Optuna-based Hyperparameter Optimization and Analytical Insights for Smart Building Metadata Classification

Yuetong Wu

August 27, 2025

Abstract

This mini-report documents my personal contribution to the UNSW Capstone Thesis project on smart building metadata classification. Building upon the BrickMIR baseline framework, I implemented an automated hyperparameter optimization strategy using the Optuna framework. The proposed approach improves robustness, recall, and generalization in multi-label classification of building sensor data. I further conducted class-wise performance analysis and distribution shift evaluation, providing insights into systematic limitations of the baseline.

1 Introduction

Smart building metadata classification is a challenging task due to label imbalance, heterogeneous distributions, and distributional shifts across buildings. Previous work has introduced various feature-engineering and boosting-based approaches to address these issues.

Our work builds upon BrickMIR [?], a minimal imbalance-tuned and ratio-based framework for Brick metadata classification, by integrating Optuna-based automated hyperparameter optimization.

2 Methods

I optimized CatBoost classifiers on a per-label basis across 94 labels. The search space included:

- `max_depth`: 6–12
- `learning_rate`: 0.05–0.3 (log-scaled)
- `n_estimators`: 50–150
- `l2_leaf_reg`: 1.0–10.0

Each label was tuned using five-fold StratifiedKFold cross-validation with Macro-F1 as the objective function. To ensure efficiency, each label was restricted to 20 trials with parallel execution enabled.

3 Results

Table 1 compares the original BrickMIR baseline against my Optuna-enhanced version.

Table 1: Performance of BrickMIR before and after Optuna optimization

Model	Accuracy	Macro-F1	Precision	Recall
Original BrickMIR	0.98540	0.50794	**0.61988**	0.50668
BrickMIR + Optuna (Mine)	**0.98588**	**0.50855**	0.61474	**0.51030**

The Optuna-enhanced pipeline achieved modest but consistent gains, particularly improving recall, which reflects increased sensitivity to minority labels.

4 Analysis

4.1 Hyperparameter Sensitivity

- **max_depth:** strongest effect, ΔF1 up to ± 0.1; best at 7–9.
- **learning_rate:** stability improved at 0.06–0.1; higher values degraded performance.
- **n_estimators:** limited effect, gains diminished beyond 200 trees.
- **l2_leaf_reg:** stabilized results, consistent at values 3–7.

4.2 Class-wise Analysis

Per-label F1 analysis across 94 classes revealed that several labels (e.g., `Cooling_Demand_Sensor`, `Heating_Demand_Sensor`, `Outside_Air_CO2_Sensor`) scored near zero, indicating inability to learn due to scarce positives.

The confusion matrix revealed systematic misclassifications, e.g., `Status` → `Command` (24,326 times), `Status` → `Alarm` (23,262 times). These suggest semantic overlap and insufficient discriminative features for certain categories.

4.3 Distribution Shift Insights

Chi-square tests showed 58/94 labels had significant distribution shifts ($p < 0.05$) between training and test sets.

- Labels with significant shifts averaged lower F1 (0.5010) vs. stable labels (0.5192).
- Worst drops: `Cooling_Demand_Setpoint`, `Discharge_Air_Dewpoint_Sensor`.
- Causes: scarcity of positives, contextual drift, feature-label misalignment, input feature distribution shifts, and lower prediction confidence.

These insights highlight structural limitations: Optuna improved recall, but imbalance and dataset drift remain unresolved.

5 Conclusion

This mini-report presents my independent contribution: integrating Optuna into BrickMIR, performing class-wise error analysis, and investigating distributional shifts. Automated hyperparameter tuning yields measurable improvements, while deeper analyses expose unresolved challenges in label scarcity and dataset drift. These findings provide a reproducible, individual extension to the group thesis and form the basis for potential future research directions.