Laboratory Practice 6 Mini Project Report BI

Title: Used Car Price Prediction

Group Members
- Nimish Agarwal (41)
- Yash Aher (42)
- Chetan Dabbe (43)
- Shreekrushna Dongare (44)

Problem Definition

Background:
The used car market is growing rapidly. With various factors like car brand, model, manufacturing year, mileage, fuel type, transmission type, etc., determining the price of a used car becomes challenging. Buyers want fair prices, while sellers aim for maximum profit. Predicting used car prices accurately helps both parties and can significantly enhance business operations for car dealers.

Problem Statement:
To develop a system that can predict the price of a used car based on various attributes using data mining techniques.

Objective:
To apply data mining methods to identify the relationship between car features and their resale value.
To build a predictive model that estimates the price of used cars.
To analyze the dataset for business insights that can be used in pricing strategy.

Required Data Mining Task:
Prediction (Supervised Learning): This task involves predicting the numerical value (used car price) based on historical labeled data.
Regression Analysis is the appropriate data mining technique for this task.

Dataset Features:
- year – Manufacturing year of the car
- manufacturer – Car brand (e.g., Toyota, Ford)
- model – Specific model name
- condition – Condition of the car (e.g., excellent, good)
- cylinders – Number of engine cylinders
- fuel – Fuel type (e.g., gas, diesel, electric)
- odometer – Total distance traveled
- transmission – Transmission type (automatic/manual)
- drive – Drive type (e.g., 4wd, fwd)
- type – Type of car (SUV, sedan, etc.)
- paint_color – Color of the vehicle
- price – Target variable (price of the car)

Implementation
Tools & Technologies:
Programming Language: Python
Libraries Used: pandas, numpy, matplotlib, seaborn, scikit-learn

Steps:
Data Cleaning: Removed null values and outliers.
Feature Engineering: Converted categorical data using Label Encoding / One-Hot Encoding.
Data Splitting: 80% training, 20% testing.

Model Used: Linear Regression, Random Forest Regressor, and XGBoost Regressor.
Evaluation Metrics: R^2 Score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

Results and Testing

Model	R^2 Score	MAE	RMSE
Linear Regression	0.68	2387	3265
Random Forest Regressor	0.82	1784	2411
XGBoost Regressor	0.85	1652	2254

Conclusion

The project successfully demonstrates the application of data mining for predictive analytics. We found that advanced regression models like XGBoost provide high accuracy in predicting used car prices. Such models can be used by online car resale platforms or dealerships to suggest fair prices to users based on car features.