Car Selling Price Prediction – Final Report

# 1. Objective

The goal was to build a predictive model that accurately estimates the selling price of cars based on their attributes such as name, km\_driven, year, fuel type, transmission, owner, max\_power, etc.

# 2. Process Overview

## a. Data Exploration & Preprocessing (EDA)

- Performed EDA using visual and statistical summaries.  
- Identified and removed irrelevant rows (e.g., test drives with no owners).  
- Detected outliers and skewed distributions (e.g., selling price).  
- Log-transformed target variable to normalize distribution.

## b. Feature Engineering

- Categorized features into numerical and categorical.  
- Applied label encoding and one-hot encoding where appropriate.  
- Scaled numerical features using standardization.

## c. Model Building

Multiple regression models were implemented and compared:  
- Linear Regression  
- Decision Tree Regressor  
- Random Forest Regressor  
- XGBoost  
- Support Vector Regressor (SVR)  
- LightGBM  
- CatBoost

## d. Hyperparameter Tuning

Used RandomizedSearchCV for:  
- Random Forest Regressor

- CatBoost

- LightGBM

- XGBoost

# 3. Evaluation Metrics Used

- R² Score  
- MAE (Mean Absolute Error)  
- MSE (Mean Squared Error)  
- RMSE (Root Mean Squared Error)

# 4. Best Model & Accuracy

The best-performing model was CatBoost, achieving the highest R² and lowest error metrics after tuning.  
A screenshot of a test results

AI-generated content may be incorrect.

# 5. Role of LLMs in the Workflow

- Exploratory Data Analysis (EDA): Used LLM to interpret patterns, suggest visuals, and guide transformation (e.g., log-transform for skewed targets).  
- Feature Engineering: LLM helped identify which variables needed encoding/scaling and created reusable code blocks for transformation pipelines.  
- Reusable Code Generation: Quickly generated loops, pipelines, and visualizations using prompts.  
- Model Evaluation: LLM helped write reusable evaluation functions and interpret results in a concise manner.

# 6. Results and Prediction



