Automobile Data Analysis And Prediction

By

M Saad Rehman

2024-MSDS-132

Project Overview

- Demonstrates a machine learning workflow from data preprocessing to model evaluation.
- Focuses on EDA, feature engineering, and performance comparison of multiple models.

File Structure

- Jupyter Notebook with the entire workflow.
- Includes code, analysis, and visualizations.

Data Loading and Cleaning

- Load datasets from CSV files.
- Analyse data
- Feature Engineering of Data

Exploratory Data Analysis (EDA)

- Statistical summaries of features.
- Data visualization with:
  - **Matplotlib** for plotting.
  - **Seaborn** for advanced visualizations.

Feature Engineering

- Encode categorical variables.
- Normalize or scale numerical features to improve model performance.

Model Training

- Machine learning models implemented:
  - **Linear Regression**
  - **Decision Trees**
  - **Random Forests**
- Models built using **scikit-learn**.

Model Evaluation

- Metrics used for comparison:
  - Accuracy, Precision, Recall, F1-score.
- Visual tools:
  - Confusion matrices.

Results & Acknowledgments

- **Results**:
  - Key insights from data visualization.
  - Strengths of each model highlighted.
- Acknowledgments:
  - Contributors and open-source libraries (Pandas, NumPy, etc.).

Thank You