

DATA SCIENCE
BOOTCAMP
(GUVI - IITM)

CAR DHEKO – USED CAR PRICE PREDICTION

A presentation is a formal or
informal communication method



PROBLEM STATEMENT

Goal:

Buying or selling a used car is challenging due to non-standard pricing. Predicting a car's fair market value using machine learning helps:

- Buyers avoid overpaying
- Sellers quote competitive prices
- Dealerships automate pricing strategy



DATASET OVERVIEW



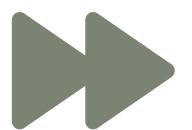
Source: Cleaned dataset with 20,000+ entries scraped from multiple marketplaces.

Key Fields:

- Categorical: brand, model, fuel_type, variant, body_type, gear_box, color, city
- Numerical: mileage, kms_driven, max_power, torque, engine_displacement, price
- Derived: car_age = 2025 - model_year

🛠 Data Cleaning Done:

- Missing values handled via median/mode
- Categorical standardization
- Redundant features removed



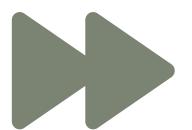
EDA & PREPROCESSING

Exploratory Data Analysis:

- Distribution plots for price, power, mileage
- Boxplots to detect and treat outliers using IQR method
- Correlation matrix to avoid multicollinearity

Preprocessing Steps:

- One-hot encoding for all categorical fields
- Feature engineering: car_age, normalized power
- Dataset split: 80% train / 20% test



MODEL BUILDING

Models Compared:

- Linear Regression, Ridge, Lasso
- Random Forest, Gradient Boosting
- **XGBoost (Final Choice)**

Why XGBoost?

- Best performance on all metrics:
 - $R^2 = 0.944$
 - $MAE = ₹98,500$
 - $RMSE = ₹2.34$ Lakhs
 - Cross-Val $R^2 = 0.912$

Final pipeline includes:

StandardScaler + XGBoost → saved as .pkl



Model	R^2	MAE (₹)	RMSE (₹)
XGBoost (Best)	0.944	98,500	234,800
Random Forest	0.937	111,000	250,000
Linear / Ridge	~0.89	167,000	331,000

XGBoost performs 20-40% better than linear models on test data.

In some outlier cases, actual price vs predicted price differs by 2L-3L – acceptable within complex used-car pricing.

STREAMLIT APP

Zen Class streamlitapp localhost:8503

GUVI | Learn to cod... Zen Class Python Full course... Excel Basics for Beg... pandas · PyPI Adobe Acrobat Solve Python | Hack... TCS iON National Q... NN SVG deep learni... vgg16 -architecture

Deploy :

Used Car Price Prediction - Car Dheko

Brand: Audi

Model: Audi A3

Variant: 35 TDI Premium

Fuel Type: Cng

Body Type: Convertibles

Transmission: Automatic

Gear Box:

Model Year: 2018

Registration Year: 2018

Engine Displacement (cc): 1200

Mileage (km/l): 18.00

Max Power (bhp): 85.00

Torque (Nm): 100.00

Next

STREAMLIT APP

Zen Class streamlitapp localhost:8503

Steering Type: EPAS

Owner Number: 1

Kilometers Driven: 30000

Predict Price

Selected Inputs Summary

Estimated Car Price: ₹ 804,627

Deploy :

This image shows a Streamlit application running in a web browser. The title bar indicates the app is titled 'streamlitapp' and is running on 'localhost:8503'. The main interface consists of several input fields: 'Steering Type' (dropdown, selected 'EPAS'), 'Owner Number' (slider, value 1), 'Kilometers Driven' (input field, value 30000), and a 'Predict Price' button. Below these inputs is a 'Selected Inputs Summary' section, which is currently empty. At the bottom, a green box displays the 'Estimated Car Price: ₹ 804,627'. A 'Deploy' button with a three-dot menu icon is located in the top right corner of the app area.

Features:

- Dropdowns with dynamic filtering (brand → model → variant)
- Numeric inputs for specs like torque, bhp, mileage
- Real-time prediction on submit
- Expandable section for reviewing inputs
- Success banner with predicted price



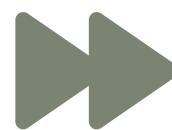
CONCLUSION & FUTURE SCOPE

What Worked:

- XGBoost performed consistently well
- Real-time app gives interactive predictions
- Streamlined pipeline allows easy deployment

Improvements & Future Scope:

- Integrate live scraped data for training refresh
- Use text-based reviews for more sentiment-based pricing
- Add image classification to adjust for car condition
- Deploy on AWS/GCP with CI/CD





THANK YOU

