Fault detection analysis of building energy consumption using Data Mining techniques

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 42 (2013) 557 – 566

Energy Procedia

The Mediterranean Green Energy Forum 2013, MGEF-13

Fault Detection Analysis of Building Energy Consumption Using Data Mining Techniques

Imran Khan[a], Alfonso Capozzoli[a,]*, Stefano Paolo Corgnati[a], Tania Cerquitelli[b]

[a]*TEBE Research Group, Department of Energy (DENERG), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

[b]*Department of Control and Computer Engineering (DAUIN) , Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Abstract

This study describes three different data mining techniques for detecting abnormal lighting energy consumption using hourly recorded energy consumption and peak demand (maximum power) data. Two outliers' detection methods are applied to each class and cluster for detecting abnormal consumption in the same data set. In each class and cluster with anomalous consumption the amount of variation from normal is determined using modified standard scores. The study will be helpful for building energy management systems to reduce operating cost and time by not having to detect faults manually or diagnose false warnings. In addition, it will be useful for developing fault detection and diagnosis model for the whole building energy consumption.

Selection and peer-review under responsibility of KES International

Energy Monitoring; Energy Consumption; Energy Performance; Data Analysis; Fault Detection; Outlier Detection

1. Introduction

Energy consumption of both residential and commercial buildings has steadily increased, reaching figures up to 40% in developed countries. Increasing demand for building services and high thermal comfort levels, together with the amount of time spent indoors, will increase the energy demand in future [1]. There is an increasing realization that many buildings do not perform as intended by their designers. Typical buildings consume 20% more energy than necessary due to faulty construction, malfunctioning equipment, incorrectly configured control systems and inappropriate operating procedures [2] and [3]. For energy optimization, the evaluation of real time building energy consumption data is a demandable and

* Corresponding author. Tel.: +39-0110904413
E-mail address: alfonso.capozzoli@polito.it.

1876-6102
Selection and peer-review under responsibility of KES International
doi:10.1016/j.egypro.2013.11.057

emerging area of building energy analysis. Despite tested for efficiency by use of different available forward models the building systems may fail to meet the performance expectations due to various faults. Poorly maintained, degraded, and improperly controlled equipment wastes an estimated 15% to 30% of energy used in commercial buildings [4] and [5]. Therefore, it is of great potential to develop automatic, quick-responding, accurate and reliable fault detection and diagnosis (FDD) schemes to ensure the optimal operations of systems in order to save energy. Energy management and control systems can collect and store massive quantities of energy consumption data. Therefore powerful and efficient tools are required to uncover valuable information from the tremendous amounts of available data and transform it into organized knowledge. Several studies have been published on methods for automatically detecting abnormal energy consumption data in buildings. J. Seem [6] presented a method for converting energy consumption data into information and accounted for weekly variation in energy consumption by grouping days of week with similar power consumption. The robust statistical method is used to determine if the energy consumption is significantly different than previous energy consumption.

In order to find the patterns in data set, classification of data is important before detecting outliers in the building energy consumption. Clustering and classification are two common techniques used in data mining for finding hidden patterns in data sets. Several classification techniques to the problem of FDD in the data generated by VAV AHU model are discussed in [7]. Some research works [6] [8] and [9] provided classification methods, including the box plots approach and pattern recognition algorithm. Liu et al. [10] used a robust statistical algorithm to detect the abnormal electricity energy consumption in building and achieved good results.

Outliers are cases that have data values that are very different from the data values for the majority of cases in the data set. Statistical-based, distance-based, deviation-based and density-based methods are discussed mainly in recent times. Many outlier procedures based on statistical theory, used extreme studentized deviate (ESD) algorithm to detect abnormal energy consumption achieving good results [6] and [10].

In this study, three different data mining techniques are used to analyze real time hourly recorded energy and power consumption data for lighting in the office building. A classification and clustering of hourly recorded data using CART, K-Means and DBSCAN algorithms respectively have been carried out. With CART and K-Means methods, the detection of evident outliers in each class and cluster has been performed using Generalized Extreme Studentized Deviate (GESD) algorithm and boxplot statistical method. In DBSCAN method the outliers are directly detected analyzing a particular cluster in which they are isolated. A comparison of aforesaid pattern recognition and clustering methods has been carried out highlighting the potentiality and the limits of the each approach for a fault detection analysis. Experimental results show the effectiveness of the proposed approaches in automatic detection of abnormal energy consumption which can improve building operators' performance by reducing time for detecting faults.

2. Building and Data Description

The case study selected for the fault detection analysis is an office building located in Rome, Italy. The building is composed of three floors and a basement connected through the larger side with a second building. The building is equipped with a monitoring system aimed at collecting energy consumption (electrical and thermal) and the environmental conditions. Moreover each room/office in the building has been equipped with a presence sensor. In the paper, experiments have been performed on a data set referred to energy consumption for artificial lighting only for the first floor. In this floor there are 13 offices of different size with a floor area ranging from 14 to 36 m^2 and two CED rooms each of about 20 m^2. Different number of fluorescent lamps (each 55 W) ranging from 4 to 8 are installed in each office/ room. In the two CED rooms 12 lamps, each 55 W, are installed. In order to identify abnormal lighting energy consumption, the features considered as dependent variables for the models are the average hourly

energy consumption and peak demand (maximum power). The office lighting energy and power consumption have been analyzed for the months of December and January.
Furthermore, the independent variables that have been recorded with an hourly time step are: people presence, number of active rooms (a room is considered active if at least one person is present), global solar radiation, time, date and day of the week. In order to verify the reliability and the effectiveness of the proposed methods two artificial faults have been created on 24th and 25th of January. In these days at the end of the working time with fewer people presence between 17:30 and 18:00 all artificial lights of the offices on the first floor have been switched on creating a peak of energy demand.
In the following first a brief theoretical description of the classification, clustering and outlier detection methods used in the work is presented. In the second part a fault detection analysis for the lighting energy consumption is performed using three different methods with the aim to compare the capability of the each method in detecting mainly the created artificial faults.

3. Classification and Regression Tree (CART)

The CART algorithm is based on Classification and Regression Trees. A CART is a binary decision tree that is constructed by splitting a parent node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample. CART can easily handle both numerical and categorical variables and useful in robust detection of outliers. A decision tree is constructed from the recorded data which can easily be converted to classification rules. CART methodology generally consists of three parts [11]:

i. Construction of maximum tree: classification tree is built in accordance with splitting rule. Each time data have to be divided into two parts with maximum homogeneity. The Gini impurity measure at a node *t* is defined as:

$$i(t) = \sum_{k \neq l} p(k/t)p(l/t) \quad (1)$$

where *k* is the index of the class and p (k|t) is the conditional probability of class *k* provided we are in node *t*.

ii. Choice of the Right Size Tree: optimizing tree size is important because maximum trees may turn out to be of very high complexity and consist of hundreds of levels. Two pruning algorithms can be used in practice: optimization by number of points in each node and cross-validation.

iii. Classification of new data: by set of questions in the tree, each of the new observations will get to one of the terminal nodes of the tree. A new observation is assigned with the dominating class/response value of terminal node, where this observation belongs to.

4. Clustering

The selected algorithms can be classified into two categories: (i) Partitioning methods and (ii) density-based methods. These methods require the definition of a metric to compute distances between objects in the dataset. In the case study analyzed, distances between objects are measured by means of the Euclidean distance computed on normalized data. Partitioning methods subdivide a dataset of *n* objects into *k* disjoint partitions, where $k < n$. The general criterion to perform partitioning assigns objects to the same cluster when they are close and to different clusters when they are far apart with respect to a particular metric. Partitioning methods are able to find only spherical-shaped clusters, unless the clusters are well separated, and are sensitive to the presence of outliers. K-Means [12] is a popular method which belongs to this category. Density-based methods are designed to deal with non-spherical shaped clusters and to be less sensitive to the presence of outliers. The objective of these methods is to identify portions of the data space characterized by a high density of objects. Density is defined as the numbers of objects which are in

a particular area of the *n* dimensional space. The general strategy is to explore the data space by growing existing clusters as long as the number of objects in their neighborhood is above a given threshold. DBSCAN [13] is the density based method considered in our case study.

4.1. Partitioning Method (K-Means)

K-Means [12] requires as input parameter *k*, the number of partitions in which the dataset should be divided. It represents each cluster with the mean value of the objects it aggregates, called centroid. The algorithm is based on an iterative procedure, preceded by a set-up phase, where *k* objects of the dataset are randomly chosen as the initial centroids. Each iteration performs two steps; in the first step, each object is assigned to the cluster whose centroid is the nearest to that object. In the second step centroids are relocated, by computing the mean of the objects within each cluster. Iterations continue until the *k* centroids do not change. K-means is effective for spherical-shaped clusters. Different cluster shapes are detected only if the clusters are well separated. Similar to other partitioning methods, k-means is sensitive to outliers and requires a prior knowledge of the number of clusters.

4.2. Density Based (DBSCAN)

DBSCAN [13] requires two input parameters, a real number *r*, and an integer number minPts, used to define a density threshold in the data space. A high density area in the data space is an *n*-dimensional sphere with radius *r* which contains at least minPts objects. DBSCAN is an iterative algorithm which iterates over the objects in the dataset, analyzing their neighborhood. If there are more than minPts objects whose distance from the considered object is less than *r*, then the object and its neighborhood originate a new cluster. DBSCAN is effective at finding clusters with arbitrary shape, and it is capable of identifying outliers as a low density area in the data space. The effectiveness of the algorithm is strongly affected by the setting of parameters *r* and minPts.

5. Outlier Detection Methods

An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs or outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data. Outliers arise because of human error, instrument error, changes in behavior of systems or faults in systems. In this study boxplot and generalized extreme studentized deviate (GESD) many-outlier algorithms have been used to detect abnormal energy consumption. Boxplot is a common statistical technique to identify the hidden patterns in a data set and can also be easily refined to identify outlier data values.

The GESD many-outlier procedure is a modified version of the extreme studentized deviate (ESD) test that was proposed by [14], which allow finding multiple outliers. In order to perform the method two parameters need to be set: the probability α of incorrectly declaring one or more false outliers, and an upper limit n_u of the expected number of potential outliers. On the basis of Carey indications [15] the potential number of the potential outliers have been evaluated finding the largest integer that satisfies the following inequality $n < 0.5(n - 1)$, where *n* is the number of observations in the data set $X : \{x_1, x_2, x_3, \ldots, x_n\}$. The method allows detecting the outlier values in a data set through the calculation and comparison of two following important parameter:

i. The *ith* extreme studentized deviate R_i, determined from:

$$R_i = \frac{|x_{e,i} - \bar{x}|}{s} \qquad (2)$$

where $x_{e,i}$, is the extreme element in set X that is furthest from the average $\bar{x}$ of elements in set X

ii. The *ith* critical value λ_i determined from:

$$\lambda_i = \frac{t_{n-i-1,p}\,(n-i)}{\sqrt{\left(n-i-1+\,t^2_{n-i-1,p}\right)(n-i+1)}} \quad (3)$$

where $t_{n-i-1,p}$ is the student's *t*-distribution with (n-i-1) degrees of freedom, and the tail area probability *p* is determined from:

$$p = 1 - \,[\alpha/2(n-i+1)] \quad (4)$$

6. Z-scores and Modified z-scores

Standard scores were used to analyze the outliers and modified z-scores were used to quantify how far and which direction an outlier is from the mean value of typical observations. The modified z-scores are defined as:

$$z_m = \frac{x_{outlier} - \bar{x}_{robust}}{s_{robust}} \quad (5)$$

where z_m is the modified standard score, $x_{outlier}$ is a raw value of an outlier, $\bar{x}_{robust}$ is the mean value of non-outliers in the data set and s_{robust} is the standard deviation of non-outliers in the data set

7. Classification and Analysis

Classification and regression tree method has been used to analyze lighting energy and power consumption data in the office building. In the data the maximum number of people and active rooms are 21 and 14 respectively. The values of solar radiation vary from 0 to 476.45 W/m^2. In average there are zero number of people and active rooms during weekends and after 21:00 weekdays. The number of people is higher during 09:00-16:00 i.e. 16 and more people. The sensitivity analysis of the data show that both energy and power have relation with other variables i.e. people, solar radiation, day and active rooms, so it can be deduced that the extreme values in both energy and power sequence graphs are not definite outliers because there are other variables too that can affect the consumption. Therefore it is important to classify the data set into similar conditions before detecting outliers.

For both energy and power separate decision trees were developed considering the day, time, active number of rooms, number of people and solar radiation as independent variables. Classification and regression tree (CART) algorithm has been used for tree growing process with maximum tree depth of 5 by using both pruning methods described in section 3. The data was divided into 9 and 10 classes for energy and power respectively. The classes constructed for energy and power were analyzed separately and the summary of the key features is given in tables 1 and 2 respectively. In these tables the solar radiation values less than 150 W/m^2 are considered lower and greater than 150 W/m^2 are considered higher.

The scatter plots for all energy classes indicate that classes 6, 9, 11, 13 and 15 are pure. Similarly for power classes 6, 7, 10, 11, 13, 14 and 18 are pure. Scatter plots for only class 3 (energy) and class 17 (power) have been presented in fig. 1. From these plots it can be presumed that abnormal energy consumption should exist in both classes.

Table 1. Energy classes and brief description of features of each class

Class	Time	People Presence	Active Room	Solar Radiation	Day
3	Mostly early morning	Zero or one	Zero or one	Mostly zero	Thu-Fri
6	18:00-21:00	Mostly < 7	Mostly < 7	zero	Weekdays
8	Weekends-all day Weekdays evenings	Zero or one	Zero or one	Zero except weekends	Weekends and Mon-Wed
9	07:00, 08:00	80 % ≤10	≤10	Lower values	Weekdays
11	06:00-07:00	zero	zero	Mostly zero	Weekends
12	06:00-07:00	zero	zero	zero	Mon-Wed
13	Different timing	Mostly ≥ 10	Mostly ≥ 10	Lower values	Weekdays
15	12:00-16:00	≥ 10	≥ 10	Higher values	Weekdays
16	09:00-11:00 and 17:00	≥ 10	≥ 10	Higher values	Weekdays

Table 2. Power classes and brief description of features of each class

Class	Time	People Presence	Active Room	Solar Radiation	Day
6	18:00-21:00	80% ≤ 5	80% ≤ 5	zero	Weekdays
7	06:00-08:00	Zero	Zero	Zero or lower values	Thu-Fri
9	Evening and early morning	Mostly zero	Mostly zero	Mostly zero	Thu-Fri
10	Weekends-all day, Weekdays early morning	Zero	Zero	Zero except weekends	Weekends and Mon-Wed
11	07:00, 08:00	Mostly ≤ 10	Mostly ≤ 10	Lower values	Weekdays
13	06:00-08:00	Zero	Zero	Mostly zero	Weekends
14	06:00-08:00	Zero	Zero	Mostly zero	Mon-Wed
15	Different timing	Almost 60 % ≥ 10	Almost 60 % ≥ 10	Lower values	Weekdays
17	09:00-17:00	Almost 70 % ≥ 10	Almost 60 % ≥ 10	Medium range	Weekdays
18	09:00-15:00	Mostly ≥ 10	Mostly ≥ 10	Higher values	Weekdays

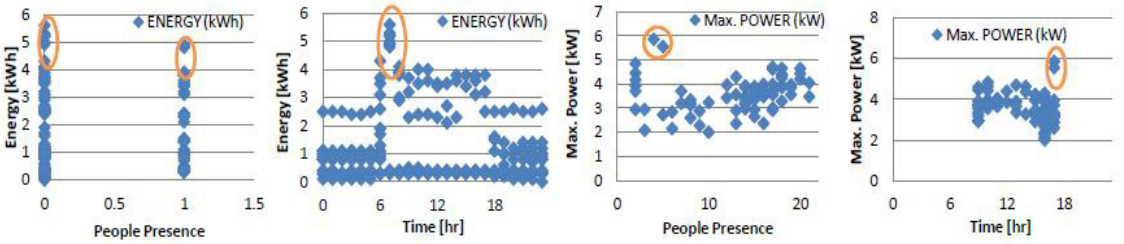

Fig. 1. Scatter plots for energy (class3) and power (class 17)- abnormal consumption is encircled

7.1. Outliers detection

The two different outliers detection methods described in section 5 were applied to each class for both energy and power. The amount of variation from normal has been determined using robust estimates of the mean and standard deviation (modified z-scores) to show the results. The outliers detected for energy and power consumption confirms that construction of classes and outlier's detection algorithms were correct. For example in class 3 (energy) the consumption is higher for zero number of people and active rooms. Most abnormal consumptions detected are in the early morning i.e., 06:00-07:00 where the energy

consumption is almost equal to that observed with 15 or more people during the working hours. Similarly in class 8 (energy) the most abnormal consumptions are at 08:00 when there are zero or few people in the building. When the average hourly energy consumption is analyzed both outliers' detection methods were not able to detect the artificial faults even though they were inserted through CART in the same class i.e. 16. The results obtained from power classes and outliers detection algorithm were quite good as they were able to detect the artificial faults present in class 17 (power). It can also be concluded that the artificial faults were associated to the abnormal maximum power consumption and not to the energy consumption. The fig. 2 shows the sequence graph of hourly recorded power consumption and modified z-score graph for class 17 (power) with two artificial outliers. It is clear from the figure that outlier detection is difficult by using sequence graph only. The outliers in individual class are mostly peak values and can be easily located, while in sequence data the same is not possible.

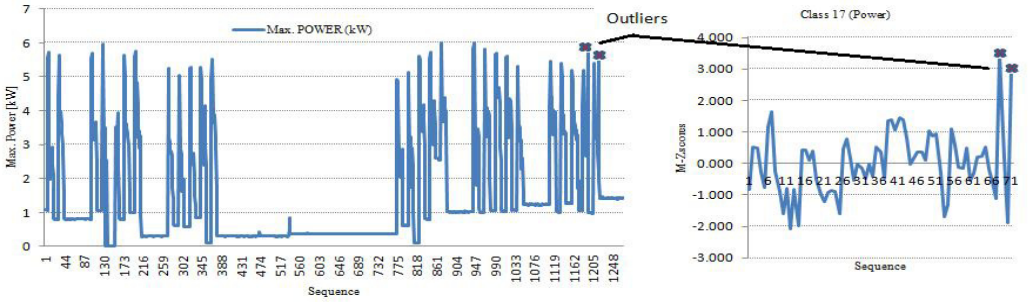

Fig. 2. Power sequence graph and M-zscores graph for class 17 showing artificial outliers

8. Clustering and Analysis

Clustering algorithms can be exploited to discover interesting correlations among monitored measures (i.e., energy, people presence, active room and solar radiations) to automatically detect which records represent the abnormal energy consumption. Both the time and day cannot be used as independent variables in clustering techniques; hence the ability to detect the fault without time pattern information with K-Means and DBSCAN has been also investigated. Since we focus only on a subset of measures used to drive the classification method, the recorded real data for both energy and power consumption have been split into daytime, night time and weekends to drive the clustering algorithms. Among the available clustering approaches we focused on the K-Means algorithm and the DBSCAN approach. Before to perform the clustering analysis recorded real data have been normalized by means of standard score (z-score) method.

8.1. K-Means application

The k-means algorithm is a popular data clustering algorithm. However, one of its drawbacks is the requirement for the number of clusters, *k*, to be specified before the algorithm is applied. In this study hierarchical clustering has been performed to define the number of clusters by applying Ward's method on the principal components score. From the agglomeration schedule by identifying the step where the "distance coefficients" makes a bigger jump, the number of clusters, *k*, has been selected. For energy clusters formation normalized values of energy, people presence, active room and solar radiation have been used, and similar for power. To detect outliers in each cluster the two outliers' detection methods (GESD and boxplot) described earlier were applied to each cluster for both energy and power. Starting from the results obtained with the classification analysis, the brief summary of clusters for both energy and power for only day time data is given in table 3 and 4 respectively.

Table 3. Brief description of energy day time clusters

Clu No	People Presence	Active Room	Solar Radiation	Standard deviation of Energy [kWh]
1	Almost 60 % ≥ 10	Almost 60 % ≥ 10	Higher values	0.6204
2	Almost 80 % ≤ 5	Almost 80 % ≤ 5	Lower values	0.9579
3	Always ≥ 7	Always ≥ 7	Different range	0.9112
4	Almost 85% ≤ 3	Almost 85% ≤ 3	Mostly zero	0.4388

Table 4. Brief description of power day time clusters

Clu No	People Presence	Active Room	Solar Radiation	Standard deviation of Power [kW]
1	Almost 60 % ≤ 5	Almost 60 % ≤ 5	Higher values	0.5945
2	Almost 95 % ≤ 5	Almost 95 % ≤ 5	Lower values	1.8911
3	Always ≥ 10	Always ≥ 10	Higher values	0.6362
4	Almost 60% ≥15	Almost 65% ≥ 10	Lower values	1.1504

The results show that most clusters were not pure and the abnormal values were disseminated. In energy clustering the artificial outliers are in the same cluster 2 while in power clustering, they are present in cluster 2 and 4. Cluster 4 is impure in both energy and power clustering. In energy cluster 4 most positive false are in the evening and few in early morning and opposite for power cluster 4. The reason to these positive false could be that some variables in the data set may have the same values as of real faults. Both outliers' detection methods were unable to detect the artificial faults present in energy cluster 2. In fig. 3 power sequence graphs for day time data and modified z-scores for cluster 4 (power) have been presented with outliers highlighted. The outliers' detection methods were able to detect the artificial fault in cluster 4 but were unable to detect in cluster 2. After thorough analysis of the results it is concluded that though the K-Means algorithm can be useful in detecting abnormal values but it is not suitable approach for robust outliers' detection.

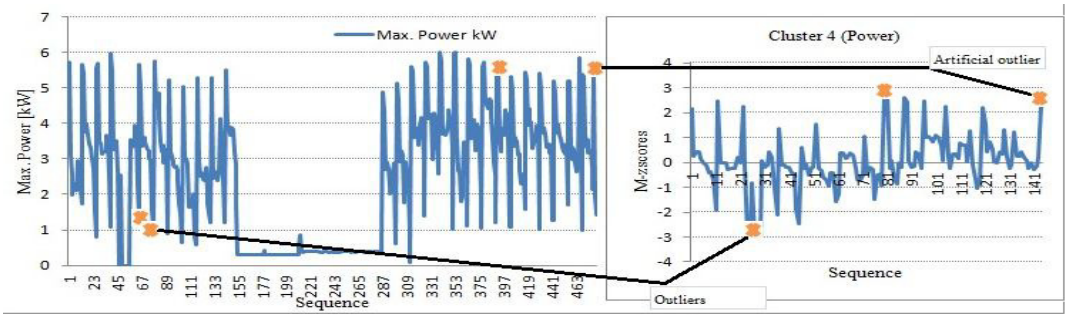

Fig 3. Power sequence graph for daytime and M-zscores graph for cluster 4 with outliers

8.2. DBSCAN application

As presented in section 4, DBSCAN requires two input parameters, *r* and minPts, and the effectiveness of the algorithm is strongly affected by the setting of these parameters. As opposed to K-Means, this method is effective in segregating the outliers. In all discovered clusters, the cluster label zero contains all points identified as outliers or noise. To set the input parameters different tests were carried out for all data (energy and power) by using different values for these parameters. The results show that by keeping the value of one parameter constant and changing the value of other parameter the discovered clusters are different. For example if the value of minPts is kept constant and the value of radius '*r*' decreases then the number of clusters and outliers increases. The results obtained from these tests were analysed and the

similarities within the clusters were investigated in order to select the appropriate values of both input parameters.

The results show that DBSCAN is able to identify clusters with the same density and with very similar data. The most benefit of DBSCAN is its ability to group all outliers (including noise) in a single cluster labelled as cluster zero. In table 5 the actual values of measures i.e. energy etc, included in cluster zero for day time energy data are given. The input parameters values are 5 and 0.7 for minPts and radius respectively. These records could be considered as abnormal energy consumption. For example for fewer numbers of people presence and active rooms (cases 1, 5, 6, 10) the energy consumption is high and opposite for cases 2, 3, 4 and 9. The DBSCAN algorithm has been effective in detecting real outliers or noise but was unable to detect the two artificial faults. That could be because of the nature of artificial faults which are strongly related to the time variable.

Table 5. Actual values of different measures included in zero cluster (minPts =5, r = 0.7)

S. No	Energy [kWh]	People Presence	Active Room	Solar Radiation [W/m2]
1	4.3	2	2	343.6
2	1.4	13	10	190.3
3	1.8	19	14	218.83
4	2.1	15	12	445
5	2.9	6	5	476.45
6	3.3	8	6	429.32
7	3.0	13	12	413.75
8	3.1	15	13	408.08
9	1.9	10	7	330.12
10	2.4	7	6	209.6

The outliers detected by all three techniques are analyzed and compared. In the table 6 some of the outliers with other measures (time, date, power, people presence, active room and solar radiation) are given. These outliers are common in at least two of the three proposed data mining techniques including the one artificial outlier (case 7). From the table it can be seen that the outliers detected are real faults i.e. the value of maximum power is high for less number of people and active room and opposite for higher number of people and active room. These results show that in general the proposed methods have been able to detect the real faults or noise.

Table 6. Data of some common outliers

S.No	Day	Date	Time	Max. Power [kW]	People Presence	Active Room	Solar Radiation [W/m2]
1	Fri	07/12/2012	08:00	5.42	7	6	38.5
2	Tue	11/12/2012	08:00	5.48	10	9	0.00
3	Fri	14/12/2012	11:00	4.84	2	2	343.6
4	Mon	14/01/2013	07:00	5.98	4	4	1.83
5	Mon	14/01/2013	14:00	1.95	19	14	218.83
6	Tue	15/01/2013	07:00	5.82	5	5	3.8
7	Fri	25/01/2013	17:00	5.55	5	4	1.75

9. Conclusions

In this paper the hourly recorded data of building energy system by exploiting three different data mining techniques for detecting anomalous energy consumption has been analyzed. In classification and K-Means clustering two outliers' detection algorithms were applied to each class and cluster respectively for detecting abnormal consumption in the same data set. While in DBSCAN the cluster label zero contains

all values identifies as outliers or noise. The results of the three approaches were analyzed and a comparison has been carried out in terms of their potentiality and the limits for a fault detection analysis. Findings from this study are:

- Classification and regression tree with GESD outliers' detection algorithm is highly accurate and correct. The method is able to detect the two artificial faults and determines more correctly if the energy consumption is significantly different from previous consumption with the similar data.
- The experimental results using K-Means approach show that though the method is able to detect some abnormal consumption including one artificial fault but is not the most suitable method for robust outliers' detection. In discovered clusters most are impure and the abnormal energy consumption values are disseminated.
- With DBSCAN algorithm both artificial faults are not detected but the method is able to identify clusters with the same density and with very similar data. The most benefit of DBSCAN is its ability to group all outliers (including noise) in a single cluster labelled as cluster zero.

In conclusion, the classification and regression tree approach with GESD outliers' method is more effective in automatically detecting the abnormal energy consumption. The clustering methods are not able to detect the faults strongly related to time variable. The study will help building energy management systems (BEMS) in preventive maintenance by tracking and detecting abnormal energy consumption in building overall energy system. Also, it will make building energy managers more productive by not having to manually detect faults or noise.

References

[1] LuisP. Lombard, J. Ortiz & C. Pout, A review on buildings energy consumption information, Energy and Buildings, 40(2008) 394-398.
[2] L. Song, M. Liu, David E. Claridge and P.Haves, Study of on-line simulation for whole building level energy consumption fault detection and optimization, Architectural Engineering 2003: Building Integration Solutions, 1-8.
[3] S. Wu, Jian Q. Sun, Cross-level fault detection and diagnosis of building HVAC systems, Building and Environment, 46(2011) 1558-1566.
[4] J Schein, Steven T. Bushby, Natascha S. Castro and John M. House, A rule-based fault detection method for air handling units, Energy and Buildings 38(2006) 1485-1492.
[5] S Katipamula, Michael R. Brambley, Methods for fault detection, diagnostics, and prognostics for building systems, A Review, Part I, HVAC&R Research 2005; 11 (1): 3-25.
[6] John E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings, 39 (2007) 52-58.
[7] J. M. House, W. Y. Lee and D. R. Shin, Classification techniques for fault detection and diagnosis of an Air-Handling Unit, CH-99-18-5, ASHRAE Transactions, 1999, 105 (1) 1987-1997.
[8] J.E. Seem, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, Energy and Buildings, 2005, 37(2) 127-139.
[9] X. Li, Chris P. Bowers and T. Schnier, Classification of energy consumption in buildings with outlier detection, Industrial Electronics- IEEE Transactions, 2010, 57 (11) 3639-3644.
[10] D Liu, Q. Chen, K. Mori and Y. Kida, A Method for detecting abnormal electricity energy consumption in buildings, Journal of Computational Information Systems, 2010, 6 (14) 4887-4895.
[11] Roman Timofeev. Classification and Regression Trees (CART) Theory and Applications, Master Thesis, Humboldt University, Berlin,2004.12.
[12] B. H. Juang, L.R. Rabiner, The segmental K-Means algorithm for estimating parameters of hidden Markov models. IEEE Transactions on acoustics, speech, and signal Processing, 1990,38(9), 1639-1641.
[13] M. Ester, H.P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, proceedings of the 2nd international conference on knowledge discovery and data mining, 226-231.
[14] B. Rosner, Percentage points for generalized esd many-outlier procedure, Technometrics.
[15] Wagner, E.E. Walters, B.A. Rosner, Resistant and test based outlier rejection: effects on Gaussian one- and two-sample inference, Technometrics 39 (3) (1997) 320–330.