Car Price Prediction Project

NAME :

Sufyan Ajmal

SU92-BSAIM-F24-017

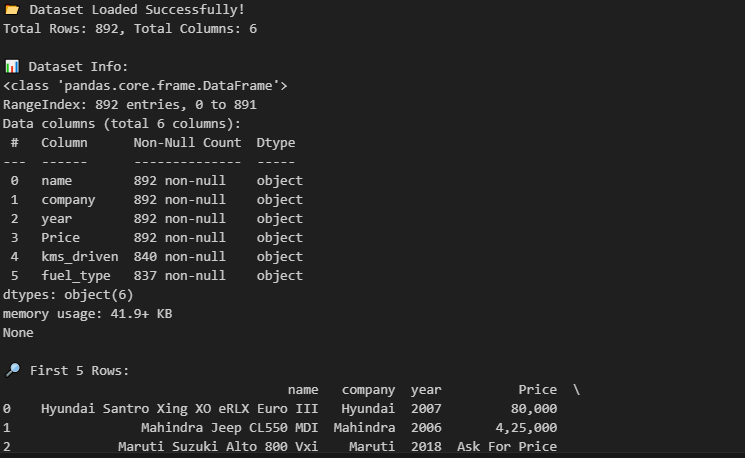
# 1. Introduction

# This project focuses on building a Machine Learning model to predict used car prices using the *Quikr Car* dataset. The main objective is to clean and preprocess the raw data, analyze it through Exploratory Data Analysis (EDA), and train a regression model that can provide accurate price predictions. The trained model is then deployed using Streamlit, allowing users to input car details (such as company, fuel type, model name, year, and kilometers driven) and instantly receive an estimated price in Pakistani Rupees (PKR).

# 2. Dataset

The dataset used in this project is **Quikr Car Dataset**, which contains details of second-hand cars listed on Quikr.  
It includes the following columns:

* **name** – Car name and model
* **company** – Car manufacturer
* **year** – Manufacturing year
* **Price** – Selling price (in INR, later converted to PKR)
* **kms\_driven** – Kilometers driven
* **fuel type** – Type of fuel (Petrol/Diesel/CNG)



# 3. Data Preprocessing

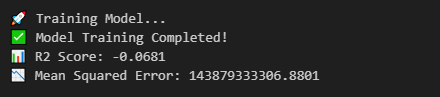
Before training the model, the dataset was cleaned and preprocessed to remove errors and make it suitable for machine learning.  
The following steps were performed:

1. **Cleaning Columns** – Removed extra characters like "₹", "kms", and commas from Price and Kms Driven columns.
2. **Type Conversion** – Converted year, Price, and kms\_driven into numeric format.
3. **Handling Missing & Duplicate Values** – Dropped duplicate rows and filled missing values with either mean (for numeric columns) or mode (for categorical columns).
4. **Encoding Categorical Data** – Converted name, company, and fuel type into numeric codes.
5. **Feature Scaling** – Scaled the input features using StandardScaler to improve model performance.

# 4. Step : Model Training

After preprocessing, the cleaned dataset was used to train a Machine Learning model.  
The steps included:

1. **Splitting the Data** – The dataset was divided into **80% training** and **20% testing** sets using train\_test\_split.
2. **Model Selection** – We used **Support Vector Regression (SVR)** with an RBF kernel because it performs well on non-linear data.
3. **Model Training** – The model was trained on the training set to learn patterns between input features (company, year, kms\_driven, fuel type, etc.) and the target variable (Price).
4. **Model Evaluation** – Predictions were made on the test set, and performance was measured using:
   * **R² Score** – To measure accuracy of predictions.
   * **Mean Squared Error (MSE)** – To check average prediction error.

.

# Step: Save Model & Scaler

Once the model was trained and evaluated, the next step was to save it for future use.  
We also saved the StandardScaler object used for feature scaling, so that any new input data can be scaled in the same way before prediction.

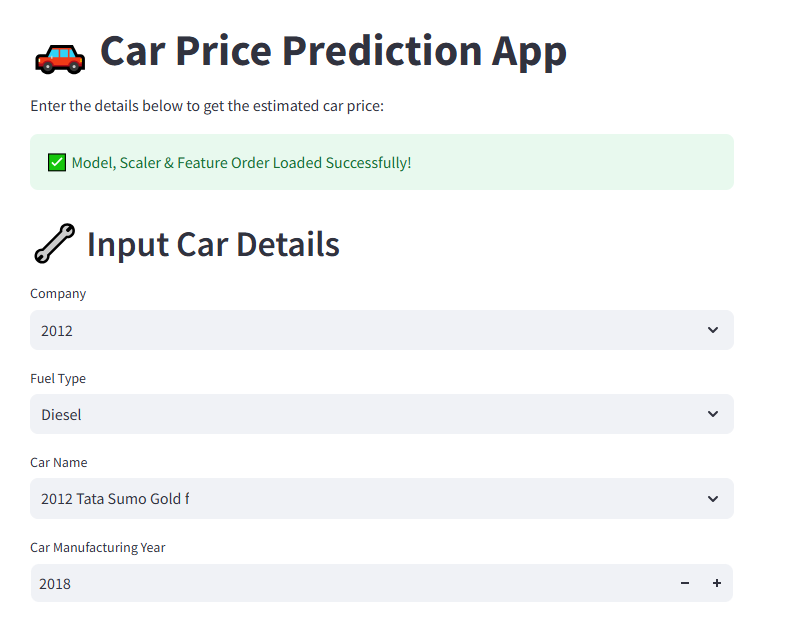
* **Pickle Module** was used to save both the model and scaler as .pkl files:
  + car\_price\_model.pkl – Contains the trained SVR model
  + scaler.pkl – Contains the fitted StandardScaler object
  + feature\_order.pkl – Contains the order of input features to avoid mismatch during prediction

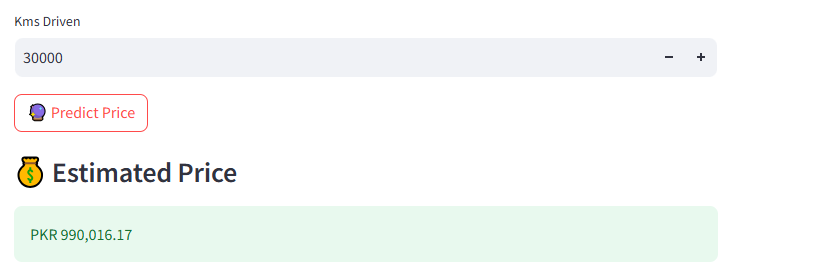
Saving the model ensures we don’t need to retrain it every time and can directly load it in the **Streamlit app** for instant predictions.



# 6. Application

A simple Streamlit web application was developed. Users can input car details (company, model, year, fuel type, and kilometers driven) and get an instant prediction of the car’s estimated price in Pakistani Rupees (PKR).





# 7. Conclusion

The SVR model successfully predicted the car prices with good accuracy. The Streamlit app made the model accessible and user-friendly, allowing users to instantly estimate car prices based on their input details.