Team Members

Group 11

1) Umut Bayar 150120043
   Mail = umutbayar@marun.edu.tr
2) Ahmet Arda Nalbant 150121004
   Mail = ardanalbant@marun.edu.tr
3) Niyazi Ozan Ateş 150121991
   Mail= nozan@marun.edu.tr
4) Kadir Pekdemir 150121069
   Mail= kadirpekdemir@marun.edu.tr
5) Murat Albayrak 150120025
   Mail= muratalbayrak@marun.edu.tr

Github Link= https://github.com/ardanlbnt/CSE4288F24_GrpX

Title: Vehicle Resale Price Prediction Using Machine Learning

Description:
This project aims to predict the selling price of used vehicles using a machine learning approach. Leveraging historical vehicle data, including attributes like make, model, mileage, and condition, we will create a predictive model to assist sellers and buyers in accurately estimating vehicle resale values.

Problem Statement and Objectives

Problem Statement:
The used car market is vast, with prices varying widely based on vehicle attributes. Buyers and sellers often struggle to estimate a fair price, leading to inefficiencies and potential dissatisfaction. This project aims to address this problem by developing a predictive model that estimates vehicle resale values based on key attributes.

Objectives:

1. To analyze and preprocess the given dataset for relevant insights.
2. To build and train a machine learning model capable of predicting the selling price of a vehicle.
3. To evaluate the performance of the model using metrics such as RMSE (Root Mean Squared Error) and R^2 score.
4. To deploy the model for real-world use, allowing users to input vehicle attributes and receive price estimates.

Dataset Description and Source

Dataset Description:

- **Attributes:** The dataset contains vehicle attributes like year, make, model, trim, body type, transmission type, condition, odometer reading, color, interior, seller details, market value (MMR), selling price, and sale date.
- **Sample Size:** Multiple rows of vehicle records.
- **Source:** https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data

Planned Methodology and Tools

Methodology:

1. **Data Cleaning and Preprocessing:** Handle missing values, outliers, and categorical data encoding.
2. **Exploratory Data Analysis (EDA):** Identify trends and relationships between attributes and selling prices.
3. **Feature Engineering:** Create new features or modify existing ones for better prediction.
4. **Model Selection:** Train and test different machine learning models such as Linear Regression, Random Forest, and XGBoost to identify the best-performing one.
5. **Evaluation:** Use metrics like RMSE and R^2 to measure the model's performance.
6. **Deployment:** Create a web-based interface for users to estimate vehicle prices using the model.

Tools:

- Python for programming.
- Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn.
- Jupyter Notebooks or Google Colab for development.

Team Roles and Responsibilities

For the preprocessing phase, two of our team members are working on it, while the rest of us are brainstorming ideas for model creation and other phases. Everyone contributes ideas and continuing with the work.

Proposed Methodologies and Timeline

Methodologies:
The project will adopt a data-driven approach to build a machine learning model for price prediction. The process includes:

1. Understanding data trends via EDA.
2. Developing a pipeline for preprocessing and feature selection.
3. Iteratively training and testing machine learning models to achieve optimal performance.

Timeline:

- **Week 1: Project Initiation**

- **Week 2: Data Preprocessing and EDA**

- **Week 3: Model Development**

- **Week 4: Model Evaluation and Optimization**

- **Week 5: Finalization and Presentation Preparation**