

**University of Maryland**

**Analysis of High-Voltage Battery Performance and Auxiliary Energy Consumption in Battery Electric Vehicles**

Version 1.0

September 15, 2024

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Version** | **Status** | **Prepared by** | **Comments** |
| September 15, 2024 | 1.0 | Initial Draft | Yang-Shun Lin | First draft for review |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Table of Contents**

[**1. Purpose** 4](#_Toc177337195)

[**2. Membership** 4](#_Toc177337196)

[**3. Project Progress** 5](#_Toc177337197)

[**3.1 Data Cleaning and Preprocessing Steps** 5](#_Toc177337198)

1. **Purpose**

The objective of this analysis is to evaluate the performance of the high-voltage battery in a BMW i3 (60 Ah) by examining real-world driving trip data from Kaggle[[1]](#footnote-1). The analysis will focus on understanding how various features affect battery performance. Specifically, the study will investigate the following features using the provided data set:

1. Environmental Data: The impact of temperature and elevation on battery performance.
2. Vehicle Data: The influence of speed and throttle position on battery efficiency and consumption.
3. Battery Data: The relationships between battery voltage, current, temperature, and State of Charge (SoC).
4. Heating Circuit Data: The effect of indoor temperature and heating power on battery performance and vehicle range.

The goal is to analyze the relationships between these features and battery performance to identify which features are most significant in affecting the high-voltage batteries of electric vehicles. By targeting the most influential features, the analysis aims to provide insights for improving battery management, optimizing vehicle range, and enhancing overall efficiency. This involves using a linear regression model to predict the working temperature, as well as employing multiple models to make predictions and compare their effectiveness.

1. **Membership**

|  |  |  |
| --- | --- | --- |
| **Name** | **UID** | **UMD Email** |
| Daniel I Lin | 121033345 | [dlin0315@umd.edu](mailto:dlin0315@umd.edu) |
| Hao-Lin Chiang | 121166357 | [beas28@umd.edu](mailto:beas28@umd.edu) |
| Yang-Shun Lin | 121287762 | [yslin227@umd.edu](mailto:yslin227@umd.edu) |
| Sheng-Hsiang Huang | 121303554 | [shhuang@umd.edu](mailto:shhuang@umd.edu) |
| Alexios C Papazoglou | 121332967 | [apapazog@umd.edu](mailto:apapazog@umd.edu) |
| Kai-Wei Hsu | 121366191 | [kwhsu@umd.edu](mailto:kwhsu@umd.edu) |

1. **Project Progress**

**3.1 Data Cleaning and Preprocessing Steps**

1. Check Imbalance Data:

* Oversampling/Undersampling: Applying these techniques is a standard approach for handling class imbalance.
* Synthetic Data Generation: SMOTE (Synthetic Minority Over-sampling Technique) and ADASYN (Adaptive Synthetic Sampling Approach) are effective methods for creating synthetic samples of the minority class.

1. Missing Values:

* Use methods like df.isnull().sum() in pandas to identify missing values. Make sure to check for missing values in all relevant columns.
* Imputation/Removal: Depending on the nature and extent of the missing data, use imputation techniques (such as mean, median, mode, or more advanced methods) or remove rows/columns with excessive missing values. Consider the impact of these actions on the dataset and the model.

1. Data Types:

* Verification: Check data types using methods like df.dtypes in pandas. Ensure numerical values are not mistakenly stored as strings and categorical data is appropriately encoded.
* Adjustment: Convert data types where necessary. For example, convert date columns from strings to datetime objects and ensure categorical variables are treated as such.

1. Outliers:

* Identification: Use methods such as statistical tests, visualization (box plots, scatter plots), or IQR (Interquartile Range) to detect outliers.
* Handling: Choose a strategy based on analysis. Removal is straightforward but may result in data loss. Capping can limit outlier values, while transformation can reduce their impact.

1. Distributions Between Datasets:

* Comparison: Use visualizations (histograms, KDE (Key Data Element) plots) or statistical tests to compare distributions between datasets. This helps identify any significant differences.
* Seasonal Effects: Address seasonal effects by segmenting data or including seasonal indicators as features.

1. Lele, A. (n.d.). Battery and heating data in real driving cycles. Kaggle. <https://www.kaggle.com/datasets/atechnohazard/battery-and-heating-data-in-real-driving-cycles?resource=download> [↑](#footnote-ref-1)