G H Patel College of Engineering & Technology

(The Charutar Vidhya Mandal (CVM) University) Bakrol, Anand

COMPUTER ENGINEERING DEPARTMENT

Project Report

On

Car Selling Price Prediction

Submitted By

Name of Student : Het Patel
Enrolment Number : 12202130501025

Name of Student : Krish Dalsaniya
Enrolment Number : 12202130501034

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

(202046702) A.Y. 2024-25

Car Price Predictor

Objective

The primary objective of this project is to develop a **machine learning-based web application** that predicts the **resale value** of a used car. This tool is designed to aid both **car buyers and sellers** in making **informed decisions** regarding the pricing of second-hand vehicles. The application considers several critical features such as:

- Car brand and model
- Year of manufacturing
- Kilometers driven
- Fuel type

By utilizing these inputs, the system delivers an estimated resale price, streamlining the pricing process in the used car market.

Dataset Used

- **Source:** Quikr.com – a popular Indian classifieds website
- **File Name:** `quikr_car.csv`
- **Size:** Approximately **8,000 rows** after cleaning

Dataset Description

The dataset was obtained by scraping listings of used cars on Quikr. It originally contained a mix of structured and unstructured data, including:

- `car name`
- `company` (brand)
- `year` of manufacture
- `fuel type`
- `price`
- `kms driven`

Preprocessing Steps

To make the data suitable for machine learning, several preprocessing steps were performed:

- **Null and irrelevant data removal** – Dropped rows with missing values and non-car entries.
- **Feature extraction** – Separated brand and model from combined text fields.
- **Data type conversion** – Converted price and kilometers driven from strings (with commas, "kms", etc.) into numeric types.
- **Encoding categorical variables** – Applied **one-hot encoding** to convert categorical features like fuel type and company into numerical format.

Model Chosen

Algorithm Used: Linear Regression

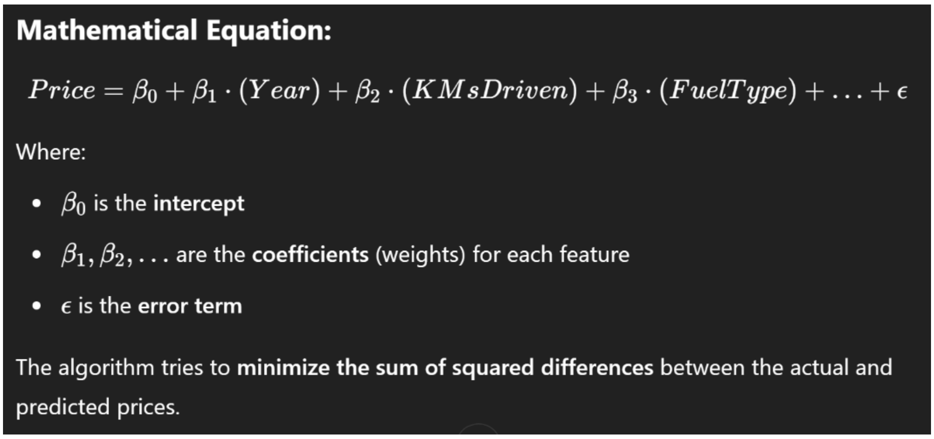
Justification:

- **Simplicity and Interpretability:** Linear Regression is easy to understand and provides interpretable coefficients for each feature.
- **Suitable for Regression Tasks:** This algorithm is designed to predict **continuous numerical values**, which makes it ideal for predicting car prices.
- **Good Baseline Performance:** When used with properly cleaned and encoded data, linear regression often provides **strong baseline results**.

Understanding Linear Regression

Linear Regression is one of the most fundamental algorithms in supervised machine learning. It works by modeling the relationship between the **independent variables** (features such as brand, year, kms driven, etc.) and the **dependent variable** (car price).

Mathematical Equation:



Mathematical Equation:

$$Price = \beta_0 + \beta_1 \cdot (Year) + \beta_2 \cdot (KMsDriven) + \beta_3 \cdot (FuelType) + \ldots + \epsilon$$

Where:

- β_0 is the **intercept**
- $\beta_1, \beta_2, \ldots$ are the **coefficients** (weights) for each feature
- ϵ is the **error term**

The algorithm tries to **minimize the sum of squared differences** between the actual and predicted prices.

Model Evaluation

- **Train/Test Split:** 80% training data, 20% test data
- **Metric Used:** R² Score (Coefficient of Determination)
- **Test R² Score: ~0.92**

Interpretation:

An R² score of **0.92** indicates that the model can explain **92% of the variability** in car prices. This reflects a **highly accurate** prediction capability, given the quality of the cleaned dataset.

Challenges & Learnings

❖ Data Cleaning:

- The raw data was unstructured.
- Required conversion of values like `"45,000 kms"` to integers.
- Removed symbols like "₹", commas, and unwanted strings.

❖ Feature Engineering:

- Separated combined fields like "Maruti Swift" into:
  - Brand = Maruti
  - Model = Swift

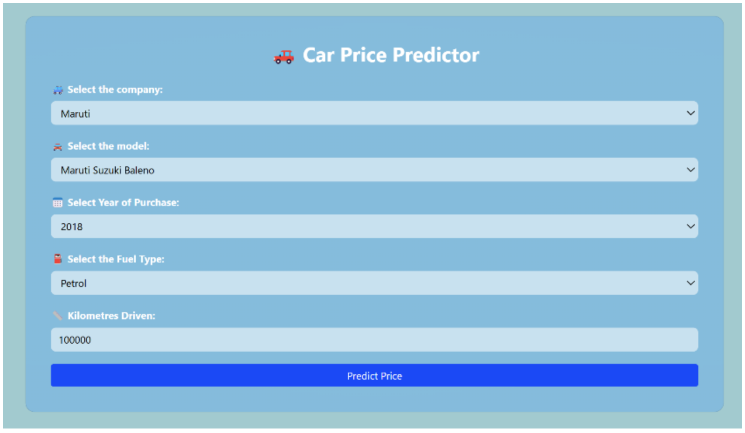
❖ Model Versioning Issues:

- Initially faced issues loading saved models due to differences in scikit-learn versions.
- Learned to use **`Pipeline` and `joblib`** to save both preprocessing and model steps together for consistent deployment.

❖ Deployment Experience:

- Deployed the model using **Flask**, a lightweight Python web framework.
- Gained hands-on experience in:
  - Integrating ML model with a web app
  - Handling user inputs from HTML forms
  - Displaying predictions in a user-friendly format.

Project Glimps



Car Price Predictor

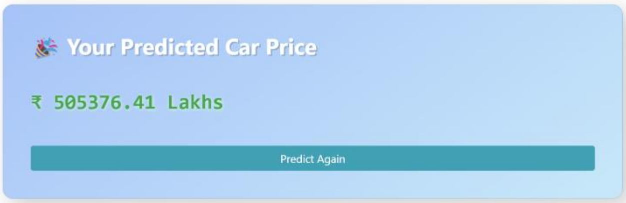
Select the company: Maruti

Select the model: Maruti Suzuki Baleno

Select Year of Purchase: 2018

Select the Fuel Type: Petrol

Kilometres Driven: 100000



Your Predicted Car Price

₹ 505376.41 Lakhs

Conclusion

The Car Price Predictor project successfully demonstrates the use of a **Linear Regression model** in solving a real-world problem. It highlights the importance of **data preprocessing**, **feature engineering**, and **model deployment** in building a complete machine learning solution. The web application can be effectively used by individuals or car dealerships to estimate fair resale values based on objective criteria.

References

1. **Quikr.com** – Source of the dataset used in the project. https://www.quikr.com
2. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
3. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html
4. Scikit-learn Documentation – Linear Regression. https://scikit-learn.org/stable/modules/linear_model.html#linear-regression
5. Flask Documentation – The Python Microframework for Building Web Applications. https://flask.palletsprojects.com/
6. Stack Overflow – Various threads related to model version compatibility and deployment troubleshooting.