



PROJECT REPORT
REAL ESTATE PRICE PREDICTION USING
ARTIFICIAL NEURAL NETWORK WITH L1
REGULARIZER AND WITHOUT REGULARIZER

ALFIN FEBRIAN
20.K1.0020

Faculty of Computer Science
Soegijapranata Catholic University

2023

ABSTRACT

In the real estate industry many real estate agencies and independent brokers facing numerous factors that can impact their business process and outcomes, resulting many agencies and independent brokers began to estimate the value of property / real estate with the help of machine learning to determine which real estate listing should be prioritized in order to be sold. In recent studies two different models of machine learning using Regression Models and Neural Networks performing price prediction on some real estate data and resulting regression models were overestimated and neural networks less. Also that the neural network insufficient on validation sample quality may cause overestimation of house prices to market prices. In this paper, provided sequence real estate data with a total 81 columns, the researcher conducts different Artificial Neural Network (ANN) where the first ANN using L1 regularization while the other ANN without L1 regularization. Before conducting different model on 2 machine learning models, the researcher did some preprocessing data using Z-score normalization and Min-max scaling also that before scaling and standardized the data, the author also done feature selection using Pearson Correlation Coefficient analysis to the 81 features so that when research only use important features. After selected the features into different multiple ranges, then the selected features of train and validation trained into 2 different Artificial Neural Network models and which we splitted the train, test, and validation data with sizes 70% - 15% - 15% respectively. The researcher expect the best model performance evaluated using RMSE, MAE, MAPE, and R-Squared was the Artificial Neural Network (ANN) with L1 regularizer that resulting evaluation values 114430, 86332, 0.4871, -2.1965 respectively. The optimum hyperparameter were 100 epochs / learning intervals, Adam as the learning algorithm, 600 batch size in each epochs, use hyperbolic tanh (Tanh) as the input activaton, logistic (Sigmoid) as the hidden activation and exponential linear unit (ELU), $1e-12$ as the alpha / learning rate per epochs, data features and target scaled and not-inverse scaled using MinMaxScaler, also we also done Pearson Correlation Coefficient to the each features with values range 0.1 to 1. The logistic (Sigmoid) activation function are causing the RMSE and MAE value bigger then the standard deviation of the target variable : 79442 because it was hard when vanishing the gradient problem.

Keyword: House pricing, Artificial Neural Network, L1 Regularizer, Optimization