UNIVERSITY OF SURREY

USED CAR PRICE PREDICTION

Planning report

Department of Computer Science

Practical Business Analytics Project (COM053)

Project Title: **Used Car Price Prediction**

By

Group Name: **WE R**

Group Members

Student Name	University Number
Anuja Wagh	6680408
Vaishnav Bharadwaj Sathya Prasad	6677359
Jhansi Lakshmi Peetha	6703174
Dewang Umesh Solanki	6690120
Aditi Shailesh Kamlakar	6699852
Ketaki Damle	KD0071

Supervisor: **Dr. Spencer Thomas**

Table of Contents

PROBLEM DEFINITION

With the onset of pandemic, environmental mindfulness and of course, with the ever-growing expenses, people are switching to budgetary, sustainable options. There is a noticeable shift in buying choices, especially when it comes to cars. Cars are an essential commodity, but cars are expensive, and more and more people are turning to used or ‘second - hand’ cars. A car is a car after all.
Thus, it becomes vital to be able to predict prices from historical data for vehicle businesses to not only adopt and devise new marketing tactics for their dynamic consumer base but also to be able to forecast customer behaviour, estimate market risks, render business intelligence solutions, and calculate financial planning. This can be achieved by building a factual model based on machine learning methodologies and practises.

Hence, we chose to analyse, predict, and visualise prices of used cars via a dataset delineating car detail and its associated price from 9 different brands. A plethora of ideas and techniques have been employed to use.

Predicting a vehicle's resale value is not an easy undertaking. It is commonly known that the value of a used automobile is determined by a number of things. The most relevant ones are generally the car's age, make (and model), origin (the manufacturer's native nation), mileage (the number of kilometres it has travelled), and horsepower.

Gasoline efficiency is particularly important because of rising fuel prices. Unfortunately, most individuals do not realise how much gasoline their automobile consumes per kilometre driven in practise. Other factors include the type of fuel it uses, the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), the safety index, the car's dimensions, number of doors, paint colour, weight, customer feedback, awards and accolades won by the auto manufacturer, the car's physical condition, whether that is a sports car, and if it has cruise control, The kind of transmission, whether it belonged to an individual or a firm, and other features like as air conditioning, sound system, power steering, cosmic wheels, and GPS navigator can all affect the price. The car's appearance and feel have a significant role on the pricing.

As we can see, the price is influenced by a variety of factors. Unfortunately, information on all these criteria is not always readily available, and the buyer must make a purchasing choice based on only a few considerations. In this study, we looked at a dataset that included some of the above-mentioned characteristics.

The dataset has been procured from Kaggle after careful consideration of its data, values, parameters, and variables and is in lieu of the project aim.

Dataset:
https://www.kaggle.com/jenniewong/used-car-price-prediction-for-all-brands-by-mlr/data

As part of our analysis, we have taken a dataset of 9 widely used car brands across Britain Markets and their prices.

Overall data set consists of 100,000 listings spread across 13 CSV files. The below parameters are taken into consideration and the values are measured with respect to each manufacturer into individual CSV files. The brands like Audi, BMW, Mercedes, Ford, Hyundai, Skoda, Toyota, Vauxhall and Volkswagen are taken into consideration for our analysis.

- Car Model – Type of car models available across 9 car brands
- Year of Manufacturing
- Model price- measured in Pounds
- Transmission - details that define type of gearbox
- Mileage - with information of miles travelled
- Fuel type - engine versions (petrol/diesel)
- Tax-applicable Road tax details measured in Pounds
- Mpg - Miles per gallon
- Engine size - that is recorded in litres

Our CSV files consist of combinational data.
Parameters like year, price, mileage, tax holds data that is numerical, decimal values in mpg & engine size, where car model, transmission & fuel type consists of descriptive values.

Business Analytics Task

1. Introduction
2. Importing dataset and exploration
3. Exploratory data analysis
4. Pre-processing for modelling
5. Model Implementation
6. Result Evaluation
7. Conclusion

Approached Models

A. Linear Regression/Multi Linear Regression:

Regression Models can deliver a strong threshold for comprehensive data exploration that may be employed as a preliminary step for subsequent approaches to be researched. In this context, we're talking about neural network models and the Random Forest Regressor.

B. Feed Forward Neural Networks (ANN):

We can resolve the issue of nonlinearity by using neural networks. The network must first be trained on the data sets using the back-propagation process. To do this, numerous network characteristics must be optimised, including the number of hidden layers, the number of hidden layer units, the kind of activation function, the learning rate, and the number of training cycles (also known as epochs).

C. Random Forest Regressor (CART):

CART may be used to solve issues in both regression and classification. The optimal split endpoints for regression are determined by lowering the squared or absolute errors. The points that best separate (Entropy/Gini) the classes are chosen as the optimal split points for classification. CART has several features, including easy-to-understand rules and automated handling of variables including selection, missing data, outliers, local effect modelling, variable interaction, and non-linear connections.
A Relative error can be used to evaluate a CART model. A Relative error near to Zero is regarded as positive, since it indicates that the method is doing superior than just forecasting the general average/median for all data.

Evaluation

We want to anticipate the price of second-hand cars from nine different brands in this study. We'd start by combining data from nine different brands, factoring, and grouping category variables, and dealing with NA values. Then, for identifying independent variables, we'll utilise the stepAIC function, which gives stepwise regression. Our strategy entails using log-transformation, the ANN technique, and CART to remove outliers from the model. We eventually arrive at a final version of the model that has the highest R2 of all of them.

Our goal is to provide output that is low in mistakes and easy to interpret. You should be able to enter custom automobile information and receive a cost breakdown.

Project plan

Week (Finish by Weekend-Sunday)	Task	Completed by
Week 6 (Completed)	**Create and Upload Project Planning Report**	**All Team Members**
Week 7	▪ **Data pre-processing** ▪ **Identifying Modelling Techniques**	▪ **Dewang and Aditi** ▪ **Ketaki and Jhansi**
Week 7	**Modelling the Data** As we are targeting to implement 3 methods, we will split into groups of 2 and work. ▪ Linear / Multilinear Regression ▪ Neural Networks (ANN) ▪ Random Forest Regressor (CART)	▪ **Aditi and Jhansi** ▪ **Anuja and Dewang** ▪ **Vaishnav and Ketaki**
Week 8	**Collect results, Analyze and Visualization** ▪ Linear/Multilinear Regression ▪ Neural Networks (ANN) ▪ Random Forest Regressor (CART)	▪ **Aditi and Vaishnav** ▪ **Anuja and Dewang** ▪ **Vaishnav and Ketaki**
Week 9	▪ **Report Writing** ▪ **Presentation Preparation**	▪ **Ketaki and Aditi** ▪ **Jhansi and Dewang**
Week 10	**Practice and Run through Presentation**	**All Team Members**

References

- Medium. 2021. *Predicting Car Prices Using Multivariate Linear Regression*. [online] Available at: <https://pareekshithkatti.medium.com/predicting-car-prices-using-multivariate-linear-regression-fcd54050c613>.

- Ripublication.com. 2021. [online] Available at: <http://ripublication.com/irph/ijict_spl/ijictv4n7spl_17.pdf>.

- Ijrte.org. 2021. [online] Available at: <https://www.ijrte.org/wp-content/uploads/papers/v8i5s/E10100285S20.pdf>.

- Aip.scitation.org. 2021. [online] Available at: <https://aip.scitation.org/doi/pdf/10.1063/1.4982530>.

- 2021. [online] Available at: <https://www.researchgate.net/publication/347520823_Car_Price_Prediction_Using_An_Artificial_Neural_Network>.