ABSTRACT

Along with the development of technology, many types of new cars were offered on the market. However, not everyone could afford to buy a new car, so many chose old cars because they were more affordable and could be resold for profit. The high demand for old cars caused this business to grow rapidly, triggering fierce competition between sellers. The main challenge for sellers was to set car prices quickly and accurately. The price of old cars was influenced by various factors such as the year of production, mileage, and type of transmission, which could vary prices. To overcome the challenges in predicting the price of used cars, this study proposed the use of the Ensemble Voting Regressor algorithm with a combination of Linear Regression and K-Nearest Neighbors Regression for predicting the pricing of used cars. Research using Ensemble Voting Regressor was not found for predicting car prices but for predicting medical waste, with R-Squared = 97%. Research for predicting car prices was found using Linear Regression with R-Squared = 98.61% and K-Nearest Neighbors Regression with R-Squared = 85%. This study used a dataset from the Kaggle platform obtained from the CarSome website, which contained price data and various specifications of used cars for sale. Data preprocessing was carried out on the dataset, such as Label Encoding and Principal Component Analysis, then the dataset was divided with a data sharing scheme of 70:30, 70:25, and 80:20 for training and testing. The results of predicting the price of old cars with the Ensemble Voting Regressor algorithm, combined with individual regression algorithms such as K-Nearest Neighbors Regression and Linear Regression, obtained the best performance using Grid Search, including the best k = 1 at a dataset ratio of 70:30 and the best k = 2 at a dataset ratio of 75:25 and 80:20. Evaluation of the prediction of old car prices was carried out using various evaluation metrics, including Mean Squared Error, Mean Absolute Error, and R-squared. Although the Ensemble Voting Regressor did not always outperform the individual regression models, the results showed that the Ensemble Voting Regressor algorithm was able to compete in predicting the price of old cars. The best evaluation results achieved by the Ensemble Voting Regressor were a Mean Squared Error of 350,934,618,153,990 on test data with a dataset ratio of 70:30, a Mean Absolute Error of 10,258,462 on test data with a dataset ratio of 80:20, and an R-Squared of 93.55% on test data with a dataset ratio of 70:30. This study showed that the Ensemble Voting Regressor could be a viable alternative in predicting the price of old cars.

Keyword: Old Car Price, Prediction, Regression, Ensemble Voting Regressor, Evaluation Metrics