Trimming outliers using trees: Winning solution of the Large-scale Energy Anomaly Detection (LEAD) competition

Chun Fu
chunfu@nus.edu.sg
National University of Singapore

Pandarasamy Arjunan
samy@bears-berkeley.sg
Berkeley Education Alliance for Research in Singapore Limited

Clayton Miller
clayton@nus.edu.sg
National University of Singapore

ABSTRACT

Prediction of building energy consumption using machine learning models has been a focal point of research for decades. However, some causes of forecast errors, particularly data quality, have not been adequately addressed, which may affect the accuracy of forecasting models and subsequent energy management. To solve the issue of data quality, a classifier that can automatically detect time series anomalies is the goal that researchers have been pursuing. Large-scale Energy Anomaly Detection (LEAD), a community competition hosted on the Kaggle platform, was created for this purpose as well as to provide a foundation for benchmarking solutions. In this competition, 200 energy time series worldwide with labeled anomalies were provided to train a classification model to predict anomalies of another 206 unseen time series. The proposed winning solution is a tree-based supervised learning anomaly classifier with ROC-AUC score as high as 0.9866 on private leaderboard. This article describes and analyzes in depth a variety of commonly employed techniques for improving the classification model. Among these strategies, feature engineering requires the most effort and dominates all other techniques; value-changing features that can represent the level of time-series variation have a particularly positive impact. Besides, the classification accuracy of solutions in the competition can serve as a benchmark for future research on supervised learning of energy anomaly detection.

CCS CONCEPTS

• **Computing methodologies** → **Anomaly detection**; **Modeling methodologies**; **Classification and regression trees**.

KEYWORDS

Building energy, Smart meter, Anomaly detection, Supervised learning, Classification

ACM Reference Format:
Chun Fu, Pandarasamy Arjunan, and Clayton Miller. 2022. Trimming outliers using trees: Winning solution of the Large-scale Energy Anomaly Detection (LEAD) competition. In *The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3563357.3566147

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
BuildSys '22, November 9–10, 2022, Boston, MA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9890-9/22/11.
https://doi.org/10.1145/3563357.3566147

1 INTRODUCTION

1.1 Data-driven anomaly detection in building energy data

Although machine learning has good potential for energy prediction, in the real world, the unstable quality of building energy data causes problems with prediction accuracy. In terms of the impact of faults on energy management, improperly maintained, faulty and degraded hardware, and improper operations waste an estimated 15 to 30 percent of energy consumption in commercial buildings [6, 12]. Some studies have demonstrated that data quality can significantly impact forecast accuracy, with the following negative consequences. If anomalous data are not correctly identified and effectively corrected, they will serve as a reference with false predictions, therefore harming the forecast's accuracy and reliability [9]. Because anomalous data can cause bias or failure to estimate parameter values, anomaly detection is becoming more crucial for energy models [13]. Studies have shown the benefits of anomaly detection in terms of energy and related cost savings. For example, it is possible that more than 10% of the energy produced in Europe is lost every year due to non-technical losses and that billions of dollars are lost every year as a result of energy theft [7, 15]. Besides, distinguishing between behavioral consumption anomalies, fraud, and unintentional consumption deviations has been identified as a current research trend in order to provide accurate feedback to end-users and energy providers [1, 5]. In terms of energy modeling, the performance of machine learning methods and the predictability of test data are also affected by abnormal energy-consumption behavior [11]. All of these papers stress the importance of anomaly detection and its effects on building operations and energy management.

To solve the aforementioned problems caused by faults in energy data, there are several past studies on anomaly detection in built environments. In terms of Heating, Ventilation, and Air Conditioning (HVAC) systems, since equipment signals and fault data can be easily collected from the system, there are a handful of studies on anomaly detection via supervised learning, such as chiller [8] and Air Handling Unit (AHU) [14]. In contrast to the air conditioning system, however, past research on detecting anomalies in energy data is relatively rare. Typically, energy data consists only of meter readings and weather data, whereas HVAC system data have comparatively complete data points (e.g., temperature, air volume, frequency, etc.). Another challenge is that energy data usually lacks fault labels, whereas HVAC systems could provide sufficient faults from signal data of equipment. All of these negative factors lead to the difficulty of developing anomaly detection models for energy data. In light of the lack of labels, a study adopted an unsupervised learning method that employed statistical pattern recognition techniques to identify anomalies in energy data [2]. Additionally, a regression-based supervised learning study uses the discrepancy between the measurement and the predicted baseline to identify anomalies [3]. However, in these energy-related studies, there is still a lack of benchmark datasets with abnormal labels for in-depth performance analysis and classification model comparison.

1.2 Introduction of Large-scale Energy Anomaly Detection (LEAD) competition

LEAD was a community prediction competition hosted on the Kaggle platform, which has been the lead platform in organizing data science competitions in recent years. In this competition, participants needed to develop accurate machine learning models to identify instances of anomalous energy consumption (point anomalies) in hourly smart meter time series over the course of the year. The competition data set is based on the energy data set used in Great Energy Predictor III competition[1] hosted by the ASHRAE organization on the Kaggle platform [10]. The training dataset contains hourly meter readings from 200 buildings throughout the entire year, with labels of either abnormal (1) or normal (0) usage. The objective of the competition is to use this training dataset to develop a machine learning model for anomaly detection and then predict anomalies in meter readings from another 206 buildings in the test dataset.

The data used for this competition is a subset of a full dataset. The full dataset consists of 1,413 smart meter data in a time series manner with annotated labels spanning across a year. An introductory article for this dataset describes the progress of annotation and offers several baseline anomaly detection models as performance benchmarks for anomaly detection [4].

This dataset was annotated with two types of anomalies: (1) point anomalies and (2) sequential or collective anomalies:

(1) Point anomaly:
A point anomaly is an instance of energy consumption that is anomalous when compared to the entire time series or its neighbors. It occurs randomly and sporadically rather than continuously.
(2) Sequential or collective anomaly:
A sequential anomaly is a collection of consecutive abnormal points that indicates an abnormal event of energy consumption. It may occur once or on a regular basis.

The evaluation metric for this competition is the Area Under Receiver Operating Characteristic Curve (AUC-ROC).

$$\text{AUC-ROC score} = \text{The area under ROC Curve} \tag{1}$$

This paper outlines the winning solution to the competition. There were 75 competitors with 600 submissions over the course of three months this year. Participated competitors come from a variety of backgrounds, including data scientists, engineers, students, etc. In addition, 14 shared notebooks containing solutions and visualizations are accessible to the public. The competition was hosted on the Kaggle platform as a community prediction competition with no prize money available for the winners[2].

2 OVERVIEW OF THE WINNING SOLUTION

The winning solution employs a framework of tree-based models that includes several tasks of data preprocessing, feature engineering, data downsampling, modeling, and post-processing (as shown in Figure 1). As the aim of this competition is to detect anomalies, data cleaning is not required during the data preprocessing stage. Instead, only the imputation of missing data and feature normalization were performed. Feature engineering took about half of the work time, but it is the most significant factor for model performance, which is also the most emphasized and crucial stage for most data competitions. During the model building stage, the first step was to solve the imbalance between normal and abnormal data by downsampling. Next, classification models were developed separately from several powerful and common tree-based models. The final submission is established by the weighted average model ensemble and post-processing.

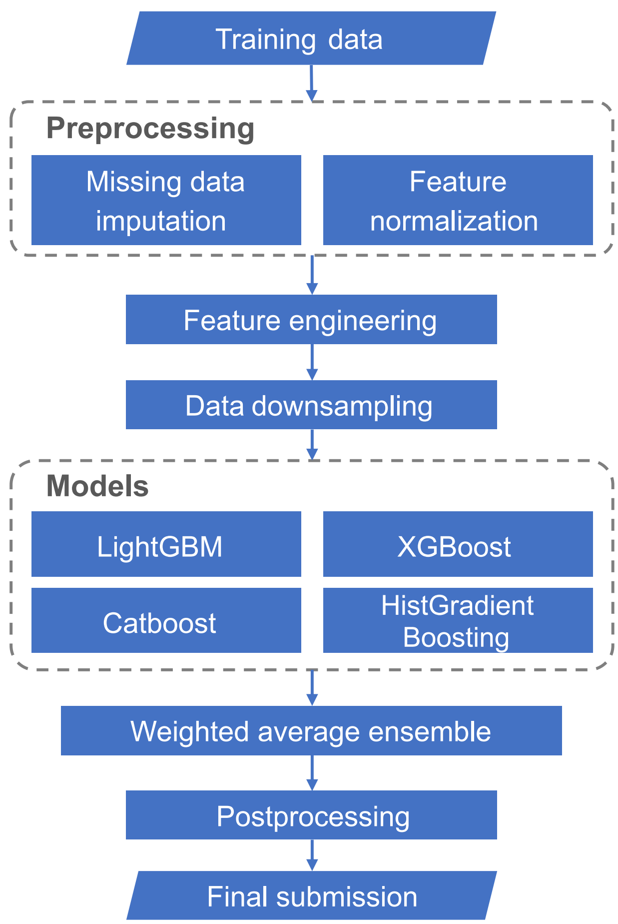

Figure 1: The overview of the proposed solution and different phases in the working pipeline

[1] https://www.kaggle.com/c/ashrae-energy-prediction

[2] https://www.kaggle.com/competitions/energy-anomaly-detection

2.1 Pre-processing

In most machine learning projects, data cleaning is crucial but time-consuming and knowledge-intensive. However, since the goal of this competition is to predict data anomalies, energy data anomalies are not cleaned up. During this phase, the only data processing work is imputing missing data. Missing values (NaN) in the provided dataset only constitute 6.2%. Hence, these missing values were replaced with the mean value for each time series. Other missing data filling methods, such as forward and backward filling, have also been tried, but the results were not as good as average-value filling.

2.2 Feature engineering

The dataset provided in this competition contains up to 57 features, some of them were original features from energy data, and some were obtained by feature engineering by the winning team in GEPIII (see Table 1). Since these features are based on building energy prediction models, they are not necessarily effective for anomaly detection. In order to strengthen the identification of anomaly detection, especially the degree of change of time series values, value-change features were added to the model features. The following subsections will elaborate on these value-change features.

Table 1: Features for developing anomaly classification model

Category	Descriptions of features
Energy use	Meter readings from power meters.
Building meta	Basic information of buildings. (e.g., site_id, building_id, primary_use, square_feet, year_built, and floor_count)
Weather data	Onsite measurements of weather conditions. (e.g., air_temperature, cloud_coverage, dew_temperature, precip_depth_1_hr, sea_level_pressure, wind_direction, and wind_speed)
Temporal feature	Derived features from timestamps. (e.g., hour, weekday, and day of year)
Target encoding feature	Average values of the target variable aggregated by category (e.g., average values grouped by building_id)
Value-change feature	Changes of time-series values in the form of difference or ratio (e.g., the increase or decrease of value compared to previous hour)

2.2.1 *Value-change features.*

Since the anomalies in this competition are (1) Point anomaly and (2) sequential anomaly, the value change of the time series will be an important feature for detection. Therefore, in the absence of this feature type in the original dataset, value-change features, which compute the value change in the form of difference and ratio, were created. Figure 2 shows how the two value changes are calculated. In addition, considering that the value change may be continuous for several timesteps and that the energy data has daily and weekly periodicity, different shift steps, from one timestep, 24 timesteps (1 day), to 168 timesteps (1 week) were also considered in the value-change features. Although the difference and ratio value changes are quite similar, the result shows that the inclusion of both features could achieve the best prediction performance, so they are both retained.

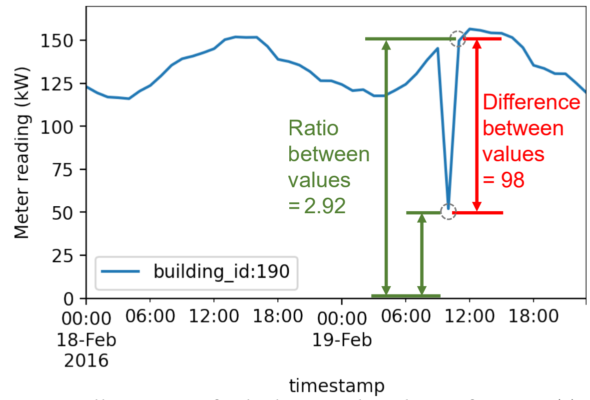

Figure 2: Illustration of calculating value-change features: (1) Value change in difference (red) and (2) Value change in ratio (green)

2.2.2 *Value change in difference.*

Difference between nearby values in time series data is one of the most intuitive characteristics of value change. A sharp change in value is likely to indicate a point anomaly; a value change of zero may indicate flatlined anomalies, which are also known as sequential anomalies. To capture these value changes, varying shifts of timesteps from 1 to 168 were included in the difference calculation. Furthermore, this competition does not require prediction into the future, so both positive and negative shift steps were included. However, if all combinations are considered, there will be up to 168*2=336 new features in total, which may adversely affect the prediction model. Therefore, only shifts within one day were fully accounted for (i.e., 1, 2, 3, and 23), while larger shifts were added at 24-hour intervals (i.e., 24, 48, 72, and 168). Equation 2 is used to calculate the value change in difference, where t is the timestamp and s is the shift of timestamps.

$$\text{Value change in difference} = X(t) - X(t-s) \tag{2}$$

2.2.3 *Value change in ratio.*

Although the aforementioned difference-based value change can effectively assist the prediction model in detecting time series anomalies, the scales between time series or different time periods sometimes vary greatly. Therefore, calculating the ratio of changes will result in a more consistent and comparable scale than calculating the difference. In the calculation of the ratio, special attention should be paid to the fact that the denominator cannot be zero, so both the numerator and the denominator are added by one, respectively, during the calculation. Equation 3 is used to calculate the value change in ratio, where t is the timestamp and s is the shift of timestamps.

$$\text{Value change in ratio} = \frac{X(t)+1}{X(t-s)+1} \tag{3}$$

2.2.4 *Other features.*

In addition to features introduced in the previous section, there were also attempts to use savgol filter for difference calculation of the smoothed values and K-means clustering of the time series. However, these created features have no apparent positive effect on anomaly detection, so this article will not elaborate on these features.

2.3 Modeling

2.3.1 Data splitting method.

Train and validation datasets were split by *building_id* to ensure valid data were unseen during training. Compared to data splitting by shuffling, the validation score generated by this method is very close to the score calculated on the test data on the leaderboard (the difference is less than 1%). Therefore, referring to the score calculated from the validation dataset, prediction performance tuning, such as feature engineering and model optimization, can be performed locally.

2.3.2 Data downsampling.

The low proportion of anomalies in this dataset, which is about 5%, caused a severe imbalance of normal and abnormal data. In order to solve the adverse effect of data imbalance on the classification model, the downsampling method for normal data is adopted so that normal and abnormal data can be balanced to occupy half of the training data.

2.3.3 Tree-based classification models.

For classification problems with tabular data, tree-based models are still the most popular and powerful choice. Among various tree-based models, a few of them are particularly popular, i.e., LightGBM, XGBoost, Catboost, and HistGradientBoosting. These models have differences in computation speed and environmental adaptability, but they all share close predictions with high accuracy. Therefore, these tree-based models are often used in Kaggle data competitions. It is particularly worth mentioning that LightGBM, as its name suggests, is as fast as light. Therefore, most people use LightGBM to model and calculate errors during feature engineering and testing of data processing strategies. In the final stage of the competition, different models will be considered for modeling and hyperparameter tuning.

2.3.4 Model ensembling.

One of the most common strategies for the final stage of data competition is model ensembling. Each model has its own strengths and weaknesses in prediction, even though performance scores between models were similar. The weighted average of the prediction values of these multivariate models with similar performance can usually improve the prediction accuracy effectively. In order to simplify the operation, the solution proposed in this paper only performs a simple average of the prediction results of the four classification models mentioned in Section 2.3.3, meaning that the weight of each prediction is 0.25 (as shown in Equation 4).

$$\text{Ensemble model prediction} = (\text{LGBM}+\text{XGB}+\text{Cat}+\text{HistGB}) * 0.25 \tag{4}$$

2.4 Post-processing

Following the completion of model prediction and model ensembling, simple rules were used for post-processing the prediction results. According to a post on the discussion board in this competition, nearly 100% of the points with *meter_reading* values of one are anomalies. Furthermore, by visualizing time-series trend of each power meter, the majority of starting and ending points of time series are not anomalies. Based on these two findings, post-processing was performed as following rules:

(1) Set prediction to 1 (abnormal) for rows with *meter_reading* values of 1
(2) Set prediction to 0 (normal) for start and end points of time series

3 RESULTS

The AUC-ROC score of the proposed solution in this competition can reach as high as 0.9866, which is far beyond the 0.9 threshold that is considered to be an excellent performance of the classifier. It also performs exceptionally well in precision, where 98.7% of the anomalies predicted by the proposed classification are correctly labeled anomalies. In terms of recall rate, 81.9% of labeled faults can be successfully detected by the classification model. The indicators that can comprehensively evaluate the performance of classification, combined with precision and recall rate, the f1 score can reach a level of about 0.89. The Confusion matrix can be seen in Figure 3, showing the number and percentage of points in each quadrant.

3.1 Importance and effects of features

Although there are numerous strategies for improving model performance in the proposed solutions, the most crucial factor that stands out in this competition is feature engineering, particularly the proposed value-change variables. Value-change variables, including two forms of difference and ratio, can effectively represent the change levels between each point and its neighboring points. In Figure 4, comparing the performance of the model before and after feature engineering, the AUC increased from 0.9311 to 0.9849, a significant increase of 5.8%. This significant improvement demonstrates the predominance of feature engineering in this competition. Additionally, Table 5 presents the top ten features of the proposed solution by sorted feature importance, exported from the trained lightGBM model. Among the existing features provided, metadata, temporal features, and target encoding features all appear in the list of the top 10 most important features and are of great significance. It is noteworthy that the proposed value-change features hold four of the top ten positions, demonstrating the crucial role of these features. The interesting fact is that the shift timesteps for these four are -1, 1, 2, and 168, showing that the most influential feature is the value change from the nearest and one week away. This might suggest that the weekly repeatability of energy data may aid in the detection of anomalies. If the value of a point is significantly different from the previous timestep or week, it is likely to be an anomaly.

3.2 Model ensembling results

In addition to the previously mentioned work of data preprocessing and feature engineering, creating diverse models for the ensemble is another useful technique that can frequently outperform a single model. The proposed solution employs four popular tree-based models with comparable good performance: LightGBM, XGBoost, Catboost, and HistGradientBoosting. Table 2 presents the individual classification performance of the four models, ranging from 0.9840 to 0.9857, with very close prediction performance. By

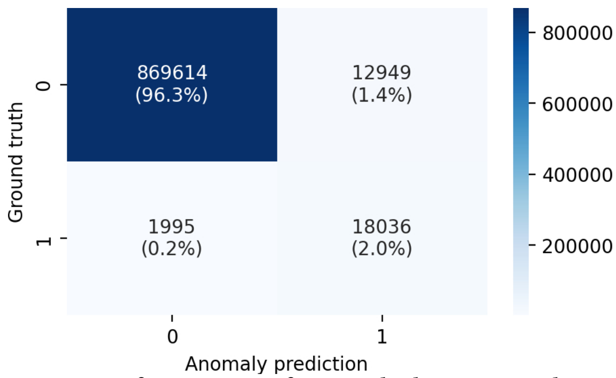

Figure 3: Confusion matrix of proposed solution on test dataset

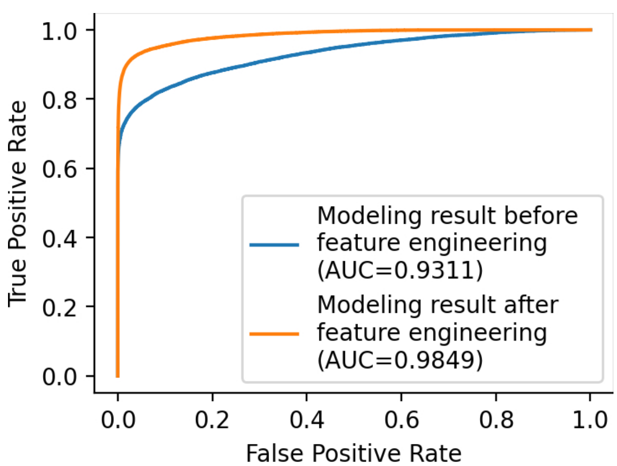

Figure 4: ROC curve and AUC-score before and after feature engineering

	feature	category	importance
0	building_id	Building meta	219
1	value_chg_ratio_1	Value-change feat.	144
2	value_chg_ratio_-1	Value-change feat.	127
3	meter_reading	Energy use	98
4	dayofyear	Temporal feat.	95
5	square_feet	Building meta	85
6	gte_building_id	Target encoding	63
7	value_chg_ratio_-168	Value-change feat.	61
8	value_chg_ratio_2	Value-change feat.	54
9	gte_meter_primary_use	Target encoding	54

Figure 5: Feature importance of the 10 most influential features exported by LightGBM

averaging the prediction results (probability of each label) of the four models, the performance can be improved to 0.9867, approximately 0.21% higher than the original average performance.

Table 2: AUC-ROC scores of tree-based models and ensemble model

Model	AUC-ROC score	
	Train	Test
LightGBM	0.9975	0.9849
XGBoost	0.9999	0.9840
Catboost	0.9999	0.9857
Hist Gradient Boosting	0.9968	0.9839
Weighted average ensemble	**0.9996**	**0.9866**

3.3 Overview of public solutions in competition

To distinguish and analyze the differences between the proposed method and other competitors' solutions, Table 3 presents the leaderboard scores (i.e., AUC-ROC score for test dataset) of public notebooks on the Kaggle platform and their model strategies. In the comparison of data preprocessing, solutions with a private score lower than 0.90 did not address the problem of data imbalance. In addition, the number of features seems to have a positive effect on classification performance. In particular, the adoption of value-changing features could effectively improve model performance. Regarding the selection of classification models, tree-based models remained the most popular option among competitors, and neither neural networks nor deep learning was used. Although the model ensemble is helpful for model performance improvement, it is not the key to improving performance in this competition.

4 DISCUSSION

This section delves deeper into the insights provided by the proposed solution. They include the proposed value-change features that distinguish the winning solution from the competition, as well as the high-performance benchmark established by the winning solution, which offers significant potential for future related research.

4.1 The significance of value-change features in capturing context in time series

In the analysis of the impact of features in Subsection 3.1 on model performance, additional value-change features can effectively assist in anomaly detection. This shows that it is crucial to provide information regarding value changes by calculating the difference between each point and its neighbors for the classification model of anomaly detection. Notably, tree-based models applied to tabular data lack the ability to self-extract features, so providing additional such value-change features is a necessary step. Although neural networks were not widely used in this competition, the success of value-change features in this competition suggests that employing a framework with context-learning capability (such as Convolutional Neural Network) could be an effective strategy.

4.2 Benchmark of supervised learning in anomaly detection of energy data

As the first anomaly detection competition for a large number of power meters on various sites and buildings, the performance of supervised learning models in this competition can serve as

Table 3: List of publicly available shared solutions and their modeling strategies

Team / Author	Public score	Private score	Preprocessing techninques	Features (count)	Modeling strategies
Proposed	0.9734	0.9866	Normalization, imputation, and downsampling	Raw, V-C (169)	Ensemble: LightGBM, XGBoost, CatBoost, Hist Gradient Boosting
Abhishek Maurya	0.8794	0.9237	Normalization, imputation, and downsampling	Raw (31)	XGBoost
Abdallah El-Sawy	0.7633	0.8189	Imputation	Raw (10)	Ensemble: KNN, DT, ET
FabioDalForno	0.7275	0.7566	Normalization, imputation	Raw, V-C (6)	Random Forest
Yoda	0.7105	0.7433	-	Raw (33)	XGBoost
shafiullah	0.6022	0.6242	Imputation	Raw (19)	XGBoost

Raw = Features from raw dataset; V-C = Value-change features

a benchmark for future research. In addition, the winning solution in this competition yielded an anomaly detection model with an extraordinarily high AUC-ROC score of 0.9866, establishing a high classification performance benchmark for anomaly detection tasks. In particular, the winning classification model achieved a high-performance level with only 200 power meters for training, accounting for 14% of the whole LEAD dataset. This shows great potential of supervised learning for anomaly detection of energy data.

5 CONCLUSION AND FUTURE WORK

This study detailed and analyzed the winning solution in the LEAD competition for anomaly detection of cross-country power meters. In addition, the proposed method revealed the significance of value-changing features and established a benchmark for future research on supervised learning in energy data anomaly detection. However, there are still possible research gaps worthy of further exploration. First, the test dataset used to evaluate model performance in this competition is a random sampling of power meters, regardless of sites or countries. However, if the train and test data were split by sites or countries, could the anomaly detection model still accurately predict anomalies for meters in unseen locations? In addition, only 200 power meters with anomaly labels were used to train the classification model, which is a relatively small number compared to the whole dataset of approximately 1,400 meters. If the number of power meters used to train the model changes, at what point does the model's performance plateau or begin to decline? Filling these gaps in the future will make the proposed solution more comprehensive and influential on building management professionals.

ACKNOWLEDGMENTS

This research was supported by both Singapore Ministry of Education (MOE) Tier 1 Grant titled *Ecological Momentary Assessment (EMA) for Built Environment Research* (A-0008301-01-00) and the Singapore National Research Foundation (NRF) through the SinBerBEST2 program.

REFERENCES

[1] Beatriz Albiero, Estevo Uyrá, Ramon Vilarino, Juliano Silva, Tales Souza, Ricardo dos Santos, Sami Yamouni, and Renato Vicente. 2019. Employing gradient boosting and anomaly detection for prediction of frauds in energy consumption. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*. SBC, 916–925.

[2] Alfonso Capozzoli, Fiorella Lauro, and Imran Khan. 2015. Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications* 42, 9 (2015), 4324–4338.

[3] William Chung. 2012. Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings. *Applied energy* 95 (2012), 45–49.

[4] Manoj Gulati and Pandarasamy Arjunan. 2022. LEAD1.0: A Large-scale Annotated Dataset for Energy Anomaly Detection in Commercial Buildings. https://doi.org/10.48550/ARXIV.2203.17256

[5] Sonal Jain, Kushan A Choksi, and Naran M Pindoriya. 2019. Rule-based classification of energy theft and anomalies in consumers load demand profile. *IET Smart Grid* 2, 4 (2019), 612–624.

[6] Srinivas Katipamula and Michael R Brambley. 2005. Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I. *Hvac&R Research* 11, 1 (2005), 3–25.

[7] Varun Badrinath Krishna, Kiryung Lee, Gabriel A Weaver, Ravishankar K Iyer, and William H Sanders. 2016. F-DETA: A framework for detecting electricity theft attacks in smart grids. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 407–418.

[8] Dan Li, Yuxun Zhou, Guoqiang Hu, and Costas J Spanos. 2016. Fault detection and diagnosis for building cooling system with a tree-structured learning method. *Energy and Buildings* 127 (2016), 540–551.

[9] Guojun Mao, LJ Duan, and S Wang. 2005. Data mining theory and algorithm.

[10] Clayton Miller, Pandarasamy Arjunan, Anjukan Kathirgamanathan, Chun Fu, Jonathan Roth, June Young Park, Chris Balbach, Krishnan Gowri, Zoltan Nagy, Anthony D Fontanini, et al. 2020. The ASHRAE Great Energy Predictor III competition: Overview and results. *Science and Technology for the Built Environment* 26, 10 (2020), 1427–1447.

[11] Clayton Miller, Bianca Picchetti, Chun Fu, and Jovan Pantelic. 2022. Limitations of machine learning for building energy prediction: ASHRAE Great Energy Predictor III Kaggle competition error analysis. *Science and Technology for the Built Environment* (2022), 1–18.

[12] Jeffrey Schein, Steven T Bushby, Natascha S Castro, and John M House. 2006. A rule-based fault detection method for air handling units. *Energy and buildings* 38, 12 (2006), 1485–1492.

[13] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. 2018. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid* 10, 3 (2018), 3125–3148.

[14] Rui Yan, Zhenjun Ma, Yang Zhao, and Georgios Kokogiannakis. 2016. A decision tree based data-driven diagnostic strategy for air handling units. *Energy and Buildings* 133 (2016), 37–45.

[15] Sook-Chin Yip, Wooi-Nee Tan, ChiaKwang Tan, Ming-Tao Gan, and KokSheik Wong. 2018. An anomaly detection framework for identifying energy theft and defective meters in smart grids. *International Journal of Electrical Power & Energy Systems* 101 (2018), 189–203.