**Car Dheko: Used Car Price Prediction**

**Project Overview**

The goal of the "Car Dheko" project is to develop an interactive web application that predicts the price of used cars based on various features. The application is designed to be used by customers and sales representatives to obtain real-time car price predictions. The project leverages machine learning algorithms to build a predictive model based on historical car data and integrates this model into a user-friendly Streamlit web app for easy interaction.

**Project Objective**

The main objective of this project is to predict the price of used cars based on input features such as make, model, year, mileage, fuel type, and more. The project integrates data processing, machine learning, and web development to create an interactive tool that can be accessed by end-users for real-time price estimation.

**Project Scope**

Using historical data on used car prices and relevant features (make, model, year, fuel type, etc.), this project aims to:

* Develop a machine learning model that accurately predicts used car prices.
* Deploy this model through a Streamlit web application that enables real-time price prediction based on user input

**Approach and Methodology**

* **Data Source and Preprocessing:**
  + **Data Source**: The dataset is sourced from CarDekho, containing historical data about used cars across multiple cities (Bangalore, Chennai, Delhi, Hyderabad, Jaipur, Kolkata).
  + **Data Preprocessing**:
    - The raw data is cleaned and transformed to remove irrelevant columns, handle missing values, and convert textual data into structured formats.
    - Specific columns are standardized, including 'price', 'engine displacement', 'mileage', 'torque', and 'gearbox type'.
    - Missing values in various features such as 'alloy wheel size' and 'RTO' are filled using domain-specific imputation methods.
    - Outliers in numeric columns are identified and removed using z-score thresholds.
* **Feature Engineering:**
  + Data Various features like 'stability features', 'Blind Spot Assistance', and 'indicators' are derived from existing columns by combining related attributes.
  + Categories such as 'fuel type', 'body type', 'transmission', 'OEM', and 'model' are encoded for use in machine learning models.
  + Non-numeric features are transformed to numeric values, where applicable, to facilitate model training
* **Model Development:**
  + **Models Used:**
    - **Linear Regression**
    - **Decision Tree Regressor**
    - **Random Forest Regressor**
    - **XGBoost Regressor**
    - **Ridge Regression**
    - **Lasso Regression**
  + **Model Evaluation:**
    - Duplicate The models are evaluated using performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared, and Mean Absolute Percentage Error (MAPE).
    - After performing hyperparameter tuning using RandomizedSearchCV, Random Forest and XGBoost models were identified as the best-performing models.
    - Random Forest model achieved an R-squared score of **0.9349** with a MAPE of **15.29%,** while XGBoost achieved an R-squared score of **0.9517** with a MAPE of **12.88%.**
* **Model Optimization:**
  + Hyperparameters for Random Forest and XGBoost were tuned using techniques such as RandomizedSearchCV to improve the model performance.
  + The best-performing models were saved as xgboost\_best\_model.pkl and random\_forest\_best\_model.pkl.
* **Streamlit Web Application:**
  + **App Development**: A Streamlit web application is developed to allow users to input car details and predict the price based on the trained model.
  + **Key Features of the Streamlit App:**
    - Users can select values from dropdown menus for categorical variables such as 'fuel\_type', 'body\_type', 'transmission', etc.
    - Continuous variables like 'mileage', 'engine displacement', and 'year' are entered using sliders.
    - After submitting the details, the app predicts the car price based on the inputs and displays the result in real-time.
  + **Model Integration**: The pre-trained XGBoost model is loaded and used for predictions.

**Technical Stack**

* **Programming Language**: Python
* **Libraries**:
  + Data Processing: Pandas, NumPy
  + Machine Learning: Scikit-learn, XGBoost
  + Web Development: Streamlit
  + Visualization: Matplotlib, Seaborn
* **Deployment**: Streamlit application for easy user access

**Model Evaluation**

* **Random Forest Model:**
  + **Best Hyperparameters**:
    - n\_estimators = 300
    - max\_depth = 30
    - min\_samples\_split = 2
  + **Model Performance**:
    - **R² (Test): 0.9349**
    - **MAPE (Test): 15.29%**
* **XGBoost Model:**
  + **Best Hyperparameters:**
    - n\_estimators = 300
    - max\_depth = 5
    - learning\_rate = 0.1
  + **Model Performance:**
    - **R² (Test): 0.9517**
    - **MAPE (Test): 12.88%**

**Results**

* The XGBoost model demonstrated the best performance in terms of prediction accuracy and minimal error. This model was selected for deployment in the Streamlit application.

**App Interface**

The **Streamlit App** allows users to interact with the model and predict car prices. The interface consists of:

* Dropdowns for selecting categorical features such as fuel type, body type, transmission, etc.
* Sliders for continuous variables like mileage, engine displacement, and the year of manufacture.
* Once the form is filled, users can submit the data and view the predicted price displayed on the screen.

**Conclusion**

The Car Dheko project successfully demonstrates the power of machine learning in predicting used car prices based on historical data. The Streamlit web application provides an interactive and user-friendly interface for real-time price prediction. With further enhancements, this tool has the potential to become a reliable solution for car buyers and sellers alike.