

1. Introduction

East Java is a province with the largest area on Java Island (Indonesia.go.id). As the largest province, East Java has potentials that can be utilized by the local governments to increase the Gross Regional Domestic Product (GRDP), respectively. Tourism is one sector that can be relied on to increase regional GRDP. In addition, globalization has made national borders obscured so that foreigners can easily travel abroad. In other words, this has led to world-famous East Java tourism and has increased the attraction of foreign tourists to visit East Java. Besides, tourism a sector to be reckoned with in increasing GRDP in East Java. Therefore, the Government as a manager should pay attention to the tourism sector, especially facilities for tourists.

However, the government has the authority and cannot carelessly in making policy decisions on the procurement of facilities. Decision making should be in accordance with the needs of the tourist area. Therefore, several factors should be considered before the procurement of tourism facilities, one of which is the arrival of foreign tourists. The number of foreign tourists who come can affect the procurement of existing facilities so that careful calculations are needed. Seeing these problems, This paper studied to predict the arrival of foreign tourists to East Java using the SVM (Support Vector Machine) method that basically applied for classifying. However, this paper improved the method in accordance with the prediction of tourists’ arrival. In addition, the advantage can classify linear and non-linear. On another hand, the data generated from this method can be an aid to support the policy for the procurement of tourism facilities in East Java.

In Indonesia, research on the predictions of foreign tourists ordinarily developed the conventional model to the hybrid method. For Example, [1] has done forecasts using the
ARIMA hybrid model to predict tourists visiting Surakarta. Besides, [2] has predicted the number of foreign tourists in the halal category using the Fourier series model. [3] have applied Generalized Regression Neural Networks to produce the best performance, [4] have modified Arima in predicting numbers. Tourist visits in Bali in 2019 in tourist visits are being far below the target. Unlike [5] which uses the fuzzy method and gets the best average error forecasting error (AFER) value. However, the use of machine learning to predict the number of tourists still needs to be studied further, nevertheless. On the other hand, [6] has begun to examine the use of big data in tourism arrivals. In addition, forecasting in the economic sector has been developed and carried out by researchers, [7] have predicted the stock index affected by the pandemic with the Kalman filter.

This paper applied a machine learning method, namely SVM (Support Vector Machine) to predict foreign tourists visiting East Java. The performance of SVM in performing classification is not in doubt. This method has been used in various countries to predict future values. [8] has used the SVM model to predict investor sentiment based on news and consequently, indicates that the SVM model has high accuracy which is useful for investors for making investment decisions. [9] has developed an SVM model that has the highest accuracy and prediction of 97.39% to predict Australian credit scoring. Meanwhile, there are [10] has applied the SVM method but collaborate using the fuzzy method, showing that the use of this method can predict with better accuracy compared to other forecasting techniques. Therefore, this paper applies the SVM Model because it produces a high level of prediction accuracy.

2. Foreign Tourist Forecasting
2.1. Tourism Forecasting
Research on tourist forecasting has been developed for hundreds of years ago. Various methods have been generated and modified in order to achieve an accurate predictive value. These methods are improved and updated every time both conventional and mixed methods (hybrid model). From the time-series side, it has undergone rapid changes, although until now some researchers still applied basic methods for prediction problems. Moreover, in Indonesia, scientific development regarding predictions, especially tourist predictions, is still lacking and needs to be further developed. Several tourist forecasting studies whose research object is Indonesia ([11, 12, 13, 14, 1, 15]). Based on these references, most of them develop the time series model, which is less able to handle data instability such as data on the tourist arrival to East Java.

The problem came up when the dataset was the inconsistency and not stable in order that it cannot be approached by conventional methods. This could be solved using machine learning methods. In general, the machine learning method is to train data continuously so that it is able to produce real data models. The more complex training data used, will result in predictions with a high level of accuracy. Several studies in other countries have applied various types of machine learning methods to predict tourist visits. [16] has combined the Kernel Extreme Learning Machine (KELM) method with internet search index data on forecasting tourist arrivals in China. [17] has tested predictions of future tourist visits in Andorra, Europe by comparing machine and deep learning methods. [18] simulates a comparison of the ARIMA method with machine learning, namely the Long Short Term Memory Neural Network (LSTM) to predict tourism flow. [19] combines the NN and SVR methods to forecast tourism demand in Spain.

2.2. Support Vector Machine for Prediction
Support Vector Machine (shown Fig.1) is a good algorithm to solve problem of classification and prediction [20, 21]. Hence, this algorithm is still widely used recently. Traditionally, SVM separated classes by finding the best hyperplane in the input space, this hyperplane was obtained by measuring margins of the two support lines between the two classes, Therefore, SVM was shown as ensure robust theoretical in terms of generalization error [22].
Figure 1. SVM Process of Separate Class

Suppose, \( \{x_1, x_2, ..., x_n\} \) was training dataset, and \( y_i \in \{-1, 1\} \) was labelled as a target of \( x \) data. Then the class of tested \( x \) data was determined based on the value of the following decision function as follow 1:

\[
f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i x_i x_d + b
\]

Which, \( x_i \) was support vector, \( ns \) was total of support vector and was the classified data. However, SVM originally is a linear classifier, yet SVM can also use for non-linear problems using two stages. First, the data projected into a new high-dimensional vector space so that the data can be separated linearly. Second, while in this new dimension, SVM looks for the optimal hyperplane in the same way as previously mentioned [23]. Overall SVM procedure can be written as follows:

**Algorithm 1:** SVM Algorithm For Training Data

1. Determined input X and target Y ; // Prediction for Training Data
2. Determined function of Kernel
3. Calculate the Kernel K matrix, which is the dot product multiplication of the input vector according to the predetermined Kernel function
4. Find the optimal solution for the \( \alpha \)
5. Find support vector
6. Get the bias obtained from \( b \)

**Algorithm 2:** SVM Algorithm For Testing Data

1. Determined input X and target Y ; // Prediction for Testing Data
2. Get the result by using the SVM Model that previously obtained in training stage

The following flowchart is presented to simply understanding of the method Fig.2. The accuracy of the model was calculated using RMSE 2 and MAPE 3. Based on theory [24, 25], MAPE Value is below 10%, it indicates Highly accurate forecasting.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y} - y)^2}
\]
3. Data and Empirical Results

3.1. Data and Case Information
This research used data of tourist arrivals at all entrances to East Java. It was obtained from Indonesia Central Statistics Agency website during the period of January 2009 to March 2019, monthly. The Dataset was divided into two group that is Training and Testing Data, 90% of dataset was used for training process means there was 110 months data from January 2009 till January 2018 for training process and the rest data from February 2018 till March 2019 used for testing process. The dataset showed that number of tourists entering East Java has increase significantly month by month even though there were slightly drop in several periods (shown in Fig.3). Furthermore, this data fluctuation indicated the importance of good algorithm (SVM) to accurately predict the next period of data to support tourism planning and development in East Java.

3.2. Result of SVM Model
The research method was divided into two stages namely training and testing as mentioned previously. The training process started with transform data into time series dataset, then train the transformed data using fine-tuned modified SVM algorithm. The SVM model was modified first by adding radial basis kernel function while fine-tuned SVM was obtained by using Grid Search Optimization (GSO) method. The GSO allowed system to train and select the best of multiple models for each different pair of 59 values of epsilon $\varepsilon$, cost $c$ and gamma $\gamma$. To validate the model, this research used k-fold cross validation by divided the dataset into 10 batch ($k=10$) and then repeated the process 10 times with different combination of data and parameter in every training and testing process. The testing process was started by put on the testing data to the SVM model that had been develop in the earlier step then calculate the result using RMSE and MAPE.

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

(3)
Figure 3. Monthly Foreign Tourist Arrivals at East Java (January 2009 - March 2019)

The result show that SVM Model has good performance to predict the number of tourist arrivals at East Java, proved by average of MAPE result was 0.191. Meanwhile the best result of this prediction was obtained when using parameter that was 0.3 of $\varepsilon$, 8192 of $c$ and 0.25 of $\gamma$ proved by the MAPE result of the SVM prediction was 0.158 or 15.8%. The SVM prediction was in a certain range between 23239 and the highest was 29846. Conversely dataset has the lowest of 14200 to the highest was around 34100. The following Table 1 show the detail of the result of all experiment.

| No | $\varepsilon$ | $c$  | $\gamma$ | RMSE  | MAPE |
|----|---------------|------|----------|-------|------|
| 1  | 0.6           | 8192 | 0.25     | 5423.141 | 0.187 |
| 2  | 0.5           | 4096 | 0.25     | 5590.594 | 0.194 |
| 3  | 0.6           | 32   | 0.2      | 6160.393 | 0.213 |
| 4  | 0.7           | 32   | 0.5      | 6266.872 | 0.218 |
| 5  | 0.6           | 512  | 0.5      | 5748.453 | 0.199 |
| 6  | 0.6           | 32768 | 0.25  | 5244.328 | 0.179 |
| 7  | 0.3           | 8192 | 0.25     | 4661.676 | 0.158 |
| 8  | 0.6           | 16384 | 0.5   | 5248.595 | 0.179 |
| 9  | 0.6           | 256  | 0.25     | 5727.708 | 0.198 |
| 10 | 0.6           | 32   | 1        | 5415.277 | 0.186 |

4. Conclusion
This research presented a prediction of tourist arrivals at East Java using Support Vector Machine. Using Indonesia Central Statistics Agency Data of tourist arrivals from January 2009 to March 2019, this research proved that SVM model with RBF Kernel could provide a good prediction with average of MAPE’s score 0.191. Meanwhile the best result was 0.158 MAPE’s Score when the parameter of $\varepsilon$ was 0.3, parameter of $c$ was 8192 and parameter of gamma was 0.25. In addition, based on the experience, unlike the data for foreign tourists from all over
Indonesia, which has always experienced a steady increase, the number of arrival tourist in East Java was fluctuate even has decreased sharply oftentimes. Thus, the good forecasting could help government to prepare the best facilities and policy to give best experience and satisfaction for all the tourist specially in East Java. Furthermore, to produce better result of forecasting, it is suggested to use additional data from the internet such as google trend data or people sentiment on twitter in the next research.

References

[1] Purwanto S, Julitta F T and Paramananda A 2019 AIPC 2114 060010
[2] Mardianto M F F, Cahyono E F, Syarifah L and Andriani P 2019 KnE Social Sciences 1093–1104
[3] Herawati S 2016 Jurnal Infotel 8 35–39
[4] Rukini R, Arini P S, Nawangsih E et al. 2015 Jurnal Ekonomi Kuantitatif Terapan
[5] Elfajar A B, Setiawan B D and Dewi C 2017 Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN 2548 964X
[6] Purnaningrum E and Ariqoh I 2019 Majalah Ekonomi 24 232–243
[7] Purnaningrum E 2020 Majalah Ekonomi 25 103–109
[8] Wang D and Zhao Y 2020 Procedia Computer Science 174 191–199
[9] Plawiai P, Abdar M and Acharya U R 2019 Applied Soft Computing 84 105740
[10] Xu X, Law R, Chen W and Tang L 2016 CAAI Transactions on Intelligence Technology 1 30–42
[11] Supriatna A, Hertini E, Saputra J, Subartini B and Robbani A A 2019 International Journal of Supply Chain Management 8 156–163
[12] Alamsyah A and Friscintia P B A 2019 Artificial neural network for indonesian tourism demand forecasting 2019 7th International Conference on Information and Communication Technology (ICoICT) (IEEE) pp 1–7
[13] Supriatna A, Lesmana E, Aridin L, Napitupulu H et al. 2019 Comparison between multiplicative holt winter and decomposition method in predicting the number of incoming international tourists to indonesia IOP Conference Series: Materials Science and Engineering vol 567 (IOP Publishing) p 012047
[14] Ete A A, Suhartono S and Atok R M 2020 Inference 3 55–63
[15] Ete A, Fitrianawati M and Arifin M 2019 Forecasting the number of tourist arrivals to batam by applying the singular spectrum analysis and the arima method First International Conference on Progressive Civil Society (ICONPROCS 2019) (Atlantis Press)
[16] Sun S, Wei Y, Tsui K L and Wang S 2019 Tourism Management 70 1–10
[17] Chen N C, Xie W, Welsch R E, Larson K and Xie J 2017 Comprehensive predictions of tourists’ next visit location based on call detail records using machine learning and deep learning methods 2017 IEEE International Congress on Big Data (BigData Congress) (IEEE) pp 1–6
[18] Li Y and Cao H 2018 Procedia Computer Science 129 277–283
[19] Claveria O, Monte E and Torra S 2016 Applied Economics Letters 23 428–431
[20] Mechelli A and Vieira S 2019 Machine Learning: Methods and Applications to Brain Disorders (Academic Press)
[21] Retico A, Bosco P, Cerello P, Fiorina E, Chincarini A and Fantacci M E 2015 Journal of Neuroimaging 25 552–563
[22] Simeone O 2017 arXiv preprint arXiv:1709.03840
[23] Athoillah M, Irawan M I and Imah E M 2015 Support vector machine with multiple kernel learning for image retrieval 2015 International Conference on Information & Communication Technology and Systems (ICTS) (IEEE) pp 17–22
[24] Moreno J J M, Pol A P, Abad A S and Blasco B C 2013 Psicothema 25 500–506
[25] Lewis C D 1982 Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting (Butterworth-Heinemann)