

**PREDICTION OF MARKET PRICE FOR USED CARS IN
SRI LANKA BASED ON MACHINE LEARNING
APPROACHES**

Colambage Deveendee Rangika Basnayaka

A dissertation submitted in partial fulfilment of the requirement for Master of Science
degree in Big Data Analytics

Department of Computing

**Informatics Institute of Technology, Sri Lanka in collaboration with
Robert Gordon University, Aberdeen, Scotland**

2023

Abstract

Due to the import limitations in Sri Lanka, vehicle price had artificially risen to a high level of unusual prices. Consequently the price of secondhand cars get doubled with the increasing demand for used cars. The majority of buyers fall victim to unscrupulous vehicle dealers who take advantage of the situation by offering unrealistic prices for used cars. On the other hand predicting secondhand automobile prices will enable consumers to sell their vehicles for a fair price. As a result price prediction for secondhand cars is very necessary for Sri Lanka in order to efficiently assess the value of the vehicle considering range of features. Although there are systems showing price lists; proper research on predicting market price based on machine learning approaches, finding significant variables and relationship between variables, visualizing data analysis still have spaces for improvements.

This study focuses on predicting market price for Sri Lankan secondhand cars using machine learning approaches. An actual Sri Lankan dataset available in Kaggle was used to develop the prediction model. A comprehensive analysis was done including data cleaning, data exploration, feature engineering and model creation. Test accuracy of Multiple Linear Regression, Random Forest, Extra Trees, Gradient Boosting, Light Gradient Boosting and Extreme Gradient Boosting models were compared and hyperparameter optimization was done to find the best model. Almost all the Regressor models gave accuracy closer to 90% except multiple linear regression model. Ensemble techniques were further used to improve the accuracy by taking simple average of predictions and the final model got 91% test accuracy.

Key Words: Secondhand Cars, Sri Lanka, Price Prediction, Machine Learning, Simple Average Ensemble Technique, Multiple Linear Regression, Random Forest, Extra Trees, Gradient Boost, Extreme Gradient Boost, Light Gradient Boost