🚙 Used Car Price Prediction System 🚙

“ Car Valuation Tool ”

Submitted by –

NAME: Jubin Mazumdar

REGISTRATION NUMBER: 12316322

ROLL NUMBER: 43

SECTION: 9P172

DECLARATION

I, Jubin Mazumdar, student of B. Tech CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 24-02-2026

Registration No.: 12316322

Name of the Student Jubin Mazumdar

CONTENT

1. Introduction

The used car market has witnessed significant growth in recent years due to the increasing demand for affordable and cost-effective transportation. Many consumers prefer purchasing used cars instead of new ones, as they offer better value for money and lower depreciation costs. However, determining the correct price of a used car is a complex task, as it depends on multiple factors such as the brand, age of the vehicle, fuel type, mileage, engine capacity, condition, and market demand.

In the current scenario, there is no standardized method for evaluating the price of used cars. Sellers often rely on personal judgment or market trends, which may lead to overpricing or underpricing. Similarly, buyers may find it difficult to assess whether a car is fairly priced. This lack of transparency and consistency creates uncertainty in the buying and selling process.

With the advancement of data-driven technologies, machine learning has emerged as a powerful tool to analyze large datasets and identify patterns that influence pricing. By utilizing historical data and relevant features, machine learning models can predict the approximate market value of a used car with improved accuracy.

This project focuses on developing a machine learning-based system that predicts the selling price of used cars. It aims to provide a reliable, efficient, and user-friendly solution that assists both buyers and sellers in making informed decisions.

2. Objective

The objective of this project is to develop a machine learning-based system that predicts the selling price of used cars based on various features such as brand, fuel type, vehicle age, kilometers driven, engine capacity, and more.

The system helps users:

- Estimate the market value of their car
- Make better buying/selling decisions
- Get instant predictions via a web interface

3. Problem Statement

Determining the correct price of a used car is a complex task in the current market due to the presence of multiple influencing factors such as brand, vehicle age, mileage, fuel type, engine capacity, and overall condition. The absence of a standardized pricing mechanism leads to inconsistency and uncertainty in valuation. Sellers often rely on personal judgment or incomplete market knowledge, which may result in overpricing or underpricing of vehicles. Similarly, buyers face difficulty in evaluating whether a car is fairly priced, leading to a lack of transparency and trust in the transaction process.

Therefore, there is a need for a systematic and data-driven approach to accurately estimate used car prices.

This project aims to address this problem by:

- Analyzing historical used car data to identify key factors affecting price
- Applying machine learning techniques to build a predictive model
- Providing accurate and consistent price estimations
- Reducing dependency on manual judgment and guesswork

4. Dataset Description

- Dataset used: CarDekho Dataset

  https://www.kaggle.com/datasets/manishkr1754/cardekho-used-car-data

- Contains features such as:
  - Brand
  - Fuel Type
  - Transmission Type
  - Seller Type
  - Vehicle Age
  - Kilometers Driven
  - Mileage

- Engine (CC)
- Max Power (bhp)
- Number of Seats
- Selling Price (target variable)

- Preprocessing steps:
  - Removed unnecessary columns (car_name, Unnamed: 0)

- train_model
  - Applied log transformation on target variable to normalize skewed data
  - Handled categorical and numerical features separately

5. Technologies Used

- Python
- Pandas & NumPy → Data processing
- Scikit-learn → Model building
- Streamlit → Web application
- Pickle → Model saving/loading

6. Machine Learning Approach

- Feature Engineering
  - Categorical features encoded using OneHotEncoder
  - Numerical features used directly

- Model Used
  - Random Forest Regressor

Reason:

- Handles non-linear relationships
- Works well with mixed data types
- Reduces overfitting using multiple trees

- Training Process
  - Data split into:
    - 80% Training
    - 20% Testing
  - Model parameters:
    - n_estimators = 100
    - max_depth = 15

7. Model Evaluation Metrics

The model is evaluated using:

- R^2 Score → Accuracy of prediction
- MAE (Mean Absolute Error) → Average prediction error
- RMSE (Root Mean Squared Error) → Penalizes large errors

These metrics ensure the model performs well on unseen data.

- Model Results:

Fig 1. Model Performance

- Interpretation:
    - The R^2 score of 0.93 indicates that the model explains 93.42% of the variance, which means high accuracy.
    - The MAE (~₹1 lakh) shows the average prediction error is reasonable for car prices.
    - The RMSE (~₹2.2 lakh) indicates some larger errors but still acceptable for real-world data.

8. System Architecture

- Training Phase
    1. Load dataset
    2. Preprocess data
    3. Encode categorical features
    4. Train Random Forest model
    5. Save:
        - model.pkl
        - encoder.pkl
        - columns.pkl

- Prediction Phase
    1. User inputs data via UI
    2. Data is encoded
    3. Model predicts log price
    4. Converted back using exponential
    5. Display final price

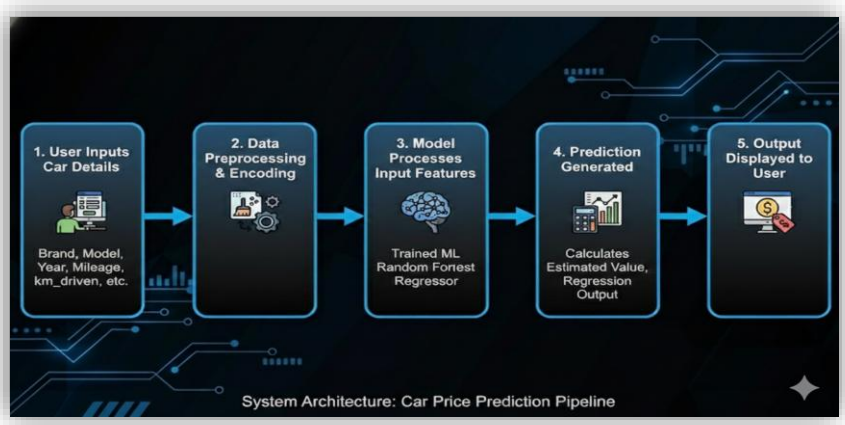

Fig 2. System Architecture

9. Web Application (Frontend)

The project includes a Streamlit-based UI app

Features:

- User-friendly interface
- Input fields:
  - Brand
  - Fuel Type
  - Transmission
  - Seller Type
  - Vehicle Age
  - KM Driven
  - Mileage
  - Engine
  - Max Power
  - Seats

- Button: "Check Value"
- Output: Predicted price in Indian format (Lakh / Crore)

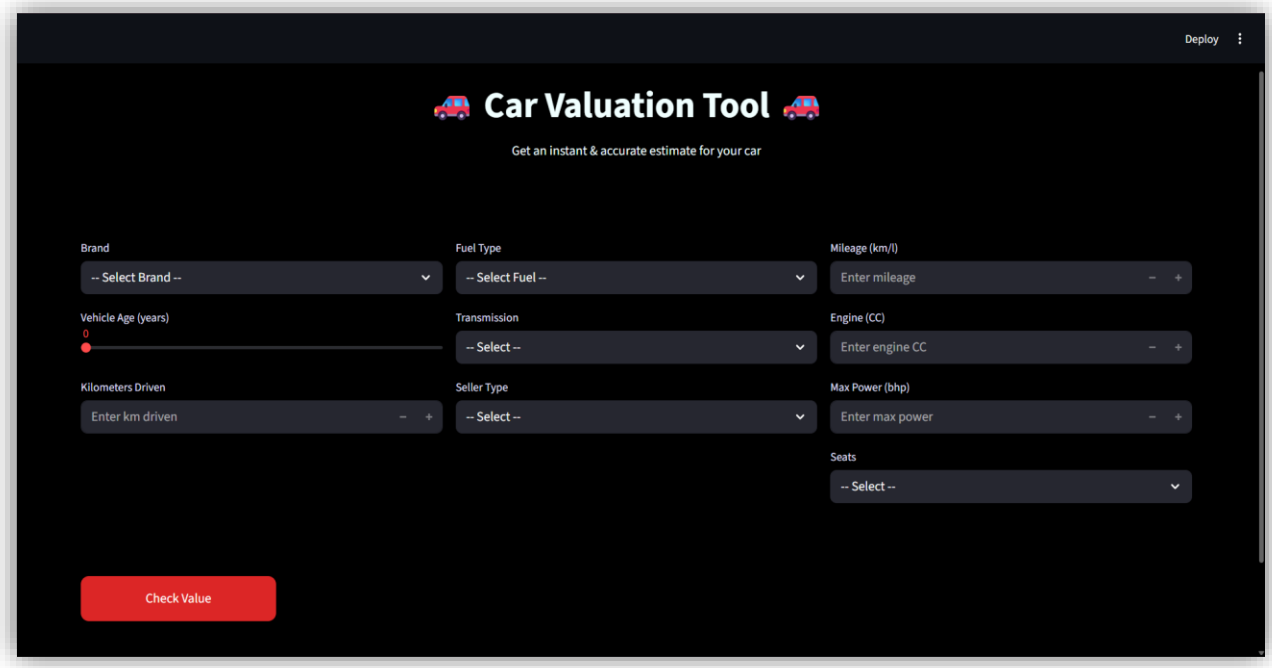

Fig 3. UI of the System

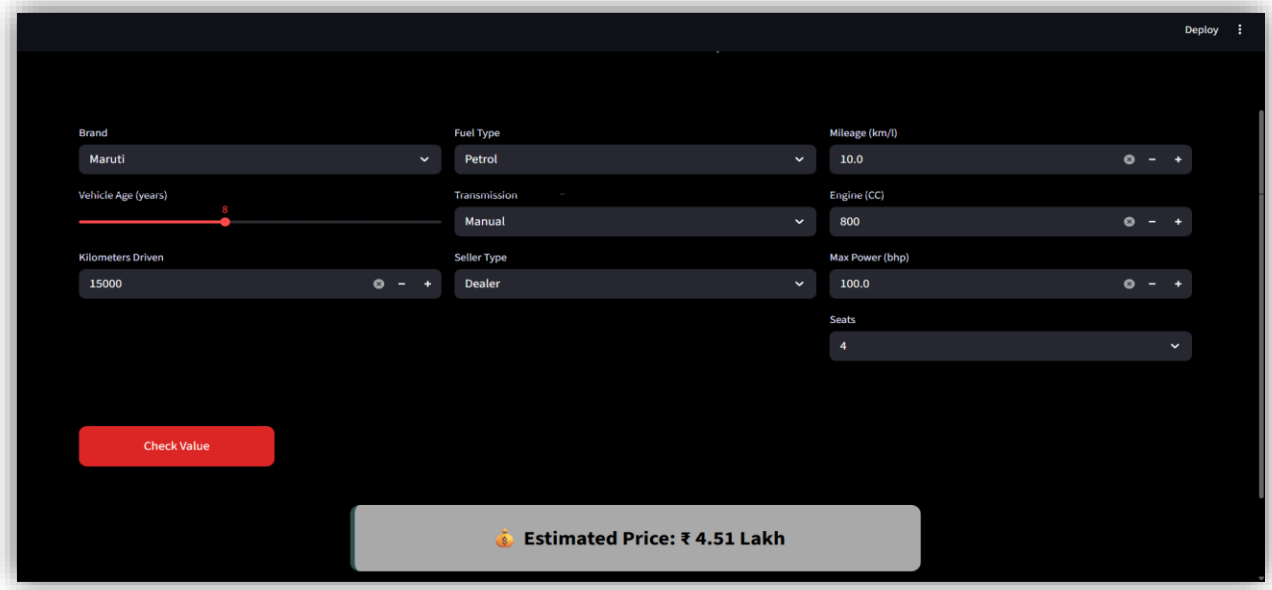

Fig 4. After the Prediction

10. Key Functionalities

- Real-time car price prediction
- Automatic feature encoding
- Log transformation handling
- Indian currency formatting (₹ Lakh / Crore)
- Error handling for missing files

11. Advantages

- Fast and accurate predictions
- Easy to use interface
- Helps both buyers and sellers
- Reduces manual price estimation errors

12. Limitations

- Limited dataset size
- Only supports selected brands
- Accuracy depends on data quality
- Does not include real-time market trends

13. Future Enhancements

- Add more brands and features
- Deploy on cloud (AWS / Heroku)
- Add image-based car evaluation
- Use advanced models (XGBoost, Deep Learning)
- Real-time data integration

14. Conclusion

This project successfully demonstrates the application of machine learning techniques in predicting used car prices with a good level of accuracy. By analyzing various features such as brand, vehicle age, fuel type, mileage, and engine specifications, the developed model can estimate the market value of a car in a reliable manner. The use of the Random Forest Regressor further enhances the prediction capability by handling complex and non-linear relationships within the data.

Additionally, the integration of the trained model with a Streamlit-based web application makes the system practical, interactive, and user-friendly. Users can easily input car details and obtain instant price predictions, which helps in making informed decisions.

Overall, this project highlights the effectiveness of data-driven approaches in solving real-world problems and provides a strong foundation for further improvements and real-time deployment in the used car market.