Case Study 3: Detecting Generator Usage from Sensor Data

Team members: Risper Audi, Edwin Owino, Andrew Kiptoo, Katherin Ncheku

Context

In many health facilities in sub-Saharan Africa, power is supplied from a combination of grid, generator, and solar sources. nLine's sensors take voltage and frequency measurements every 2 minutes but we don't know which power source is active at a given time.

This creates a challenge: which power source is currently supplying the load?

The ability to do so unlocks rich performance metrics — from solar PV and generator usage time (linked to fuel cost and carbon emissions) to evaluating the effectiveness of solar investments.

nLine sensors were deployed in different rooms in a number of hospitals in Sierra Leone and have been collecting data on voltage and frequency.

Your Challenge

> ***You'll take on the role of a data scientist at nLine tasked with improving the existing rule-based power-source classification algorithm. You will experiment (reasonably) with different Machine Learning models, and select which one works best for the power source detection problem. Using the selected model, you will label each two-minute time window with the most likely power source (solar, grid, generator). Finally, you will determine solar, generator, and grid usage durations for each of the rooms in the hospitals.***

Data Provided

You'll receive time series voltage and frequency data (two-minute resolution) from rooms in selected healthcare facilities in Sierra Leone along with the known power-source configurations supplying the room(e.g., "grid + solar", "grid + gen"). This data will also contain the power source detected to be supplying the room at each time window, a detection done by the rule-based algorithm. Treat the detected power source column as the ground truth against which you will compare the results of your ML-based algorithm.

You will also receive a description of the current rule-based power-source detection method.

Guiding Questions

- Understand and review the rule-based power-source detection algorithm. What are its strengths and shortcomings?
- Visually explore the frequency and voltage data. Are there any distinct statistical or visual features that can be used to distinguish between power sources?

- Feature engineering: What derived features(from voltage and frequency) can be used to determine the power source supplying the load at any time?
- List different supervised Machine Learning models that can be applied to this power-source detection problem and explain why each would be appropriate.
- Train the ML models using the provided dataset and perform model evaluation.
- Report on which one performs best.
- What would a hybrid ML + rule-based version of this algorithm look like?
- Outlining next steps: how to improve accuracy, handle edge cases, etc.

Deliverables

- A Jupyter notebook with analysis, visualizations, model training, evaluation, selection.
- A detailed report responding to the guiding questions.
- A slide deck to present your work to nLine.