Farah Makkawi

In this project, I was responsible for the **data preparation, exploratory data analysis, and Decision Tree model training** using the Electric Vehicle Population dataset.

My contributions include:

- Loading the dataset into Google Colab
- Cleaning and preparing the data by selecting relevant columns and removing missing values
- Performing exploratory data analysis (EDA) using simple visualizations to understand electric range and vehicle type distributions
- Encoding categorical variables to make the data suitable for machine learning
- Splitting the dataset into training and testing sets
- Training a Decision Tree classification model
- Evaluating the model using accuracy
- Predicting class probabilities for the test data

Noah

Short Interpretation of the Classification Report

For every type of electric vehicle, the model's precision, recall, and F1-score are displayed in the classification report. Strong memory and precision for that category are demonstrated by the Decision Tree's ability to identify the majority class. Nonetheless, the minority class performs worse, indicating that there may be some class imbalance in the sample. The model's overall accuracy shows that it accurately identifies the majority

of EVs; nevertheless, class balancing strategies or more sophisticated models could be used in the future to enhance forecasts for underrepresented classes.

Decision Tree Interpretation

The model's classification views based on vehicle features are displayed in the decision tree visualization. The most major breaks occur at the top of the tree, demonstrating that features such as electric range and car manufacturer play a key influence in determining whether a vehicle is a BEV or PHEV. More characteristics refine the choice as the tree branches downward, resulting in more precise classifications. This demonstrates that the model's primary source of information for differentiating between EV kinds is electric range.

https://docs.google.com/document/d/1x6ivo0XSuuIJwv1qJ6Jcltc4RrKuizu8Jv8_muPQDkA/edit?usp=sharing

Project Overview

This project uses the Electric Vehicle Population dataset to build a Decision Tree classification model. The goal of the project is to explore electric vehicle data, understand important features through exploratory data analysis, and use machine learning to predict the type of electric vehicle based on several characteristics.

Dataset

The dataset used in this project is the Electric Vehicle Population Data, which contains information about registered electric vehicles, including model year, manufacturer, electric range, base MSRP, and vehicle type. The dataset includes well over 1,000 rows, meeting the project requirements.

Research Questions

- Can vehicle features such as model year, electric range, and MSRP be used to predict the type of electric vehicle?
- Which features are most important when classifying electric vehicle types?

Methods

The project was completed using Python in Google Colab and includes the following steps:

- Data loading and cleaning using Pandas and NumPy
- Exploratory Data Analysis (EDA) using visualizations
- Encoding categorical variables for machine learning
- Splitting the data into training and testing sets
- Training a Decision Tree classification model using scikit-learn
- Evaluating the model using accuracy and class probability predictions
- Visualizing the trained Decision Tree

Tools and Libraries

- Python
- Google Colab
- Pandas
- NumPy
- Matplotlib
- scikit-learn