Assignment: Battery Analytics + Machine Learning

Dataset Description

The provided dataset contains **687,212 high-frequency measurement records** collected from a lithium-ion battery across multiple charge–discharge cycles. Each row represents a single time-ordered measurement containing:

- **Voltage (mV):** Instantaneous voltage in millivolts.
- **Current (mA):** Current in milliamps.
    - Positive values: Charging
    - Negative values: Discharging
    - Zero (or near-zero): Idle
- **Capacity (mAh):** Accumulated charge or discharge capacity at that moment.
- **Cycles:** The cycle count at that timestamp (ranges approximately from cycle **5 to 104**).
- **Status:** Textual operational label from the logging system (e.g., “Con-C Discharge”). Candidates should rely primarily on the **Current** column to determine operational mode rather than the Status text.

The dataset represents continuous operation over many cycles and is suitable for:

1. **Operational mode analytics** (Charging / Discharging / Idle).
2. **Predictive modelling** to estimate cycle number based on electrical parameters and engineered features.

Section 1: Analytics Problem

Objective

Analyse the dataset to extract insights based on different operational modes (**Charging, Discharging, Idle**).

Tasks

1. **Total Number of Sessions**
    - Identify and count all sessions for each mode.
    - A session refers to a continuous period where the battery stays in the same mode.
2. **Total Duration**
    - Compute the total time spent in each mode.
    - If timestamps are absent, assume each row represents one uniform time interval (state this clearly).
3. **Average Parameter Values**
    - For each mode, compute averages for relevant parameters, such as:
        - Voltage
        - Current
        - Capacity
        - Any derived features you create
4. **Additional Parameters (Can add more)**
   Identify and analyse any additional relevant metrics. Examples include:
    - Power (Voltage × Current)
    - Rate of change (dV/dt, dI/dt)
    - Energy estimate (integration of power)
    - C-rate approximation
    - Capacity drop patterns across cycles
    - Any anomaly or outlier behaviour

Visualisations (Mandatory)

Create clear, well-labelled visualisations that explain your findings. Examples:

- Time-series plots
- Histograms
- Box plots
- Scatter plots
- Mode-wise comparison charts

Bonus points for creative, insightful visualisations that bring out hidden patterns.

Section 2: Machine Learning Problem

Objective

Develop predictive models to estimate the battery **Cycle Number** using the columns:

- Current
- Voltage
- Capacity
- Cycles (target variable)
  You must engineer additional features to improve prediction quality.

Steps

1. Data Preprocessing

- Handle missing values, outliers, noise, and inconsistent readings.

- Document all decisions: dropping, imputing, smoothing, clipping, etc.

2. Feature Engineering

Create additional features that may help predict cycles. Potential ideas:

- Power = Voltage × Current
- Rolling mean/std of Voltage, Current, Capacity
- dV/dt and dI/dt
- Charge/discharge amplitude features
- Mode flag as categorical variable
- Interaction features (V×C, I×C)
- Session-level aggregates (min/max/mean/std values per session)

Explain why each engineered feature might help.

3. Model Development

Train and compare the following models:

- Linear Regression
- Decision Tree
- Random Forest
- Gradient Boosting Machines (XGBoost/LightGBM allowed)
- Feedforward Neural Network (simple MLP)

4. Hyperparameter Tuning

Use **Grid Search** or **Random Search**.

5. Cross-Validation

Use K-fold or time-series cross-validation.
Ensure **no data leakage**, especially because cycles increase over time.

6. Evaluation Metrics

Report, compare, and interpret:

- MAE
- MSE
- RMSE
- R^2 score

7. Model Comparison

Compare all models and clearly explain:

- Which performed best and why
- Trade-offs (speed, complexity, generalisation, stability)
- Feature importance insights

Submission Requirements

1. Report (PDF)

Your PDF report must include:

- Introduction & objectives
- Dataset description
- Methodology & full approach
- Section 1 analytics findings with visualisations
- Feature engineering and data preprocessing explanation
- ML model development details
- Hyperparameter tuning process

- Cross-validation methodology
- Evaluation results with plots/tables
- Conclusions and key insights

2. Code

Submit **Notebook(s)** or **Python scripts** containing:

- Clean, readable, modular code
- Comments explaining critical steps
- All plots reproducibly generated
- All models trained with visible hyperparameters

Closing Note

We look forward to reviewing your submission and potentially welcoming you to our team.

Good luck.

Emo.energy