Prediction of House Prices in South DKI Jakarta to Avoid Loan Failure Using Method Support Vector Machine

Bagus Bayu Kencana[1,a)], Silvester Dian Handy Permana[2,b)], Ade Syahputra[3,c)]
[1,2,3]*Department of Informatics, Universitas Trilogi, Jakarta, Indonesia*

ABSTRACT: House is a property asset that experiences an increase in price every year. Such a high price must be paid in installments to make it easier to get a house by means of a Home Ownership Credit (KPR) through a bank. All of this will not work without a third party for home appraisal through the Public Appraisal Service Office (KJPP). The need for KJPP to estimate the price of land when KPR. KJPP is an independent institution in charge of estimating the selling price of land and buildings prior to a decision for a mortgage. KJPP ordered the appraisal to go into the field in estimating the price of land and buildings. However, the appraisal in the field can intervene between the seller and the buyer by negotiating the price of the land and buildings. This is done to get a mortgage with a cheap down payment. The result of this intervention is the occurrence of default by the mortgage recipient. However, in property appraisal there is the possibility of errors in the home assessment to be used as a credit reference. As a result, the selling value of the house does not match which is a high selling value. So that the credit paid by the bank becomes high and home buyers have to pay a down payment and the installments become large. This study uses Machine Learning with the algorithm method SVM (Support Vector Machine). This method is used to study how patterns affect the selling price of land, in order to get a prediction of house prices so that they are not easily fooled by price games. The variables that have been determined include the location, certificate and physical condition of the building. The results of this study are expected to predict land prices using the SVM method. So it is hoped that prospective home buyers who want to apply for a mortgage can get the correct value.

Keywords: Land Price, SVM (Support Vector Machine), Prediction

I. INTRODUCTION

[1]Houses have a strategic role in urban or regional development. Rapid urban development and higher population growth are one of the supporters of the increasing need for housing. The need for housing has increased along with increasing population growth. The Annual Population taken from the Central Statistics Agency (2020), the Jabodetabek area has an increase of 1.6% per year or about three hundred thousand per year within a period of 5 years. The increase in population growth resulted in the need for less land for settlement and resulted in land prices rising, with rising house prices some people took advantage of the inappropriate selling value.

[2]DKI Jakarta Province is the center of government and also a business center, therefore there is only vacant land or vacant land to be built to be used as housing or offices, but the prices offered can soar due to the lack of public knowledge in selling land and building prices. In 2020 the Central Statistics Agency which is a Non-Ministerial Government Institution that is directly responsible to the President. Conduct annual publications on People's Welfare Statistics for DKI Jakarta Province, providing information and descriptions of the state of people's welfare. In the publication, 45.04% of the houses are self-owned, 37.71% are rented or contracted, 16.08% are rent-free, these are houses owned by other people or owned by relatives or parents and the last 1, 17% are official houses, more than 50% of the data received are not owned by their own homes.

[3]The number of people who do not have their own houses, one of the efforts to encourage housing demand is to provide housing provided by the government or the private sector. Home Ownership Credit is an effort to increase home ownership for Low-Income Communities in the face of rising house prices. The houses built are included in the category of Housing Financing

Liquidity Facility or more often referred to as subsidized houses, which encourage someone to own a house in a housing complex so that they no longer need to look for land to build a house. Live alone and build your own house. Buying the type or shape of the house according to the ability to buy the house.

[4]The Public Appraisal Service Office is taken from the regulation of the Minister of Finance NUMBER 125/PMK.01/2008 article 1 paragraph 5 The Public Appraisal Service Office is a business entity that has received a business license from the Minister as a forum for public appraisers in providing their services. In article 2 paragraph 1 it is explained that the service fields that can be assessed include the property appraisal service sector and also the business appraiser service sector, in property appraisal explains in more detail in article 2 paragraph 2, appraisal services include, among others, land and buildings and their accessories.

[5]Basically, the price is assessed by two factors, the first factor is a measurable factor (tangible factor). Where these factors are based on a strategic location, there are certificates and the physical condition of the building, then the second factor is intangible (intangible factor). Many of these unmeasured factors are caused by socio-psychological aspects, such as the length of time occupying the building, the convenience of the building according to the sellerand also the security of the area.

[6]Based on the explanation above, there are problems faced in providing house prices through intangible factors or unmeasured factors. Seeing the community's need for houses, the appraisal is very high, which is appointed to enter the field to intervene against sellers and buyers. As a result, the selling value of the house does not match which is a high selling value. So that the credit paid by the bank becomes high and home buyers have to pay a down payment and the installments become large. This results in losses to the bank and also potential home buyers so that they fail to pay.

[7]In the research that wants to be made related to house price prediction using the SVM method, the SVM algorithm is one of the methods pattern recognition, and is still in its early stages, in the field of pattern recognition, a series of advanced concepts. Through the evaluation results, this method has experienced rapid development and the development of the latest pattern recognition capabilities. SVM is a machine learning method based on the principle of

Structural Risk Minimization (SRM) which aims to find the hyperplane best that separates two categories in the input space. The SVM method is also amethod deep learning that is usually used for classifications such asclassification support vector and regression or support vector regression. The concept of this method is more mature and clear in mathematics compared to other methods of classification technology. In addition, this method can also solve the problem of classification and linear and non-linear regression. In this case, the method used to classify existing data.

II. RESEARCH

[8]Methodology   The   research methodology used in this study includes data collection,   data   preprocessing,   SVM implementation, prediction results, and evaluation. However, in this study, the main method used is the Support Vector Machine method. The SVM method in the world of machine learning is also included in an Artificial neural network which is also known as the Support Vector Network which is   a   supervised   method   related   to   learning algorithms to analyze a linear and non-linear data pattern used for classification and regression.
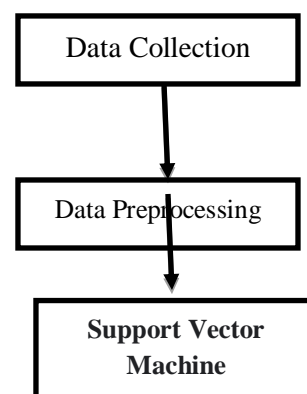

Figure 1 Research Methodology

2.1 Data Collection

[9]Collection stage was carried out to obtain the information needed to conduct research. The data collected is the Building Permit (IMB) in South Jakarta. Data collection was obtained directly from Tebet District, South Jakarta. The data used in this study consisted of 140 data. The data collected has several variables, namely building location, land area, building area, SHM, physical condition of the building. The data is expected to be able to support the system in research so that the results obtained are more accurate by using methods machine learning.

Before carrying out the process of using predictions, the data will be adjusted to the needs of the research. Researchers use Microsoft Excel to determine the test label whether the data get the results of the accuracy or not of continuity in predicting house prices.

2.2 Data Preprocessing

[10]Selection phase is carried out to select data for research needs. Datasets are the main material in machine learning training as knowledge of the program. Researchers determine data that includes several variables such as building location, land area, building area, SHM, physical condition of the building. The data that has been selected will be carried out in the next stage, namely classification using the Support Vector Machine algorithm.

2.3 Support Vector Machine

[11]Support Vector Machine (SVM) is a method in supervised learning that is usually used for classification (such as Support Vector Classification) and regression (Support Vector Regression). In classification modeling, SVM has a more mature and clearer mathematical concept than other classification techniques. SVM can also solve classification and regression problems with linear and non-linear.

[12]SVM is used to find the hyper plane best by maximizing the distance between classes. Hyperplane is a function that can be used for the separator between classes. In 2-D the function used for classification between classes is called line whereas, the function used for classification between classes in 3-D is called plane similarly, while the function used for classification in higher dimensional class spaces is called hyperplane.

Researchers who want to use the formula in this study, can see the separator between the two categories in the equation to use the following formula to make predictions to get the results of the accuracy or not of continuity in predicting house prices.

$$T(w) = \frac{1}{2}\|\vec{w}\|^2$$

$$\frac{1}{2}\|w\|^2 = \frac{1}{2}(w_1^2 + w_2^2)$$

[13]In the equation is an equation, where it is used to find the hyperplane optimal to separate the two classes completely in the input space, as shown on the graph of the sample data on, by obtaining two separate classes. There is a condition in the data input space and the formula to find the

hyperplane using this formula is used to compare the two margins of the two categories. To experiment with the existing data with the condition to n Use the equation:

$$yi(xi \cdot w + b) - 1 \geq 0, i = 1,2,3,\ldots,n$$
$$yi(x2\,x1 \cdots w1 + w2 + b) \geq 1$$

Based on formula used to determine the condition of the hyperplane next, four equations can be found, where the equation is obtained based on the conditions in the equation.

$$(w_1 + w_2 + b\ 1) \text{ for } y1 = 1, x1 = 1, x2 = 1$$
$$(-w_1 + w_{21} - b) \text{ for } y2 = -1, x1 = 1, x2 = -1$$
$$(w_1 - w_2 - b\ 1) \text{ for } y3 = -1, x1 = -1, x2 = 1$$
$$(w_1 + w_{21} - b) \text{ for } y1 = -1, x1 = -1, x2 = -1$$

[14] In the equations that have been found, then by calculating every existing equation to get the value of each existing variable, by adding up each equation, starting from equations 1 and 2, then we get:

$$(w_1 + w_2 + b) \geq 1$$
$$(-w_1 + w_2 - b) \geq 1+$$
$$2w_2 = 2w_2 = 1$$

[15]Can be seen from the above calculation that any calculations based on similarities can be seen that the equation is not worth , and both classes can decide that the calculation is positive. In the next equation, the equation used is the same as before, and in the next equation, the calculation between equations 1 and 3 is as follows.

$$(w_1 + w_2 + b) \geq 1$$
$$(w_1 - w_2 - b) \geq 1+$$
$$2w_1 = 2w_1 = 1$$

The following equation with the same method lies in the equations 2 and 3, resulting in the following equation:

$$(-w_1 + w_2 - b) \geq 1$$
$$(w_1 - w_2 - b) \geq 1+$$
$$-2b = 2b = -1$$

Based on the above equation can be drawn a conclusion that leads to the separation of all kinds of hyperplane with high enough precision in theinput space space is in the equation:

$$W_1 \cdot X_1 + W_2 \cdot X_2 + b = 0X_1 + X_2 - 1 = 0X_2 = 1 - X_1$$

Given
W_1 = First class
W_2 = second class
B = Angle between the two margins
X = kernel in determining the results

III. RESULT AND DISCUSSSION

This study used sample data, namely 140 datasets. Then the dataset is divided into two data subsets, namely training data and variable test data used in the process of testing the SVM algorithm. The variables used in the prediction are building area, land area, SHM, physical condition, location. At this stage of testing the researchers used google colab. After testing the prediction accuracy level, here are the results of testing the data using Google Colab, the test is carried out by dividing 140 data into 119 training data and 21 test data. The results of the identification using the existing input parameters, for the initial test of the system model with test data of 98 data have an average accuracy of 97% for the initial modeling of the system. The error value measures the difference in the predicted sample values of 3%.

the maximum average accuracy according to the process in the method used is related to the Support Vector Machine method. For further research it can be suggested to be applied with other algorithms to improve performance and determine the most accurate algorithm results, and further research on Support Vector Machine algorithms must be considered in terms of parameters or datasets, more trials of Support Vector Machines, and additional parameters are needed to optimal results.
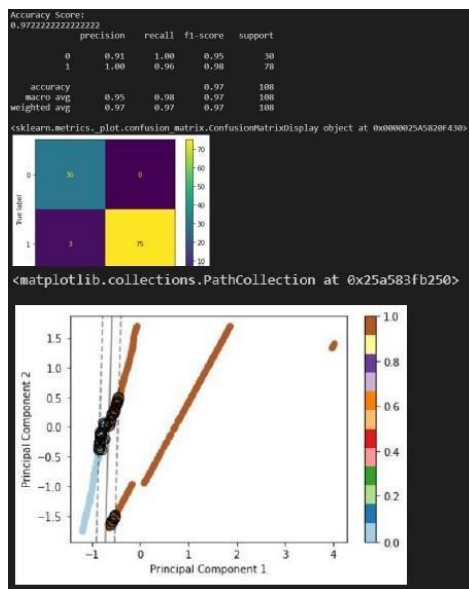

Figure 2 Prediction Results

IV. CONCLUSION

Based on the results of the research conducted, these results can be said that in testing predictions using the Support Vector Machine on google colab has the highest result in the accuracy value by testing the dataset of 85% (119 data) training and 15% test data (21 data) with an accuracy value of 97%. These results indicate that the Support Vector Machine algorithm has the performance of all existing data can be separated perfectly by the system using a hyperplane and forms a margin to determine which data is a support vector into two classes, namely successful and failed to support the prediction results obtained have an average value.

REFERENCES

[1] E. Ermanto and E. Humaeroh, "Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Untuk Menentukan Kelayakan Kredit Rumah Bersubsidi," *J. SIGMA*, vol. 10, no. 4, pp. 17–134, 2020.

[2] V. K. Rana and T. M. Venkata Suryanarayana, "Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands," *Remote Sens. Appl. Soc. Environ.*, vol. 19, p. 100351, 2020, doi: 10.1016/j.rsase.2020.100351.

[3] D. Wang and Y. Zhao, "Using News to Predict Investor Sentiment: Based on SVM Model," *Procedia Comput. Sci.*, vol. 174, no. 2019, pp. 191–199, 2020, doi: 10.1016/j.procs.2020.06.074.

[4] A. F. Kamara, E. Chen, Q. Liu, and Z. Pan, "A hybrid neural network for predicting Days on Market a measure of liquidity in real estate industry," *Knowledge-Based Syst.*, vol. 208, p. 106417, 2020, doi: 10.1016/j.knosys.2020.106417.

[5] J. Liu, C. Xing, and Q. Zhang, "House price, fertility rates and reproductive intentions," *China Econ. Rev.*, vol. 62, p. 101496, 2020, doi: 10.1016/j.chieco.2020.101496.

[6] H. Yan, J. Zhang, S. S. Rahman, N. Zhou, and Y. Suo, "Predicting permeability changes with injecting CO_2 in coal seams during CO_2 geological sequestration: A comparative study among six SVM-based hybrid models," *Sci. Total Environ.*, vol. 705, p. 135941, 2020, doi: 10.1016/j.scitotenv.2019.135941.

[7] P. Hukum and A. Silvianna, "Law , Development & Justice Review Law , Development & Justice Review," no. 9, pp. 70–83, 2020.

[8] Y. Niu, Y. Shang, and Y. Tian, "Multi-view SVM Classification with Feature Selection," *Procedia Comput. Sci.*, vol. 162, pp. 405–412, 2019, doi: 10.1016/j.procs.2019.12.004.

[9] C. A. Moreira, E. A. Philot, A. N. Lima, and A. L. Scott, "Predicting regions prone to protein aggregation based on SVM algorithm," *Appl. Math. Comput.*, vol. 359, pp. 502–511, 2019, doi: 10.1016/j.amc.2019.04.015.

[10] C. Day, "House prices post-GFC: More household debt for longer," *Econ. Anal. Policy*, vol. 64, pp. 91–102, 2019, doi: 10.1016/j.eap.2019.07.007.

[11] R. K. Pace and S. Zhu, "The influence of house, seller, and locational factors on the probability of sale," *J. Hous. Econ.*, vol. 43, pp. 72–82, 2019, doi: 10.1016/j.jhe.2018.09.009.

[12] H. Peng, Y. Shao, K. Chen, Y. Deng, and C. Xue, "Research on Multi-class Fruits Recognition Based on Machine Vision and SVM," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 817–821, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.08.094 .

[13] A. Hosseinzadeh, A. Moeinaddini, and A. Ghasemzadeh, "Investigating factors affecting severity of large truck-involved crashes: Comparison of the SVM and random parameter logit model," *J. Safety Res.*, 2021.

[14] K. Shao, W. Fu, J. Tan, and K. Wang, "Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing," *Measurement*, vol. 173, p. 108580, 2021.

[15] S. N. Singh and A. Mohapatra, "Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model," *Appl. Soft Comput.*, p. 107730, 2021.