ALLA SAI ADITHYA REDDY

UX02894

Prediction of Used Car Price using Machine Learning

Problem statement:

The increased prices of modern technology led to the new cars becoming more expensive. Additionally, during the Covid-19 pandemic epidemic, the lack of public transportation and fear of infection force people to rely on their own means of transportation. However, the increased demand for used cars led some car dealers to overcharge clients by putting their prices higher than usual. The need to develop a model that can forecast the cost of used cars by taking into account the various characteristics and costs of other automobiles already on the market in the nation arises in order to aid consumers in understanding market trends and prices for used cars.

In this study, we applied a variety of machine learning techniques, including decision trees, random forest regression, These algorithms can predict the price of used cars based on various features that are unique to buyers.

Data source:

. https://www.kaggle.com/datasets/jpayne/852k-used-car-listings

. https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho

. https://www.kaggle.com/datasets/tugberkkaran/used-car-listings-features-and-prices-carscom

And still have some data sets.

 Data preparation

By managing missing values, duplicates, and outliers, clean up the data.

Transform data by encoding categorical variables and scaling numerical features, for example.

EDA: Exploratory Data Analysis

To understand the data, perform EDA.

Visualize the patterns, relationships, and distributions of the data.

Determine the connections between the variables.

To enhance the performance of the model, add new features or modify those that already present.

Analyse the data :

Linear Regression, Random Forest and Gradient Boost were our baseline methods. For most of the model implementations, the open-source Scikit-Learn package will be used.

Plan for the project:

Each week the task is divided and should be complted

Week 1-2 : data collecting and cleaning

Week 3-4 : data preprocessing and quality checking

Week 5-6 : analysing the data with machine learning algorithms

Week 7: try to predict and visualse the data

This is the basic overview of my project and I will try to add some more details while doing the project.