

ABSTRACT

Property is one of the main factors in a country's economy and contributes significantly to economic growth. On the other hand, property prices, such as housing prices, are among the most commonly discussed topics in society. Therefore, automating house price prediction can provide various benefits, one of which is offering insights into the property market condition. The main challenge lies in the numerous factors that influence a house's price, such as location, available facilities, and physical characteristics of the building. Hence, an effective method for automated house price prediction is required, and one potential approach is through machine learning. This study aims to develop a house price prediction model using Random Forest, one of the machine learning algorithms. DBSCAN is also utilized in this study to perform outlier cleaning. The study process began with data collection from a property advertisement website. Several data preprocessing stages were then performed, including data cleaning to handle missing values and outliers, and data transformation through encoding. Finally, hyperparameter tuning was conducted to obtain the best combination of parameters for the model. The results show that the Random Forest model achieved an RMSE of 413,103,734.9797, an MAE of 217,470,524.2122, and an R-Squared of 0.8911 during the training phase. Alternatively, in the testing phase the Random Forest model produced evaluation metrics of RMSE at 2,813,791,506.0763, MAE at 454,268,842.7966, and R-squared at 0.3448. These results indicate that the Random Forest model experienced a decline in performance during the testing phase compared to the training phase.

Keywords : Machine Learning, Price Prediction, Random Forest, DBSCAN