PROC. INTERNAT. CONF. SCI. ENGIN.
Volume 3, April 2020 | Pages: 323-327

ISSN 2597-5250
E-ISSN 2598-232X

Support Vector Machine Predictive Analysis Implementation: Case Study of Tax Revenue in Government of South Lampung

Arina Ashfa Fikriya*, Sanny Hikmawati**
Department of Electrical Engineering and Information Technology, Faculty of Engineering, Gadjah Mada University,
Jl. Grafika No. 2 Senolowo Sinduadi Mlati Yogyakarta 55281, Indonesia. Tel. +62-838-34648911, Fax. +62-274-6492191.
Email: arinaashfa2018@mail.ugm.ac.id*, sanny.hikmawati@mail.ugm.ac.id**

Abstract. Tax is potential revenue that is used by the government as a source of funding to run the government. One of the conditions needed to implement decentralization is the availability of sources of local revenue. Some local revenue sources are hotel and restaurant taxes. Issues raised in this research are that the number of hotel and restaurant visitors changes every year, resulting in fluctuations in the amount of tax. On the other hand, the Government of South Lampung had difficulty in predicting the target of hotel and restaurant tax revenue when arranging a revenue budget. This is due to the absence of formula to calculate the potential tax revenue accurately, resulting in a lack of strategic management for local revenue improvement. Now, Business Intelligence is becoming a trend. Many sectors use Business Intelligence to analyze and prepare new strategies and improve performance. Thus, it is necessary to use Business Intelligence to predict the potential of hotel and restaurant tax revenue so that the Government of South Lampung can develop appropriate strategies to improve local tax revenue and minimize tax reduction. The method used is a predictive analysis using the Support Vector Machine (SVM). The result of this study is expected to be taken into consideration for South Lampung Local Government Revenue Service, in particular for the determination of the target of the hotel and restaurant tax sector in the coming year.

Keywords: Business Intelligence, Predictive Analysis, SVM, Taxes, Hotel, Restaurant.

INTRODUCTION

Tax is a compulsory levy paid by someone based on regulations used for public purposes. Each region has the authority and responsibility for regional development based on regional autonomy policy. Local taxes are the main source of local revenue. Local governments must set targets to see the achievement of revenue. Target calculation system which is applied only based on last year's revenue target plus 10-20%. The calculation is less accurate to find out the potential increase in Local Revenue to result in revenue targets not shows its potential. Therefore a prediction is needed to know the potential of Local Revenue (LKPD, 2016-2018).

Fluctuations in hotel and restaurant revenue affect the Local Revenue South Lampung Regency. An increase in regional revenue is a positive thing because the funds can be used for regional development, especially public facilities such as hospitals and highways. However, on the contrary, a decrease in regional revenue can cause obstacles for local governments to carry out regional development due to a lack of funds. If regional development does not go well, it will have a harmful impact on the general public. For example, they are reducing the number of tourists who come to South Lampung because of inadequate facilities.

Fluctuations the number of restaurant and hotel taxes must be followed by proper planning so that regional development can run effectively in accordance with budgeted funds. Therefore, forecasting can be the right solution for estimating the amount of hotel and restaurant tax revenue for the coming year. If there is an increase, it can be used as additional revenue to increase regional development South Lampung Regency. But if there is a decline, the government can re-plan by reviewing the causes of revenue decline.

Many sectors use Business Intelligence to analyze and prepare new strategies and improve performance. Several benefits can be obtained if any organizations implement BI. First, increasing the value of organizational data and information. Through BI, all data and information can be integrated so as to produce complete decision making. Second, the data and information generated easily accessible and more easily understood. Third, facilitate monitoring of organizational performance. Fourth, BI produces comprehensive information. Fifth, support well-informed workers

Thus, it is necessary to use Business Intelligence to predict the potential of hotel and restaurant tax revenue so that the Government of South Lampung can develop appropriate strategies to improve local tax revenue and minimize tax reduction.

Meanwhile, forecasting hotel and restaurant tax revenue very useful as planning and decision making in South Lampung Government. One forecasting method that has excellent performance is the Support Vector Machine (SVM) method (Budiarti ,2017). The SVM classification uses a hyperplane function that aims to separate the two pattern classes.

MATERIALS AND METHODS

Study area

The dataset used in this research is tax revenue data in South Lampung Region, Indonesia, from 2016 to 2018. The dataset used for training is tax revenue data from 2016 to 2017. While the dataset used for testing is tax revenue data from 2017 to 2018. The dataset contains nine fields, and there are the date, month, number of evidence, description, type of tax, sub-type of tax, debit, credit, and saldo. The total of the data is 14495, with 3017 rows in 2016, 4747 in 2017, and 6729 rows in 2018. The data that will be analyzed are only two types of tax. There are hotel and restaurant tax based on the data taken from regional revenue data in South Lampung Region. The tools used to analyze descriptive and predictive analysis in this study are Microsoft excel and Rapidminer.

Procedures

The procedural steps carried out in this study can be seen in Figure 1.

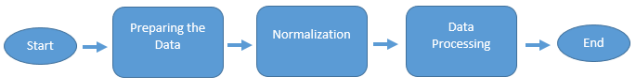

Figure 1. Research Flowchart

- ***Preparing the data***

The data obtained still needs to be reprocessed based on the amount of revenue per month and year. The data is processed into training data and testing data. The processed data generates revenue input from 2016 to 2017, which is then used as a training dataset. While the data from 2017 to 2018 used as testing data. Training and testing data are grouped into two, input data and target data. Input data starts from month 1 to 12, while target data starts at month 13.

- ***Normalization***

Then the data is being normalized. The process of data normalization is done by making existing data into smaller values to optimize the computational process. The data is normalized to be in a certain range, which is between 0 to 1. Data transformation is better if it is done in smaller intervals, such as [0.1 , 0.9]. The value of the data uses a very large rupiah value. This will have an impact on the computational process. In this study, the normalization process uses the following formula () below:

$$X' = \frac{0.8\,(X-b)}{(a-b)} + 0.1 \qquad (1)$$

X' : Normalized data
X : Original data
a : Maximum value of original data
b : Minimum value of original data

- ***Data Processing***

The method approach used is descriptive and predictive analysis. The predictive analysis is done by using the Support Vector Machine method. After getting the result from SVM analysis, the result is being denormalization to get Rupiah value.

Data analysis

The data is analyzed using the Support Vector Machine (SVM) method. SVM is a technique used to make a prediction (Santosa, 2007). Characteristics SVM explained as follows:

1. SVM is a linear classifier.
2. Pattern recognition is done by transforming data in the input space to that higher dimension space, and optimization is done on space the new vector. That matter differentiates SVM from the solution pattern recognition in general.
3. Implement a Structural Risk Minimization (SRM) strategy.
4. SVM is basically only able to handle the classification of two classes.

Many data mining or techniques machine learning developed for linearity cases. The resulting algorithm is also limited to linear cases. Therefore, it can use the kernel function. The kernel functions used in SVM are a dot, linear, polinomyal, radial, annova, etc. The kernel functions are used depending on the data. The cross-validation method can be used in kernel selection. An alternative approach for "train and test" which is often used in some cases called n-fold cross-validation (Bramer, 2007), by testing the number of errors in the data test (Santosa, 2007).

N-folds Cross Validation is one of the techniques for validation that is very popular and suitable for a limited number of sample data to carry out the classification process; the data is divided into training and testing. In n-fold cross-validation, data (D) is divided into n subsets of D1, D2, D3, ..., Dn with the same amount. The data used for training is the n-1 subset data, which is combined together and then applied to the subset data as a testing result. This process is repeated as many as n subsets.N-folds commonly used are 3, 5, 10, and 20 (Bolon et al. 2015).

Some of the advantages of the SVM method are as follows:

1. Generalization: Generalization is defined as the ability of a method to classify a pattern, which does not include data used in the learning phase of that method.
2. Feasibility: SVM can be implemented relatively easily because the process of determining support vectors can be formulated in the QP problem (Nugroho et al. 2003).

There are several kernels in SVM, including linear, polynomial, RBF, and sigmoid. Kernel selection affects the level accuracy, as well as root, mean squared error (RMSE). The Kernel used in this research is the polynomial kernel. This is because the polynomial kernel is used in the data that class boundaries are non-linear or overlapping (Kancherla et al. 2019). The dataset used is suitable to use this kernel, which is proven with a level of accuracy and RMSE smaller than using other kernels.

The success of a prediction is measured not only by results an accurate and optimal trial but also from error (Santosa, 2007). The following is the error used in measurement, namely Root Mean Squared Error (RMSE).

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(Yi-\hat{Y})^2}}{n} \quad (2)$$

Yi : Original data
$\hat{Y}$: Prediction data
n : the amount of data

RESULTS AND DISCUSSION

Descriptive analysis

Descriptive analysis is done to get visible results based on available information. The purpose of descriptive analysis is to simplify the data so that readers will easily see information from data visualization. Descriptive analysis needs to be done before forecasting to find out trends.

a) The highest average hotel tax revenue is in July and December. This happened because of the school holidays, so the amount of hotel revenue was high during the month. The graph of hotel tax revenue can be seen in Figure 2

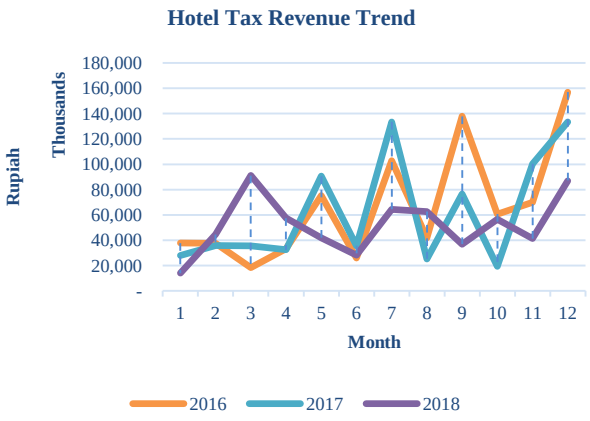

Figure 2. Hotel Tax Revenue Trend Chart.

b) The highest average restaurant tax revenue is also found in July and December, such as hotel tax revenue. The graph of hotel tax revenue can be seen in Figure 3

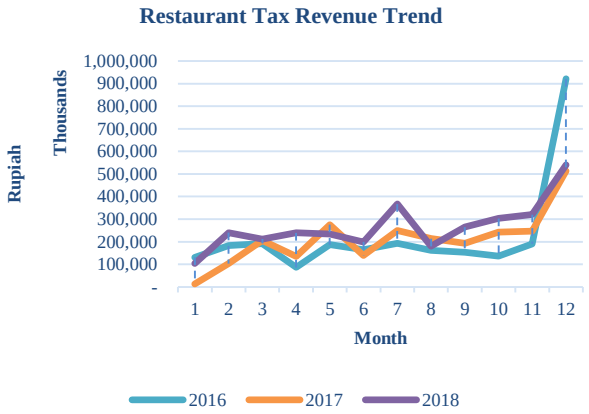

Figure 3. Restaurant Tax Revenue Trend Chart.

c) The realization of hotel taxes in 2016 exceeded the target, while in 2017 and 2018, the realization did not reach the target. Besides, revenue realization decreased from 2016 to 2018, but the tax revenue target continued to increase from 2016 to 2018. This shows that the target and revenue are not synchronous, and there is no evaluation of the decline in tax revenue. The graph of the target and realization of hotel tax revenue can be seen in Figure 4. The realization of hotel tax revenue that did not reach the target can be seen in Figure 5 and Figure 6.

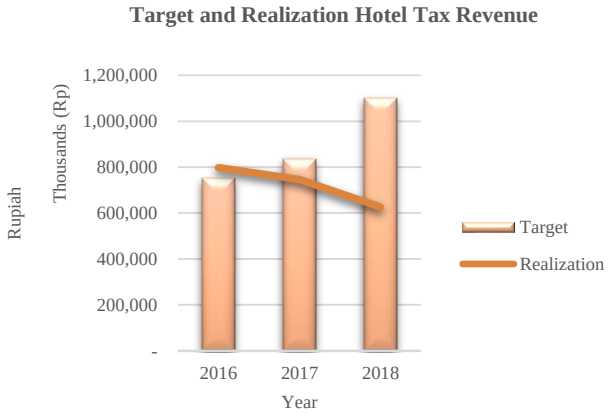

Figure 4. Target and Realization of Hotel Tax Revenue.

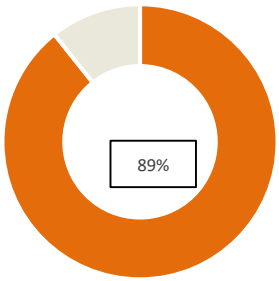

Figure 5. Percentage of Hotel Tax Revenue Realization 2017.

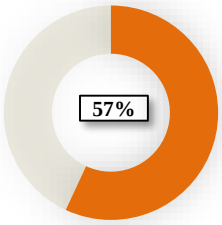

Figure 6. Percentage of Hotel Tax Revenue Realization 2018.

d) The realization of the restaurant tax is higher than hotel taxes. Restaurant tax revenue exceeded the target in 2016 but did not reach the target in 2017 and 2018. Realization of revenue decline from 2016 to 2017. This shows that the regional government did not evaluate tax revenue so that the target set did not reflect the potential revenue to be obtained. The graph of the target and realization of hotel tax revenue can be seen in Figure 7. The realization of restaurant tax revenue that did not reach the target can be seen in Figure 8 and Figure 9.

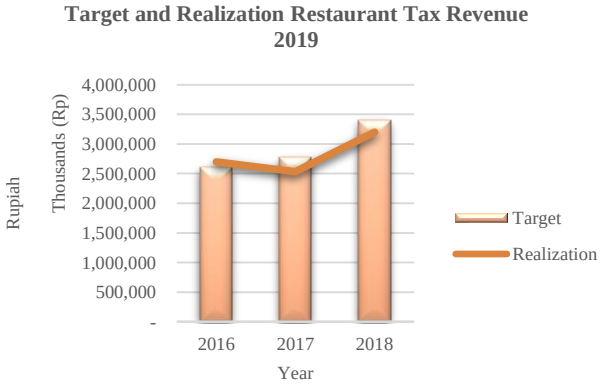

Figure 7. Target and Realization of Restaurant Tax Revenue.

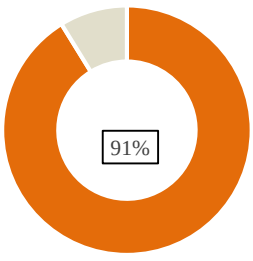

Figure 8. Percentage of Restaurant Tax Revenue Realization 2017.

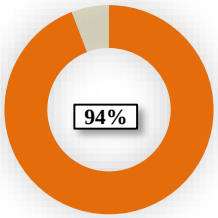

Figure 9. Percentage of Restaurant Tax Revenue Realization 2018

Predictive analysis

Predictive analysis in this study was conducted to determine trends in the amount of hotel and restaurant tax revenue in the following month. The data is processed by the SVM method using the Polynomial kernel. This method has a lower Root Mean Square Error than other methods for dataset trends of hotel and restaurant tax revenue in South Lampung Regency. The results of the predictive analysis are explained below.

a) Hotel Tax Revenue Forecast
Forecasting results using SVM with the Polynomial kernel show the RMSE value is 0,200. The result of the forecasting shown in Figure 10.

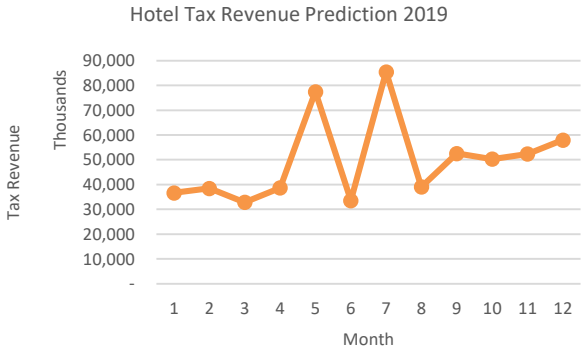

Figure 10. Hotel Tax Revenue Prediction.

b) Restaurant Tax Revenue Forecast
Forecasting results using SVM with the Polynomial kernel show the RMSE value is 0,226. The result of the forecasting shown in Figure 11.

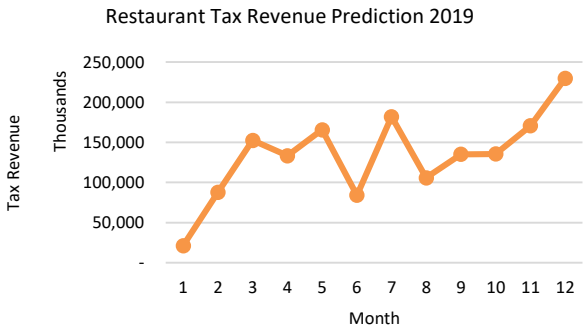

Figure 11. RestaurantTax Revenue Prediction.

Discussion

Based on the prediction results of hotel tax revenues, it is known that quite high revenue was found in May and July, while restaurant tax revenue was quite high found in July and December. This revenue prediction can be used by organizations to find out the potential revenue that will occur in the coming year and the period with the highest revenue so that organizations can maximize tax collection at that time.

From the results of descriptive analysis, it is known that the hotel tax revenue in 2018 did not reach the target, and its realization was only 57%. The target set is too high. For January to June 2019, the realization of hotel tax revenue is around 302 million, and the predicted results show around 257 million. While in 2018, the realization is around 278 million. It shows that the realization of 2018, the realization of 2019, and the prediction of 2019 are not much different. The estimated total hotel revenue in 2019 is 596 million, while the tax revenue target in 2019 is 1,2 billion. This shows that the predicted results are closer to realization when compared to the targets set by the government. A comparison of actual and predicted hotel tax revenue data can be seen in Figure 12.

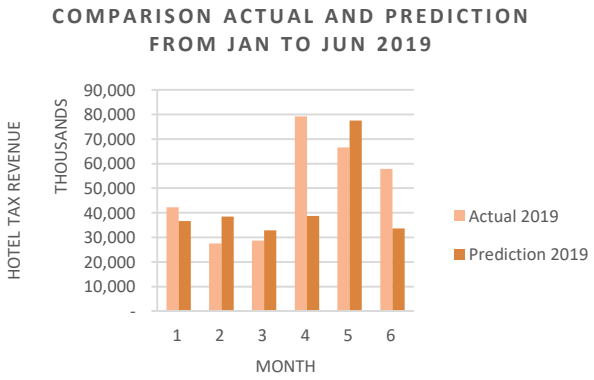

Figure 12. Hotel tax Revenue Comparison.

The results of the comparison of actual hotel tax revenue data from January to June 2019 show that the forecast results and the original data are not much different. It shows that the results of forecasting can be used as a basis for determining targets in the coming year so that the targets set shows the revenue potential and can be achieved.

Also, from the above discussion, it is known that local governments set tax revenue targets from the previous year's target plus 10-20%. If this continues, setting targets that are too high can result in targets not being achieved and showing poor performance. Until now, the calculation process of revenue target by local governments is far from expected. This problem certainly has an impact on various development programs that were budgeted. For example, activities that must be carried out now must be postponed for the following year because the tax realization was not achieved properly.

The results of the study are expected to be a consideration for the South Lampung Government to determine the target of receiving tax revenue, especially from the hotel tax sector realistically in the coming year.

CONCLUSION

Many sectors use Business Intelligence to analyze and prepare new strategies and improve performance. Thus, it is necessary to use Business Intelligence to predict the potential of hotel and restaurant tax revenue so that the Government of South Lampung can develop appropriate strategies to improve local tax revenue and minimize tax reduction. The method used is a predictive analysis using the Support Vector Machine (SVM). The result of this study is expected to be taken into consideration for South Lampung Local Government Revenue Service, in particular for the determination of the target of the hotel and restaurant tax sector in the coming year.

REFERENCES

Bolon CV, Sanchez MN et al. 2014. A Review of Microarray Datasets and Applied Feature Selection Methods. Information Sciences. 282: 111–135.

Bramer M. 2007. Principles of Data Mining. London: Springer.

Budiarti RPN. 2017. Klasifikasi Air Sungai Berbasis Kombinasi Teknologi IOT-Big Data Menggunakan SVM. Institut Teknologi Sepuluh November. Surabaya.

Kancherla D et al. 2019. Effect of Different Kernels on the Performance of an SVM Based Classification. International Journal of Recent Technology and Engineering (IJRTE). ISSN: 2277-3878. Vol.7 Issue 584.

Laporan Keuangan Pemerintah Daerah (LKPD) Kabupaten Lampung Selatan TA 2016, TA 2017, TA 2018.

Nugroho, AS et al. 2003. Support Vector Machine dan Aplikasinya Dalam Bioinformatika.

P. Studi, T. Informatika, J. T. Informatika, F. Sains, D. A. N. Teknologi, and U. S. Dharma, "Deteksi Outlier Pada Data Campuran Numerik Dan Kategorikal Menggunakan Algoritma Enhanced Class Outlier Distance Based (Ecodb) Algoritma Enhanced Class Outlier Distance Based (Ecodb)."

Priddy KL, Paul EK. 2005. Artificial Neural Networks: An Introduction, Washington: SPIE-The International Society for Optical Engineering

Santosa B. 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.

THIS PAGE INTENTIONALLY LEFT BLANK