A second-hand car price prediction system with additional trends and insights

Kai Broux (r0745431), Luigi Guerriero (r0760424)

UHasselt - KULeuven

INTRODUCTION

The production of cars has been steadily increasing in the past decade and this has given rise to the used car market. The emergence of online second-hand car portals in the United Kingdom and Europe has facilitated the need for both the customer and the seller to be better informed about trends and patterns that determine the value of a used car [1]. A second-hand car price prediction system is thus required to effectively determine the worthiness of the car using a variety of features.

GOAL and METHOD

The goal of our project is to make a price prediction system for used cars and to extract interesting information and insights. The price prediction model and insights can be extracted from a dataset of different used cars. The model and insights could be then later used as a tool for a second-hand car retailer to give insights to potential customers shopping for a second-hand car. Moreover, this price prediction model and insights tool may also come in handy for a customer looking to sell his/her car.

To do this, we will use the dataset from Aditya (https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes). This dataset contains data from 100 000 used cars from the UK which are divided into separate brands. Each brand is divided into a separate file and contains information about the model, year, price, transmission, mileage, fuel type, road tax, mpg and engine size.

A first main point of interest is to predict the price of a car using linear regression with multiple variables (model, year, fuel type, ...). Second hand retailers would hereby give a price prediction to customers searching for a specific car model with fuel type etc or give an estimate trade-in value for the customer.

Another interesting pattern to research is the influence of the taxes on different brands or models. This could lead to a possible conclusion that a popular car brand could be inelastic based on the tax.

Additionally, it is interesting to predict if a specific car model with a given price is rather manual or automatic. This can be a useful insight for consumers who are searching for a specific car and want to know if the cars they would find would rather have a manual or automatic transmission. This could be predicted by using logistic regression. The same principle can be used to predict if a car uses petrol or diesel. Customers could use this to choose the cheapest option.

Finally, it is important to note that other similar predictions could be studied that are not explicitly mentioned in our method.

REFERENCES

[1] P. Venkatasubbu and G. Mukkesh, "Used Cars Price Prediction using Supervised Learning Techniques," *International Journal of Engineering and Advanced Technology (IJEAT),* vol. 2019, no. 3, pp. 216-223, 2019.