USED VEHICLE PRICE PREDICTION SYSTEM

By

Sandusha Marasinghe

W1699682 / 2017820

Supervised by

Ms. Fathima Farhath

A dissertation submitted in partial fulfilment of the requirement for Bachelor of Engineering (Hons) degree in Software Engineering

Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2021

Abstract

With the increase of vehicle imports, used vehicle market has become popular and interesting. With the market popularity and no prior knowledge about used vehicles, many people get fraud to the vehicle dealers and vehicle buyers. Therefore, this becomes an important and interesting problem. In my research, I proposed a system to predict used vehicle price for the Sri Lankan vehicle market. Therefore, I decided to implement a web application that can predict the price of used vehicles. As first priority, I collected the dataset from Ikman.lk, which is a website that can sell and buy used vehicles. I was able to collect 21510 rows and after several phases of data cleaning process there was 17677 rows. In the data preprocessing phase I have used label encoding as well as one hot encoding for the categorical variables. Before selecting an appropriate algorithm for the model, I have used three algorithms and did a comparative study on each other performance. XGBoost algorithm performed better than linear regression and random forest. XGBoost training and testing results were 97.14% and 91.72%. The best-performed algorithm optimized with Bayesian optimization and compared the result. Then the web application built with flask using the prediction model which gave the best result.

Keywords: Machine Learning, Random Forest, XGBoost , Bayesian Optimization