**Project Summary**

**Title:** Car Popularity Prediction

**Objective:** The objective of this project is to predict the popularity of cars based on various features using machine learning techniques.

**Technologies and Tools Used:**

1. **Programming Language:** Python
2. **Libraries and Frameworks:**
   * **Pandas:** For data manipulation and analysis.
   * **Matplotlib:** For data visualization.
   * **Scikit-learn:** For machine learning algorithms and model building.
   * **Django:** For building the web application (mentioned but not fully detailed in the provided files).

**Dataset:**

* **train.csv:** This file contains the training data with features and their corresponding popularity labels.
* **test.csv:** This file contains the test data with features for which we need to predict the popularity.
* **prediction.csv:** This file stores the predicted popularity for the test data.

**Detailed Explanation:**

1. **Data Preprocessing:**
   * The data preprocessing script (datapreprocess.py) reads the training data from train.csv.
   * It visualizes the data distribution and relationships between features and the target variable (popularity) using various plots:
     + Histogram of car popularity.
     + Boxplot of car popularity.
     + Scatter plots showing the relationship between each feature and car popularity.
   * It checks for null values in the dataset to ensure data cleanliness, but no explicit handling of null values (like imputation or removal) is performed in the script.
2. **Machine Learning Model:**
   * The model fitting script (model\_fit.py) reads the training data from train.csv and test data from test.csv.
   * It uses the SVC (Support Vector Classification) algorithm from the Scikit-learn library.
     + **SVC (Support Vector Classification):**
       - **Algorithm Type:** Classification
       - **Kernel Used:** RBF (Radial Basis Function)
       - **Hyperparameter C:** 300 (Controls the trade-off between achieving a low training error and a low testing error)
   * The script trains the SVC model using the training data features and their corresponding popularity labels.
   * After training, it predicts the popularity for the test data features and saves the predictions in prediction.csv.
3. **Data Visualization:**
   * The data visualization script (datapreprocess.py) generates various plots to understand the data distribution and relationships. This step helps in:
     + Understanding the spread and central tendency of the popularity variable.
     + Identifying potential outliers.
     + Visualizing the correlation between features and the target variable.

**Key Points to Mention in Interviews:**

* **Objective:** Predicting car popularity based on features like buying price, maintenance cost, number of doors, number of seats, luggage boot size, and safety rating.
* **Technologies Used:** Python, Pandas, Matplotlib, Scikit-learn, Django.
* **Machine Learning Algorithm:** Support Vector Classification (SVC) with RBF kernel.
* **Data Preprocessing:** Included data visualization to understand the data distribution and relationships, and checking for null values.
* **Model Training and Prediction:** Trained an SVC model with RBF kernel on the training data and predicted the popularity for the test data.
* **Outcome:** Successfully predicted car popularity and saved the results in prediction.csv.