Title
Unsupervised Learning for Online AbnormalityDetection in Smart Meter Data

Authors
Aligholian, Armin
Farajollahi, Mohammad
Mohsenian-Rad, Hamed

Publication Date
2019-08-01

Peer reviewed

Unsupervised Learning for Online Abnormality Detection in Smart Meter Data

Armin Aligholian, Mohammad Farajollahi, and Hamed Mohsenian-Rad
Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA
e-mails: aalig002@ucr.edu, mfara006@ucr.edu and hamed@ece.ucr.edu

***Abstract*—The analysis of abnormalities in smart meter data has applications in load forecasting, cyber security, fault detection, electricity theft detection, demand response, etc. Abnormality is broadly defined in this paper as any unusual electricity consumption *instance* or *trend* that falls outside of the normal usage patterns for each load, whether in terms of magnitude, time of usage, etc. Unusual electricity consumption can have different signatures and different duration of time. This paper aims to evaluate the performance of four unsupervised machine learning methods for abnormality detection on real-world smart meter data, namely prediction-based regression, prediction-based neural network, clustered-based, and projection-based methods. Different types of features, such as load-based, contextual, and environmental, are investigated to construct the data-driven models. It is shown that different abnormality detection methods have different ability for detecting different types of abnormalities; and their performance depends on the set of features used for training the method. Accordingly, different types of features are scrutinized for each abnormality detection method.**

***Keywords*: Data-driven analysis, abnormality detection, smart meter data, feature selection, unsupervised learning.**

I. Introduction

The deployment of smart meters has provided system operators with an unprecedented level of visibility over the distribution networks and customer loads, with a multitude of applications, c.f. [1]. However, it is now a challenge to handle the tremendous growth in the volume and velocity of data that is generated by the smart meters in the load sector. Therefore, it is necessary to find ways to the extract the most useful parts of the data and transform them to actionable information.

This paper aims to conduct a data-driven study of smart meters using real-world data streams from Pecan Street project in Austin, TX [2] to identify load abnormality. Abnormality in the context of this paper is broadly defined as any unusual electricity consumption *instance* or *trend* that falls outside of the normal power consumption patterns for a load or load sector, whether in terms of magnitude, time, duration, etc. [3]. The analysis of abnormalities in smart meter data streams is of great interest to several applications, such as load forecasting [4], cyber attack detection [5], fault and outage detection [6], electricity theft detection [7], demand response [4], etc.

Our approach in this paper is based on machine learning. Accordingly, this study is in its broad sense comparable with those in [8]–[11]. A deep semi-supervised convolutional neural network with confidence sampling is proposed in [8]. Also, a supervised ensemble-based method with sliding window is proposed in [9]. However, when it comes to abnormality detection, we must deal with an inherently unsupervised learning problem because abnormalities do *not* have a known paradigm; they are rather determined in comparison with the history of data. Therefore, the usage of unsupervised learning methods are more practical than supervised and semi-supervised methods. In [10], a prediction-based unsupervised abnormality detection method is proposed that comprises a dynamic regression model and an adaptive abnormality threshold. In [11], an unsupervised clustering-based algorithm on the low-dimensional dissimilarity matrix is used to detect irregular power consumption. However, these two studies do not consider the role of feature selection in training their model. They also only consider specific types of abnormality.

The above open problems are addressed in this paper, where the focus is on *unsupervised* machine learning methods. The main contributions in this paper can be summarized as follows:

1) This paper provides a systematic comparative study of four different *unsupervised* machine learning methods to understand how different methods can best serve to detect different types of abnormalities in real-world smart meter data. Specifically, we examine load prediction regression-based, load prediction neural-network-based, clustered-based, and projection-based methods for abnormality detection and compare their performance.
2) Different features are investigated for different methods to obtain the *best combination of features* for each method. An important conclusion is that, when it comes to historical load features, they are useful in prediction-based methods for the purpose of finding abnormal load *trends*; while they are also useful in cluster-based methods for the purpose of finding abnormal load *instances*. In addition, cluster-based methods can use a proper combination of historical load features, contextual features, and environmental features to simultaneously identify both abnormal load trends and abnormal load instances.
3) To speed up detection, all methods are implemented in *online* mode, where the models are updated upon the arrival of new data. To the best of our knowledge, this is the first study to address the application of Isolation Forest (IF) and Lightweight On-line Detector of Anomalies (LODA), as two computationally efficient online methods, to do abnormality detection in smart meter data. The models are adjusted to meet the needs of the data-driven application domain in this paper.

II. Methodology

In this section, we describe the unsupervised online abnormality detection approach, based on four different methods.

A. Feature Selection

Broadly speaking, the features of electricity power consumption data can be categorized into three generic groups: load-based features, contextual features, and environmental features. It is vital to identify the right choices of features within each category that are most informative with respect to the specific problem and the specific data-driven method.

1) Load Based Features: These features account for the power consumption of residential household in different time steps. They are obtained from historical power consumption data with different time lags. The followings are the set of load based features that we consider in this study:

$$\begin{aligned} L^t &= \{P^t, P_Y^t, P_W^t, P_M^t\}, \\ L^w &= \{P^{t-24}, P^{t-23}, \cdots, P^{t-1}\}, \end{aligned} \tag{1}$$

where L^t is the set of historical load data at time t. In this set, P^t is power consumption at time t, P_Y^t is power consumption yesterday at time t, P_W^t is power consumption in the last two weeks at time t, and P_M^t is the mean of power consumption at time t. As for L^w, it is the set of previous 24 hours.

2) Contextual Features: These features are not specific to power consumption, but they do have indirect impact on power consumption. Time of the day, day of the week, weekends versus weekdays flag, holidays, and season of the year are instances of contextual information, as listed below:

$$C = \{T_d^t, D_w^t, W_s^t, H^t, S^t\}. \tag{2}$$

3) Environmental Features: Electricity consumption in some appliances such as heating ventilation and air conditioning (HVAC) systems depend on some environmental features such as temperature. Therefore, total power consumption of household are affected by these features which in this study considered as set E and comprises of temperature ($Temp^t$) and humidity (Hum^t) factors as illustrated below:

$$E = \{Temp^t, Hum^t\}. \tag{3}$$

These three features can be correlated. This may affect the quality of the learning process. Thus, we must study the effect of different feature combinations on each detection method, in order to customize features with respect to each model.

B. Abnormality Detection Techniques

Importantly, the abnormality detection problem does *not* have a known paradigm; therefore, it is inherently an unsupervised learning problem. This is more so when it comes to smart meter data, because we must *explore* the type of abnormalities that may arise in such data streams, along with the potential applications of detecting such different abnormalities. It should be added that, for analyzing "unusual" load patterns of costumers, we do not have specific pre-determined labels.

Online unsupervised learning methods are updated as soon as they see new data; thus, *they can learn new patterns and the changes in trends*, such as due to seasonal changes. Moreover, these methods can be implemented in the real time in order to detect abnormalities quickly. In this study, we implement four unsupervised online abnormality detection methods:

1) Load Prediction with Regression (LPBSVR): This method works based on the comparison between the predicted and the actual power consumption. Accordingly, it is required to be built upon a prediction method. In this method, the prediction of power consumption is done using Support Vector Regression (SVR). SVR is a regression model which tries to minimize errors associated with the support vectors, so the prediction model is trained based on the outliers [12]. Therefore, this method is suitable for abnormality detection, where the abnormal patterns are treated as outliers.

Based on the obtained regression model using SVR, the residual can be calculated as the difference between real electricity consumption of each data and the predicted data in each time slots. These residuals are characterized with a probability distribution function (PDF), which can be used to detect the outliers data. For example, given that the PDF of residuals is a normal distribution with mean μ and variance σ, the data that falls outside of the $[\mu - 3\sigma, \mu + 3\sigma]$ span can be considered as outliers, i.e., abnormality data and outliers.

2) Load Prediction with Neural Network (LPBNN): As far as sole load prediction is concerned, neural network (NN) is proven to be a powerful data-driven tool, e.g., see [13]. Therefore, they can be used to develop a load prediction-based abnormality detection method. Other than the method of prediction, LPBNN is similar to LPBSVR. but the prediction is now done using a neural network [14]. We examined different neural network configurations, with 1 to seven hidden layers, three to thirteen nodes in each hidden layer, and both Relu and sigmoid as activation functions. The best results are then used in terms of prediction accuracy, with respect to MSE.

For each new reading from the smart meters, it first passes to the trained NN model to predict power consumption. Next, the residual is obtained; and if it is out of the mentioned span, then it is labeled as abnormal. The model is updated after making decision for each new data: if the new data is labeled as normal, then it is used as a new training data to update the NN model and residual PDF; otherwise, i.e., if the new data is abnormal; then it is reported and is not used to update the NN model. This process exactly implement for the regression based model which in that case SVR model are updated.

3) Clustered Based Method: In this method, the whole set of available data is clustered into two sets of "abnormal data" and "normal data". Some examples of cluster base methods are K Nearest Neighborhood (KNN) [15] and Local Outlier Factor (LOF) [16]. However, in this paper, we use the isolated forest (IF) method [17]. The basic idea of IF is to isolate instances without calculating any type of distance among measurements. This helps enhance computational time and online detection. IF utilizes two main characteristics when it comes to abnormalities: a) abnormal data is very rare; b) certain features of abnormal data are very different from those of normal data. Clustering is done using binary tree clustering. Because of susceptibility to isolation, anomalies tend to be isolated closer to the root of the binary tree. There are two reasons for that: first, instances with obvious feature value are tend to be divided in the early partitioning process; Second, in different parts which contains anomalies, less anomalies creates fewer partitions which causes shorter paths in the tree.

4) Projection Based Method: Input space is projected to a subspace by using projection vectors mainly for dimension reduction. Principle Component Analysis (PCA) and LODA [3] are two common projection based methods for abnormality detection. In this study, we use LODA, due to its computational efficiency for random sparse projection. It is based on ensemble of some random sparse projection of feature vector.

The projected values of a training data with respect to a projection vector w_i, give us a one-dimensional set which is used to obtain a histogram for each set. Hence, we have k one-dimensional histogram from training data. The LODA output can be defined as negative log-likelihood of the sample data:

$$f(x) = -\frac{1}{k}\sum_{i=1}^{k} \log\, \hat{p}_i(x^T w_i). \tag{4}$$

A higher value of $f(x)$ indicates a lower probability of the sample being abnormal. Also, $\hat{p}_i$ denotes the respective probability of projected value of vector w_i and input x. For online abnormality detection of each input data, the projection value in each w_i is calculated and the respective $\hat{p}_i$ is found by it's histogram. Therefore, the LODA output of input data can be found based on trained histograms. By comparing this value with a certain threshold, the input data is labeled as normal or abnormal data. If the input data is recognized as abnormal, then it is used to update the histograms.

The number of bins in LODA for each histogram is calculated through the following optimization problem [3]:

$$\begin{aligned} &\underset{W,b}{\textbf{maximize}} \sum_{i=1}^{b} n_i \log \frac{bn_i}{N} - \left[b - 1 + (\log\, b)^{2.5}\right] \\ &\textbf{subject to} \quad N = \sum_{i=1}^{b} n_i, \end{aligned} \tag{5}$$

where N is the total number of samples, n_i is the number of samples in the i^{th} bin, and b is the total number of bins. This optimization problem is solved for each histogram individually. As another important parameter in LODA, the number of sparse projection vectors is calculated as:

$$\hat{\sigma}_k = \frac{1}{N}\sum_{i=1}^{N} |f_{k+1}(x_i) - f_k(x_i)|. \tag{6}$$

where k is the number of histograms and the optimum value of this parameter can be determined by equation as arg $\underset{k}{min}\ \frac{\hat{\sigma}_k}{\hat{\sigma}_1}$.

III. Case Studies

The test cases in this paper are based on the smart meter data from Pecan Street project in Austin, TX [2]. The collected data is for 92 consecutive days for five households with resolution of 15 minutes. After pre-processing and cleansing, the data is divided into 70% train data and 30% test data, respectively.

Recall from Section III that abnormality detection is inherently an unsupervised learning process. Therefore, since our data has no per-determined labels for abnormality, we need to obtain a benchmark to define unusual electricity consumption.

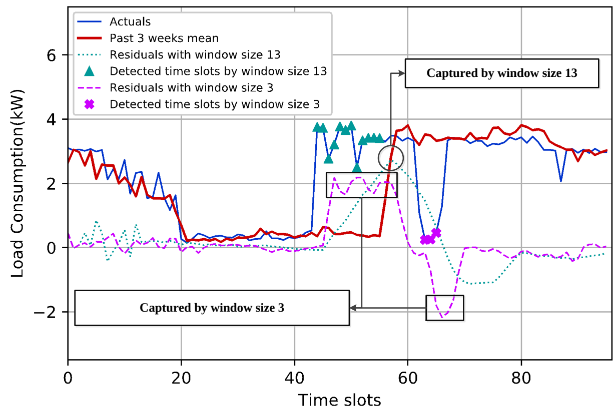

Fig. 1. Abnormality benchmark based on statistical model

A. Defining Abnormality in Electricity Consumption

Unusual electricity consumption can have different signatures and different duration of time. In order to capture abnormalities of different lengths, we use *moving windows of different sizes* on recent data and compare the data in the most recent window with those in the previous windows. By examining various experimental data in the database, we figured out that the power consumption data over the last *three-weeks* could efficiently show the trend of historical data. Therefore, by taking the average of each three-weeks period of data, we can construct a model to capture the trends of data over time. By conducting such comparison, we obtain a set of residual data for any window size and then fit a normal distribution function to the residuals with mean μ_p and standard deviation σ_p. These residuals are sum of the residuals across all individual time slots within each window.

For any residual that deviates from $3\sigma_p$, we consider it as an unusual trend or abnormality data. For the points which are determined in several moving windows, we consider the *largest consecutive window*. Such window can be obtained by analyzing residual curves related to each window size. After that, we select the curve with one peak in the detected points. Essentially, the peak point among all detected abnormalities is the size of the largest window that detected the abnormality.

Fig. 1 shows two consecutive abnormalities on one day which are captured by *two different window size*. This figure also depicts electricity consumption of the day, mean of the same day for the last three-weeks, and the residuals for the two detection window sizes. Note that, the window with size 3 detect the whole span of time slots $44-57$ but as the window with size 13 is a larger detector, and we have one peak (rather than several peaks in the window with size 3) in the detected span, we choose the later window as the period of abnormality.

B. Comparing Four Methods

All the proposed methods are applied on the available test data. In the "base case" all the three categories of features, i.e., load related, contextual and environmental, are applied to the models to detect abnormalities. Fig. 2 shows the performance of the four methods for part of the test data. The benchmark

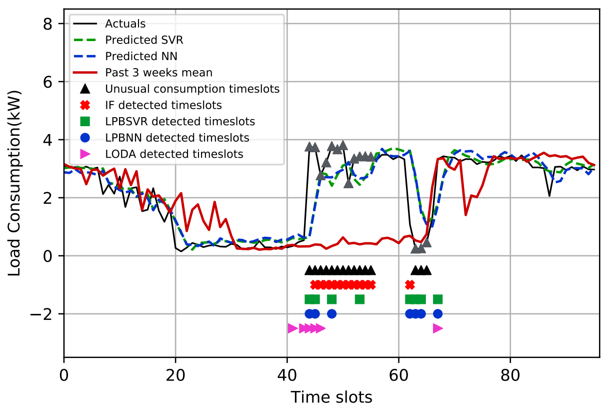

Fig. 2. Comparison of different abnormality detection methods with benchmark in the base case when all features are utilized.

abnormalities are marked on the data curves using triangles. The detected abnormalities time slots for each methods have been illustrated with different shapes below the curves.

To compare the four methods, we use the Matthews Correlation Coefficient (MCC) which is defined as [18]:

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+TF)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

Here, TP, TN, FP, FN are true positive (correctly identified), true negative (correctly rejected), false positive (incorrectly identified) and false negative (incorrectly rejected), respectively. MCC is broadly used as a measure of accuracy in binary classification, which essentially includes abnormality detection as a special case. MCC score is between -1 (the worst performance) and 1 (the best performance).

Based on the above results, IF has the best performance with MCC equal to 0.81; while MCC for LODA, LPBSVR and LPBNN is 0.54, 0.49 and 0.47, respectively. In the base case, all features are used which gives the best prediction result, i.e. the lowest MSE. However, despite having good prediction performance, LPBSVR and LPBNN have poor performance in detecting abnormality. This is due to comparing the consumption with its prediction, not with the previous consumption trends, which are different at the unusual benchmark points.

On the other hand, IF and LODA consider all features to detect abnormalities rather than conducting prediction. IF detects many points as abnormalities even more than benchmark. This may derive the fact that these points are different in certain feature from the usual trends, such as humidity, temperature, higher consumption in these time slots or even holiday flag.

C. Feature Selection and Sensitivity Analysis

In this section, we examined the impact of different features on the performance of each method. The result of the simulation are summarized in Table I. It is worth mentioning that all methods have been examined with different thresholds and the best MCC result is reported for each method. Also, the MSE of the prediction based methods is given in the Table II.

TABLE I
THE METHODS ACCURACY IN FEATURES SELECTION SCENARIOS

Feature scenarios	IF	LODA	LPBSVR	LPBNN
Whole features	0.8132	0.5447	0.493	0.4751
L^t	0.79156	0.7313	0.9276	0.7288
$L^t + L^w$	0.7666	0.7313	0.3206	0.2273
$L^t + C$	0.7924	0.3467	0.9276	0.7662
$L^t + C + L^w$	0.8783	0.7976	0.43102	0.5204
$L^t + C + E$	0.9196	0.43427	0.8852	0.6874

TABLE II
MSE OF PREDICTION BASED METHODS WITH RESPECT TO DIFFERENT FEATURE COMBINATION

Feature scenarios	LPBSVR	LPBNN
Whole features	0.2741	0.3016
L^t	0.6516	0.6375
$L^t + L^w$	0.2723	0.3893
$L^t + C$	0.7501	0.6770
$L^t + C + L^w$	1.3376	0.3778
$L^t + C + E$	0.7839	0.6547

Recall from Section III-B that, while the use of all features can improve prediction in prediction-based methods, it does not necessarily improve abnormality detection. This issue is better understood in Fig. 3, where features L^t, C and L^w are used. Despite having higher MSE, LPBSVR and LPBNN have lower MCC. In prediction-based methods, we need a prediction (dashed green curve) close to the previous consumption trend (red curve) and not necessarily close to the real power consumption (blue curve). In fact, for prediction-based methods, those features that simulate the previous consumption trends serve better for abnormality detection. Fig. 4 shows the results for the case where only the L^t features are used. Here, prediction-based methods appropriately simulate the previous consumption trend (red curve) as its expected value (dashed green curve) for the actual load. Conversely, for more accurate predictions, the better descriptive and complementary features should be used. It is worth mentioning that, NN works better than SVR in most cases in predicting the electricity consumption. However, LPBSVR has better performance than LPBNN in detecting abnormality in most cases, except when L^t, L^w and C are all utilized. This is because SVR is trained based on the outliers, which helps in abnormality detection.

Another observation is that the performance of IF is *not* sensitive to L^w, i.e., window features. This is due to the cohesion of these features which are related and close to each other. As a result, IF cannot divide them suitably in the trained model (trained trees). In contrast, if we substitute the window features with the environmental features, the best performance for IF is happened. This shows that the tree nodes in IF are sensitive to the environmental features. Understanding the cause of unusual consumption can be derived by examining diverse features which is out of the scope of this study. Analyzing simulation results shown that prediction based method generally is more accurate for the abnormalities with very high or very low magnitude. Conversely, other two methods, specially IF, are more accurate for unusual trends which may does not have a high peak power consumption.

Since LODA depends on random projection, many repetition must be considered to test its performance. The result

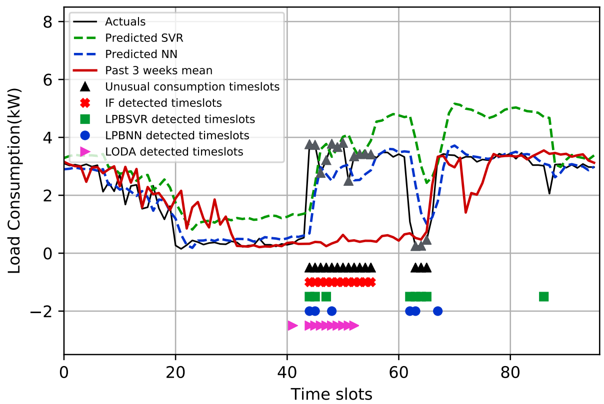

Fig. 3. Comparison with benchmark based on $L^t + C + L^w$ features

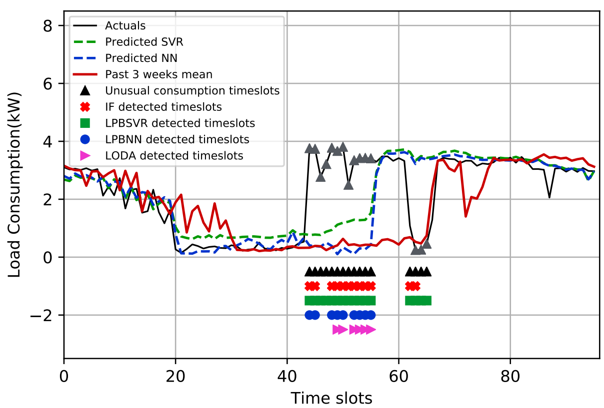

Fig. 4. Comparison with benchmark based on L^t features

with L^t, C and E as features in different Houses and time slots shows that LODA tends to detect power consumption which are really close to zero. In other words, LODA is sensitive to the very low power consumption periods. However, it has an acceptable performance compared to other methods when L^t is used or it is accompanied with L^t features.

IV. CONCLUSIONS

In this study, we examine the performance of four online unsupervised machine learning abnormality detection methods to detect abnormalities in smart meter data. The real-world data traces are used for this purpose. Four key conclusions are made. First, it is observed that, in general, i.e., when all available features are considered, clustering-based methods, such as IF, have a better performance that prediction-based and projection-based methods. Second, prediction-based methods gain their best performance when the prediction model simulates the previous consumption trend accurately rather than following the upcoming real-time electricity consumption. Third, simulation results show that prediction based methods generally are more accurate for the abnormalities with very high or very low magnitude. Forth, projection-based methods, such as LODA, do not show promising performance for abnormality detection in smart meter data; however, LODA can demonstrate a slightly better performance through a better feature selection when only a certain subset of available features are utilized. In this regard, when it comes to the detection of abnormalities in smart meter data, it is better to customize the features for each method individually, despite the fact that the common practice in the previous literature is to consider the same set of features for all methods.

ACKNOWLEDGEMENT

We thank Scott Hinson and Suzanne Russo from Pecan Street. This work is supported in part by NSF grant 1253516 and its INTERN supplemental fund, NSF grant 1462530, CEC grant EPIC-16-033, and UCOP grant LFR-18-548175

REFERENCES

[1] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” *IEEE Trans. on Smart Grid*, accepted for publication, March 2018.

[2] "pecan street, real energy, real customers in real time". [Online]. Available: http://www.pecanstreet.org/energy/

[3] T. Pevný, “Loda: Lightweight on-line detector of anomalies,” *Machine Learning*, vol. 102, no. 2, pp. 275 – 304, Feb 2016.

[4] A. Tascikaraoglu and B. M. Sanandaji, “Short-term residential electric load forecasting: A compressive spatio-temporal approach,” *Energy and Buildings*, vol. 111, pp. 380 – 392, 2016.

[5] M. N. Kurt, O. .E.Ogundijo, C. Li, and X. Wang, “Online cyber-attack detection in smart grid: A reinforcement learning approach,” 2018. [Online]. Available: http://arxiv.org/abs/1809.05258

[6] Y. Liao, Y. Weng, C. w. Tan, and R. Rajagopal, “Urban distribution grid line outage identification,” in *Proc. of Probabilistic Methods Applied to Power Systems (PMAPS), 2016 International Conference on*, Beijing, China, 2016, pp. 1 – 8.

[7] P. Jokar, N. Arianpoo, and V. C. Leung, “Electricity theft detection in AMI using customers’ consumption patterns.” *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216 – 226, jan 2016.

[8] N. L. Tasfi, W. A. Higashino, K. Grolinger, and M. A. M. Capretz, “Deep neural networks with confidence sampling for electrical anomaly detection,” in *Proc. of 2017 IEEE International Conference on Internet of Things*, Exeter, UK, Jun 2017, pp. 1038–1045.

[9] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. Capretz, and G. Bitsuamlak, “An ensemble learning framework for anomaly detection in building energy consumption,” *Energy and Buildings*, vol. 144, pp. 191 – 206, 2017.

[10] J. LUO, T. HONG, and M. YUE", “Real-time anomaly detection for very short-term load forecasting,” *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 235 – 243, Mar 2018.

[11] A. Saad and N. Sisworahardjo, “Data analytics-based anomaly detection in smart distribution network,” in *Proc. of International Conference on High Voltage Engineering and Power Systems*, Sanur, Indonesia, oct 2017, pp. 1 – 5.

[12] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, “Support vector regression machines,” in *Advances in Neural Information Processing Systems 9*, may 1997, pp. 155 – 161.

[13] Y. Liang, D. Niu, and W. C. Hong, “Short term load forecasting based on feature extraction and improved general regression neural network model,” *Energy*, vol. 166, pp. 653 – 663, 2019.

[14] S. S. Haykin,, *Kalman Filtering and Neural Networks*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[15] E. Knorr and R. T. Ng, “Finding intentional knowledge of distance-based outliers.” in *Proc. of the international conference on very large data bases*, Jun 1999, pp. 211–222.

[16] M. Breunig, a. R. T. N. H. P. Kriegel, and v. n. p. y. o. a. J. Sander, booktitle=in Proc. of the 2000 ACM SIGMOD international conference on Management of data, “LOF: identifying density-based local outliers.”

[17] F. T. Liu, K. M. Ting, and Z. H. Zhou, “Isolation-based anomaly detection,” *ACM Trans. on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, p. 3, 2012.

[18] D. M. W. Powers, “Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37 – 63, 2011.