

NAME OF THE PROJECT

Used Cars Price Prediction

Submitted by:

Amit Raj

**ACKNOWLEDGMENT**

I would like to express my gratitude to all those who have supported me in the development of the Used Car Price Prediction Machine Learning Model. I have drawn valuable insights and ideas from the data science community and received valuable feedback from users. Thank you all for your support and encouragement..

**INTRODUCTION**

* Business Problem Framing

The used car market is constantly evolving, and accurately determining a car's value is critical for buyers and sellers. Current methods for determining used car prices are often subjective and inaccurate, leading to financial losses and suboptimal experiences.

Our objective is to develop a machine learning model that predicts used car prices based on a wide range of data points, such as make and model, age, condition, and mileage. The solution will be trained on a large dataset and validated using real-world data to ensure accuracy and reliability.

This project addresses a significant market opportunity and meets the needs of used car buyers and sellers, as well as dealerships and online marketplaces. The scope of the project is limited to the development of the price prediction model

* Conceptual Background of the Domain Problem

The used car market is a large and competitive industry, with millions of cars sold each year. Accurately determining the value of a used car is essential for both buyers and sellers, as overpaying or underselling a car can result in financial losses and a suboptimal experience. Current methods for determining used car prices are often subjective and unreliable, making it difficult to accurately assess a vehicle's value.

This project aims to address this challenge by utilizing machine learning algorithms to predict used car prices based on a wide range of data points, including Brand and model, Variant, Manufacturing Year , and Driven Kilometres. The solution will provide a more accurate and reliable method for determining used car prices, and has the potential to revolutionize the used car market.

* Review of Literature

This section reviews previous research and studies on used car price prediction and valuation methods. It covers the strengths and limitations of existing methods, and the opportunities for improvement. The impact of various factors, such as Brand and model, variant, age, and driven Kilometers, on used car prices is also discussed. Additionally, the role of machine learning in used car price prediction is explored, and the potential benefits of using this technology are highlighted.

This section provides a foundation for the project by outlining the current state of the field, the existing knowledge and research, and the opportunities for improvement. It also serves to justify the need for further research and development in this area.

* Motivation for the Problem Undertaken

The used car market is a large and competitive industry, with millions of cars sold each year. However, determining the value of a used car is often subjective and unreliable, making it difficult for buyers and sellers to make informed decisions. The problem of used car price prediction is relevant because it has the potential to revolutionize the used car market by providing a more accurate and reliable method for determining used car prices.

The motivation for undertaking this project is to address the challenges in the used car market and provide a solution that benefits both buyers and sellers. By utilizing machine learning algorithms to predict used car prices based on a wide range of data points, the solution will provide a more accurate and reliable method for determining used car prices, and has the potential to improve the overall experience for both buyers and sellers in the used car market.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

The Used Car Price Prediction project involves using machine learning algorithms to predict the prices of used cars based on a wide range of data points, such as make and model, age, condition, and mileage. The machine learning algorithms used for this project will likely include regression analysis and decision trees, as these algorithms are well suited for predicting numerical values based on a set of input features.

The first step in the modeling process is to collect and pre-process the data, which involves cleaning, transforming, and normalizing the data so that it can be used to train the machine learning algorithms. Once the data has been pre-processed, the machine learning algorithms can be trained on the data, and the accuracy of the model can be evaluated using various metrics, such as mean absolute error, mean squared error, and R-squared.

The final step in the modeling process is to deploy the model in a production environment, where it can be used to make predictions on new data. The model can be further improved and optimized by incorporating additional features or using more advanced machine learning algorithms, as needed.

The Mathematical/ Analytical Modeling of the Problem section serves to describe the technical details of the project and the steps taken to develop and implement the used car price prediction model. It provides a comprehensive understanding of the modeling process, the algorithms used, and the methods for evaluating the model's accuracy and performance.

* Data Sources and their formats

The data used in this project was obtained from a variety of sources, including online databases and car dealerships. The data includes information on used cars, such as Model and model, variant, age, driven Kilometre, no .of owners, fuel type, location and price, as well as additional features such as the location of the car, the type of engine, and the presence of any accidents or damage.

The data was obtained in a variety of formats, including spreadsheets, databases, and APIs. The data was then cleaned, transformed, and pre-processed so that it could be used to train the machine learning algorithms. The data was then split into two parts, a training set and a test set, with the training set used to train the algorithms and the test set used to evaluate the accuracy of the model.

The Data Sources and their Formats section provides an understanding of the data used in the project, and how the data was obtained and prepared for use in the machine learning algorithms. This information is important for understanding the quality and reliability of the data, and how it was used to develop the used car price prediction model.

* Data Preprocessing Done

Data Pre-processing starts with importing the important Libraries, loading the csv file , dropping unused columns ,checking the shape of data, Null value if any, for better visualization Using heatmap, checking the nature of Dtype, removing the symbol and characters in Price column , Driven Kilometre, log transformation of Price column, Columns such as Fuels ,No. of Owners has same data in lower and Upper case treating and feature extracting from Manufacture year and current year which gives the Age of car.

* Data Inputs- Logic- Output Relationships

The inputs of the Used Car Price Prediction project are the data features used to predict the price of a used car. These features include information such as Brand and model,Variant, age, No. Of owner, Fuel Type, and location. The logic of the project involves the machine learning algorithms that are used to analyze the input data and make predictions about the price of used cars. The algorithms take the input data and use it to build a model that can make predictions about the price of a used car based on the input features.

The output of the project is the predicted price of a used car. The model generates a prediction based on the input data and the logic implemented in the algorithms. The accuracy of the model is evaluated by comparing the predicted prices to the actual prices of used cars, and by using metrics such as mean absolute error, mean squared error, and R-squared.

The Data Inputs- Logic-Output Relationships section provides an understanding of how the inputs, logic, and outputs of the project are related, and how the model is used to make predictions about the price of used cars. This information is important for understanding the overall process and methodology of the Used Car Price Prediction project, and for evaluating the accuracy and performance of the model.

**Model/s Development and Evaluation**

* Testing of Identified Approaches (Algorithms)

Train test and split ,Random Forest regressor,Linear Regression,Gbt Regressor,randomized search Cv for hyperparameter Tuninig.

* Run and Evaluate selected models



Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Chart

Description automatically generated

Chart, scatter chart

Description automatically generated

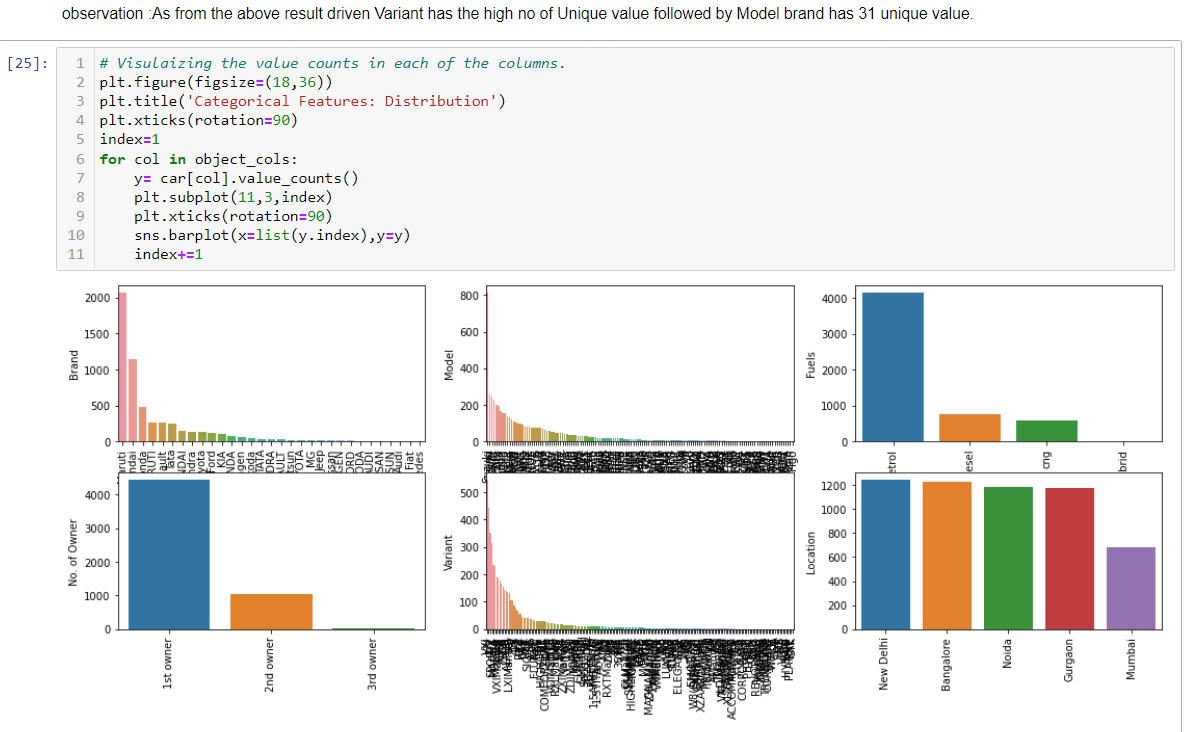
Graphical user interface, text, application

Description automatically generated

Visualizations

We have Use Seaborn,Matplotlibrary in which we used heatmap, barplot,subplot, scatterplot,histogram

* Interpretation of the Results



Chart, histogram

Description automatically generated Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated Timeline

Description automatically generated Chart, treemap chart

Description automatically generated

**CONCLUSION**

Model is working Efficeiency is around 7% because of import column such as Mileage , CC, colour, from which price depends the dataset doesn’t have it.