End-to-End Car Price Prediction

Author: Abdelrahman Solii

Date: 2025-12-07

Project Abstract & Introduction

This project aims to predict car resale prices using machine learning, addressing the used car market's inconsistent pricing and lack of transparency. We develop a model that accurately estimates car prices and includes a web interface for real-time predictions.

Automate Estimation

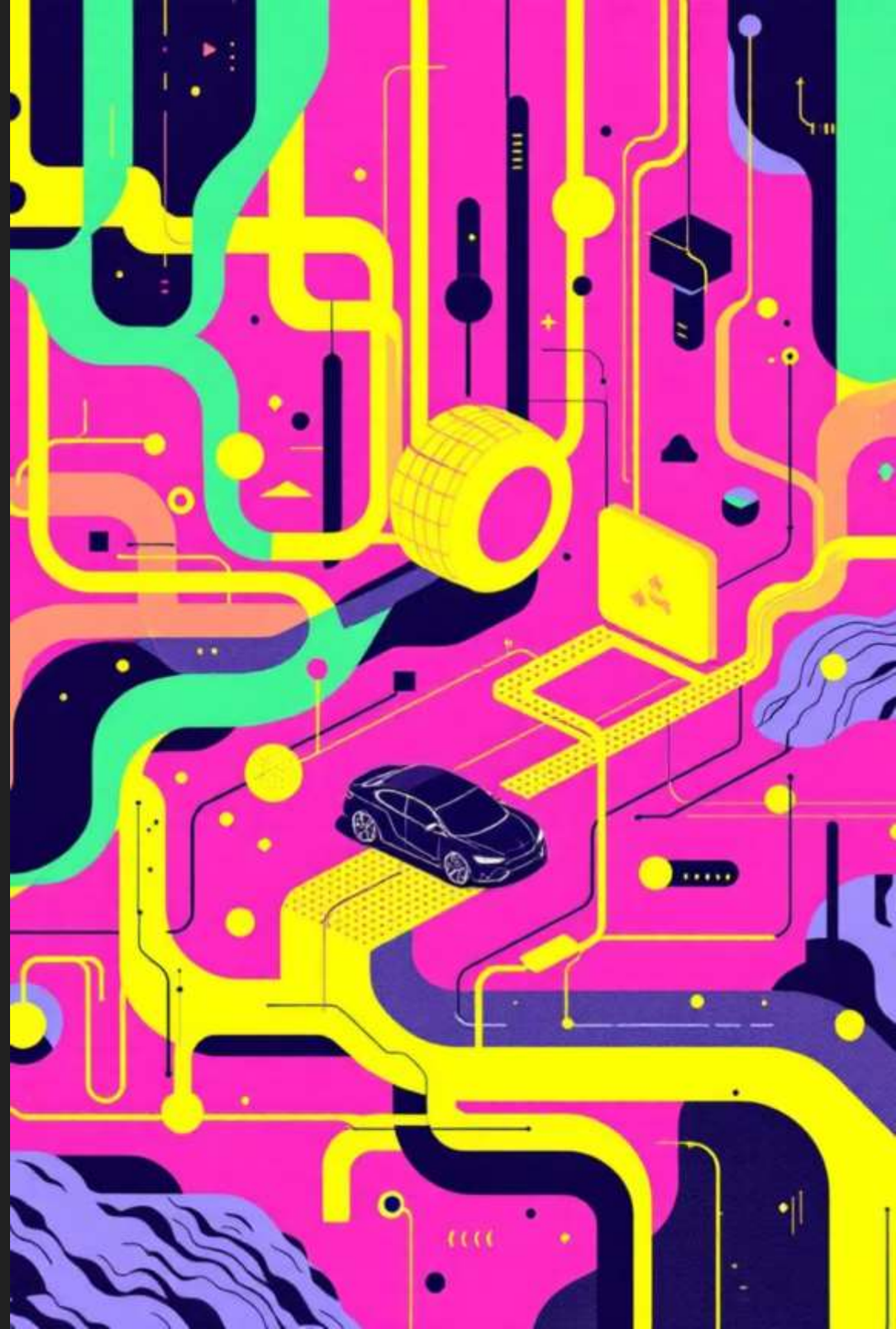
Provide accurate and reliable price predictions.

Analyze Datasets

Understand features impacting car prices.

Deploy Model

User-friendly web interface for predictions.



Dataset Overview

Our dataset contains car features crucial for price prediction, collected from online listings and historical sales data.

manufacturer	Car brand (e.g., Toyota, Hyundai)
model	Car model name
mileage	Distance driven in km
engine_volume	Engine size in liters
car_age	Age in years
price	Selling price (target variable)

Data Preprocessing Steps

Before modeling, data underwent rigorous preprocessing to ensure accuracy and consistency.

01

Missing Values

Handled using mean/mode imputation.

02

Duplicate Records

Removed to avoid bias in the model.

03

Feature Engineering

Calculated `car_age` and `price_per_mileage`.

04

Categorical Encoding

Used LabelEncoder for various features.

05

Train/Test Split

80% training, 20% testing for model validation.

Exploratory Data Analysis (EDA)

Key observations from data visualization revealed significant correlations impacting car prices.

1

Price Distribution

Average car price: $18,559, median: $13,172.

2

Manufacturer Impact

BMW, Mercedes show higher mean prices.

3

Mileage vs. Price

Negative correlation; higher mileage means lower price.

4

Engine & Power

Positive correlation with price.

5

Fuel Type & Category

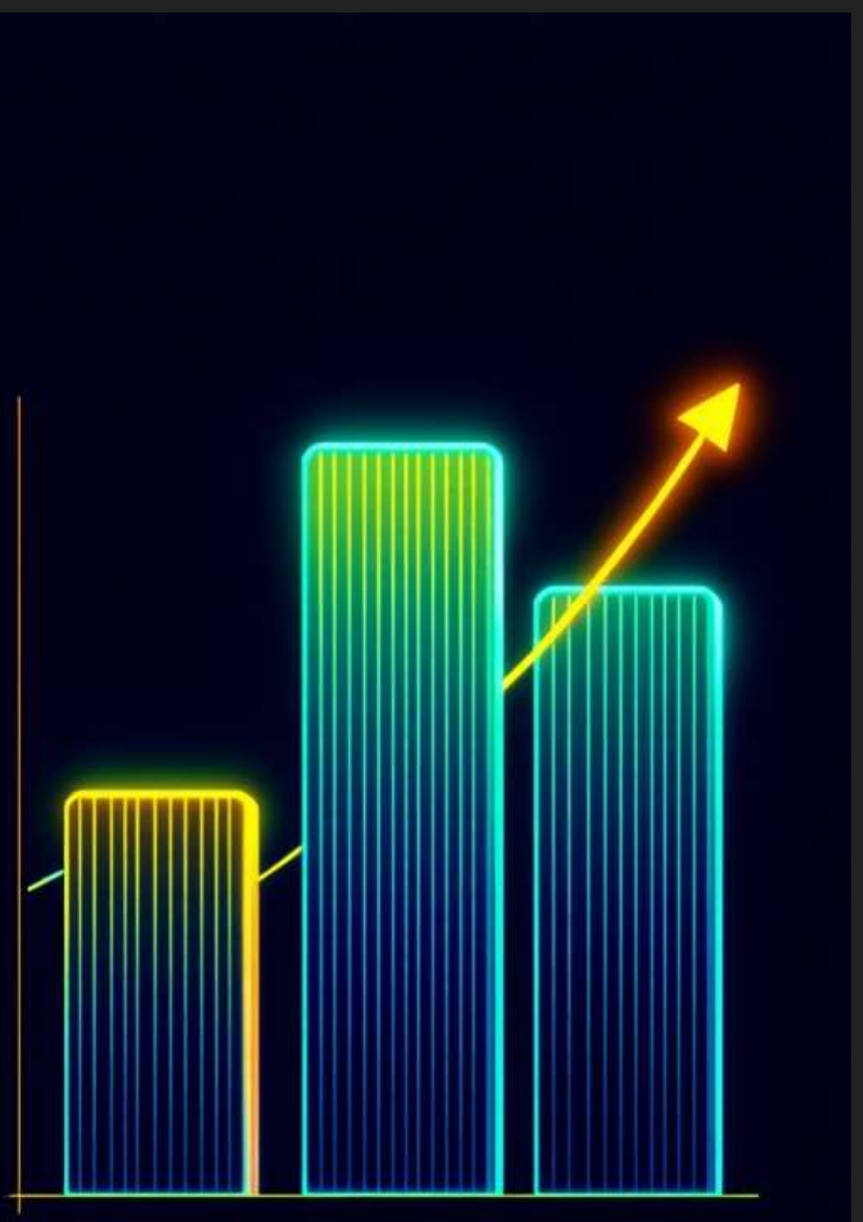
SUVs and diesel cars tend to be more expensive.

Model Selection & Training

Several regression models were tested, with CatBoost Regressor emerging as the top performer.

CatBoost Regressor	0.9998	166.75
XGBoost Regressor	0.9982	469.53

The CatBoost Regressor was selected for its superior accuracy and stability, trained with default parameters and early stopping.



Model Evaluation

The CatBoost model demonstrated exceptional performance on the test set.

0.998	166.75	941.57
R^2 Score	RMSE	RMSE
Indicates excellent fit to data.	Low error in price prediction.	Another measure of prediction accuracy.

Top features influencing price include mileage, engine volume, car age, and manufacturer/model mean prices.

Deployment Strategy

The trained model is deployed for real-time predictions using Streamlit and Gunicorn, ensuring accessibility and scalability.

Install Dependencies

```
pip install -r requirements.tx
```

Run Locally

```
python car_pred_price.py
```

Via Gunicorn

gunicorn car_pred_price:app --bind 0.0.0.0:5000

This application predicts car prices based on various features. Fill in all the details on the right and click "Predict Price".

Predict Price

About

This model was trained on historical car data to predict market prices.

The prediction is based on features like manufacturer, model, mileage, age, and other specifications.

Deploy

Car Price Prediction

Classic Cars Collection

Modern Luxury Cars

Enter Car Details

Manufacturer: TOYOTA

Model: RX 450

Category: Sedan

Leather interior: Yes

Drive wheels: Front

Wheel: Left wheel

Color: Black

Engine volume: 0.8

Fuel type: Petrol

Gear box type: Automatic

Cylinders: 2

Airbags: 5

Car Age (years): 5

Levy: 500

Mileage (km): 50000

Engine Power: 100.00

Additional Features

Price per Mileage (auto-calculated): 50.00

Model Mean Price: 25000

Brand Mean Price: 20000

Doors: 04-May

View Input Data

	manufacturer	model	category	leather_interior	fuel_type	mileage	gear_box_type	drive_wheels	wheel	color	levy	engine_volume	cylinders	airbags	car_age
0	49	890	9	1	5	50000	0	1	0	1	500	0.8	2	5	5

Predict Price

About

This model was trained on historical car data to predict market prices.

The prediction is based on features like manufacturer, model, mileage, age, and other specifications.

Prediction Result

Prediction Result

PREDICTED CAR PRICE

$45,659

Based on provided specifications

Car Age	Mileage	Engine Power
5 years	50,000 km	100.0 HP

Note: This prediction is based on historical market data. Actual price may vary based on condition, location, and market trends.

Conclusion & Future Work

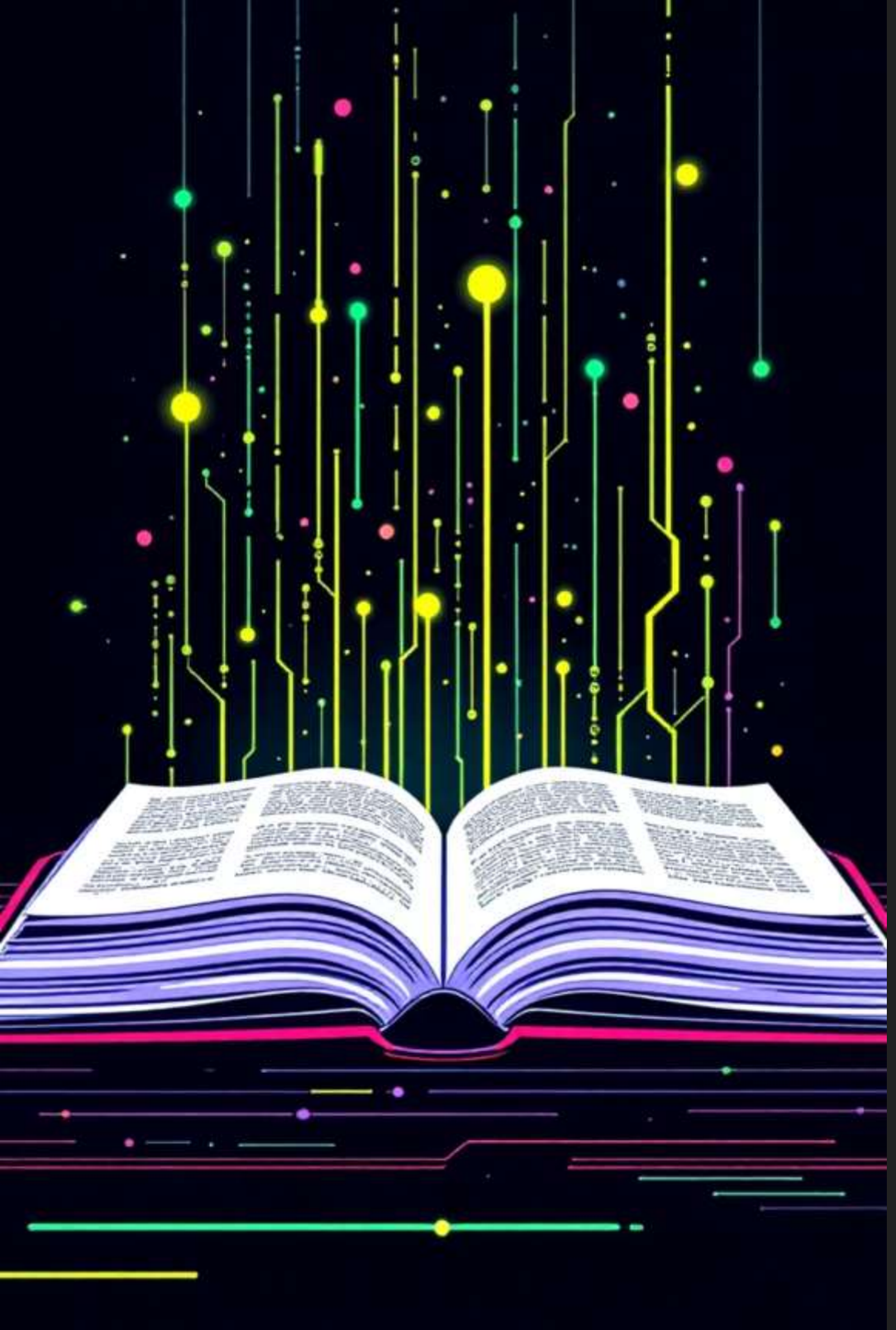
The CatBoost model provides highly accurate car price predictions, driven by key factors like mileage and engine volume.

Conclusion

- Highly accurate predictions with CatBoost.
- Key factors: mileage, engine volume, car age, manufacturer/model averages.

Future Work

- Incorporate location, demand trends, accident history.
- Develop a full web platform with user accounts.
- Explore ensemble models for enhanced accuracy.



References

Further information and resources can be found through these references:

- **Scikit-learn Documentation**

  https://scikit-learn.org/

- **CatBoost Documentation**

  https://catboost.ai/

- **Used Car Price Prediction Examples**

  Analytics Vidhya