INNOVAITE Hackathon

Machine Learning Energy Disaggregation Report

Team:
Ayan Sultanov · Zhansaya Akhmetova · Jakob Lutz · Martina Russo

---

1. Cross-Validation Strategy

To assess model generalization to **unseen households**, we adopted a **Leave-One-House-Out (LOHO)** cross-validation strategy.
In each fold, one household was held out for validation while the remaining households were used for training.

This evaluation setup closely reflects real-world **Non-Intrusive Load Monitoring (NILM)** deployment, where a model trained on known households must disaggregate appliance consumption for a new household with different appliances, noise characteristics, and usage patterns.

---

2. Models Evaluated

We evaluated two families of gradient-boosting models:

- **LightGBM hybrid model**
- **XGBoost hybrid model**

Each model family was tested under multiple output formulations:

1. **Pure regression**
   Direct prediction of refrigerator power consumption (in watts).
2. **Hard hybrid gating**
   Regression output gated by a binary ON/OFF classifier.
3. **Soft hybrid gating**
   Regression output scaled by the predicted probability of the fridge being ON.
4. **Blended hybrid (optional)**
   Averaging predictions from multiple output heads to improve stability.

For each LOHO fold, performance was evaluated using **Mean Absolute Error (MAE)** on the held-out household.

---

3. Cross-Validation Results

Across the four validation households, **XGBoost with hard hybrid gating** achieved the best overall performance, with an average **MAE of 28.43 W**.
The best-performing LightGBM configuration (also hard hybrid) reached an average **MAE of 31.10 W**.

These results indicate that:

- Explicitly modeling **ON/OFF dynamics** significantly improves disaggregation accuracy.
- XGBoost is slightly more effective at capturing **nonlinear refrigerator behavior** in this dataset.

Performance varied substantially across households.
Both models performed best on **clean households** (e.g., *home_4*), where refrigerator cycles were clearly separable from other loads.
Conversely, performance degraded on **noisier households** (notably *home_5*), where overlapping appliance activity and unstable mains patterns made disaggregation more challenging.

This variability highlights the importance of robust ON/OFF detection, feature smoothing, and temporal consistency in NILM settings.

---

4. Final Model Selection

The LOHO analysis motivated the use of:

- **Hybrid model architectures**
- **Threshold tuning**
- **Model ensembling**

These techniques were incorporated into the final model used for test-set inference in order to stabilize predictions and improve generalization across heterogeneous households.

---

5. Suggestions for Improving Performance

Several directions could further improve refrigerator disaggregation accuracy:

5.1 Additional Data Sources

- **Higher-frequency sampling** (1–5 seconds) to better capture sharp ON/OFF transitions.

- **Environmental data** (ambient or outdoor temperature) to anticipate compressor cycle length.
- **Door-open sensors** to directly indicate likely cooling events.
- **Appliance metadata** (efficiency class, age, model) to differentiate household-specific behavior.

5.2 Enhanced Feature Engineering

- **Cycle-aware segmentation** to explicitly model OFF → ON → steady ON → OFF states.
- **Frequency-domain features** (e.g., short-window FFTs) to capture compressor harmonics.
- **Appliance-context features** to account for interference from high-power devices.

5.3 Alternative Model Classes

Beyond gradient boosting, several model families are promising:

- **CatBoost**, which performs well on correlated tabular features.
- **Deep learning approaches** such as Seq2Point CNNs, CNN–LSTM hybrids, and transformer-based models, which have shown strong performance in NILM research by learning temporal motifs directly.

With additional time, deeper exploration of these models could likely outperform the current approaches.
A **hybrid ensemble** combining boosting models with deep learning could further leverage the strengths of both paradigms.

5.4 Post-Processing Improvements

- Replacing hard gating with **clipped soft hybrid gating** to avoid prediction collapse.
- **Temporal smoothing** (moving average or Gaussian filters) to remove unrealistic noise.
- **Confidence-weighted blending** between LightGBM and XGBoost predictions.

5.5 Training Methodology

- Hyperparameter optimization using **Optuna** or similar frameworks.
- Improved balance between ON and OFF cycles to mitigate class imbalance.
- Household-specific normalization to reduce bias.
- Data augmentation through noise injection or synthetic cycle generation.

---

6. Visualization of Results

A qualitative evaluation was planned through time-series plots comparing predicted refrigerator power consumption to ground-truth signals.

The intended approach was to:

- Apply LightGBM, XGBoost, and ensemble models to validation segments.
- Overlay predicted fridge power on the true signal over time.
- Visually inspect ON/OFF cycles, compressor peaks, and transition timing.

Such plots provide intuitive insight into model behavior, revealing systematic delays, underestimation, or noise artifacts, and typically demonstrate the smoothing and accuracy benefits of ensembling over individual models.

---

7. Conclusion

This project demonstrates that hybrid regression–classification models, combined with careful validation across households, can effectively address refrigerator energy disaggregation under irregular smart-meter sampling.
The results highlight both the strengths and limitations of gradient-boosting approaches in NILM settings and point toward clear directions for future improvements.