Used Car Price Prediction Project

Project Overview

The **Used Car Price Prediction Project** aims to develop a machine learning model to predict the prices of used cars based on their features. The project involves several key stages: data collection, data preprocessing, model development, and deployment of an interactive web application using Streamlit. This tool will help both customers and sales representatives estimate car prices accurately.

Project Workflow

1. Data Collection

Source Files:

- **CSV Files:**
    - bangalore_cars.csv
    - chennai_cars.csv
    - delhi_cars.csv
    - hyderabad_cars.csv
    - jaipur_cars.csv
    - kolkata_cars.csv

It contains details about used cars, including their features and prices.

2. Data Cleaning and Preparation

Steps Performed:

- **Handling Missing Values:** Addressed missing data in the dataset to ensure completeness.
- **Standardising Data Formats:** Removing commas, special characters, and other non-numeric symbols from numerical columns.
- **Feature Engineering:** Created new features to enhance model performance.
- **Data Transformation:** Converted categorical features into numerical values using techniques such as label encoding.
- **Normalization:** Scaled numerical features to ensure consistency by Min-Max scaling.

3. Exploratory Data Analysis (EDA)

Steps Performed:

- **Data Visualization:** Created visualization to identify patterns and correlations by using scatter plots, histograms, box plots, and correlation heatmaps.
- **Feature Selection:** Identified important features that significantly impact the car prices.

4. Model Development

Approach:

- **Model Selection:** Evaluated various regression models, including:
  - **Linear Regression**
  - **Decision Trees**
  - **Random Forest**
  - **Gradient Boosting Machines**
  - KNN
- **Model Training:** Trained the chosen models using the prepared dataset.
- **Model Evaluation:** Assessed model performance using metrics such as:
  - **Mean Absolute Error (MAE)**
  - **Mean Squared Error (MSE)**
  - **R-squared**
- **Hyper parameter Tuning:** Found optimized model parameters to improve performance by grid search method
- **Model Selection:** Chosen Random Forest machine learning algorithm for price prediction

5. Deployment

Streamlit Application:

- **Development:** Created a Streamlit web application to allow users to input car features and receive price predictions.

6. Documentation and Testing

Documentation:

- **Code Documentation:** Included comments and explanations within the code for clarity.

Tools and Technologies Used

- **Python & Pandas:** For data cleaning and pre-processing.
- **Scikit-learn:** For developing and evaluating machine learning models.
- **Streamlit:** For creating and deploying the web application.
- **Jupyter Notebook:** For performing exploratory data analysis (EDA) and model development.

Installation and Setup

Prerequisites:

- Python with the following packages: Pandas, Scikit-learn, Streamlit.
- Jupyter Notebook for development and testing.

Steps:

1. **Clone the Repository:**

```bash
Copy code
git clone https://github.com/Shabanabacker/used-car-price-prediction.git
```

2. **Run the Streamlit Application:**
   - Navigate to the project directory:

```bash
Copy code
cd used-car-price-prediction/src
```

   - Start the Streamlit application:

```bash
Copy code
streamlit run cardekho_usedcar_priceprediction.py
```

3. **Testing:**
   - Open the web application in a web browser to verify its functionality.

Repository Structure

```bash
Copy code
used-car-price-prediction/
│
├── data/
│   ├── raw/ # Contains the raw data files
│
├── notebooks/
│   ├── cardekhonew.ipynb # Jupyter notebook for EDA and feature
engineering
│
├── src/
│   ├── cardekho_usedcar_priceprediction.py # Script for model development
and training
│
│
├── README.md
└── Used_Car_Price_Prediction_Project.pdf # Detailed project documentation
```

Conclusion

This project showcases the application of machine learning techniques to predict used car prices based on various features. The interactive Streamlit application developed provides a practical tool for price estimation, benefiting both customers and sales representatives. By leveraging data-driven insights, the project aims to streamline the pricing process and enhance the overall customer experience.