README for Car Price Prediction Project

Welcome to the Car Price Prediction Project! This guide explains how to set up, run, and interact with the project. The project predicts used car prices using machine learning models and provides an interactive Shiny app for user-friendly predictions.

Project Overview

This project uses a cleaned dataset of used car listings to build predictive models, including:

1. **Linear Regression**
2. **Random Forest**
3. **XGBoost**
4. **Stacked Model (combination of Random Forest and XGBoost)**

The final step includes a **Shiny app** where users can input car details and get a predicted price.

Requirements

Before running the project, ensure you have the following installed:

- **R and RStudio** (latest version recommended)
- Required R packages: dplyr, ggplot2, caret, randomForest, xgboost, shiny

---

Project Files

- Data Cleaning.R: Cleans the raw dataset and saves the cleaned dataset.
- Visualize.R: Generates visualizations for data exploration.
- Model Cleaning.R: Prepares and filters the dataset for modeling.
- Linear Regression.R: Builds and evaluates linear regression models.
- Random Forests.R: Trains and evaluates the Random Forest model.
- XGBoost.R: Trains the XGBoost model and builds the Stacked Model.
- input.R: Script for user-defined input prediction.
- app.R: Runs the interactive Shiny app.

Steps to Run the Project

Step 1: Data Preparation

1. Since the original dataset is too large to upload to GitHub, you will start with a **cleaned dataset**.
    - Download the cleaned dataset (cleaned_vehicles.csv) and save it in your working directory.
    - Skip the Data Cleaning.R file and start directly with the Visualize.R script.

Step 2: Visualizations

1. Open the Visualize.R file in RStudio.
2. Run the script to generate data visualizations (e.g., price distribution, price vs. odometer).
3. These visualizations help you understand the dataset better.

Step 3: Model Training

1. Run the following files in order to build and evaluate the models:
    - Model Cleaning.R: Prepares the dataset and filters for specific manufacturers and models.
    - Linear Regression.R: Builds linear regression models and evaluates their performance.
    - Random Forests.R: Trains the Random Forest model and visualizes feature importance.
    - XGBoost.R: Trains the XGBoost model, builds the Stacked Model, and evaluates their performance.

Each file saves intermediate results and models for use in subsequent steps.

Step 4: User Input Prediction

Before running the Shiny app, define your car details and test predictions:

1. Open the input.R file in RStudio.

2. Replace the placeholders for **year**, **manufacturer**, **odometer**, and **model** with your desired values:
3. year <- 2018
4. manufacturer <- "Ford"
5. odometer <- 30000
6. model <- "F-150"
7. Run the input.R file to get a predicted price using the Stacked Model.
8. Note the predicted price as it will appear in the console.

Step 5: Shiny App

Once you've tested predictions, you can run the Shiny app:

1. Open the app.R file in RStudio.
2. Run the entire script to start the Shiny app.
3. A browser window will open with the Car Price Prediction app.
4. Use the dropdowns to select the **manufacturer** and **model** and input the **year** and **odometer value**.
5. Click **Predict Price** to get a real-time price prediction.

Important Notes

1. **Cleaned Dataset**:
    - Since the raw dataset is too larger and hard to provided, start with the cleaned dataset (cleaned_vehicles.csv).
    - If you want to clean the raw dataset yourself, ensure you have it saved as vehicles.csv and run Data Cleaning.R first.
2. **Order of Execution**:
    - Always run input.R before using the Shiny app. This ensures all models and variables are defined correctly.
3. **Custom Inputs**:

- You can test predictions for any car's details using the input.R script before launching the app.

Troubleshooting

1. **Missing Files**:
    - Ensure all scripts are in the same directory and you've downloaded the cleaned dataset.
2. **Errors in Shiny App**:
    - Verify that you've run input.R before launching the Shiny app.
    - Check that your input values match the expected format (e.g., year is numeric, model is valid for the selected manufacturer).

Contact

If you encounter issues or have questions, feel free to reach out at zhangzubin1997@gmail.com. Happy coding!