**Analyzing Selling Price of used Cars**

**Introduction:**

**1.** In the modern era, the second-hand car market has grown rapidly. Consumers

often look for used cars that offer value for money. However, estimating a car's

correct resale value is not always straightforward due to various influencing factors

such as brand, age, fuel type, kilometers driven, etc.

**2.** This project aims to analyze a dataset of used cars and predict the selling price

based on key attributes using data analysis and machine learning techniques in

Python.

**Objectives:**

**1.** To clean and preprocess the used car dataset.

**2.** To perform exploratory data analysis (EDA) to understand important factors

affecting the selling price.

**3.** To build regression models to predict the selling price of used cars.

**4.** To evaluate model performance and determine the best-performing algorithm.

**Methodology:**

**1. Data Collection**

A dataset containing used car listings will be collected (e.g., from Kaggle or an open

dataset platform).

**The dataset typically includes features like:**

Car Name / Brand

Year

Present Price

Kilometers Driven

Fuel Type

Seller Type

Transmission

Owner

Selling Price

**2.Data Preprocessing**

Handle missing values and duplicates.

Convert categorical variables into numerical form (label encoding / one-hot

encoding).

Feature engineering (e.g., calculating car age).

**3. Exploratory Data Analysis (EDA)**

Visualize feature relationships using Matplotlib and Seaborn.

Analyze correlation of variables with the selling price.

Detect and handle outliers.

**4. Model Building**

Split the data into training and test sets.

Apply regression algorithms such as:

Linear Regression

Decision Tree Regressor

Random Forest Regressor

Tune hyperparameters using GridSearchCV or RandomizedSearchCV.

**5. Model Evaluation**

Evaluate model accuracy using metrics such as:

R² Score

Mean Absolute Error (MAE)

Mean Squared Error (MSE)

Compare model performance and select the best one.

**Tools & Technologies Used:**

**Tool/Library: Purpose:**

**Python** Programming Language

**Pandas, NumPy** Data Manipulation

**Matplotlib, Seaborn** Data Visualization

**Scikit-learn** ML Model Building and Evaluation

**Jupyter Notebook** Development Environment

**Expected Outcome:**

**1.** A working Python-based tool to analyze and predict the selling price of used cars.

**2.** Insights into key factors influencing car prices.

**3.** Trained machine learning model with satisfactory accuracy for real-world

Prediction.

**Advanced Features:**

**1. Feature Importance Analysis**

Use models like Random Forest or XGBoost to extract and rank the most important

features affecting selling price.

**2. Model Explainability with SHAP / LIME**

Apply SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model

agnostic Explanations) to explain how each feature contributes to individual

predictions.

**3. Cross-Validation & Hyperparameter Tuning**

Implement k-fold cross-validation for more robust evaluation.

Use GridSearchCV or RandomizedSearchCV to fine-tune models and improve

accuracy.

**4. Interactive Web App using Streamlit / Flask**

Build an interactive UI where users can input car features and get the predicted

selling price.

**5. Data Versioning & Pipeline Management**

Use MLFlow or DVC (Data Version Control) to track experiments, model versions,

and datasets.