



Efficient feature selection and prediction for automobile dataset

Sobana, S.

*Faculty of Applied Science
University of Vavuniya
sobanaselvaratnam@gmail.com*

Yogarajah, B.

*Faculty of Applied Science
University of Vavuniya
yoganbala@yahoo.com*

ABSTRACT

Machine learning is currently the most commonly used branch of computer science, and it can guide better decisions and intelligent actions for high-value predictions in real-time. Researchers can use machine learning model predictions to make conclusions about the expected outcomes of a research question based on past data. The primary goal of this study is to discover the factors that influence car prices and to forecast car prices by evaluating them. A considerable number of distinct features are examined for reliable and accurate prediction. Using a combined technique of LASSO and stepwise selection regression algorithms, the influencing characteristics of automobile pricing have been determined. To build a model for predicting the price of automobile dataset, UC Irvine machine learning repository data have been used. The experimental results show that automobile price's optimal characteristics are width, engine size, city mpg, stroke, make, aspiration, body style, number of doors, and drive wheels. The training accuracy was found out to be 92%, and the testing accuracy was 87% using multiple linear regression. LASSO is significant in the minimization of prediction errors, and it offers models with high prediction accuracy. Stepwise selection can manage high amounts of predictors, and it is faster than other automatic model selection methods in machine learning. Therefore, the proposed approach helps select the most influencing features to predict the price of automobiles effectively.

Keywords: Automobile price prediction, Feature selection, LASSO, Stepwise selection.