**Car Price Prediction using Machine Learning**

**1. Introduction**

The goal of this project is to predict the selling price of used cars based on various factors such as brand, model, fuel type, transmission, and other features. Using **machine learning algorithms**, we analyze historical car data to develop an accurate predictive model

**2. Dataset Overview**

The dataset contains information about used cars, including the following attributes:

* **Year** – Manufacturing year of the car
* **Present\_Price** – Current market price of the car (in lakhs)
* **Selling\_Price** – Selling price of the car (in lakhs)
* **Kms\_Driven** – Total kilometers driven by the car
* **Fuel\_Type** – Type of fuel used (Petrol, Diesel, CNG)
* **Transmission** – Type of transmission (Manual or Automatic)
* **Owner** – Number of previous owners (0 for first owner, 1 for second, etc.)

The dataset is pre-processed to remove missing values and categorical variables are converted into numerical values.

**3. Technologies Used**

The following libraries and tools were used in this project:

* **Python** – Programming language
* **Pandas** – Data manipulation and analysis
* **NumPy** – Numerical computations
* **Matplotlib & Seaborn** – Data visualization
* **Scikit-Learn** – Machine learning models(Random Forst Regressor)

**4. Data Preprocessing**

Before training the machine learning model, the dataset undergoes several preprocessing steps:

* **Handling Missing Data** – Checking and handling any null values in the dataset
* **Categorical Encoding** – Using **One-Hot Encoding** to convert categorical values into numerical format
* **Feature Scaling** – Normalizing features for better model performance
* **Splitting Data** – Dividing the dataset into **Training (80%) and Testing (20%)** sets

**5. Model Selection & Training**

Several machine learning models were trained and evaluated:

* **Linear Regression**
* **Decision Tree Regressor**
* **Random Forest Regressor**

Among these, the **Random Forest Regressor** performed the best in terms of accuracy.

**6. Model Evaluation**

To assess the performance of our models, we used the following metrics:

* **Mean Absolute Error (MAE)** – Measures the average absolute difference between actual and predicted values.
* **Mean Squared Error (MSE)** – Measures the average squared difference between actual and predicted values.
* **R² Score** – Indicates how well the model explains the variance in data (closer to 1 is better).

| **Model** | **R² Score** |
| --- | --- |
| **Linear Regression** | 0.84 |
| **Decision Tree Regressor** | 0.90 |
| **Random Forest Regressor** | 0.96 (Best) |

Since **Random Forest Regressor** had the highest R² Score (0.96), it was selected as the final model.

**7. How to Run the Project?**

**Step 1: Clone the Repository**

git clone https://github.com/yourusername/car-price-prediction.git

cd car-price-prediction

**Step 2: Install Required Libraries**

pip install -r requirements.txt

**Step 3: Run Jupyter Notebook**

jupyter notebook

**Step 4: Train & Test the Model**

Run the notebook to train the model and check predictions.

**8. Future Improvements**

To further improve the project, the following enhancements can be made:  
✅ **Hyperparameter Tuning** to optimize model performance  
✅ **Incorporating Deep Learning models** for better predictions  
✅ **Deploying the model as a web application**

**9. Conclusion**

This project successfully predicted car selling prices using **machine learning techniques**. The **Random Forest Regressor** model provided the best accuracy. Future improvements may include **deep learning models** and **hyperparameter tuning** for even better predictions.

**10. Contact & Contribution**

📧 **Email:** [ansarmazhar477@gmail.com](mailto:ansarmazhar477@gmail.com)  
🔗 **GitHub:** https://github.com/ansar-mazhar  
Feel free to contribute or fork this repository! 🚀