Anomaly Detection in Smart Meter Data for Preventing Potential Smart Grid Imbalance

Rituka Jaiswal*
rituka.jaiswal@uis.no
University Of Stavanger, Norway
Stavanger, Norway

Fadwa Maatug
University Of Stavanger, Norway
Stavanger, Norway
fadwamaatug@gmail.com

Reggie Davidrajuh
University Of Stavanger, Norway
Stavanger, Norway
reggie.davidrajuh@uis.no

Chunming Rong
University Of Stavanger, Norway
Stavanger, Norway
chunming.rong@uis.no

ABSTRACT

The households and buildings use almost one-third of the total energy consumption among all the power consumption sources. This trend is continuing to rise as more and more buildings install smart meter sensors and connect to Smart Grids and Micro Grids. Smart Grids use sensors and ICT technologies to prevent outages, power imbalance and minimize power wastage. Faults in appliances (like air conditioner duct leakage), abnormal appliances usage (like leaving heating iron, stoves on after usage), sensor faults and abnormal consumer behavior can lead to power outages. Studying the power consumption pattern of houses can lead to a substantial reduction in power wastage which can save millions of dollars. Research works also show that detecting such anomalies can result in preventing outages and save around 20% of power. In this work, we propose an anomaly detection approach for smart meter data for an open data set of houses from Ausgrid Corporation Australia, which is the largest distributor of electricity on Australia's east coast, providing power to 1.8 million consumers. The power consumption of a house is affected by various factors such as weather and temperature conditions, daily, weekly, yearly seasonality and, holidays. We propose an efficient machine learning-based algorithm to forecast and label power data with anomalies in the first part of this paper. In the second part, after generating the data set with anomaly labels, an efficient machine learning based classification method is proposed to classify power consumption data as either anomalous or normal. We achieve a G-mean score of 97.3% for the proposed classification algorithm. The run time of these classification models is also measured which is within 70 seconds. We performed our experiments on a low capacity Fog device rather than on a Cloud server.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Temporal reasoning**.

AICCC '21, December 17–19, 2021, Kyoto, Japan

ACM ISBN 978-1-4503-8416-2/21/12…$15.00
https://doi.org/10.1145/3508259.3508281

KEYWORDS

Smart Grid, Power outage, Facebook Prophet, Time series forecasting, Imbalanced data classification

ACM Reference Format:
Rituka Jaiswal, Fadwa Maatug, Reggie Davidrajuh, and Chunming Rong. 2021. Anomaly Detection in Smart Meter Data for Preventing Potential Smart Grid Imbalance. In *2021 4th Artificial Intelligence and Cloud Computing Conference (AICCC '21), December 17–19, 2021, Kyoto, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3508259.3508281

1 INTRODUCTION

The recent smart grid is an enhanced electrical network infrastructure that incorporates advanced metering and information communication technologies in order to increase the efficiency, reliability, and security of the power grid. Additionally, it enables the collection, transmission, and storage of real-time power data [1]. The smart grid is crucial for the economic growth, energy structure adjustment, and climate change adaptation, all of which can lead to energy savings and pollution mitigation [25]. Preventing and detecting power anomalies has become more necessary in today's global power system. Each year, the economy losses hundreds of millions of dollars caused by power wastage and outages [23, 24]. With the growth and popularity of Internet of things, demand and supply complexity, monitoring and forecasting of power is critical for power companies in terms of power generation, scheduling and dispatching. It benefits power consumers by allowing them to enhance their power usage schedules and thus reduce their costs. Additionally, power suppliers can detect abnormal meter readings caused by unforeseen meter failures, intentional meter manipulations, or users' unusual consumption behaviors [29, 31] and thus, reduce their costs. Discovering unusual meter measurements, appliance faults and, unexpected consumer behavior is known as Anomaly detection or outlier detection. Anomaly detection has been widely applied in a variety of applications, such as fraud detection and fault detection in safety-critical systems, medical diagnosis, etc. Anomaly detection can be applied to smart meter data to assist power consumers in identifying abnormal activities, such as faulty appliances, forgotten turned on appliances. Anomaly feedback will alert the consumers to reduce their energy consumption or replace inefficient appliances. Importantly, anomaly detection will assist power suppliers and utility companies in preventing outages, identifying energy wastage and unnoticed meter faults. It will help them establish a baseline for more precise demand-response management [20] for their clients.

1.1 The need for anomaly detection for power consumption data

In Smart Grids, smart meter records the power consumed by household appliances. The power consumption of every household is highly dynamic. It is dependent on various factors like outside temperature, Electric Vehicle charging needs, consumers daily, weekly and yearly usage patterns, holiday events, etc. Therefore, the power supply should adapt rapidly as per the changing demands [30]. Demand response management (DRM) try to reduce the power demand by the consumers, by, reducing the peak power demand from the demand side and, thus, preventing power outages and emergencies. Despite that, power outages are common as, the demand side data is not analysed and forecasted for anomalies. There have been numerous incidents of power outages across the world in the past [13]. During the winter storm in Texas in February 2021, nearly 5 million people lost power [6]. Also, in June 2021, due to the high power demand and insufficient supply, California grid operators suggested consumers to charge their electric vehicles during off-peak hours. In New York, for the first time power outages hit several neighborhoods due to high temperature rise and officials sent an emergency mobile alert to consumers urging them to save electricity [32]. To avoid these power outages, it becomes highly important to forecast the power anomalies. It can maintain Smart Grid balance and Demand Response Management can be achieved effectively. We release the labelled anomaly dataset as an open dataset for anyone to use and make our code and dataset available on GitHub.[1]. At the same time, we propose an efficient anomaly classification approach and verify it on our released anomaly labelled dataset.

1.2 Paper Structure

This paper consists of six sections. Section 2 will introduce the literature review of anomaly detection approaches. In Section 3, we describe the methodology to forecast power anomalies using the power domain knowledge applied on the Prophet algorithm. Also, we propose a classification method to classify future anomalies in the same section. In Section 4, we present the experimental results of the anomaly forecasting approach. In section 5, we present the classification approach experimental results and, finally, conclusions and future directions are presented in section 6.

2 RELATED WORKS

Statistical, clustering, and data mining approaches are the commonly used techniques for discovering abnormal consumption behaviors [17]. Liu and Nielsen [19] proposed a model for anomaly detection, where they first applied a mixture of supervised learning algorithm called Periodic Auto-Regression with eXogenous variables (PARX) and Gaussian statistical distribution to detect the anomalies on historical consumption data. Even though their results have identified certain anomalies based on temperature change, their suggested method could not detect long term daily and weekly seasonality, and, abnormal holiday behaviors. Fathnia et al.[9] provided a method that combined the Local Outlier Factor LOF index and Ordering Points To Identify the Clustering Structure OPTICS density-based algorithm to detect the unusual nature of the data. Their strategy recognised outliers as transmitted faults in the smart meter data to the control center. They reasoned that cyberattacks caused these failures. While the research findings demonstrated the efficiency of the proposed technique, they have not considered unnatural users activities or inefficient appliances causing anomalies. On the same lines, Zhang et al. [33] introduced an adaptive method to detect abnormalities in the smart meter data. They labelled the data set using the Gaussian Mixture Model Linear Discriminant Analysis GMM-LDA algorithm, which clusters some data sets to obtain the optimal feature representation of normal and abnormal patterns. Thereafter, they used the Particle Swarm Optimization Support Vector Machine (PSO-SVM) classifier to learn from the labelled data and predict future unlabelled data. Their approach also could not detect unexpected patterns due to special events and, could be complex to detect strong seasonal behaviors. In the research [16], the authors experimented with two approaches for detecting abnormalities in real and generated synthetic datasets. They started with a statistical technique, which uses Standard Deviation to identify extremes. The second technique utilised the K-Nearest Neighbor KNN clustering algorithm to distinguish anomalous from normal data using the point-to-point distance measure. Also, research work by Janetzko et al. [17] introduced two methods that adapted to the seasonality of the energy consumption data. The first approach estimated the error rate using weighted prediction, where they gave more substantial influence of the current measurements than older ones. The latter approach used similarity-based anomaly detection after transforming the daily pattern into the frequency domain by Fourier transformation. Additionally, they provided a range of time series visualisation techniques for resulting anomaly scores that aid in analysing and comprehending energy consumption behaviour. Research work [14] propose a prediction based approach for finding anomalies in the power consumption data. The drawback of this approach is that their model does not identify all the seasonality patterns and therefore, misses some of the anomalies. Nevertheless, a combined drawback of all the above papers is that they have not considered some anomalies, by not considering one or more of these factors: all kinds of seasonality impact, holidays effect, long term seasonality and trend change, did not scaled well for the data. This could have significant impact on the grid balance and will effect decision making for power companies. In light of the shortcomings of the mentioned prior works, we present a simple and highly effective technique to detect a wide range of anomalies considering the strong seasonality and could overcome all the shortcomings mentioned above.

3 METHODOLOGY

Prophet, is a Facebook-developed open source library mainly used to analyze time-series data and excels at forecasting highly periodic data [26]. We identify the anomalies by using the Prophet model's confidence intervals. Any data point that lies outside the interval is considered as an anomaly. Further, since zero power consumption could indicate sensor fault, we classified all zero values in the data

[1]https://github.com/ritu2012/Anomaly-Detection-Code-and-Anomaly-Labelled-Dataset-on-Ausgrid-Data

as anomaly. After classifying the data into two categories (positive class indicating anomaly, negative class indicating normal data), we get an unbalanced data set. Thus, in the second phase of our study, we propose the best classification model that classifies the power consumption data as either normal or anomaly. We evaluate our models performance using Sensitivity, Specificity, G-mean, F-score, AUC values and model run-time metrics. This approach can be tested on any anomaly labelled dataset. In below subsections, we describe our approaches in detail.

3.1 The Prophet Method

The Prophet is a Generalize Additive Model(GAM) that Facebook has devised and is available as an open-source library. It is a decomposable time series model [11] with three main model components: trend, seasonality, and holidays. It provides with the ability to make fast, time series predictions with good accuracy using simple intuitive parameters and has support for customizing seasonality, holidays and behavior based on the domain knowledge (power domain in our case) [8, 18].

The Prophet core is an additive regression model that accommodates three components as indicated by the following mathematical notation [10, 12, 27].

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t) \tag{1}$$

where,

- $g(t)$ is the trend function that represents the linear or logistic non-periodic changes in power consumption readings obtained at evenly time-space intervals.
- $s(t)$ is a seasonal component formed using Fourier series, and it captures the historical data periodic changes, which can be daily, weekly, monthly or yearly seasonality. In our case, we identified daily, weekdays, weekends, yearly seasonality, so models regression equation will be [28]:

$$\begin{aligned} y(t) = g(t) + s(t)_{daily} + s(t)_{weekdays} + s(t)_{weekends} \\ +s(t)_{yearly} + h(t) + \epsilon(t) \end{aligned} \tag{2}$$

where,

- $h(t)$ represents the holidays during the year and can be provided by the user or unusual expected times that happen irregularly.
- $\epsilon(t)$ indicates an independent error term, which is assumed to be normally distributed.

We considered an uncertainty interval by taking the *99-th* quantile of the posterior predictive distribution, which has a 99% chance of containing the parameter's actual value [21]. Besides, any value that lies outside the interval is identified as an anomaly. We define the uncertainty interval $\boldsymbol{I}$ as:

$$I = [\hat{y}_{lower}, \hat{y}_{upper}] \tag{3}$$

We established the following decision rule for an original data sample y [3]:

$$\begin{cases} \text{y is normal, if } y \in I \\ \text{y is anomaly, if } y \notin I \end{cases} \tag{4}$$

3.2 Anomaly Classification Approach

In the two-class (binary) imbalanced classification problem, the minority (underrepresented) group is typically referred to as the positive class, while the dominant group is the negative class [2]. The classification algorithm's performance is often assessed by comparing the predicted class labels to the real ones. The model is trained on the training set and then evaluated on the holdout test set. The standard evaluation metrics, such as Accuracy (which is the percentage of accurately categorised samples), can not be used for the classification problem because in an imbalanced classification problem, the majority class has more data compared to the minority class and therefore, the model is biased towards the majority class [2, 4].

Before presenting the evaluation metrics for the classification task that we used, we need to understand the Confusion Matrix for binary classification tasks. The confusion matrix can provide more insight into the model performance and gives information about the correctly and incorrectly predicted labels [4]. It has four possibilities: A true positive(TP) is an outcome where the model correctly predicts the positive class. Similarly, a true negative(TN) is an outcome where the model correctly predicts the negative class. A false positive(FP) is an outcome where the model incorrectly predicts the positive class. And a false negative(FN) is an outcome where the model incorrectly predicts the negative class. In our power anomaly context, False Negatives should be focused, since it shows how many true anomalies are predicted as normal samples and, we want to make it as low as possible.

- **Sensitivity** indicates the accuracy with which the positive class was anticipated and it is suitable when the objective is to minimise false negatives.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

- **Specificity** indicates the accuracy with which the negative class was anticipated and it is suitable when the objective is to minimise false positives.

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

- **G-mean** combines Sensitivity and Specificity into a single score that takes both into consideration.

$$G - mean = \sqrt{Sensitivity \times Specificity} \tag{7}$$

- **ROC Curves** The Receiver Operating Characteristic (ROC) curve is a graphical evaluation technique that widely used for summarising classifier performance over a range of true positive and false positive error rates trade-offs. ROC demonstrates that the true positive rate cannot be increased without raising the false positive rate for any classifier. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. The higher the AUC score, the more accurate the classifier.

4 EXPERIMENTAL RESULTS

The primary purpose of this study is to detect anomalies in power consumption data and find the best model to classify future power

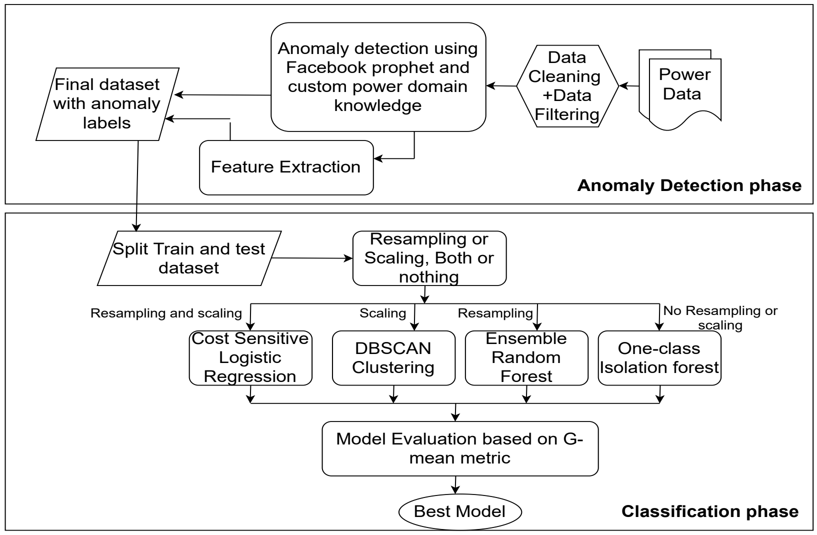

Figure 1: Workflow Anomaly detection and classification approach.

anomalies in an imbalanced data. In this section, we present the experiments results. Figure 1 is a graphical presentation of the workflow, divided into two stages. As shown in the figure, the first phase starts with collecting and doing exploratory data analysis, followed by data cleaning and feature extraction. The first phase's outcome is a data set that contains the extracted features and the labelled data with anomalies. After this point, the second phase begins by splitting the dataset into two sets, Training and Testing. Following this, the data standardization is done. After that, we use four ML models: Logistic Regression, DBSCAN, Ensemble RF and, Isolation Forest to identify the best model based on metrics.

4.1 Ausgrid Dataset

We used the open dataset by the largest distributor of electricity on Australia's east coast called Ausgrid. The data was collected from 300 randomly selected consumers between 1 July 2010 and 30 June 2013, sampled at half-hour interval. The Ausgrid corporation used three separate meter recording devices for three different categories Gross Generation GG, General Consumption GC, Controlled Load CL. The first meter (GG) records the solar power generated by the solar PV units placed at the rooftop of each household. The second meter records the daily power consumption (GC) in (KWh) for each consumer. And, the last meter recorded the power consumption with water heating placed in some of the houses by offering a monetary incentive [22]. The dataset contains 54 columns. These are columns: consumer IDs, Postcode, Generator Capacity, Consumption Category and Date, followed by 48 columns of half-hour power meter data. We used consumption category GC, since we are interested in finding anomalies in the power consumption.

4.1.1 Data Filtering. To demonstrate the model and approach, we picked only ten consumers to represent a small community and minimise the time required for the experimentation for all 300 consumers. We analysed the Postcode feature and selected the area with the most consumers. Our files contain postcodes ranging from 2008 to 2330, which are of New South Wales (NSW) state. Additionally, as seen in Figure 2, the largest area is located between Newcastle and Sydney; and its postcode value is 2259. We selected the top ten consumers from this postcode. The ten selected consumers' IDs are 7, 29, 30, 64, 155, 160, 184, 202, 206 and, 215.

4.1.2 Exploratory Data Analysis. To obtain a summary of the energy consumption of the consumers, we started by aggregating the data annually using the summation method for each consumer. Figure 3 shows the overall households consumption over two years. It indicates how the individual consumers behaviour varies from other consumers. The power consumption was highest for the year 2012, and this could be due to comparatively low temperatures during winter. This consumption behavior can potentially cause grid imbalance/outages if not forecasted in advance.

Furthermore, we aggregated all the consumers in one and found out the total power consumed for all the four seasons of the year. In Australia, the peak Winter period months are (June-July-August), the Spring period includes (September-October-November), the Summer period includes (December-January-February), and the Autumn period includes (March-April-May). Figure 4 shows the peak power consumption in seasons of 2011, 2012 and 2013. People tend to use air conditioning in the summer and heating in the winter. The winter of 2012 recorded the maximum power consumption while the winter of 2011 recorded the lowest consumption value.

Further, the consumer no. 7's total consumption was aggregated for a week for a period of one year. Figure 5 shows the plot, which shows that the highest consumption was in month July (winter season), while the lowest demand was in September and October (autumn) months. Figure 6 the power consumption aggregated for 24 hours for the seven days of a week for consumer 7. The darker the color blue, the higher the demand.

As seen there is a higher variation in power consumption throughout the 24 hours of the day. Higher power consumption occurs during the evenings between 5 p.m and 9 p.m, reaching more that 1000 KWh or more at 6 p.m; while it decreases at least by 70 per cent (from 1000 to 300 KWh) after midnight and until 6 a.m.

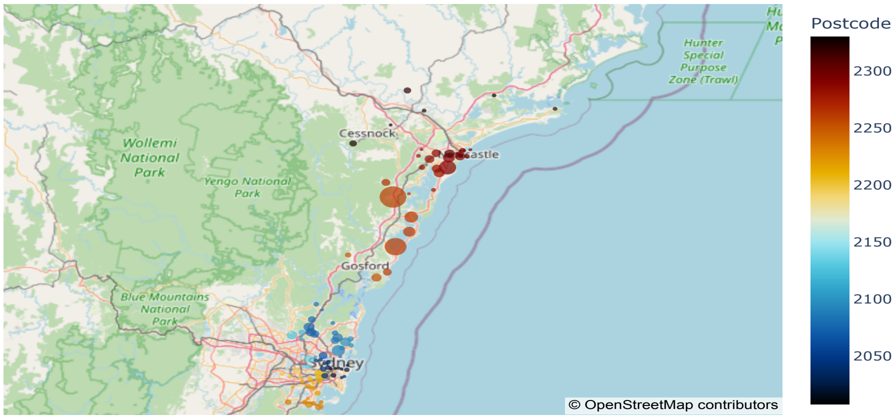

Figure 2: The distribution map of Ausgrid. The circles represent postcode regions covered in the 300-consumer dataset.

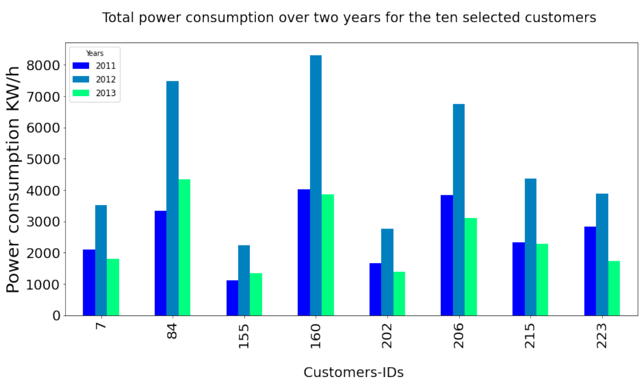

Figure 3: Total power consumption over two years for ten consumers.

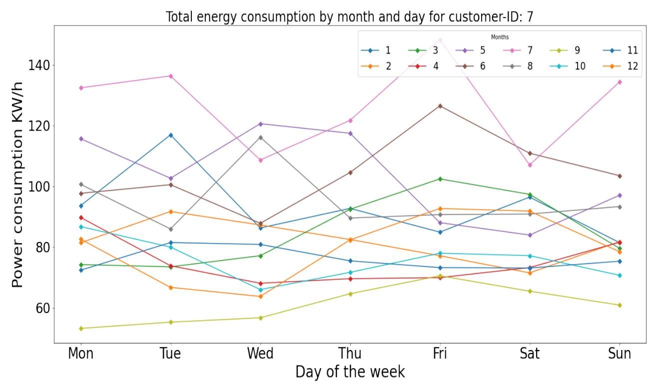

Figure 5: Total power consumption aggregated for week for a period of one year for consumer number 7

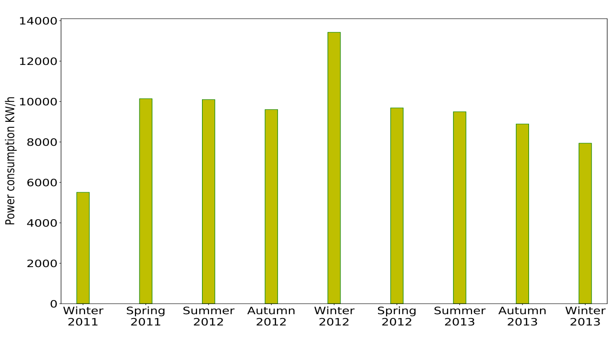

Figure 4: Total peak power consumption over the four seasons.

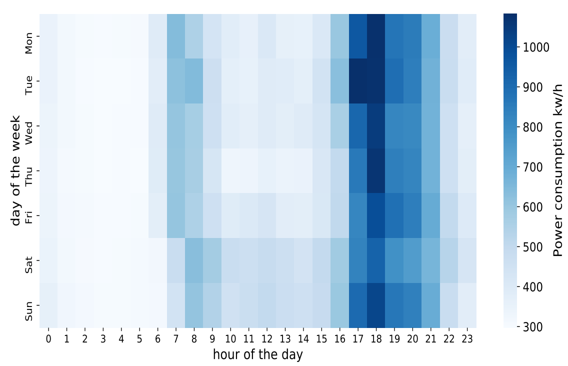

Figure 6: Power consumption aggregated for 24 hours for the seven days of a week.

4.1.3 Feature Extraction. Feature extraction is essential to ensure that our ML algorithms perform optimally. Through the data exploration and analysis step describe above, we were able to derive the following important features: *hour-of-day*, *day-of-week*, *month-of-year*, and *year*.

4.2 Anomaly Identification

We employed the *Prophet* method to analyse and identify anomalies by applying the anomaly decision rules mentioned in previous

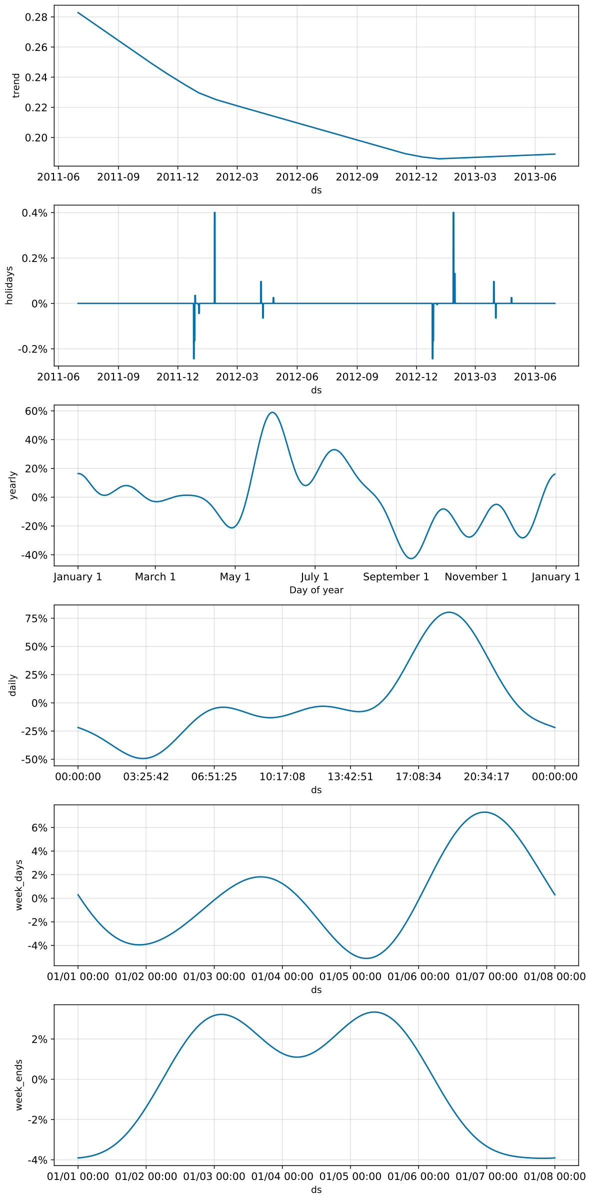

Figure 7: Trend, daily, weekdays, weekends and yearly seasonality and, holidays pattern.

sections. After fitting the model, predictions are done for a period referred to as the "horizon". The "cutoff" points were defined as follows:

The "initial" training period is chosen to be three times the length of the "horizon" forecasting period, and its size is required to be large enough to capture the seasonality being studied. Since the yearly seasonality is included in the model, the initial period must be at least one year in length [7]. Cutoffs were made every half a "horizon"[7]. With the range of two years data and taking the above considerations, we determined an initial training period as 18 months, a horizon forecasting period as six months and a single cutoff point of "01-12-2012". The Prophet provides a performance metrics tool that can be used to select the hyperparameter combination set with the lowest evaluation measure. This utility analyses each hyper-parameter combination forecast and calculates a valuable statistic on the prediction performance, the root mean squared error (RMSE) score. To tune the hyper-parameters, we created a dictionary of all rational hyperparameter combinations. The parameters included in the tuning are: **change_point_range**, **changepoint_prior_scale**, **seasonality_prior_scale**, **holiday_prior_scale** and, **seasonality_mode**. After tuning the hyper-parameters, we created custom functions for weekday and weekends seasonalities. Finally, we forecasted the power consumption.

The year wise trend changes is visualised in Fig 7, starting with a higher overall power consumption in 2011 and downward growth reaching year 2013. The holidays component is shown in part two of the figure 7, while in part 3 of the same figure, the yearly seasonality is shown, part 4 shows the daily seasonality and the weekends and weekdays consumption pattern is shown in part 5 and 6 of the figure 7.

For the Prophet model, we chose a 99% uncertainty interval to account for any unexpected events requiring additional energy. Figure 8 shows the result of the Prophet tuned model with anomalies for consumer-ID 7. As seen in the figure the yellow region represents the uncertainty interval, and the black dots denotes the actual power consumption for consumer-ID 7 during a a period of six months. The green circles represent anomalies with a diameter proportionate to their distance from the interval range $[\hat{y}_{lower}, \hat{y}_{upper}]$.

We exhibit a subset of the data to illustrate the abnormalities for shorter time periods in greater depth. Figure 9 depicts actual consumption data over three months; on the other hand, figure 10 depicts anomalies over the same time period.

5 ANOMALY CLASSIFICATION APPROACH AND EXPERIMENTAL RESULTS

Figure 11 depicts how the data is disproportionately distributed among the two classes.

We have 96.6% values for the negative class and 3.4% for the positive class. To sample the data for accurate classification, we used the SMOTE-ENN class from Scikit-imbalanced-learn library. The SMOTE function generates new instances of the minority class, while ENN reduces noisy points along with class borders. We used the sampling techniques only with the train set and not the test set. After applying the sampling techniques, the balanced distribution is shown in figure 12 which shows a significant improvement over the imbalanced dataset.

We examined four effective models with the balanced training set. These are Logistic Regression, DBSCAN, Random forest and Isolation forest. Three runs of 10-fold stratified cross-validation are run to validate the models. The F1-score is computed in each iteration of cross-validation. Subsequently, the four models are compared by the model's G-mean scores and their execution times. DBSCAN model from the Scikit-learn library requires tuning of two critical parameters, namely eps and min_samples. Additionally, the default metric parameter option is set to "euclidean", which uses the

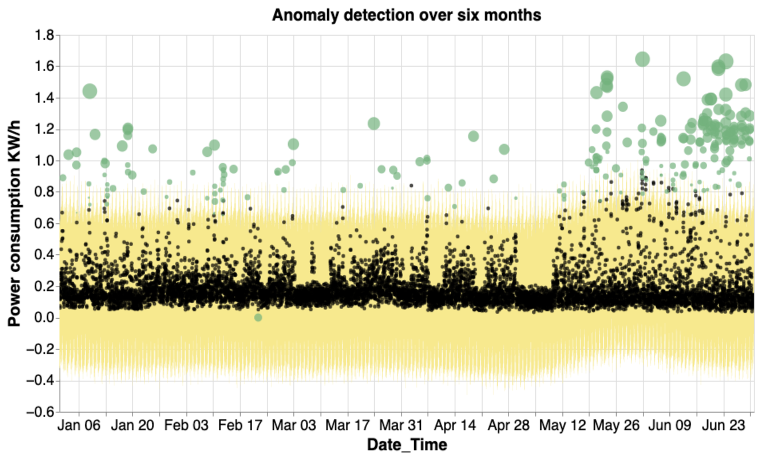

Figure 8: Actual and Anomalous power consumption over a period of six months for consumer number 7.

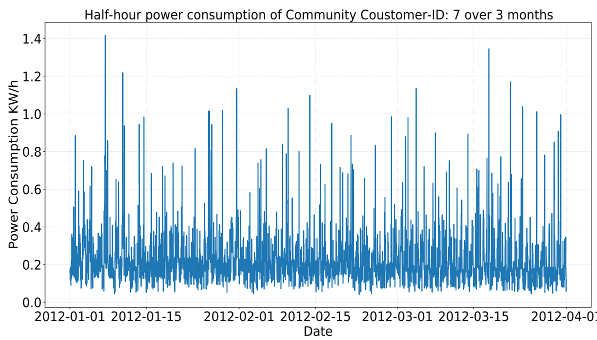

Figure 9: Actual power consumption for three months.

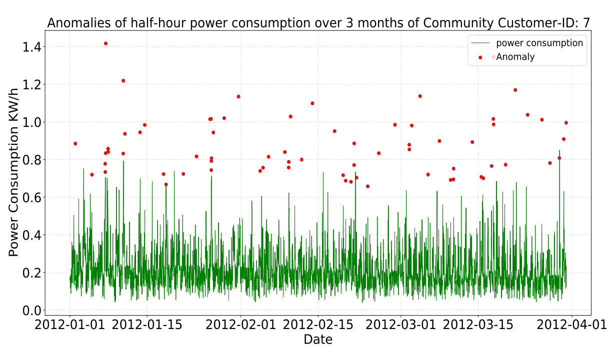

Figure 10: Actual and Anomalous power consumption for a period of three months

Euclidean distance measure to determine the distance between the samples. The RandomForestClassifier class from Scikit-learn is used for implementing the ensemble random forest algorithm. Also, the IsolationForest isolates data points by randomly choosing a feature

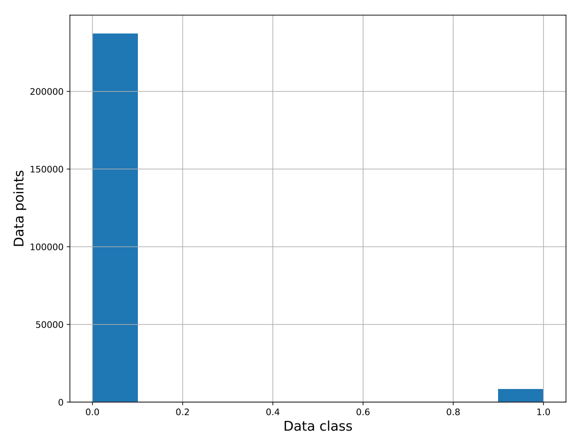

Figure 11: Imbalanced data of Actual and Anomalous power consumption.

and a split-value in a range between [minimum value, maximum value] of the selected feature. Further, the number of splitting required to isolate a sample is set to the path's length from the root to the terminal node. Additionally, random partitioning significantly reduces the pathways taken by anomalies, and any specific sample with shorter path lengths is considered to be anomalous [5]. We applied Grid search and Bayesian optimisation for hyperparameter optimization for our two models: logistic regression and DBSCAN.

5.1 Classification Models Evaluation

To evaluate the models, we created a function called evaluation that computes all the evaluation measures discussed in section 3.2. We started by using the Confusion Matrix metric to compute and return the True Negative TN, False Positive FP, False Negative FN, and True Positive TP values. After that, the Sensitivity, Specificity,

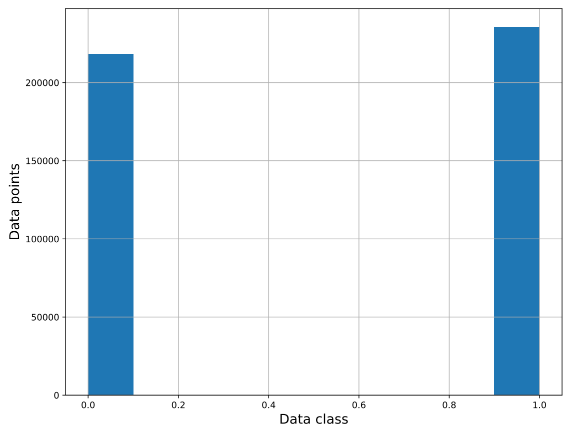

Figure 12: Balanced data distribution of Actual and Anomalous power consumption after applying sampling techniques on imbalanced data.

G-mean, True Positive Rate TPR and, False Positive Rate FPR scores, F1 and F2 scores were computed.

We used the roc_curve and roc_auc_score functions from Scikit-learn library. The first function enables us to visualise the ROC curve of our models, which calculated the probabilities of the estimated class labels (all the models except DBSCAN). The second function computes the area under the ROC curve, whose scores are presented on the ROC plot to have a clear visual evaluation of the models.

The Sensitivity, Specificity, G-mean and F1 scores are shown in table 1. As seen, ensemble RF model had the highest Sensitivity, Specificity, and G-mean score and F-1 scores among all the models. Further, the DBSCAN and iForest also produced almost comparable results slightly less than ensemble RF model. Figure 13 shows the true positive rate TPR and false positive rates FPR for each of the four models.

Models	Sensitivity	Specificity	G-Mean	F1-score
Logistic Regression	0.911	0.932	0.921	0.946
DBSCAN	0.921	0.896	0.908	0.924
Ensemble RF	**0.972**	**0.975**	**0.973**	**0.978**
One-class iForest	0.926	0.895	0.91	0.924

Table 1: Sensitivity, Specificity, G-mean and F1 scores for the four models.

The capability of ensemble RF to detect true positives (TPR) is the highest, while its capability to misclassify negatives (FPR) is the lowest among the four models. Further, seen in the figure, DBSCAN and iForest performed equally and had a higher TPR score than the Logistic regression model. Figure 14 demonstrates that ensemble RF outperforms the other models and has the highest F2-score.

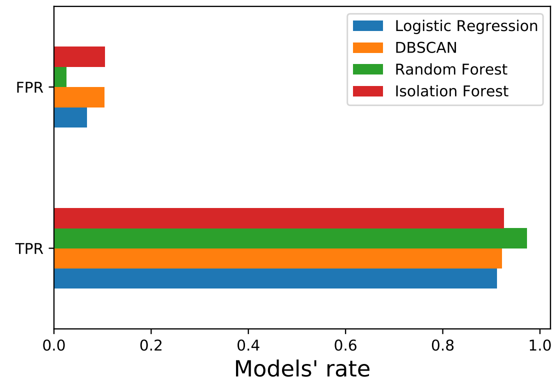

Figure 13: Comparison between True Positive Rates and False Positive Rates for the models.

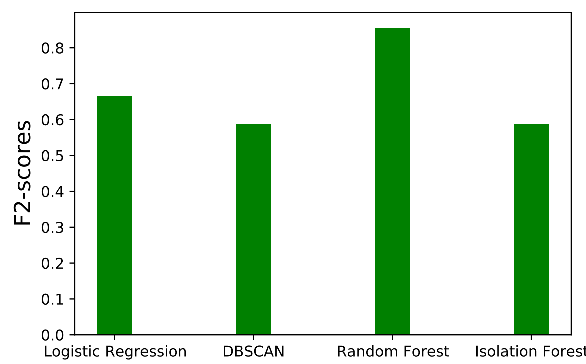

Figure 14: F2-scores for the models.

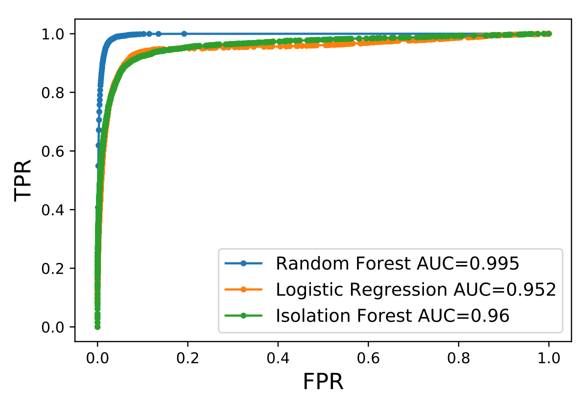

Figure 15: ROC curve and AUC values for the models.

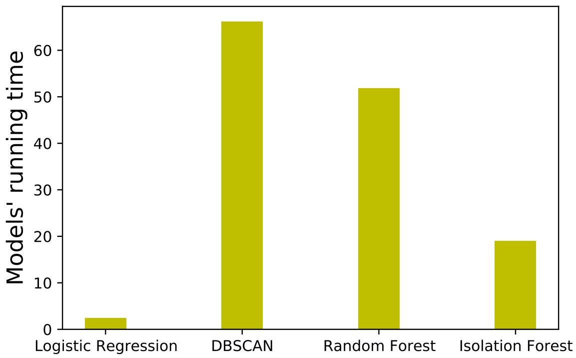

Figure 16: Run time on the Fog device for the models.

The receiver operating characteristic (ROC) curve is utilized to evaluate the models (except DBSCAN). The ROC curves and AUC values for the Random Forest, iForest, and logistic regression models are shown in figure 15. We observe that the RF produces highest score from the figure, whereas logistic regression has the lowest score. We also measured the run time of all these models which is shown in figure 16. We notice that logistic regression had the least execution time, while DBSCAN took the longest time to execute. To summarize, as seen by the classification metric results, except for the model run-time, the ensemble random forest model outperformed all other models and had the best classification accuracy.

We performed the experiments on a low computing capacity Fog device and not on high computing Cloud server. Fog device are suitable for processing huge sensor data, as the number of smart meters increases from thousands to millions, the current state-of-the-art centralized data processing architecture will no longer be sustainable with such a big data explosion [15]. For the evaluation of the models, we used a **Fog device** with a capacity of 2 GHz Quad-Core Intel Core i5 processor and 16 GB RAM.

6 CONCLUSION AND FUTURE WORK

In this paper, we discussed the need of effective power forecasting for anomaly detection to prevent potential power outages and maintain Grid balance. We proposed a effective and simple anomaly detection approach using Facebook Prophet library. The approach is fast, scalable and, the parameters are easily interpretable. This approach is first time used on a real-world power dataset and, anomaly labels are created. We release the anomaly dataset for anyone to use and research. Also, we proposed a machine learning based classification approach to classify power anomalies which achieved a G-mean score of 97.3 percent. For the future work, we plan to utilize and analyze the Ausgrid dataset's solar power generation data since, the houses have the provision of producing clear renewable power through solar panels. We plan to predict the local power produced by the solar panels. with accurate estimate of local power production, the community can sustain better on the power produced and, operate in MicroGrid mode. This can reduce the dependency of the community on the main Grid and help maintain power balance.

REFERENCES

[1] Fadi Al-Turjman and Mohammad Abujubbeh. 2019. IoT-enabled smart grid via SM: An overview. *Future Generation Computer Systems* 96 (2019), 579–590.

[2] Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. 2018. *Learning From Imbalanced Data Sets*. Springer, Cham, Switzerland. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1920612&scope=site

[3] Anastasios Bellas, Charles Bouveyron, Marie Cottrell, and Jerome Lacaille. 2014. Anomaly Detection Based on Confidence Intervals Using SOM with an Application to Health Monitoring. *arXiv:1508.04154 [stat]* 295 (2014), 145–155. https://doi.org/10.1007/978-3-319-07695-9_14 arXiv: 1508.04154.

[4] Jason Brownlee. 2021. *Imbalanced Classification with Python* (v1.3 ed.). Jason Brownlee. https://machinelearningmastery.com/imbalanced-classification-with-python/

[5] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake Vanderplas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. 2013. API design for machine learning software: experiences from the scikit-learn project. *arXiv:1309.0238 [cs]* (Sept. 2013). http://arxiv.org/abs/1309.0238 arXiv: 1309.0238.

[6] James Doss-Gollin, David J Farnham, Upmanu Lall, and Vijay Modi. 2021. How unprecedented was the February 2021 Texas cold snap? *Environmental Research Letters* 16, 6 (jun 2021), 064056. https://doi.org/10.1088/1748-9326/ac0278

[7] Open Source Facebook. 2021. Quick Start. http://facebook.github.io/prophet/docs/quick_start.html

[8] Wen-Xiang Fang, Po-Chao Lan, Wan-Rung Lin, Hsiao-Chen Chang, Hai-Yen Chang, and Yi-Hsien Wang. 2019. Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: the Morgan Taiwan Index. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. 1–2. https://doi.org/10.1109/ISPACS48206.2019.8986377 ISSN: 2642-3529.

[9] Froogh Fathnia, Farid Fathnia, and D. B. Mohammad Hossein Javidi. 2017. Detection of anomalies in smart meter data: A density-based approach. In *2017 Smart Grid Conference (SGC)*. 1–6. https://doi.org/10.1109/SGC.2017.8308852

[10] Cynthia Freeman, Jonathan Merriman, Ian Beaver, and Abdullah Mueen. 2019. Experimental Comparison of Online Anomaly Detection Algorithms. In *Experimental Comparison of Online Anomaly Detection Algorithms*. Artificial Intelligence Research Society Conference, Sarasota, Florida, USA.

[11] Andrew C Harvey and Simon Peters. 1990. Estimation procedures for structural time series models. *Journal of forecasting* 9, 2 (1990), 89–108.

[12] Md. Mehedi Hasan Shawon, Sumaiya Akter, Md. Kamrul Islam, Sabbir Ahmed, and Md. Mosaddequr Rahman. 2020. Forecasting PV Panel Output Using Prophet Time Series Machine Learning Model. In *2020 IEEE REGION 10 CONFERENCE (TENCON)*. 1141–1144. https://doi.org/10.1109/TENCON50793.2020.9293751 ISSN: 2159-3450.

[13] Michael F Hordeski. 2020. *Emergency and Backup Power Sources:: Preparing for Blackouts and Brownouts*. River Publishers.

[14] Rituka Jaiswal, Antorweep Chakravorty, and Chunming Rong. 2020. Distributed Fog Computing Architecture for Real-Time Anomaly Detection in Smart Meter Data. In *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*. 1–8. https://doi.org/10.1109/BigDataService49289.2020.00009

[15] Rituka Jaiswal, Reggie Davidrajuh, and Chunming Rong. 2020. Fog Computing for Realizing Smart Neighborhoods in Smart Grids. *Computers* 9, 3 (2020). https://doi.org/10.3390/computers9030076

[16] Vikramaditya Jakkula and Diane Cook. 2010. Outlier detection in smart environment structured power datasets. In *2010 sixth international conference on intelligent environments*. IEEE, 29–33.

[17] Halldór Janetzko, Florian Stoffel, Sebastian Mittelstädt, and Daniel A. Keim. 2014. Anomaly detection for visual analytics of power consumption data. *Computers & Graphics* 38 (2014), 27–37. https://doi.org/10.1016/j.cag.2013.10.006

[18] Bineet Kumar Jha and Shilpa Pande. 2021. Time Series Forecasting Model for Supermarket Sales using FB-Prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. 547–554. https://doi.org/10.1109/ICCMC51019.2021.9418033

[19] Xiufeng Liu and Per Sieverts Nielsen. 2016. Regression-based online anomaly detection for smart grid data. *arXiv preprint arXiv:1606.05781* (2016).

[20] Matteo Muratori and Giorgio Rizzoni. 2015. Residential demand response: Dynamic energy management and time-varying electricity pricing. *IEEE Transactions on Power systems* 31, 2 (2015), 1108–1117.

[21] Jeremy Oakley. 2021. *Chapter 5 Interval estimates and confidence intervals | MAS113 Part 2: Data Science*. BOOKDOWN. http://www.jeremy-oakley.staff.shef.ac.uk/mas113/notes/interval-estimates-and-confidence-intervals.html

[22] Elizabeth L. Ratnam, Steven R. Weller, Christopher M. Kellett, and Alan T. Murray. 2017. Residential load and rooftop PV generation: an Australian distribution network dataset. *International Journal of Sustainable Energy* 36, 8 (Sept. 2017), 787–806. https://doi.org/10.1080/14786451.2015.1100196

[23] D.; Feng M. Y.; Llana P.; Quartararo L. Roth, K. W.; Westphalen. 2005. *Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential.* Retrieved October 4, 2021 from https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB2006100567.xhtml

[24] Thomas B Smith. 2004. Electricity theft: a comparative analysis. *Energy policy* 32, 18 (2004), 2067–2076.

[25] Jennie C Stephens, Elizabeth J Wilson, Tarla R Peterson, and James Meadowcroft. 2013. Getting smart? climate change and the electric grid. *Challenges* 4, 2 (2013), 201–216.

[26] Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician* 72, 1 (2018), 37–45.

[27] Karthick Thiyagarajan, Sarath Kodagoda, Nalika Ulapane, and Mukesh Prasad. 2020. A Temporal Forecasting Driven Approach Using Facebook's Prophet Method for Anomaly Detection in Sewer Air Temperature Sensor System. In *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. 25–30. https://doi.org/10.1109/ICIEA48937.2020.9248142 ISSN: 2158-2297.

[28] Ritchie Vink. 2018. Build Facebook's Prophet in PyMC3; Bayesian time series analyis with Generalized Additive Models - Ritchie Vink. https://www.ritchievink.com/blog/2018/10/09/build-facebooks-prophet-in-pymc3-bayesian-time-series-analyis-with-generalized-additive-models/

[29] Xiaohui Wang, Ting Zhao, He Liu, and Rong He. 2019. Power consumption predicting and anomaly detection based on long short-term memory neural network. In *2019 IEEE 4th international conference on cloud computing and big data analysis (ICCCBDA)*. IEEE, 487–491.

[30] Ye Yan, Yi Qian, Hamid Sharif, and David Tipper. 2013. A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements and Challenges. *IEEE Communications Surveys Tutorials* 15, 1 (2013), 5–20. https://doi.org/10.1109/SURV.2012.021312.00034

[31] Ibrahim Yilmaz and Ambareen Siraj. 2021. Avoiding occupancy detection from smart meter using adversarial machine learning. *IEEE Access* 9 (2021), 35411–35430.

[32] Mihir Zaveri and Ashley Wong. 2021 (accessed July 12, 2021). 'Conserve Energy': New York City Begs Residents to Help Avoid Outages. https://www.nytimes.com/2021/06/30/nyregion/nyc-energy-alert-heatwave.html.

[33] Leping Zhang, Lu Wan, Yong Xiao, Shuangquan Li, and Chengpeng Zhu. 2019. Anomaly Detection method of Smart Meters data based on GMM-LDA clustering feature Learning and PSO Support Vector Machine. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. 2407–2412. https://doi.org/10.1109/iSPEC48194.2019.8974989