**📄 Car Price Prediction – Full Project Report**

**🔷 Project Title: Car Price Prediction using Machine Learning**

**👨‍🎓 Author Details**

* **Name:** Arjav Jain
* **Course:** BCA
* **University:** Teerthanker Mahaveer University, Moradabad
* **Email:** [arjavjain062@gmail.com](mailto:arjavjain062@gmail.com)
* **GitHub:** [github.com/arjav527](https://github.com/arjav527)

**🎯 Objective**

The goal of this project is to build a machine learning model that predicts the **resale price** (selling price) of a used car based on various factors such as present price, car age, kilometers driven, fuel type, and more.

Predicting used car prices accurately is crucial for both **buyers** and **sellers** to make informed decisions and ensure fair deals.

**📂 Dataset Description**

* **Source:** Provided in CSV format
* **Total Records:** 301
* **Total Features (Columns):** 9

**Key Columns:**

| **Column Name** | **Description** |
| --- | --- |
| Car\_Name | Name of the car (categorical, dropped) |
| Year | Manufacturing year of the car |
| Selling\_Price | Target variable - resale price of the car |
| Present\_Price | Original price when car was bought new |
| Driven\_kms | Total distance driven |
| Fuel\_Type | Type of fuel (Petrol / Diesel / CNG) |
| Selling\_type | Dealer or Individual |
| Transmission | Manual or Automatic |
| Owner | Number of previous owners |

**🧹 Data Preprocessing**

* **Dropped Car\_Name** column as it is too specific and non-numeric.
* **Created new feature Car\_Age** = 2020 - Year.
* **Dropped Year** after extracting age.
* **Label Encoding** applied to Fuel\_Type, Selling\_type, Transmission.
* Checked and confirmed **no missing values** in the dataset.

**📊 Exploratory Data Analysis (EDA)**

**🔸 Distribution Observations:**

* Most cars are **petrol-based** and have **manual transmission**.
* Resale prices are **right-skewed** (most cars sell below ₹5 lakhs).
* Newer cars and those with lower mileage tend to sell at higher prices.

**🔸 Correlation Insights:**

* **Present\_Price** has strong positive correlation with **Selling\_Price**.
* **Car\_Age** and **Driven\_kms** have negative correlation with **Selling\_Price**.

**📈 Visualizations**

* **Correlation Heatmap**: Reveals feature relationships
* **Selling\_Price vs Present\_Price**: Strong upward trend
* **Car\_Age vs Selling\_Price**: Older cars sell for less
* **Actual vs Predicted Plot**: Random Forest fits closely
* **Residuals Distribution**: Random Forest has low and balanced error
* **Feature Importance Plot**: Present\_Price, Car\_Age, and Transmission most important

**🤖 Machine Learning Models**

| **Model** | **R² Score** | **Mean Absolute Error (MAE)** |
| --- | --- | --- |
| Linear Regression | ~0.85 | ~1.2 Lakhs |
| Random Forest | ~0.96 | ~0.5 Lakhs |

➡️ **Random Forest Regressor** was selected as the final model due to its superior accuracy and robustness.

**✅ Conclusion**

* The model effectively predicts the selling price of cars with high accuracy.
* The most impactful features are: Present\_Price, Car\_Age, Transmission.
* The project demonstrates a complete ML workflow: preprocessing, EDA, modeling, evaluation, and visualization.

**🔮 Future Scope**

* 🔧 **Deploy the model** via Streamlit or Flask to make it interactive.
* 📈 Add more features like brand, insurance, servicing history.
* 📷 Include image-based pricing using deep learning (CNN).
* 🧪 Tune the Random Forest with GridSearchCV for even better results.

**🗃️ Tools and Libraries Used**

* Python
* Pandas, NumPy
* Seaborn, Matplotlib
* Scikit-learn
* Jupyter Notebook

**📁 Folder Structure (GitHub Recommended)**

car-price-prediction/

│

├── Car\_Price\_Prediction.ipynb # Jupyter notebook

├── Car\_Price\_Prediction\_Report.pdf # Final report (optional)

├── requirements.txt # Python dependencies

├── README.md # GitHub documentation

└── dataset/

└── car\_data.csv

**👨‍💻 About the Author**

I'm Arjav Jain, a passionate data science student currently pursuing BCA. I enjoy building real-world machine learning projects and sharing them on GitHub and LinkedIn. Let's connect and grow together in the world of data and AI!