

**A PREDICTIVE MODELLING APPROACH TO
DETERMINE THE PRICES OF USED VEHICLES IN
SRI LANKA**

Maggona Gurunselage Lahiru Roshan Anthony Perera

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

Department of Business
Informatic Institute of Technology, Sri Lanka
In Collaboration with
Robert Gordon University, Aberdeen, Scotland, UK

2022

Abstract

Vehicle industry is in a stage where most of the dealers have run out of business. With covid coming through and foreign reserves going low, the government decided to stop importing vehicles. This issue led to stage where the remainder of vehicles' prices have started sky rocketing. This issue has come to a crucial stage where most of the people are trying to sell their own vehicles and others have given up on the hope of buying vehicles. The problem led to conducting this study was to identify a method to predict the prices of the vehicles to make the prices standardized. Then the main aims would be to identify the variables that would determine the price along with gathering a dataset before the price hike. Variables for this study has been covered under "Performance", "Appearance" and "usage". Based on these, the data was collected under 10 variables where the variable named "users" had to be taken off due to lack of data. The dataset was gathered through a certain extensive period to have enough data for the study. Once the dataset has been gathered, the descriptive analytics have been conducted to gain insights about the dataset at hand. This was done through one of the most efficient visualization tools named as power BI. Descriptive analytics showed the importance and the relationship between the independent and dependent variables. Once the insights have been gained, the data needed to be uploaded to google colab which has been used as the platform for compiling. When using python there are always a certain set of libraries and functions that need to be used to perform various tasks. The main task is to create the models where the study has used Linear Regression, Polynomial Regression, Lasso Regression, Ridge regression, Decision Tree Regression and Random Forest Regression. 6 different techniques have been imported using sklearn library to perform these models. But before building up the models, there a certain aspect that the researchers look into. Identifying the multicollinearity are Identifying the impact of the independent variables towards the dependent variables are key factors before building up the models. Once the most suitable variables were selected, the data has been separated into train and test and then the models were created. Models have been evaluated using the R squared score, Root mean squared Error and Mean absolute Error. 6 models have scored 6 different set of scores for the above evaluation methods. The model with the highest accuracy was the Random Forest and the model with the

lowest accuracy was the Linear Regression. Once the models have been compared, the study has been ended leaving a door opened for the study to be conducted further and develop a software or web-based application for the benefit of the people. Finally, the research question and the problem statement have been justified with the study while satisfying the aims of the research. If this study can be implemented as a software or a web-based application, it will be beneficial for not just the buyers and sellers but also many other parties.

Acknowledgement

In preparation of my research, I had to take the help and guidance from some respected persons, who deserve my deepest gratitude. First and foremost, praises and thanks to God, the Almighty, for his showers of blessings throughout my research work to complete the research successfully.

I am extremely grateful to my parents for their love, prayers, caring, understanding and sacrifices for educating and preparing me for my future and for the continuous support given to complete this research work as well as the regular encouragement in every step to make me in present stage.

As the completion of this assignment gave me much pleasure, I would like to express my deep and sincere gratitude to my research supervisor, Dr. Alroy Mascranghe, for giving me the opportunity to do this research and providing invaluable guidance throughout numerous consultations. His dynamism, vision, sincerity, and motivation has deeply inspired me. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under his guidance. I am extremely grateful for what he has offered me.

I would like to say a big thank you to Ms. Asini Handugoda who helped me not only with collecting data over a period.

I would also like to express my appreciation to my colleagues for their moral support throughout this project and to all those who have directly and indirectly guided me in completing this assignment.