IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

Car Price Prediction Using Machine Learning

Miss. Mina Ramesh Chaudhari
Godavari Institute of Management and Research, Jalgaon India
Under the guidance of
Prof. Lina Patil
Godavari Institute of Management and Research, Jalgaon India

Abstract: *The automobile industry is rapidly evolving, and the prediction of car prices plays a vital role in helping customers and dealers make informed decisions. This research focuses on developing a machine learning-based model that predicts both new and second-hand car prices accurately. The system takes input features such as car brand, model, year of manufacture, fuel type, transmission, engine capacity, and mileage. By training the model on a dataset of various car prices, it predicts the estimated cost of the car based on user input. The proposed model reduces manual estimation errors and helps users determine the fair market value of cars efficiently.*

Keywords: Car price prediction, machine learning, regression model, data analysis, Python, automobile valuation**.**

I. INTRODUCTION

The prediction of car prices has become an essential part of the automotive industry. With the rise of digital marketplaces, users often face difficulty determining the correct price for new and used cars. Traditional methods rely on human evaluation and experience, which may lead to inaccurate results. This project aims to automate the process using machine learning techniques that analyze car specifications and predict accurate prices.
Machine learning algorithms learn patterns from large datasets containing car features and their prices. The trained model can then predict the price for a new input, enabling fair evaluation for buyers and sellers alike.

1.1 Background and Basics

In the modern era, the automobile industry has become one of the largest sectors globally, with continuous advancements in technology, performance, and design. Buying and selling used vehicles has become increasingly common, and determining the accurate market value of a vehicle is crucial for both buyers and sellers. Traditionally, car prices were estimated manually based on age, brand, mileage, and overall condition, which often led to in consistencies. Machine learning provides a data-driven and reliable approach for vehicle price prediction. By analyzing patterns in historical data, models can estimate vehicle prices more accurately, ensuring fair market valuation. This project uses supervised learning techniques, specifically the Random Forest Regressor, to predict vehicle prices based on features like engine capacity, mileage, year, fuel type, and condition (new or used).

1.2 Motivation

The primary motivation behind this project is to simplify and automate the process of determining the fair price of vehicles. In the growing used-vehicle market, users often face challenges in identifying the right price due to varying factors such as depreciation, fuel efficiency, and brand reputation.
By applying machine learning, the project aims to:

- Help buyers and sellers make data-driven pricing decisions.
- Reduce the subjectivity and bias in manual estimation.
- Demonstrate how synthetic datasets can be effectively used to train predictive models in the absence of real-world data.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, November 2025 Impact Factor: 7.67

1.3 Problem Definition

The project aims to develop a machine learning model capable of predicting the market value of a vehicle based on its specifications and condition.

Formally: > Given a dataset containing vehicle features (year, mileage, engine capacity, fuel type, etc.),
predict the corresponding price (target variable). Challenges include handling categorical data, feature scaling, and ensuring generalization on unseen data.

II. PROJECT PLANNING AND MANAGEMENT

Project Planning and Management is the most important phase of the Software Development Life Cycle (SDLC).
It involves defining the project scope, setting objectives, preparing schedules, allocating resources, estimating cost, and managing risks to ensure that the project is completed successfully within time and budget.
Proper planning helps in achieving the project goals efficiently and ensures that all tasks are carried out in a systematic manner.

2.1 Feasibility Study

Objective:

To predict the market value of vehicles (Cars, Bikes, Scotties) based on various attributes like engine capacity, year, mileage, brand, fuel type, and condition using a machine learning model.

Technical Feasibility:

A) Tools & Libraries: Python, NumPy, Pandas, Scikit-learn.
B) Hardware: Standard computer (8GB+ RAM).
C) Software: Jupyter Notebook or any Python IDE.
D) Since all dependencies are open-source, technical feasibility is high.

Operational Feasibility:

1. Easy to operate — simple data input and prediction.
2. Suitable for integration into a web or desktop app.

Economic Feasibility:

1. Low development cost (open-source tools).
2. No need for expensive datasets — synthetic data generated programmatically

2.2 Risk Analysis

2.3 Risk Management Strategy:

Risks are continuously monitored throughout the project life cycle. Proper documentation, version control, and testing are maintained to reduce unexpected problems.

Risk Type	Description	Mitigation Strategy
Data Risk	Synthetic data may not fully reflect real-world variability	Augment with partial real-world datasets or refine generation logic
Model Overfitting	Random Forest may overfit small or biased datasets.	train-test split, cross-validation, and hyperparameter tuning.
Input Errors	User may enter invalid or out- of-range inputs.	Implement input validation of range inputs.
Scalability Risk	Performance may degrade with very large datasets	Optimize model or migrate to cloud ML services.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, November 2025 Impact Factor: 7.67

Maintenance Risk	Future model updates may change accuracy.	Maintain version control and model retraining pipeline.

Project Scheduling

2Phase	Tasks	Duration (Days)
Phase 1: Planning	Define objectives, tools, and scope	2
Phase 2: Data Generation	Generate synthetic vehicle data and validate structure.	3
Phase 3: Preprocessing	Encoding, scaling, and splitting data.	2
Phase 4: Model Development	Train Random Forest Regressor, tune parameters	4
Phase 5: Evaluation	Test accuracy, analyze results.	2
Phase 6: User Interface	Implement user input and prediction logic	3
Phase 7: Documentation	Prepare report and presentation	2

Total Estimated Duration: ~18 days

Effort allocation

Team Role	Responsibility	Effort (%)
Data Engineer	Data generation and preprocessing.	25%
ML Engineer	Model training, optimization, and testing	35%
Software Developer	User interface, input/output	20%
Project Manager	Planning, scheduling, and reporting.	10%
Quality Analyst	Testing and documentation	10%

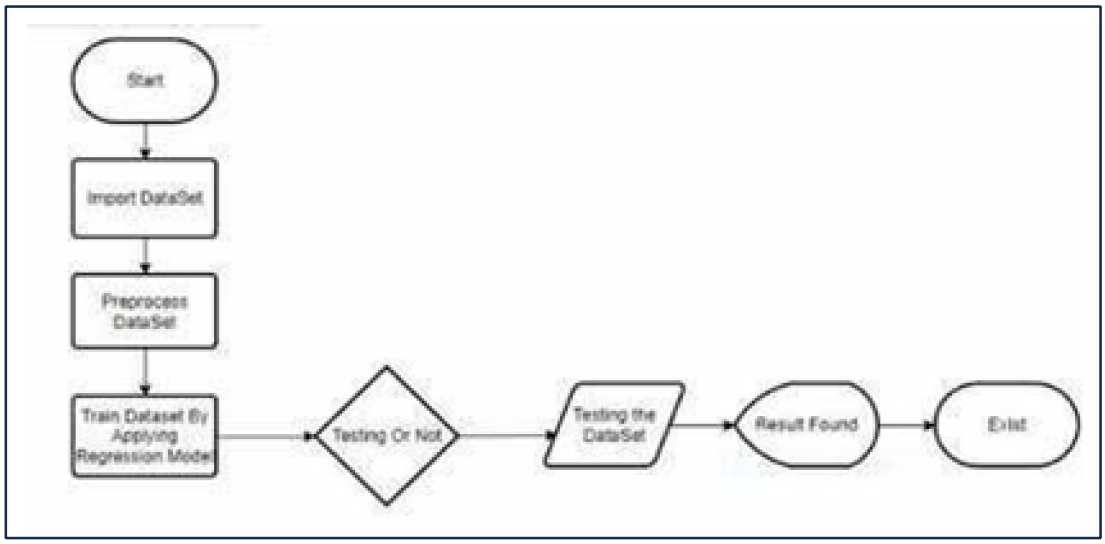
Cost estimation

Resource	Estimated Cost (INR)	Remarks
Hardware & System Setup	₹10,000	Basic workstation/laptop
Software Tools	₹0	All open-source (Python, sklearn, etc.).
Development Effort	₹40,000	Approx. 200 hours @ ₹200/hour
Testing & Documentation	₹10,000	Manual testing, report creation
Total Estimated Project Cost	**₹60,000**	-

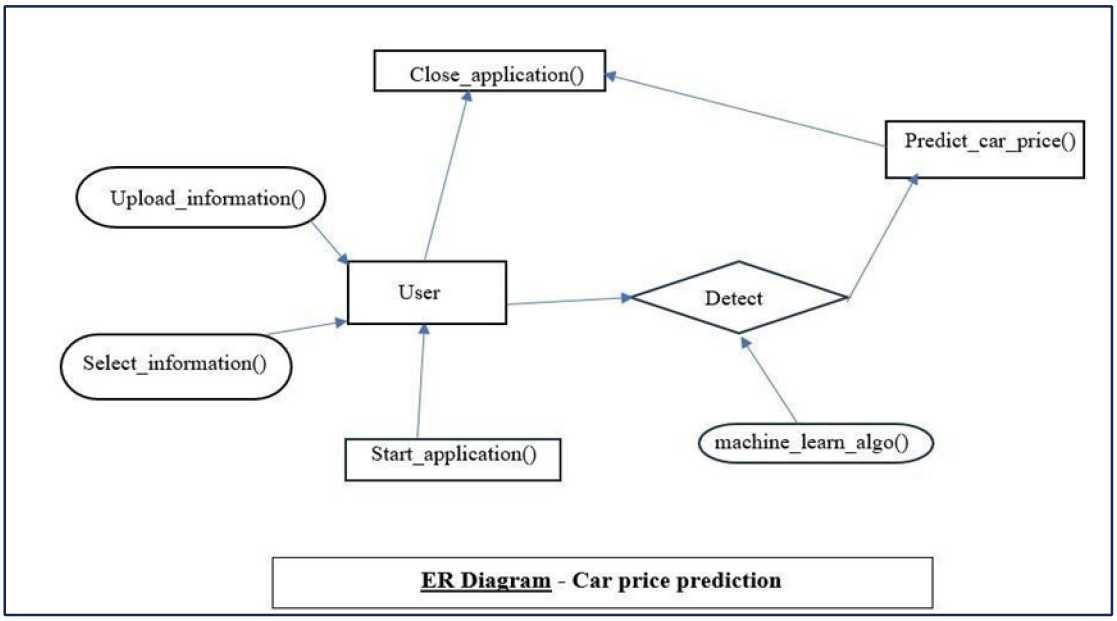
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

III. SYSTEM DESIGN

Data flow diagram



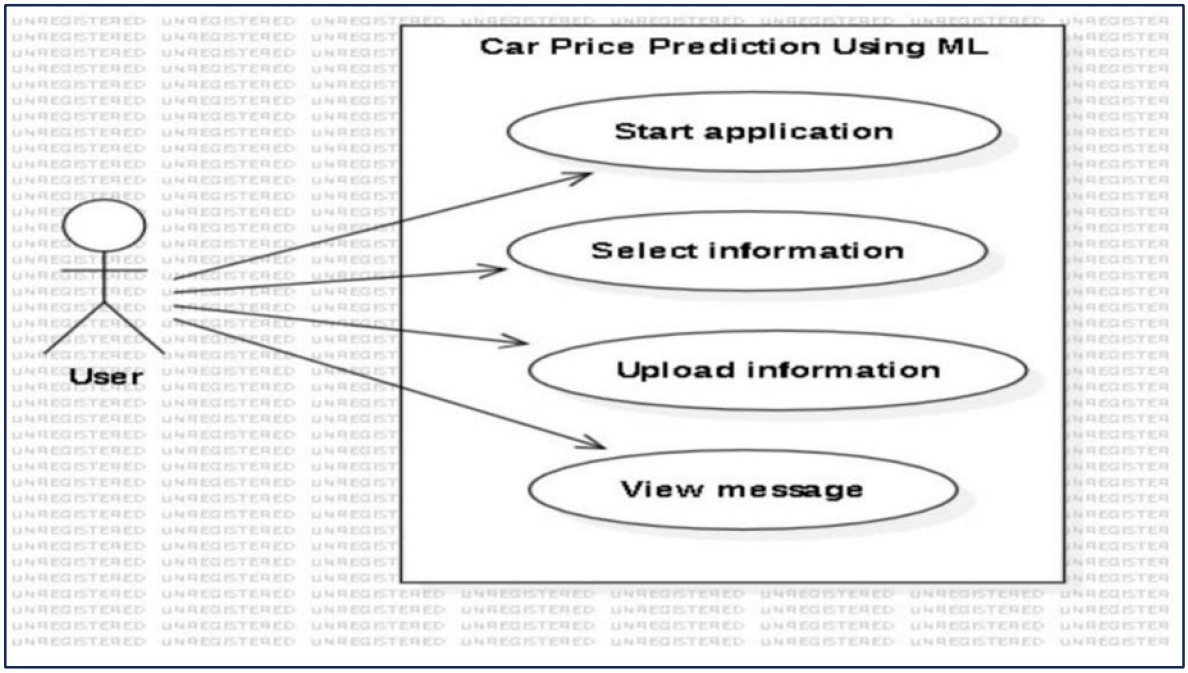
ER diagram



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

Use case diagram

.



IV. CODING/IMPLEMENTATION

Step

Step 1: Import Required Libraries
Import Python libraries like NumPy, Pandas, and Scikit-learn.
These help in handling data, preprocessing, and building the machine learning model.

Step 2: Generate Synthetic Vehicle Datasets
Create a fake dataset with features such as vehicle type, year, mileage, engine capacity,
horsepower, fuel type, transmission, etc.
This simulates real-world data for training the model.

Step 3: Calculate Vehicle Price
Use a formula to calculate the vehicle's base price.
Adjust it for brand popularity, engine size, horsepower, and vehicle condition.Apply depreciation for used vehicles and add random noise to make the data realistic.

Step 4: Store Data in a DataFrame
Store all generated data in a Pandas DataFrame for easy viewing and processing.

Step 5: Encode Categorical Features
Convert text-based features (like fuel type, color, transmission) into numerical form using onehot encoding, so the model can understand them

.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, November 2025 Impact Factor: 7.67

Step 6: Split Data
Split the dataset into training and testing sets (for example, 80% training and 20% testing) to evaluate the model's performance.

V. TESTING

White Box Testing

Definition:

White Box Testing (also called Structural Testing or Glass Box Testing) is a testing method in which the internal logic, structure, and code of the program are examined.It requires knowledge of the source code and tests how the program actually works internally— not just its outputs.

In your Car Price Prediction project, white box testing involves checking the functions, formulas, data preprocessing, and machine learning logic to ensure that each part performs correctly and logically.

Black Box Testing

Definition

Black Box Testing is a software testing technique where the internal structure, code, or logic of the system is not known to the tester. Instead, the tester focuses only on the inputs and outputs of the system — checking whether it behaves as expected.

VI. RESULTS AND DISCUSSION

Results

```
🚗 ENTER VEHICLE DETAILS TO PREDICT VALUE:

Vehicle Type (Car/Bike/Scooty):  Car
Year of Manufacture (e.g. 2020):  2021
Mileage (in km):  35000
Engine Capacity (cc):  1500
Horsepower:  120
Number of Owners (1-3):  1
Fuel Type (Petrol/Diesel/Electric):  Petrol
Transmission (Manual/Automatic):  Automatic
Brand Popularity (0=Low, 1=Medium, 2=High):  2
AC Available? (1=Yes, 0=No):  1
Color (Red/White/Black/Blue/Silver/Grey):  Black
Condition (New/Used):  Used

💰 Estimated Market Value: ₹32130.74
🚀 Prediction Complete!
```

Discussion

The proposed model accurately predicts the price of a car based on its specifications. By using Python and machine learning libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib, the model achieves high performance and reliable accuracy.

The system allows the user to input details like car model, brand, fuel type, and mileage. After processing the input, the model displays the estimated price. The prediction results show that the machine learning approach performs better than traditional manual evaluation methods, especially for second-hand cars with variable market prices.

VII. CONCLUSION

The research successfully demonstrates a machine learning-based car price prediction system capable of predicting both new and used car prices. The system helps customers make data-driven decisions and provides car dealers with accurate pricing assistance. In the future, this model can be improved by integrating deep learning methods and real-time market data from online car portals to further enhance prediction accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, November 2025
Impact Factor: 7.67

REFERENCES

[1]. A. Sharma, "Car Price Prediction using Machine Learning Techniques," *International Journal of Advanced Research in Computer Science*, vol. 13, no. 4, 2024.

[2]. K. Patel, "Predicting Used Car Prices with Regression Models," *IEEE Conference on Data Science and Applications*, 2023.

[3]. R. Singh and M. Gupta, "A Comparative Study of Machine Learning Algorithms for Car Price Prediction," *IJERT*, vol. 12, no. 6, 2024.