Evaluating High-Voltage Battery Performance in the BMW i3 (60 Ah)

Group 8

2024/12/17

Project Overview

- Objective
  - Analyze real-world data to predict and optimize battery performance by focusing on:
    a) **Environmental factors**
    b) **Vehicle dynamics**
    c) **Heating circuit efficiency**
- Related Variables
  - Key Features
    a) **Environmental Data:** Temperature, Elevation
    b) **Vehicle Data:** Speed, Throttle Position
    c) **Heating Circuit Data:** Indoor Temp, Heating Power
  - Target Outcome: Predict the **State of Charge (SoC) Difference** (Battery Consumption)
- Approach
  - **Data Preprocessing:** Cleaning and transforming features for modeling
  - **Modeling:** Use **Linear Models** and **Tree-Based Models**
  - **Performance Comparison:** Evaluate models for actionable insights

Model Performance Summary

1. **Linear Models**
   1) Linear Regression: **Training MSE** = 0.0004 | **Test MSE** = 0.0006 | **R^2** = 0.7768
   2) Lasso Regression: **Training MSE** = 0.0014 | **Test MSE** = 0.0003 | **R^2** = 0.8797
   3) Ridge Regression: **Training MSE** = 0.0005 | **Test MSE** = 0.0004 | **R^2** = 0.8644
2. **Tree-Based Models**
   1) Decision Tree:
      a) **Training MSE** = 0.0000 | **Test MSE** = 0.0064 | **R^2** = -0.4816
      b) **Feature Importance: Distance [km]** = 0.924999
   2) Random Forest:
      a) **Training MSE** = 0.0002 | **Test MSE** = 0.0018 | **R^2** = 0.5764
      b) **Feature Importance: Distance [km]** = 0.816695
3. Key Insights

- **Linear Methods** clearly outperform tree-based methods: **Lower Test MSE** (< 0.0006) and **higher R^2** (> 77%)
- Data exhibits strong **linear relationships** between predictors and response
- **Distance** is the most influential feature in both model types, aligning with theoretical expectations

Analysis Results Summary

1. **SOC Variability:**

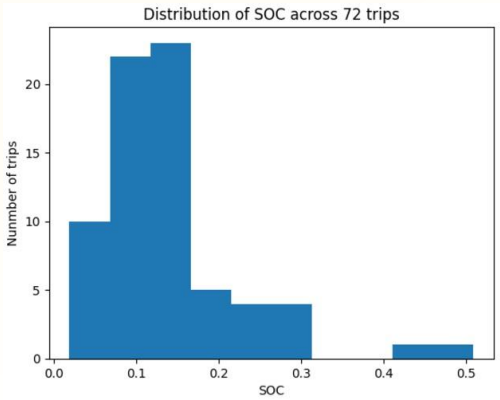
   The chart (on the left) shows that the variable of interest, SOC, has significant variations, allowing for a meaningful prediction exercise using the features available in the data.

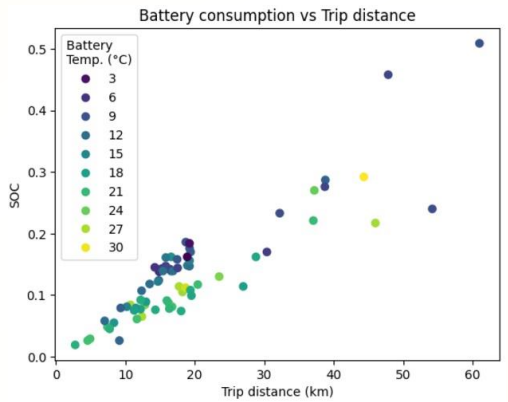
2. **Distance and SOC Relationship:**

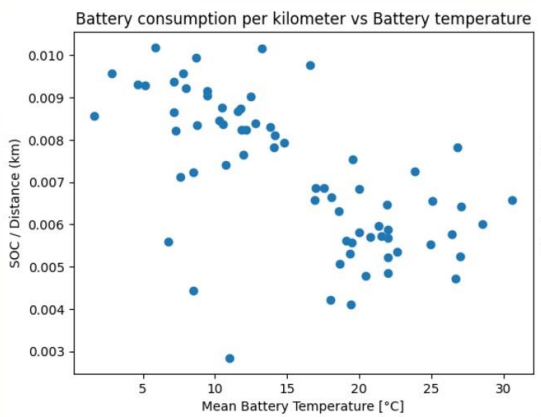
   The chart (second from the left) demonstrates a linear relationship between distance and battery consumption, with each additional mile requiring a similar amount of energy (SOC/km). This supports the earlier finding that linear models outperform tree-based models, with distance being the most important feature (0.92 feature importance).

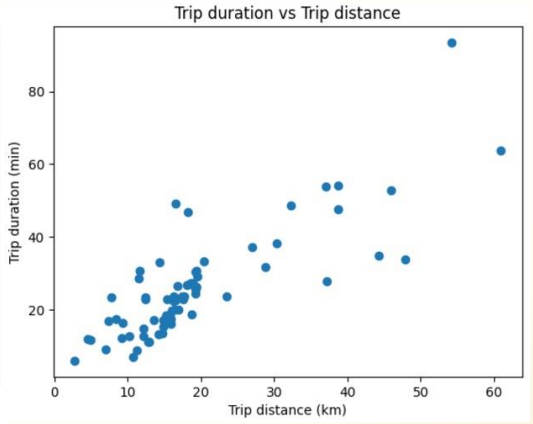
3. **Duration's Limited Importance:**

   Finally, while trip duration should influence SOC, feature importance analysis (second from the right) shows its importance is minimal (0.02). This is due to high correlation between distance and duration, with distance already capturing most of the relationship with SOC (importance > 0.92, the right one).









Acknowledgment of Team Members - Group 8

Daniel I Lin
UID: 121033345
Email: dlin0315@umd.edu

Sheng-Hsiang Huang
UID: 121303554
Email: shhuang@umd.edu

Hao-Lin Chiang
UID: 121166357
Email: beas28@umd.edu

Alexios C Papazoglou
UID: 121332967
Email: apapazog@umd.edu

Yang-Shun Lin
UID: 121287762
Email: yslin227@umd.edu

Kai-Wei Hsu
UID: 121366191
Email: kwhsu@umd.edu

Thank you for your attention!
Group 8 - University of Maryland