

ABSTRAK

Mobil menjadi salah satu kendaraan yang tingkat pembeliannya juga ramai di Indonesia ataupun dunia. Sayangnya, harga mobil yang tinggi menjadikan banyak orang memilih untuk membeli mobil bekas dengan kondisi yang masih baik dibandingkan harus membeli mobil baru. Hal ini dilakukan karena mobil bekas dengan spesifikasi yang masih bagus dapat digunakan dengan baik dengan harga pembelian yang *affordable*. Untuk memudahkan pembeli dalam memprediksi harga mobil bekas, dilakukan analisis prediksi harga mobil menggunakan metode *Multiple Linear Regression* dan *Random Forest*. Penelitian ini menggunakan 16 variabel yang terbagi menjadi variabel dependen dan variabel independen. Data yang digunakan sebanyak 309 data yang kemudian terbagi menjadi *train* dan *test* dengan perbandingan 80:20. Dengan menggunakan *Multiple Linear Regression* didapatkan hasil kesimpulan *overfitting* karena variabel independen tidak linear terhadap variabel dependen. Sehingga analisis dilanjutkan menggunakan *Random Forest* yang sifatnya *robust* terhadap ketidaklinearan data atau dapat dikatakan bahwa *Random Forest* merupakan metode *non-linear regression*. Berdasarkan penelitian, nilai R-Squared untuk data *train* dan *test* masing-masing sebesar 41% dan 71%. Sementara, nilai MAPE untuk masing-masing data *train* dan *test* sebesar 46% dan 71%.

Kata kunci : Mobil Bekas, *Multiple Linear Regression*, *Random Forest*

ABSTRACT

Cars are one of the vehicles whose purchase rate is also high in Indonesia and the world. Unfortunately, high car prices mean that many people choose to buy used cars in good condition rather than buying new cars. This is done because used cars with good specifications can be used well at an affordable purchase price. To make it easier for buyers to predict used car prices, a car price prediction analysis was carried out using the *Multiple Linear Regression* and Random Forest methods. This research uses 16 variables which are divided into dependent variables and independent variables. The data used was 309 data which was then divided into train and test with a ratio of 80:20. By using *Multiple Linear Regression*, the conclusion is overfitting because the independent variable is not linear with respect to the dependent variable. So the analysis continues using Random Forest which is robust against data nonlinearities or it can be said that Random Forest is a non-linear regression method. Based on research, the R-Squared value for train and test data is 41% and 71% respectively. Meanwhile, the MAPE value for each train and test data is 46% and 71%.

Keywords: Used Cars, Multiple Linear Regression, Random Forest