Used Car Price Prediction

Hanu Akshita Manepalli

Northeastern University

Used Cars Price Predictions

The main objective of this project us to predict the prices of used cars using Machine Learning methods. The dataset that will be used is from this link https://raw.githubusercontent.com/insaid2018/Term-1/master/Data/Projects/car\_sales.csv. The dataset consist of more than 9000 rows and 10 columns. All features are independent except for ‘price’ which is the target feature. ‘Used car price’, the target feature is a continuous variable therefore this project is a regression analysis. Explanatory Data Analysis will be used before working on the main task, price prediction.Supervised Machine Learning techniques (regression models) such as Linear Regression, Random Forest Regression and Support Vector Machine Regression will be used.

The feature engineering and selection techniques will be used to select the most important and necessary features in the dataset. One of the instances of feature correlation techniques will be correlation calculation between each independent feature and the target variable. The feature engineering techniques will ensure the data is normally distributed. Besides, dimensionality reduction methods such Principal Component Analysis (PCA) or Truncated Singular Value Division (Truncated SVD) may be used to reduce the number of features. The dataset will then be split into train and test datasets.

The transformed data will be fitted to the regression models. Evaluation such as RMSE and R2 score will be calculated. Cross validation techniques such Grid Search CV or K Fold cross validation will be used to detect overfitting and fine-tune the model to find the best parameters for the model.

To best of my knowledge, this analysis has not been done by anyone before on this specific dataset. Similar analysis may have been done but on different dataset.