

# Predictive Modeling and Interactive Visualization Dashboard for Car Price Forecasting

## Author

**Sanjana Rajesh Pisal**

master's in data Analytics

University of Illinois Springfield

Guided by: Prof. Liang Kong

## Project Overview

This project aims to build a machine learning system to predict used car prices using structured data, and to deliver meaningful insights through a Power BI dashboard.

The end goal is to assist car buyers, sellers, and analysts by offering:

- Accurate car price predictions
- Visual exploration of pricing patterns and key influencers
- A business-ready, interpretable model selection pipeline

## Problem Statement

Traditional pricing methods are either manual or rule-based and fail to capture the complexity of modern car features. The goal here is:

- To build a robust, scalable ML model for predicting car prices.
- To complement this with an **interactive dashboard** for insights and decision support.

## Dataset

- Source: [Kaggle - Car Price Prediction Dataset](#)
- Rows: 205
- Features: 26 (including brand, fuel type, engine size, horsepower, etc.)
- Target: price

## Machine Learning Models Used

Six regression models were trained and compared:

1. **Linear Regression**
2. **Ridge Regression**
3. **Lasso Regression**
4. **Decision Tree Regressor**
5. **Random Forest Regressor - *Best Performer***
6. **Gradient Boosting Regressor**

### Evaluation Metrics:

- $R^2$  Score
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Cross-validation score (CV)

A **composite scoring system** was used to objectively compare model performance.

### Best Model: Random Forest Regressor

- **Test  $R^2$ :** ~0.96
- **Cross-Validation Score:** ~0.89
- **MAE (Top 30 samples):** 1356.65
- Chosen for its accuracy, generalization, and ability to handle outliers and feature interactions.

### Power BI Dashboard

- Created to visualize car price trends and model output
- Features:
  - Brand-wise price distribution

- Filters: Fuel type, body type, drive wheels
- Donut & bar charts for average prices
- Line plots to show trend by brand and year

### Technologies Used

- **Python (Jupyter Notebook)**
  - pandas, numpy, sklearn, seaborn, matplotlib
- **Machine Learning**
  - Regression Models (Linear, Tree-based, Ensemble)
- **Power BI**
  - For dashboard and dynamic visualizations

### Results & Conclusion

- **Random Forest** outperformed all models.
- Created a **fully interpretable dashboard** to aid real-world business decision-making.
- Delivered a complete ML pipeline, from data loading to prediction and visualization.

### References

- [Kaggle Dataset](#)

Dashboard: -

