DSML Capstone Project: Final Report

Project Title

Electric Vehicle Population Analysis & Classification

Objective

To analyze and predict whether a vehicle is a Battery Electric Vehicle (BEV) or Plug-in Hybrid Electric Vehicle (PHEV) using features such as make, model year, electric range, and MSRP.

Dataset

Washington State Electric Vehicle Population dataset (publicly available).

Phase 1: Problem Definition

- Defined the classification goal
- Selected key attributes relevant to predicting EV type

Phase 2: Data Cleaning (Excel)

- Removed duplicates and missing values
- Standardized text columns
- Converted categorical fields to consistent format

Phase 3: Visualization (Power BI)

- Developed a dashboard with KPIs, charts, and interactive slicers
- Insights into vehicle types by region, range, and manufacturer

Phase 4: Model Building (Python)

- Trained Logistic Regression and Decision Tree models
- Preprocessed data and split into training/testing sets
- Evaluated using accuracy, confusion matrix, and classification report

Performance Evaluation

- Logistic Regression Accuracy: (based on run)
- Decision Tree Accuracy: (based on run)
- Visualized confusion matrix and model comparison

Visualizations Included

- Distribution of electric range
- Class count (BEV vs PHEV)
- Correlation heatmap
- Top manufacturers by range
- Accuracy comparison of models

Tools Used

Excel, Power BI, Python (pandas, sklearn, seaborn, matplotlib), Jupyter Notebook

Conclusion

A Decision Tree model performed well in classifying EV types. The dataset offered rich insights into the characteristics of BEVs vs PHEVs. This project demonstrated a full DSML pipeline from cleaning and visualization to modeling and evaluation.