

# Car price Prediction System

Group A



Business Understanding

Data Understanding

Data Preparation

Modelling

Evaluation

Deployment

# Business Understanding

Project Domain: Automotive Secondary Market

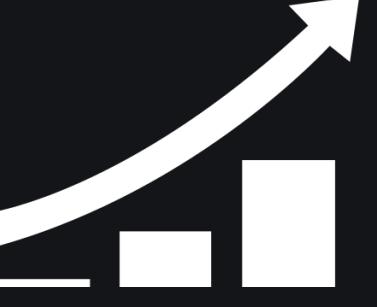
- Focused on Used Car Pricing & Valuation.

ML Relevance

- Why ML? The relationship between car features (Mileage, Age, Engine Type) and Price is complex and dynamic.

It is a Supervised Regression Task (Predicting a continuous numerical value: Price).

# Data Understanding



## Collect, Describe, and Explore Data.

Identify how the car's characteristics (features) translate into its market Price (Target Variable).

## Feature Relevance

Intrinsic Value : Company Name, model name, engine capacity

Depreciation : car age, mileage

Market Factors : location, assembly, registration status, color

## Key Data Challenges

High Cardinality: Need efficient handling for variables with many unique values (e.g., model\_name, location).

Categorical Data: categorical columns require encoding (e.g., One-Hot or Label Encoding).

# Data Preparation

## Data Cleaning (No Issues Found)

- **Missing Values:** No null values were detected across any of the features.
- **Duplicates:** No duplicate rows were found, ensuring unique data points.

## Data Formatting (Categorical Encoding)

features that are text-based (e.g., `body_type`, `color`).

One-Hot Encoding (OHE) was applied to all categorical columns.

## Feature Engineering

**Problem:** The `model_year` column is not directly useful for depreciation models.

**Solution:** Created a new, highly relevant feature: `car_age` (Current Year - Model Year).

**Result:** The original `model_year` column was deleted, simplifying the dataset while retaining critical information.

• Good when you want feature selection and also want to reduce overfitting.

# Modelling

Problem Type: Regression

- Goal: Predict a continuous numerical value (Car Price).

## Linear Regression

establishes a baseline performance.

## Lasso Regression

Like Ridge but can remove unimportant features by setting their coefficients to zero.

## ElasticNet

Builds decision trees one after another, each one fixing the mistakes of the previous one.

## KNN Regressor

Predicts price based on similarity to other cars in the dataset.

## Ridge Regression

Same as linear regression but adds a penalty to prevent overfitting.

## Random Forest

Reduces variance by averaging multiple trees).

## Gradient Boosting

Builds decision trees one after another, each one fixing the mistakes of the previous one.

Machine Learning Algorithms

# Evaluation

MAE, MSE, RMAE, R2, ACCURACY

## MODEL EVALUATION RESULTS:

	Model	MAE	MSE	RMSE	R2	\
2	Random Forest	224559.010583	4.440858e+11	6.663976e+05	0.950078	
3	Gradient Boosting	254983.673685	4.556314e+11	6.750048e+05	0.948780	
4	KNN	223609.016839	6.422327e+11	8.013942e+05	0.927803	
1	Ridge Regression	593482.567472	2.717167e+12	1.648383e+06	0.694550	
0	Linear Regression	588597.700386	2.722484e+12	1.649995e+06	0.693952	
6	Lasso	588689.869311	2.724997e+12	1.650756e+06	0.693669	
5	ElasticNet	651567.287795	4.702536e+12	2.168533e+06	0.471364	

## Accuracy

2	87.698746
3	84.692364
4	88.782757
1	56.365382
0	56.717274
6	56.716195
5	61.567201

# Deployment

## Streamlit

Python framework that lets you turn your machine-learning code into a real, interactive web app — with almost zero web-development skills.

### Car Price Prediction App

Enter Car Details

Company Name	Model Name
Adam	3
Engine Type	Body Type
Diesel	Cross Over
Transmission Type	Location
Automatic	Balochistan
Mileage (km)	Engine Capacity (cc)
50000	1300
Assembly	Registration Status
Imported	Registered

# conclusion

This project successfully demonstrated the end-to-end application of the CRISP-DM methodology. We began with the business challenge of subjective car pricing and delivered a highly accurate, deployed solution powered by the Random Forest Regressor.

we will use real-time user feedback from the Streamlit app to refine our data, incorporate new features, and ensure our car price predictions remain accurate and best-in-class."

# Thank you