

Project Title: CarDekho Used Car Price Prediction
Technologies: Data Cleaning, Exploratory Data Analysis (EDA), Visualization and Machine Learning
Domain: Automobile

Problem Statement:

The primary objective of this project is to create a data science solution for predicting used car prices accurately by analyzing a diverse dataset including car model, no. of owners, age, mileage, fuel type, kilometers driven, features and location. The aim is to build a machine learning model that offers users to find current valuations for used cars.

Data Understanding

The Dataset contains multiple excel files, each represents its city, columns in each excel gives you an overview of each car, its details, specification and available features.

Data Collected From: [CarDekho](#)

Dataset Link: [Dataset](#)

Feature Description Link: [Features](#)

Approach:

- Import data from all excel files
- Examine the structure of each dataset component (New Car Detail, New Car Overview, etc.).

- Check for missing values, outliers, data types and other statistical inference.
- Data Preprocessing:
 - Handle Missing Values: Impute or remove missing values appropriately.
 - Feature Engineering: Extract relevant information from features like age, mileage, and others.
 - Encode categorical variables using suitable techniques.
 - Normalization/Scaling: Scale numerical features to bring them to a comparable range.
- Exploratory Data Analysis: Create visualizations to understand the distribution of target variables (used car prices) and relationships between features.
- Choose regression models suitable for predicting continuous values
- Model Evaluation: Use suitable metrics
- Fine-tune Hyperparameters: Optimize model hyperparameters to improve performance.
- Feature Importance: Analyze feature importance to understand which features contribute most to the predictions.

Submission:

1. Provide a well-commented Python file (*.py) containing the complete code for the project, organized into sections for data preprocessing, model training, and evaluation.
2. Document the entire project, including data sources, preprocessing steps, model selection and evaluation metric.

Note:

After completion of all the task you need to create a PowerPoint presentation that should contain the:

1. Problem Statement

2. Tools Used

3. Approaches

4. EDA Insights

5. Model Selection

6. Model Performance

Project Evaluation metrics:

- Project evaluation will be done in the live session and have to showcase the approaches done to complete the project
- You are supposed to write a code in a modular fashion (in functional blocks)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.(Mandatory)
- You have to keep your GitHub repo public so that anyone can check yourcode.(Mandatory)
- Proper readme file you have to maintain for any project development(Mandatory)
- Follow the coding standards: <https://www.python.org/dev/peps/pep-0008/>
- You should include basic workflow and execution of the entire project in the readme file on GitHub