

Car Price Prediction – Project Report

Overview

In this project, I built a car price prediction model using Linear Regression. The idea was to predict the selling price of a used car based on various factors like brand, year, mileage, fuel type, etc. This is a common real-world problem, and solving it helped me understand the complete flow of a machine learning pipeline—from data cleaning to model evaluation.

1. Dataset and Preprocessing

The dataset had features like:

- Car Brand & Model
- Year of Manufacture
- Mileage
- Fuel Type
- Transmission
- Owner Type
- Location
- Selling Price (target variable)

I handled missing values by dropping them (as they were not many). Then, I used Label Encoding for the categorical columns and scaled the numeric ones using StandardScaler. Finally, I split the data into training and testing sets in an 80:20 ratio.

2. Exploratory Data Analysis (EDA)

- I checked the distribution of the target variable (Selling Price), and it was right-skewed.
- Heatmaps showed that Year and Mileage had strong correlations with price.
- Boxplots made it clear that fuel type and transmission also had an effect on price.

This step helped me understand which features were more important.

3. Model Building

I used the Linear Regression model from scikit-learn. After training it on the training data, I predicted prices for the test set.

4. Evaluation Metrics

Here are the evaluation results:

- **MAE (Mean Absolute Error):** around 35,000
 - **MSE (Mean Squared Error):** quite large due to outliers
 - **RMSE (Root Mean Squared Error):** more interpretable than MSE
 - **R² Score:** around 0.78 (shows a good fit overall)
-

5. Interpretation

- Features like **Year of Manufacture**, **Mileage**, and **Fuel Type** had a strong impact.
 - Since Linear Regression is a basic model, it worked okay, but I noticed slight underfitting in some cases—probably because of outliers or non-linear patterns.
 - I think trying other models like Random Forest or XGBoost could give better results.
-

Conclusion

This project gave me hands-on experience with EDA, preprocessing, model training, and evaluation. It also helped me understand where Linear Regression works well and where it doesn't. In future improvements, I'd try more advanced models and look into removing outliers for better accuracy.