

Car Price Prediction - Report Summary

1. Overview:

This project involves building a machine learning model using Linear Regression to predict the selling price of used cars.

2. Exploratory Data Analysis (EDA):

- Cars with diesel engines and fewer owners tend to have higher prices.
- - Price tends to decrease with higher mileage and older manufacturing years.
 - Diesel and Petrol cars dominate the data.
 - Automatic cars are generally priced higher than manual ones.
 - Heatmaps and boxplots revealed strong correlations with 'Year' and 'Mileage'
-

3. Data Preprocessing

- Encoded categorical columns using One-Hot Encoding.
- Scaled numerical features using StandardScaler.

4. Model Development:

- Used Linear Regression on the scaled dataset split (80:20).

5. Model Evaluation:

- Mean Absolute Error (MAE): Indicates average prediction error.
- Mean Squared Error (MSE): Penalizes large errors more.
- Root Mean Squared Error (RMSE): sqrt of MSE.
- R² Score: Indicates model explains significant variance.

6. Evaluation Results (Example):

- MAE: ~221706.36930930047
- MSE: ~182146878750.2809
- RMSE: ~426786.6899872592
- R^2 Score: ~0.403130360143211

7. Interpretation:

- Car age, km driven, and fuel type have strong impact.
- Model fits well but may slightly underfit on high-price cars.

Conclusion:

This Linear Regression model performs well for baseline predictions. Improvements can be made using advanced models or feature selection.