Name – Mohit Kumar

Entry Number – 2023EE10862

LOHUM Summer Intern Problem: Analysis Report

This report outlines the analysis of cell-level electrical test data from three distinct **Original Equipment Manufacturers (OEMs)**, each following either a 5-step or 7-step charge-discharge protocol. The core objectives were to characterize the data, identify an unknown feature, compare cell behaviour and degradation across manufacturers, and build predictive models for **State of Health (SOH)** and **State of Power (SOP)** for a selected OEM.

Data Handling and Feature Identification

All thirty text files (ten per OEM) were processed and combined into per-OEM data frames, with columns appropriately named and converted to numeric types.

Feature 10

A key task was to identify the electrical quantity represented by **Feature 10**. By comparing it against the product of instantaneous voltage and discharge capacity (V x Ah), it was concluded that Feature 10 is the **cumulative discharge energy in milli-watt-hours (mWh)**. This feature is monotonically increasing during discharge.

Targets and Predictive Features

The modelling targets, SOH and SOP, were defined based on rated values inferred directly from the test data:

- **Rated Capacity (mAh):** The maximum discharge capacity observed in the OEM dataset.
- **Rated Power (W):** The maximum power observed in the OEM dataset.
- **SOH:** The ratio of the maximum discharge capacity during the **main discharge step** to the rated capacity. The main discharge step is chosen as the one with the largest final discharge capacity.
- **SOP:** The ratio of the maximum power during the main discharge step to the rated power.

Predictive features were extracted from the **initial fraction** of the main discharge step (the first 10% of samples) to simulate a quick health check. These features included:

- Mean voltage and standard deviation of voltage.
- Maximum current.
- Mean power.
- Mean absolute value of dV/dQ, which is sensitive to internal resistance and aging.

Comparative OEM Analysis

The analysis revealed significant differences both within and across manufacturers.

Intra-OEM Variation

Within each OEM, there was a **noticeable spread** in final discharge capacity and voltage profiles, suggesting variation in SOH for nominally similar cells. **Weaker cells** tended to have voltage-capacity curves that dropped earlier or showed a stronger slope, indicating higher internal resistance or degradation.

Inter-OEM Differences

The average profiles across the three OEMs differed fundamentally, suggesting varying **chemistry and design choices**:

- **OEM 1:** Exhibits relatively flat voltage plateaus, consistent with an **LFP-like chemistry**.
- **OEM 2:** Shows more sloped curves with distinct shoulders, typical of high-energy **NMC-type chemistries**.
- **OEM 3:** Tends to have lower average capacity and more pronounced slopes, potentially indicating an **older or more degraded population**.

This highlights that any second-life evaluation framework must be **chemistry-aware**.

SOH and SOP Predictive Modelling (OEM 2 Example)

OEM 2 was selected for modelling due to the wide spread in its cells final capacity and power, providing a robust test case.

The pipeline involved computing rated values, identifying the main discharge step, calculating SOH and SOP targets, and extracting the early-discharge features.

A **Random Forest Regressor** was chosen for both SOH and SOP prediction due to its ability to handle non-linear relationships. Performance was assessed using **5-fold cross-validation** with the following metrics:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)

For SOH -

- Root Mean Square Error (RMSE) - 0.45370956966005
- Mean Absolute Error (MAE) - 0.45370956966005
- Mean Absolute Percentage Error (MAPE) - 1687.52085966398

For SOP -

- Root Mean Square Error (RMSE) - 0.34246253229974
- Mean Absolute Error (MAE) - 0.34246253229974
- Mean Absolute Percentage Error (MAPE) - 275.188968574587