**CAR PRICE PREDICTION USING MACHINE LEARNING**

Internship Project

Domain: Data Science

Company: CodeAlpha

Submitted By: Aiswarya k

Internship Duration: Oct 1 – Oct 31

Submitted To: CodeAlpha

**1. Introduction**  
Car price prediction is an important application of machine learning in the automotive industry. The selling price of a car depends on multiple factors such as brand goodwill, features, mileage, horsepower, and more. This project focuses on building a predictive model to estimate the selling price of used cars based on historical data.

**2. Objective**- To predict the selling price of used cars accurately.  
- To analyze the key factors affecting car prices.  
- To implement a machine learning model for regression analysis.

**3. Dataset Description**The dataset contains historical data of cars including the following features:

Feature | Description | Type  
--------|-------------|-----  
Car\_Name | Name/brand of the car | Categorical  
Year | Year of manufacture | Numerical  
Selling\_Price | Selling price (Target) | Numerical  
Present\_Price | Original price | Numerical  
Driven\_kms | Distance driven | Numerical  
Fuel\_Type | Petrol/Diesel/CNG | Categorical  
Seller\_Type | Dealer/Individual | Categorical  
Transmission | Manual/Automatic | Categorical  
Owner | Number of previous owners | Numerical

Note: Bikes and motorcycles were filtered out for this project to focus on cars only.

**4. Data Preprocessing**- Removed bikes and motorcycles from the dataset.  
- Handled missing values by removing null entries.  
- Encoded categorical variables using one-hot encoding.  
- Created a new feature `Car\_Age = 2025 - Year`.  
- Dropped original columns `Year` and `Car\_Name` after feature engineering.

**5. Exploratory Data Analysis (EDA)**  
- Distribution of `Selling\_Price` was visualized using a histogram.  
- Correlation heatmap revealed strong correlation between `Present\_Price`, `Car\_Age`, and `Selling\_Price`.  
- Feature importance analysis showed that `Present\_Price`, `Car\_Age`, and `Driven\_kms` are most influential in predicting car prices.

**6. Model Building**- Split the dataset into training (80%) and testing (20%) sets.  
- Standardized features using `StandardScaler` (optional for Random Forest).  
- Trained a Random Forest Regressor with 200 trees (`n\_estimators=200`).

**7. Model Evaluation**Metrics on the test set:  
- Mean Absolute Error (MAE): X.XX  
- Root Mean Squared Error (RMSE): X.XX  
- R² Score: X.XX

Feature importance plot shows that `Present\_Price` and `Car\_Age` have the highest influence on car price predictions.

**8. Conclusion**- Successfully built a Random Forest model to predict used car prices.  
- Found that original price, car age, and kilometers driven are the most important factors.  
- The model can help car sellers and buyers make informed decisions.

**9. Future Work**- Test other regression models such as XGBoost or Gradient Boosting for improved accuracy.  
- Hyperparameter tuning to optimize model performance.  
- Incorporate more features like horsepower, number of airbags, brand reputation, etc.

**10. References**1. CodeAlpha Internship Guidelines  
2. Kaggle – Used Car Dataset  
3. Géron, A., Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd Edition