ISSN:3026-0442

Proceedings of the 1st International Conference on Education, Science Technology And Health (ICONESTH 2023 Universitas Bina Bangsa Getsempena, Dec 12-14, 2023, Banda Aceh, Indonesia)

IMPLEMENTATION OF SUPPORT VECTOR REGRESSION FOR PREDICTING LAWYER CHARGES USING CLOUD COMPUTING IN GOOGLE COLAB

Ilmiyati Sari[1], Rifki Kosasih[2], Dina Indarti[3]
[1,2,3]Gunadarma University, Jawa Barat, Indonesia

***** Corresponding email: *ilmiyati@staff.gunadrma.ac.id*

ABSTRACT

In court, the role of the lawyer is needed by the defendant in carrying out the defense. In various legal cases, lawyers are able to assist and accompany their clients until the judge's decision is rendered. However, lawyer charge can vary depending on several factors. This research proposes 3 factors that can affect attorney fees. i.e., length of sentence, type of case and distance of lawyer's office to Jakarta district court. The data used in this study is 100 judge's decision data originating from the website of the Jakarta district court. Then the data is divided into two parts with 80 data for training data and 20 for test data. In this study, we predict the lawyer charge by implementing Support Vector Regression (SVR) method in Google Colab. Based on results of this research, we found that the value of Mean Square Error (MSE) was 0.401, the value of Mean Absolute Error (MAE) was 0.347, and the value of Root Mean Square Error (RMSE) was 0.633.

Keywords: *lawyer, Jakarta district court, judge's decision, Support Vector Regression.*

INTRODUCTION

Lawyers have an important role in every process of the criminal justice system. A person or a corporation that is experiencing legal problems requires professional assistance in the field of law due to their limited knowledge of criminal cases. In several cases in the settlement of legal problems, other legal problems often arise. The presence of a lawyer can explain everything related to

the law and will ensure that the client understands the situation being faced. Laws can be complicated no matter the type of problem or situation.

Therefore, it is important for clients to have a lawyer who can help with the problems they are facing. However, we are often hampered by the problem of determining the fees that clients must pay to lawyers. Romine (1977) asserted that there is a significant level of uncertainty in determining the fees for legal services.

In the present era of big data and artificial intelligence, attorney fees can be predicted by considering fundamental factors in building a predictive model. The predictions of attorney fees can be utilized to standardize the minimum margins for legal assistance costs. Given the development of applications that can connect individuals in need of legal assistance with attorneys, these fee predictions can serve as a reference for the community when seeking legal representation.

Support vector regression (SVR) is one of machine learning algorithm that has similar framework as support vector machine (SVM), but for regression problem (Keyvanpour & Shirzad, 2022). Keyvanpour & Shirzad (2022) implement SVM in agriculture. This algorithm is also used in various fields such as forecasting stock prices (Chen & Hou, 2017), prediction water consumption (Setiyowati & Alamsyah, 2019), forecasting number of passanger (Purnama & Setianingsih, 2020). Therefore, in this research, we propose to predict lawyer charges machine learning approaches such as support vector regression.

METHODS

To predict lawyer charges, the following steps are needed as in Fig 1. The first step, we collect datasets, find the factors that can affect the lawyer charges, divide the data into training dataset and testing dataset, use support regression model to predict lawyer charges and we evaluate the model by calculating Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE).

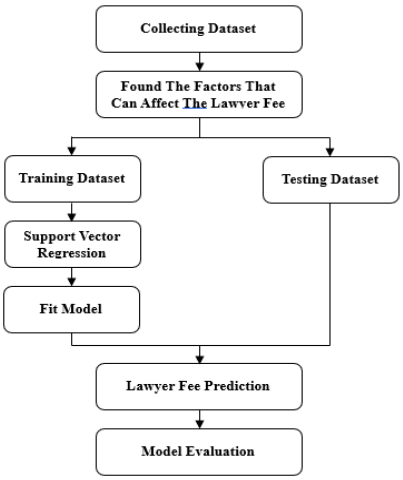

Figure 1. Stages of This Research

A. Dataset

We collect a dataset of judge decisions at trials in the Jakarta district court. We use 100 legal case data obtained from the website of the Jakarta district court. We divide the data randomly into 80 for train dataset and 20 for test dataset.

B. The Factors That Can Affect The Lawyer Fee

In this study, we propose three factors that can affect lawyer charge, namely the type of case being faced, the length of time the sentence is related to the level of difficulty of a case and the distance the lawyer's office takes to the court.

In this study, the types of cases used consisted of 5 categories i.e., 1 for narcotics cases, 2 for corruption cases, 3 for murder cases, 4 for theft cases and 5 for currency counterfeiting cases.

Furthermore, the length of sentence factor is divided into 4 categories i.e., 1 for cases with a sentence duration of less than 2 years, 2 for cases with a sentence duration of between 2 to 5 years, 3 for cases with a sentence duration of between 5 to 10 years and 4 for cases with a sentence of more than 10 years.

The third factor is the distance from the lawyer's office to the court. In this research, we use the services of a partner attorney who is located at Jl. H. R. Rasuna Said, Kuningan Jakarta. While the district courts used are 5 district courts in the Jakarta area i.e., North Jakarta district court, central Jakarta district court, south Jakarta district court, west Jakarta district court and west Jakarta district court.

C. Support Vector Regression

SVR is explicitly defined as an optimization issue by first establishing a convex ε -insensitive loss function to be reduced and then locating the flattest tube that contains the loss function (Chen & Gao, 2017). Then, using suitable numerical optimization methods, the convex optimization, which has a unique solution, is solved.

We build a hyperplane represented by a support vector $= (\xi_1^*, \xi_2, \xi_3, \xi_4^*, \xi_5)$, which is a training sample that is outside or on the border of the area as shown in Fig 2.

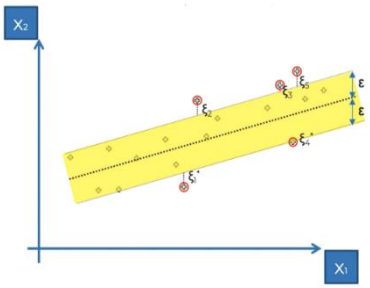

Figure 2. One Dimensional Linear of (SVR)

The training data was taken as $\{(x_1, y_1), \cdots, (x_l, y_l)\} \subseteq \mathfrak{N} \times \mathbb{R}$, where $\mathfrak{N}$ denotes the space of the input pattern, for example $\mathbb{R}^d$. SVR is a convex optimization problem with variable slack ξ_i^*, ξ_i, as written in equation (1),

$$minimize \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \qquad (1)$$

$$subject\ to \begin{cases} y_i - (w \cdot x_1) - b = \varepsilon + \xi_i \\ (w \cdot x_1) + b - y_i = \varepsilon + \xi_i^*, \\ \xi_i,\ \xi_i^* \geq 0 \end{cases}$$

Where $\|\cdot\|$ is euclidean norm, $\boldsymbol{w} \in \mathfrak{N}$, and $b \in \mathbb{R}$. A regularization parameter C defines the trade-off between overfitting and complexity model (Guo & Feng, 2017). We use ε-SVR with ε -intensive loss function, as shown in equation (2).

$$\lceil \xi \rceil_\varepsilon := \begin{cases} 0 & if\ \lceil \xi \rceil \leq \varepsilon \\ \lceil \xi \rceil - \varepsilon & else \end{cases} \qquad (2)$$

D. Fit Model

Making a model that fits the data is needed so that the model can predict well. One way to find out whether the model fits the data is to calculate the value of R^2 as in equation (3) (Kasuya, 2019),

$$R^2 = \frac{\left(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (3)$$

The R^2 value describes whether the model used can represent the data (Ratnawati, 2021). After the fit model is done, we perform predictions by using testing dataset. Then we evaluate the model.

E. Evaluation Model

We evaluate the model in predicting attorney fees. We use MSE in equation (4), MAE in (5) and RMSE in (6) (Tang, et. al, 2016).

$$MSE = \frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}{n}, \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}{n}}, \quad (5)$$

$$MAE = \frac{\sum_{i=1}^{N}|x_i-\hat{x}_i|}{n}, \quad (6)$$

with n is number of testing dataset, x_i is actual value and $\hat{x}_i$ is prediction value.

RESULTS AND DISCUSSION

Table 1 shows 5 of 100 data. According to Table 1, we have three independent variables and one dependent variable. The independent variables are is length of sentence, type of case and distance from the lawyer's office to the Jakarta district court. The dependent variable is the lawyer fee. We run this program in Google Colab using phyton.

Table 1. 5 of 100 data

No.	*length of sentence*	*type of case*	*Distance (Km)*	*Lawyer Fee (in Million Rupiah)*
1.	2	1	7	7.5
2.	3	1	14	9.5
3.	1	1	16	8
4.	3	1	7	8
5.	2	1	14	9

SVR method has a parameter value of C. We use the value of R^2 to find out whether the model can fit the data. We calculate R^2 value for parameter C from 1 to 1000. The calculation results are displayed in a scatter plot as shown in Fig 3.

Based on Figure 3, when the parameter value C = 1 an R^2 value of 0.796 is obtained, for a C = 2 value an R^2 value of 0.931 is obtained, for a C = 2 value

an R^2 value of 0.966 is obtained. The greater the C value, the higher the R^2 value. When the value of C = 1000, the value of R^2 is 0.996.

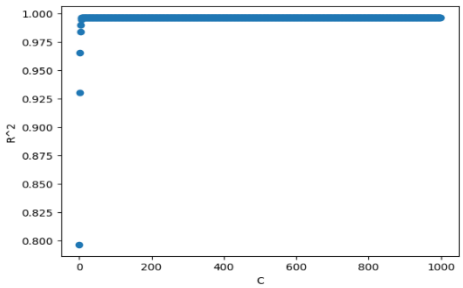

Figure 3. R^2 value for C = 1 to C = 1000

The results of the prediction of lawyer fees can be seen in Table 2. In Table 2, lawyer charge predictions were obtained by Support Vector Regression method according to 5 of 10 data in Table 1. Furthermore, we evaluate the model by calculate MSE, MAE and RMSE values using (4), (5) and (6). Based on the calculations, the MSE value is 0.401, the MAE value is 0.347 and the RMSE value is 0.633. Based on these results, the MSE, MAE and RMSE values have values close to 0 which indicates the model's performance is good in predicting data.

Table 2. Lawyer Charge Prediction

No.	Actual Lawyer Charge	Prediction of Lawyer Charge
1.	8	7.90016064
2.	9	9.09998363
3.	10.5	10.40019849
4.	8	8.09994697
5.	12.5	12.41673465

CONCLUSION

This study aims to predict lawyer charges using the SVR. The data used is data on the judge decision on the case at hand, which comes from the Jakarta district court website. We use three independent variables such as length of punishment, types of case and distance from lawyer's office to the Jakarta district court.

Before predicting, we do model fit by calculating the value of R^2. Based on the calculation, the R^2 value is 0.9964. These results indicate that approximately 99.64% of the model can represent the data. Then predictions are made using test data and model evaluation is carried out. Based on the

evaluation results, the value of the MSE of testing set is 0.401, the value of MAE of testing set is 0.347 and the value of RMSE of testing set is 0.633. Based on these results, the MSE, MAE and RMSE values have values close to 0 which indicates the model's performance is good in predicting data.

For further research, the data used is multiplied and the independent variables used are also added. As a comparison, predictions are then made using other methods such as other regression methods.

ACKNOWLEDGEMENT

This research was funded by PTUPT 2023 Research Grant, Ministry of Education, Culture, Research and Technology, Indonesia to Gunadarma University No. 073/E5/PG.02.00.PL/2023 in 12 April 2023.

REFERENCES

Chen, J., Chen, H., Huo, Y., & Gao, W. (2017). Application of svr models in stock index forecast based on different parameter search methods. *Open Journal of Statistics*, *7*(02), 194.

Guo, Q., Feng, Y., Sun, X., & Zhang, L. (2017). Power demand forecasting and application based on SVR. *Procedia computer science*, *122*, 269-275.

Kasuya, E. (2019). *On the use of r and r squared in correlation and regression* (Vol. 34, No. 1, pp. 235-236). Hoboken, USA: John Wiley & Sons, Inc..

Keyvanpour, M. R., & Shirzad, M. B. (2022). Machine learning techniques for agricultural image recognition. In *Application of Machine Learning in Agriculture* (pp. 283-305). Academic Press.

Purnama, D. I., & Setianingsih, S. (2020). Support vector regression (SVR) model for forecasting number of passengers on domestic flights at Sultan Hasanudin airport Makassar. *Jurnal Matematika, Statistika dan Komputasi*, *16*(3), 391-403.

Ratnawati, W. (2021). The Effect of Work Communication And Work Discipline on Employee Performance at PT Si Cepat Ekspres Indonesia. *Journal of Research in Business, Economics, and Education*, *3*(5), 140-148.

Romine, W. (1977). Legal fees: Gross overcharging by an attorney warranting disciplinary action. *J. Legal Prof.*, *2*, 119.

Setiyowati, D., Alamsyah, A., & Muslim, M. A. (2019). Water Consumption Prediction of Semarang Water Utilities using Support Vector Regression Radial Basic Function Kernel Method. *Journal of Advances in Information Systems and Technology*, *1*(1), 55-63.

Tang, J., Zou, Y., Ash, J., Zhang, S., Liu, F., & Wang, Y. (2016). Travel time estimation using freeway point detector data based on evolving fuzzy neural inference system. *PloS one*, *11*(2), e0147263.