A Machine Learning-based Anomaly Detection Framework for Building Electricity Consumption Data

A Machine Learning-based Anomaly Detection Framework for Building Electricity Consumption Data

Lorenzo Mascali[a], Daniele Salvatore Schiera[a], Simone Eiraudo[a], Luca Barbierato[a], Roberta Giannantonio[b], Edoardo Patti[a], Lorenzo Bottaccioli[a], Andrea Lanzini[a]

[a]*Energy Center Lab, Politecnico di Torino, name.surname@polito.it, Torino, 10138, Italy,*
[b]*TIM S.p.A, name.surname@telecomitalia.it, Milan, 20123, Italy,*

Abstract

A suboptimal management or system malfunction can often lead to abnormal energy consumption in buildings, which results in a significant waste of energy. For this reason, the adoption of advanced monitoring systems, based on Machine Learning (ML) and visualization techniques, is crucial to avoid possible deviations from the baseline energy consumption. However, the historical data on which analyses are based generally do not report the occurrence of anomalies. Therefore, the application of supervised ML techniques is limited and unsupervised approaches are favoured. Moreover, domain experts find most ML techniques hard to interpret and thus find it difficult to contextualize anomalies. To overcome these issues, this work proposes a machine learning-based Anomaly Detection Framework (ADF) that involves the use of two complementary semi-supervised ML applications to obtain a highly interpretable and accurate detection of anomalies. Both techniques use Symbolic Aggregate approXimation (SAX) encoding to extract the most relevant information from load profiles. The aim of the first approach is to maximize the interpretability of the definition and distinction between anomalous and normal behavior. This is achieved using a Classification And Regression Tree (CART), albeit at the expense of a coarser output granularity. The second approach exploits a Multi-Layer Perceptron (MLP) algorithm to obtain a higher and more accurate output resolution, although it leads to a less interpretable definition of any anomalous behavior. The ADF has been applied to a real case study using electricity consumption data provided by a large telecommunications service provider. The results show that combining both ML models enhances the accuracy and interpretability of the detected anomalies.

Keywords: Anomaly Detection, Electricity Consumption, Machine Learning, Semi-Supervised, Synthetic Ground Truth, Symbolic Aggregate Approximation, Smart Meter
PACS: 0000, 1111
2000 MSC: 0000, 1111

List of Acronyms

ADF	Anomaly Detection Framework
ARIMA	Autoregressive Integrated Moving Average
CART	Classification And Regression Tree
CO	Central Office
KPI	Key Performance Indicator
LOF	Local Outlier Factor
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MIA	Mean Index Adequacy
ML	Machine Learning
MLP	Multi-Layer Perceptron
NDA	Non-Disclosure Agreements
NN	Neural Network
PAA	Piecewise Aggregate Approximation
PCA	Principal Component Analysis
PDF	Probability Density Function
RNN	Recurrent Neural Network
SAX	Symbolic Aggregate approXimation
SGT	Synthetic Ground Truth
SOM	Self-Organizing Map
SVR	Support Vector Regression
TLC	Telecommunication
WCSS	Within Cluster Sum of Squares

1. Introduction

Buildings contribute to global energy consumption to a significant extent, with around one-third of electricity final use and about a quarter of global CO_2 emissions being attributed to their operation [1]. Commercial and industrial buildings were found to account for 8 % of the total energy consumption in 2018 [2]. Therefore, enhancing energy efficiency and a rational resource management in these sectors are imperative to reduce the energy footprint and greenhouse gas emissions.

Academic research has focused on two primary approaches to tackle the aforementioned issues: *i)* retrofitting building envelopes and systems, and *ii)* optimizing the control and management of building energy systems. Moreover, a significant portion of the energy inefficiencies of buildings results from a poor energy management rather than inefficient building en-

velopes and systems [3]. Therefore, improving building energy monitoring systems is essential in order to address the aforementioned challenges [4]. Such monitoring systems can be classified into two main types: low-level monitoring systems and high-level monitoring systems. Low-level monitoring systems oversee energy consumption on a machine-by-machine basis and exploit dedicated sensors. However, this approach requires highly specialized applications in each individual building, thus resulting in high costs. On the other hand, high-level monitoring systems exploit energy consumption data from smart meters,thereby allowing for a more generalized approach that is not intrusive or specific to individual buildings. However, in order to obtain an effective high-level monitoring system, the consumption data analysis and anomaly alerting shouldbe accurate and interpretable and should inform on how the consumption diverges from typical responses to external drivers. Moreover, the user interface plays a fundamental role, especially for high-level monitoring systems. These systems are not directly interfaced with the sensors or actuators of the plant and cannot act directly on the control systems. Furthermore, identifying any abnormal behavior within historical consumption time series can be a challenging task since there are no clear boundaries to distinguish such behavior from normal behavior. Therefore, it is crucial to offer additional assistance to analysts / domain experts / energy managers, apart from simply alerting them about anomalies. This result to be fundamental to persuade stakeholders about the need to take action to restore consumptions to normal levels.

This work extends a previous work of ours [5], by proposing an Anomaly Detection Framework that is based on combining two different semi-supervised ML approaches to emphasize the accuracy, interoperability and graphical representation of any detected anomalies. The method [5] is here extended to extract the Synthetic Ground Truth (SGT) from a raw dataset by improving the accuracy of anomaly detection algorithms through two semi-supervised model trainings. The first developed algorithm, named SAX-CART, uses SAX encoding as a method to extract relevant features of load profiles and a Classification And Regression Tree (CART) model to reconstruct profiles from exogenous variables and define instances classified incorrectly as anomalous. The key feature of the SAX-CART approach is the interpretability of the model outputs, albeit at the expense of accuracy. The second algorithm, named SAX-MLP, uses an MLP that is trained in estimating electricity consumption and defines anomalies by imposing an upper limit on the magnitude of the errors committed by the model beyond which observations are considered anomalous. In this case, the use of SAX encoding is reserved for only the steps related to extracting the SGT from the dataset. This approach, unlike the SAX-CART approach, is aimed at privileging the accuracy of model outputs rather than their interpretability. This proposed framework introduces a novel approach to data selection, which has the aim of extracting SGT from raw time series data. In order to extract the SGT, we applied the Pareto principle to the appropriately clustered dataset. The Pareto principle, also known as the 80/20 rule, allows to distinguish the portion of useful and generalizable information from the noise present in the dataset. Furthermore, the study conducts a comprehensive comparison between two distinct types of algorithms, highlighting their precision performance, and evaluating the type of output with a focus on extracting useful information for identifying the possible causes of the detected anomalies. Another significant contribution lies in the application of the proposed framework to a real-world case study, thereby extending the applicability of the adopted methodologies to the analyzed scenario and enhancing the framework's significance and robustness. The data used in this study originate from smart meters installed in various buildings of a prominent telecommunication service provider in Italy. For Non-Disclosure Agreements (NDA) reasons, the organization restricts the dissemination of research data. The remainder of this work is structured as follows. Section 2 introduces the anomaly detection process in building electricity consumption data and presents a review of the relevant solutions in the literature. Section 3 outlines the proposed Anomaly Detection Framework in detail. Section 4 reports the experimental results obtained by analyzing a real dataset consisting of hourly measurements of the aggregate power data of the aforementioned Telecommunication (TLC) stations. Finally, Section 5 presents the concluding remarks and discusses potential avenues for future works.

2. Related work

Improving energy monitoring systems in buildings, especially in the highly energy-intensive commercial and industrial sectors, is of crucial importance to increase energy efficiency and reduce operating costs. One effective method that can be used to identify abnormal energy consumption patterns in various locations and at various times involvess the use of anomaly detection and visualization techniques. These techniques help a energy manager interpret and contextualize patterns and anomalies. They also allow a better understanding of the root causes and more effective actions to be introduced. It is worth noting that, in the context of power consumption data, it is fundamental to exploit domain knowledge in order to define what outliers actually represent real anomalies of the system being monitored and what one instead represent only inconsistencies in the dataset. Indeed, anomalies can be classified on the basis of three different characteristics: *i)* punctual anomalies, *ii)* collective anomalies, and *iii)* contextual anomalies [6]. Punctual anomalies are anomalies that are characterised by the involvement of only a few observations, if not just a sin- gle one, within a dataset, which deviate significantly from the characteristic value ranges assumed by the other observations. Collective anomalies are instead characterized by a set of observations which, taken individually, do not constitute an anomaly but which, when observed in a comprehensive way, do not conform to the typical patterns found in the dataset. Punctual anomalies and collective anomalies are deduced exclusively through the analysis of historical series. On the other hand, contextual anomalies can be deduced by correlating a monitored variable with the explanatory variables. For instance, energy consumption patterns may be deemed anomalous if they do not reflect

the typical correlation with temperature or if they do not conform to seasonality or weekly scheduling patterns or more generally exogenous variables that represent drivers of energy consumption. Defining anomalies through context in power consumption data is the most robust and advantageous approach, as it can provide initial elements that can then be used to investigate the causes of the anomalies.

2.1. Anomaly detection techniques

The choice of the algorithm to use for anomaly detection depends to a great extent on the dataset that is available for training, and particularly on the availability of labels for anomalous observations [6]. Three categories of algorithms can be identified: supervised, unsupervised, and semi-supervised. Supervised algorithms exploit the labels assigned to each dataset element to learn the difference between normal and anomalous behaviors. A high performance is generally guaranteed, but acquiring a labelled dataset requires a significant effort and a significant expenditure of resources. Moreover, the performances are closely related to the preprocessing of the dataset, as errors in the labels can lead to ineffective models. In addition, it is often a challenging task to identify anomalies in complex systems during a manual labelling process, and labelled data are not always readily available. For these reasons, to the best of our knowledge, there are currently no studies in the literature that have utilized a supervised approach in this particular research field.

A new and innovative way of tackling the challenge of using labelled datasets is by resorting semi-supervised algorithms. This approach has not been widely explored yet and is often referred to as unsupervised [7]. The substantial difference between the two approaches lies in how the training dataset is preprocessed. The training dataset in semi-supervised approaches undergoes filtering and any anomalous behavior is removed. This filtering can greatly improve the robustness of a model by training it on ground truth [8]. For example, a semi-supervised techniques was exploited in [9] to train a Long Short-Term Memory (LSTM) on a filtered dataset. Although this method leads to appreciable results, the adopted filter relies only on a statistical approach that is more suitable for removing noise from observations than detecting real anomalies. In [10], a hybrid approach is proposed that cascades an unsupervised algorithm to label the dataset and a supervised Two-Class Boosted Decision Tree algorithm in order to accurately classify anomalous observations. In [11], the authors used the Self-Organizing Map (SOM) to identify consumption profiles characterized by anomalous consumption and leveraged on the obtained results to enhance load forecasting through a Neural Network (NN).

Finally, unsupervised anomaly detection algorithms use unlabelled datasets and are based on the implicit assumption that normal instances are much more frequent than anomalies in the tested data [12, 13]. If this assumption is verified, the model that is trained by these algorithms can accurately model normal behavior but may encounter significant errors when dealing with anomalous behavior. Unsupervised approaches are undoubtedly the most widespread in the literature. A significant amount of research has focused on autoregressive techniques. For example, the authors of [13], the authors used an Autoregressive Integrated Moving Average (ARIMA) model to identify anomalous load profiles in an office building. An approach based on a Recurrent Neural Network (RNN) was tested in [14] to predict consumptions and define consumption patterns that did not conform to patterns forecast as anomalous. An approach based on LSTM was tested in [15] through the artificial introduction of anomalies into a dataset, and then considering observations that were certainly anomalous. An improvement in model precision, compared to an ARIMA-type model was observed. Finally, an anomaly detection approach, based on a Support Vector Regression (SVR) model, was presented in [12] and a decomposition of the time series was exploited to auto-correlate the consumption time series with specific characteristics of the load profiles. Several studies in the literature have shown the effectiveness of clustering algorithms in identifying anomalous profiles. In [16], a k-medoids algorithm was adopted to identify homogeneous consumption clusters, within which anomalous observations were identified through a Local Outlier Factor (LOF)-type algorithm. The same algorithm was also adopted in [17], although in this case, the time series was transformed into a frequency domain before being processed by the clustering algorithm. A Principal Component Analysis (PCA)-type dimensionality reduction technique was used in [18] coupled with a clustering techniques to identify anomalous consumption profiles.

In conclusion, it is worth noting that autoregressive and clustering algorithms rely solely on the analysis of the recurrence of consumption patterns. Therefore, they are able to identify anomalies of both punctual and collective type but are not effective in detecting contextual anomalies. This leads to difficulties in interpreting the anomalies identified by these algorithms. In fact, in order to develop more interpretable algorithms, it is crucial to identify the key factors that drive consumption and to define anomalous behaviors in relation to such factors. Therefore, non-autoregressive algorithms should be introduced along with efficient visualization systems to obtain better results.

2.2. Anomaly visualization techniques

Anomaly detection models are frequently used as information systems by energy managers in the industrial or commercial sectors. However, detecting an anomaly without contextualizing it does not always involve a physical intervention to the system. Therefore, the algorithms of these models should not only differentiate between normal and anomalous behaviors but should also provide clear and dependable information to the stakeholders.

A technique that can be used to aid energy managers in making sense of anomaly detection model results is to integrate these models with visualization tools. This approach allows a more comprehensive understanding of the data to be obtained. The importance of visual analysis was emphasized In [19], because the absence of pre-labelled datasets and the need to define anomalous behavior, not only with respect to the data but also in the context in which it occurs, make it essential to develop approaches that can provide a reacon to consider an observation as being anomalous. Moreover, it is of fundamental im-

portance to develop an effective graphical representation that enables an analyst to evaluate the anomalies identified by the algorithm on basis of their domain knowledge. The above was also supported in [20], which emphasized the importance of using visualization tools to support analysts in analyzing unlabelled datasets. The work conducted in [21, 22] described practical applications of time series visualization systems aimed at improving the monitoring capabilities of energy managers. The role of the user was found to have a strategic level of importance in [23], in which outputs were evaluated through satisfaction questionnaires, which were administered to a sample of energy managers.

A complementary approach to the use of graphic visualization to improve the interpretability of anomaly detection systems is the adoption of algorithms that provide simple and self-explanatory outputs. One ML method which, according to the literature, is highly interpretable, is the decision tree technique, which include CART algorithms. This type of algorithm divides the domain of input variables to identify regions of the domain that are characterized by the highest possible purity of the output variable. The partition rules of the domain are explicit and can therefore be easily extracted from the model. Moreover, CART models are able to handle multiple variables and were adopted in [24, 25] to identify homogeneous consumption conditions. Multiple types of CART visualization were described In [26], and the possibility of analysts interacting easily with the rules extracted from the models and, consequently, of deducing new knowledge was highlighted. The possibility of structuring an anomaly detection tool, based on decision trees, was demonstrated In [27]. The methodology was based on the use of k-means to identify typical consumption profiles and, subsequently, CART was used to determine the external conditions that determined the consumption patterns. However, clustering algorithms exploit the daily periodicity of a time series to categorize consumption. Therefore, they can only provide information on whether a particular day is anomalous or not, without specifying the exact time of the anomalous observation, and a further analysis is required to obtain such details. A solution to this limitation was identified in [28, 29] where SAX encoding was applied to time series of building consumptions. SAX encoding [30] allows consumption profiles to be categorized at a sub-daily level, thus enabling greater detail to be obtained in identifying anomalous periods. SAX transformation allows the profiles of the dataset to be grouped according to the word through which they are encoded [31]. In literature, it is possible to identify two different types of approaches aimed at improving the encoding of SAX. The first approach, found in [32, 33, 34], involves extracting a larger number of features from the upstream data set before the SAX encoding. While this approach allows for a more accurate description of the data, it significantly increases the complexity of the encoding, thereby resulting in a clear reduction in the possibility of implicitly defining meaningful clusters within the dataset. On the other hand, the second approach aims to optimize the encoding itself, making it more flexible and adaptable to the dataset under examination [28, 35]. Additionally, the SAX methodology introduces another innovative aspect, that is a self-determination of the parameters that characterize its application [36]. However, the methodology is not generally applicable, because it uses an objective function extracted from the knowledge context to provide the best clustering of the dataset.

2.3. Proposed contribution

This paper tackles the challenge of applying anomaly detection algorithms to various types of anomalies while ensuring high interpretability. An Anomaly Detection Framework, in which two semi-supervised algorithms based on the SGT extraction process introduced in [5], are combined and expanded to address this issue. The purpose of these algorithms is on identifying anomalies in historical electricity consumption data, by means of contextual logic with a good accuracy. The adopted algorithms within the proposed framework fell on a CART decision tree and a MLP neural network. The decision tree was adopted due to its ability to provide explicit correlations between inputs and outputs once trained. On the other hand, the neural network was chosen to explore the complex and non-linear relationships between inputs and outputs. Although the correlations in the case of the NN may not be directly extractable and interpretable, it's more accurate output will allow for the identification of anomalous periods with greater confidence. The main contributions and novelties of this paper, compared with our previous work [5], are:

- A further improvement is made to the SAX encoding algorithm through the use of partitive clustering as a means of fine-tuning the definition of the consumption intervals, which in turn will result in a more accurate fit with the dataset.
- A framework is formulated to train anomaly detection algorithms through a semi-supervised approach, which was consequently applied to a real case study.
- An anomaly detection method based on decision tree, has been applied. This approach was structured with the objective of extracting more information than than just the anomaly notification. In fact, CART allows visualising the domain of existence of the variables that characterize the predicted level of consumption and the investigation of the context of the occurrence of the anomalous behavior.
- A second anomaly detection algorithm, based on a non-self-regressive NN, is introduced. This approach was chosen as it offers a finer granularity than the SAX-CART approach.
- The two structured anomaly detection models have been compared and cooperative strategies necessary for the combined use of the two models have been identified in order to include the accuracy of the NN wiht the interpretability of the CART models.

3. Anomaly Detection Framework

This paper proposes an ADF that is based on two semi-supervised ML techniques, named SAX-CART and SAX-MLP,

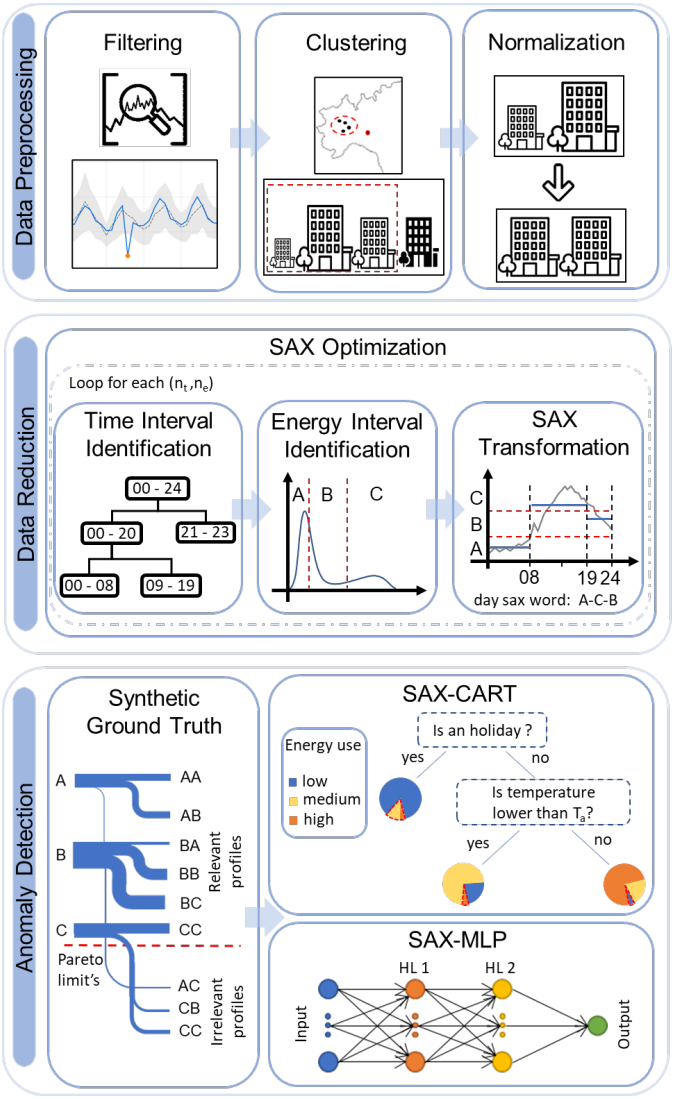

Figure 1: Workflow representation of the proposed Anomaly Detection Framework.

to identify anomalous consumption profiles in the time series of a building portfolio. The workflow of the methodology is illustrated in Fig. 1 and comprises three primary modules: *i) Dataset Preprocessing*, *ii) Data Reduction*, and *iii) Anomaly Detection*. The modules are described in the following subsections.

3.1. Data Preprocessing

In order to detect anomalies, the ADF uses certain raw data, such as the electricity consumption data of buildings collected at the smart-meter level, information on buildings related to their final uses and their geographical positions, and meteorological data related to the building sites. Electricity consumption time series often suffer from several inconsistencies, for example, they may contain outliers or missing values. For this reason, the data need to be preprocessed so that they can be used by the ML algorithms of the ADF.

The *Data Preprocessing* module consists of three distinct tasks, that are: *Filtering*, *Clustering* and *Normalization*. The aim of these tasks is to remove outliers by selecting a uniform subset of the electricity consumption dataset. This ensures that the subset is ready for use by subsequent modules.

Consumption data extracted directly from smart meters often contain some inconsistencies, such as observations characterized by out-of-scale consumption. These can be attributed to errors made during the data collection or to short maintenance activities. Therefore a *Filtering* task is introduced to make the methodology more robust and exclude these types of outliers, which are characterized by a short duration and high deviation, as they are irrelevant for the detection of anomalies. During this process, the historical time series of each building is scanned by applying a moving window of 15 days' width, which means 360 observations, and identifying those measurements that deviate from the mean by more than 1.5 of the standard deviation of the sample as outliers. If the identified outliers have a duration of less than 3 hours, the filtered observations are excluded and subsequently reconstructed by means of linear interpolation. Conversely, if the filtered period lasts longer than 3 hours, it is considered unprofitable to exclude these observations in the subsequent analysis. By following this procedure, it is possible to eliminate statistical anomalies from the dataset. However, collective anomalies cannot be excluded, and they need to be further analyzed through the methodology that is presented in Section 3.3.

The *Clustering* task groups the homogeneous buildings that have to be used for the following steps. Homogeneity is defined by two factors: the final use of the building and its geographic proximity. Indeed, the amount of energy a building uses is closely linked to its purpose and the surrounding weather conditions. Therefore, the factors that drive energy consumption and building systems schedules are often shared among buildings with similar purposes and weather conditions. Creating homogeneous categories to make further inferences is crucial in this regard. In this case, an ML algorithm and sample checks were utilized to ensure the correctness of the labelled final uses, following the approach described in [37]. A radius of 15 km was defined for each available meteorological station to define the geographical proximity, and it was assumed that the ambient conditions was constant at each point inside the circle. This is, of course, a simplification that does not consider specific irregularities. These aspects are more relevant for some variables, such as irradiance, due to the presence of clouds or the horizon shade profile of a building, which can vary significantly within an area considered homogeneous. Therefore, this information is excluded from the dataset, and only the temperature and relative humidity are considered, for which the hypothesis of homogeneity has been verified. In fact, if any non-uniform variables across geographic areas were to be included in the study, the dataset would be inconsistent. These variables do not reflect the actual conditions to which each building is subjected and, therefore, would produce an unequal performance comparison.

After identifying the homogeneous groups, the *Normalization* task is performed to facilitate the comparison of buildings of various sizes and electricity consumption levels and to adapt the input data to the ML techniques. The min-max normalization type is chosen to maintain the original shape of the con-

sumption of each building. The normalized power consumption (P_{norm}) is calculated through the following mathematical formulation:

$$P_{norm}(t) = \frac{P(t) - P_{min}}{P_{max} - P_{min}} \quad (1)$$

where $P(t)$ is the power consumption at time t, while the maximum and minimum power (P_{max} and P_{min}, respectively) are calculated for each building considering the entire available time series.

3.2. Data Reduction

The *Data Reduction* module, shown in Fig. 1, involves four primary tasks: *i) Time Interval Identification*, *ii) Energy Interval Identification*, *iii) SAX Transformation*, and *iv) SAX Optimization*.

The aim of the *Time Interval Identification* task is to represent the daily load profiles using a variable number of n_t time periods. This task is generally accomplished by dividing the hours of the day into n_t equal time intervals. However, this method is not always effective since the load profiles may have periods in which the load varies suddenly and periods in which the consumption remains constant. To address this issue, a CART is applied to identify variable-width temporal windows by optimally defining their boundaries, as proposed in [28]. However, the selection of the number of intervals n_t is constrained by the CART pruning functions. In order to tackle this problem, this paper proposes to determining the n_t intervals in the *SAX Optimization* task through a sensitivity analysis that involves the entire SAX encoding process.

The *Energy Interval Identification* task considers each *time split case* as input and calculates the mean value of the electricity consumption pattern for all the n_t time periods. This task reduces all the daily consumption patterns to n_t mean power consumption values. The Probability Density Function (PDF) of the mean values of the homogeneous group is then computed. Different *energy split cases* are generated for each *time split case* by dividing the PDF into a variable number n_e of equally probable energy intervals. The mean calculation helps to reduce the effect of large time intervals that have a constant consumption, which could cause unbalanced weights in the *Energy Interval Identification* task. A fair probability of the intervals is ensured by calculating the quantiles of the dataset population. This procedure therefore generates balanced energy intervals. However, to ensure equiprobability, it results that the power intervals are significantly skewed for any data that show distributions that deviate significantly from a Gaussian-type distribution. An additional step has been introduced to mitigate this issue, using a k-means algorithm to find a domain subdivision by minimizing the internal variance of each cluster. The k-means algorithm is initialized by providing the homogeneous clusters that were previously identified through quantiles as the starting clusters. This procedure is crucial as it removes the randomness of the process and allows the prior subdivision to be optimized with respect to the variance of the data.

The *SAX Transformation* task converts each daily load profile into words composed of n_t letters using a dictionary of n_e characters, which represent the (n_t, n_e) combination. Indeed, the daily load profiles are grouped in a particular word to determine a cluster.

The *SAX Optimization* exploits the three aforementioned tasks to identify the optimal combination of time and energy intervals among the (n_t, n_e) combinations. After conducting multiple tests, it was determined that the optimal domain to search for the ideal combination of (n_t, n_e) lies within the $[3, 7]$ range. This ensures an exhaustive search without any unnecessary burden on the optimization process. The optimal combination is determined by comparing each combination using the Mean Index Adequacy (MIA) index [38] in accordance with the elbow method. The MIA index expresses the goodness of the cluster (n_t, n_e) being examined as the average of the intra-cluster homogeneity, as known as the Within Cluster Sum of Squares (WCSS). The MIA index is expressed by Equation (2)

$$MIA = \frac{1}{N_k} \sum_{k=1}^{N_k} WCSS_k \quad (2)$$

where the N_k is the number of clusters corresponding to the number of different words detected by the SAX. $WCSS_k$ is calculated for each cluster k and it is equal to the daily average of the Euclidean distance of each profile from the cluster centroid, as expressed in (3)

$$WCSS_k = \frac{1}{N_{d_k}} \sum_{d_k}^{N_{d_k}} \left(\frac{1}{24} \sum_{t=1}^{24} |c_{k,t} - P_{d_k,t}|^2 \right)^{1/2} \quad (3)$$

where N_{d_k} is the number of days classified by SAX within the same cluster, i.e. distinguished by the same sequence of letters, while $c_{k,t}$ and $P_{d_k,t}$ are the hourly power values of the centroid of the k-th cluster and the daily load profile, d_k, belonging to the cluster, respectively.

3.3. Anomaly Detection

After defining the optimal combination of time windows and consumption intervals (n_t, n_e), the next step is to move on to the *Anomaly Detection* module shown in Fig 1. This module includes a preliminary task, that is, *SGT Identification*. This task involves using a purely statistical approach to exclude any profiles that are not significant for the structure under examination from the dataset. SGT is used for the subsequent applications of the semi-supervised ML approaches to detect anomalies, i.e., the *SAX-CART* and the *SAX-MLP* algorithms.

3.3.1. Synthetic Ground Truth

The *SGT Identification* task applies the Pareto principle [39] to eliminate less significant clusters from the optimal combination (n_t, n_e) identified in the previous module. This principle is commonly adopted for quality analysis in such fields as Total Quality Management, Six Sigma, and ISO9000 [40]. In this paper, the methodology developed in a previous work [5] is extended and included in the proposed ADF. The Pareto principle has been employed to assess the significance of a load profile. Therefore, a dataset is ordered on the basis of the recurrence of the SAX encoding, i.e., the magnitude of each cluster. Indeed,

20% of the clusters with the most recurrent words are expected to contain around 80% of the dataset, thus representing significant load profiles of the analyzed buildings. On the other hand, the remaining part of the dataset is distributed over minor clusters, which are sparse and could probably contain anomalous observations. These infrequent profiles are marked as outliers and removed from the dataset, as their inclusion in the training of ML models could worsen the accuracy of the models. Although this approach yields noticeable results, it suffers from certain limitations, in terms of defining the occurrence of a consumption anomaly. Nevertheless, it allows somewhat insignificant data to be excluded although it does not provide a robust definition of the occurrence of consumption anomaly.

The *SGT Identification* task provides the SAX profiles of days judged reliable for the semi-supervised training of the anomaly detection models as output. This output can be used directly by the SAX-CART algorithm, which was specifically designed to handle inputs encoded through SAX. The *SGT Identification* task can insted be exploited to exclusively extract the dates of the days belonging to SGT by re-adopting the original sampling of the dataset. This second approach is generalizable to various anomaly detection algorithms and was adopted to train the SAX-MLP algorithm.

3.3.2. SAX-CART

To provide a clear definition of anomalous consumption, the implementation of the CART algorithm by scikit-learn [41] has been adopted. This type of ML algorithm provides a substantial advantage in anomaly detection over other types of approaches, such as NN. Indeed, although NN often achieves higher accuracy values, it does not provide any information about the correlation between the input and output variables, which lead to a more difficult understanding of the phenomena. CART, on the other hand, makes the correlations identified between the dependent and independent variables explicit in a clear and intelligible way. The proposed framework adopts a decision tree that is trained for each time window as a classifier. The model predicts the level of consumption for each day by identifying the correct SAX coding. Table 1 reports the input variables used for the n_t trees. The variables are divided into categorical variables (i.e., *Day Type*, *Seasons*, *Months*) and numeric ones (i.e., *Outdoor Temperature* and the *Relative Humidity*).

Table 1: Input variables used for the CART models.

Type	Name	Value / Unit
Categorical	Day Type	0 Weekdays, 1 Saturdays, 2 Sundays and Holidays
	Season	0 Winter, 1 Spring, 2 Summer, 3 Fall
	Month	[1,12], where 1 represents January and 12 represents December
Numerical	Outdoor Temperature	[$^{\circ}C$]
	Relative Humidity	[%]

The numerical variables are transformed through Piecewise Aggregate Approximation (PAA) coding. They are averaged over the same time windows that were identified during the *Time Interval Identification* task. The partitioning of the dataset into a training sample (70%) and a validation sample (30%) ensures homogeneity of the two samples as it imposed that the proportion of symbols remains the same within the two samples.

A grid search, with cross-validation of the trained models, is implemented to optimize the selection of the hyperparameters. Table 2 reports the hyperparameters that were tested by the grid search algorithm. The *Criterion* defines the objective function through which the model is trained. The objective functions that were tested for this type of analysis were the gini and entropy indexes. The *Class weight* hyperparameter modifies the calculation of the objective function, particularly when it is set to "balanced", thereby allowing the SAX symbols to be weighted according to their probability of occurrence. In this way, single errors made for less frequent classes are penalized more, which sometimes leads to a better balance in the tree structure. Conversely, in strongly unbalanced cases, if the "balanced" parameter were not adopted, the tree algorithm could result in the misclassification of less densely populated consumption classes. In other words, the rarest classes would never be predicted at any leaf node and would therefore be incorrectly classified as anomalies. The remaining three hyperparameters tested do not modify the tree structure but do contribute to tree pruning. The *max depth* and *min impurity decrease* parameters define the maximum number of tree subdivisions and the minimum value of the Gini, or of the Entropy, index improvement following a split, respectively. The *CCP-*α parameter or Cost Complexity Pruning, which is similar to *min impurity decrease*, limits the introduction of a new split to those that exceed a minimum improvement value of the objective function. However, in the case of *CCP-*α, the improvement will not be made in absolute terms but weighed according to the complexity reached by the tree, i.e., the number of leaf nodes.

Table 2: Hyperparameters tested for The CART optimization

Hyperparameter	Tested value
Criterion	[gini, entropy]
Class weight	[None, balanced]
Max depth	[2, 3, 4, 5, 6]
Min impurity decrease	[0.05, 0.1, 0.15, 0.2, 0.25, 0.3]
CCP-α	[0.1, 0.15, 0.2, ..., 0.4, 0.45, 0.5]

Once the tuning of the model parameters has been performed, a robust tool that is capable of identifying the level of consumption within each time window is obtained. However, the accuracy of the model predictions is not linked directly to its ability to detect anomalies. In fact, any level of consumption that differs from the one predicted by the leaf node is not necessarily defined as anomalous, because this condition would be overly strict in many cases. Conversely, anomalous consumption levels are defined as those that occur with a frequency of less than 20% within a leaf node, on the basis of specific boundary conditions. This approach allows anomalies to be defined according to the context in which they are detected.

3.3.3. SAX-MLP

The second approach used to detect abnormal consumptions exploits the effectiveness of Multiple Layer Perceptron MLP algorithms to estimate electricity consumption through exogenous variables. In particular, the MLP algorithm developed by scikit-learn [41] has been adopted. In order to combine the results of the two ML approaches, the inputs provided to the MLP algorithm are the same as those provided to the previously described SAX-CART algorithm. The only additional variable provided to MLP is the time of day. However, it can be observed that this variable was also used intrinsically in the SAX-CART methodology to identify the SAX time windows. The input data to the NN are preprocessed on the basis of the data type, distinguishing between categorical data, such as the day of the week and month, and numerical data. Categorical data cannot be directly introduced into an NN, and the one-hot encoding technique is therfore used. Numerical data are standardized through a z-score transformation to improve the stability of the model at different input variable scales.

The anomalies in the SAX-MLP approach are defined on the basis of the assumption that the algorithm makes more pronounced errors in estimating anomalous consumptions than normal consumption. The definition of anomalies through this assumption is further supported by the fact that, during learning, the NN is exposed exclusively to the dataset extracted by means of the ground-truth identification process [22]. All the observations for which the NN commits a larger error than 3 times the standard deviation obtained by the model during the training phase are considered anomalous [17], as expressed in Equation (4)

$$\Delta = \max(0, |Y - \hat{Y}| - 3\sigma) \tag{4}$$

where Δ is the score of the anomaly, Y is the electricity consumption recorded by the meter, $\hat{Y}$ is the electricity consumption estimated by the NN, and σ is the standard deviation of the errors $(Y - \hat{Y})$ made by the model on the training dataset. This formulation makes it possible to exploit the predictions of the NN to detect the occurrence of abnormal consumptions.

The proposed SAX-MLP algorithm exploits a simple NN topology that is based on two hidden layers. Various hyperparameters were tested to identify an optimal configuration of the NN, as reported in Table 3.

The loss function is the function by which the model is evaluated and it influences the optimization algorithm by assigning values to the weights and biases. The loss function adopted is the Mean Squared Error. The reason for selecting this function was to highlight how sensitive the algorithm is to significant point errors. This approach is necessary to create a model that can accurately depict the dynamic nature of the real system. The used optimization algorithm is Adam, which uses the gradient calculation of the loss function to find the minimum point along its slope. The batch size parameter determines the number of elements after which the loss function is calculated and the weights and biases of the network are updated. This is a fundamental parameter that affects the learning time and stability of the network results to a great extent. The learning rate represents the percentage of weights that are updated at each epoch. When the learning rate value is low, the learning process slows down, which means it requires more epochs for the problem to reach convergence. On the other hand, if this parameter has high values, the model may converge quickly, but it could lead to low-quality solutions due to the presence of local minima. Finally, the decay hyperparameter is closely related to the Learning Rate (LR), as it modifies the latter by updating it after each epoch e through Equation (5)

Table 3: Hyperparameter tested for the MLP optimization

Hyperparameters	Tested value
Loss function	MSE
Optimizer	adam
Initialization mode	uniform
Activation	ReLu
Decay	[0, 1e-6, 1e-5, 1e-4, 1e-3]
Dropout rate	[0.0, 0.1, 0.12, 0.14, 0.2]
Learning rate	[0.0001, 0.001, 0.01, 0.1, 0.3]
Neurons HL1	[26, 50, 100, 150]
Neurons HL2	[25, 45, 65]
Batch size	[24, 168]

$$LR = LR_{e=0} * \frac{1}{1 + Decay * e} \tag{5}$$

Therefore, high learning rates are permitted during the initial epochs, and this allows a faster learning. However, lower learning rates are also possible in the final learning stages, and these result in better refinement. The drop-out parameter is introduced to encourage the NN to adopt low values for the weights of each connection, thereby avoiding overfitting the dataset. A percentage of neurons that is equal to the parameter value is randomly excluded during each epoch through this parameter. This practice prevents a neuron from acquiring excessive importance within the network and makes learning more diffused. Finally, the HL1 and HL2 parameter Neurons establish the different combinations of the number of neurons between the two hidden layers.

The dataset is split by respecting the proportions used for training (70%) and for testing (30%) . The two samples are split in a semi-random manner by imposing that the two samples are balanced with respect to the months to which the observations belong. This expediency reduces the risk of inducing seasonality-related bias into the model. All the possible combinations of the hyperparameters listed in Table 3 were tested by applying the cross-validation. The latter was achieved by dividing the training dataset into three folders and using two folders alternately for model training and one for testing. The optimal combination of hyperparameters was defined on the basis of the Mean Absolute Percentage Error (MAPE) calculated from the three splits of the designated test folder.

4. Experimental Results and Analysis

The proposed Anomaly Detection Framework (ADF) ha been applied to a real case study of around 1,000 building distributed throughout the Italian territory from the nationa TLC provider. The dataset consists of a collection of record taken from smart meters at an hourly resolution. Moreover, th dataset contains a unique identifier for each building, the fina use (e.g., Central Office (CO), data center, radio base station or offices), and the geographic location. It is worth noting tha the data were all anonymized to preserve private and confiden tial information about the TLC provider, due to the presenc of NDA. Therefore, all the identifiers were masked, and th consumption profiles were normalized. In order to present th dataset, within the limits of NDA agreements, the following statistical parameters were reported. The mean value of normalized annual consumption 0.289 suggested that, on average, the CO exhibits relatively low consumption levels. This can be attributed to two main factors: limited full load operation hours of the power plants and the distribution of power plant sizes within the population. The median value of 0.192 is noticeably lower than the mean, indicating an asymmetrical distribution towards lower consumption values. This suggested that a significant proportion of power plants had consumption levels below the mean, contributing to the skewness of the distribution. The first percentile (Q1) value of 0.0744 was significantly closer to the median than the third percentile (Q3) value of 0.506. This further confirmed the asymmetry in the consumption distribution, with a larger spread of power plants falling between the median and the third quartile compared to those with the lowest consumption. The dataset presented was accompanied by a meteorological dataset, which included hourly data on the temperature and, relative humidity, was also used.

As described in Section 3, and depicted in Fig. 1, the first step of the ADF involves the *Data Preprocessing* module, which performs the filtering, clustering and normalization of the dataset. Fig. 2 depicts the typical point anomalies identified in the dataset. After conducting further investigations and consulting domain experts, two primary reasonswere identified for these anomalies: *i)* scheduled maintenance interventions and *ii)* errors in the hourly power accounting. The conducted verifications confirmedthat the filtering task led to the expected results, which means that any anomalies that were representative of a malfunctioning of the systems were not excluded from the analysis. Fig. 2 *(a)* shows a typical case of anomalies related to the data acquisition system. An anomaly related to scheduled maintenance is shown in Fig. 2 *(b)*, which is characterized by a specific shape that indicates a sudden reduction and an increase in load over intervals of a maximum of three hours. These types of anomalies do not constitute opportunities for energy savings or energy management optimization, and they were therefore deemed of little interest for the subsequent analyses and consequently excluded from the dataset.

The dataset was clustered to identify several building groups to which the anomaly detection process was applied. A total of 160 homogeneous building groups were identified on the basis of their geographical location and the final use of the building.

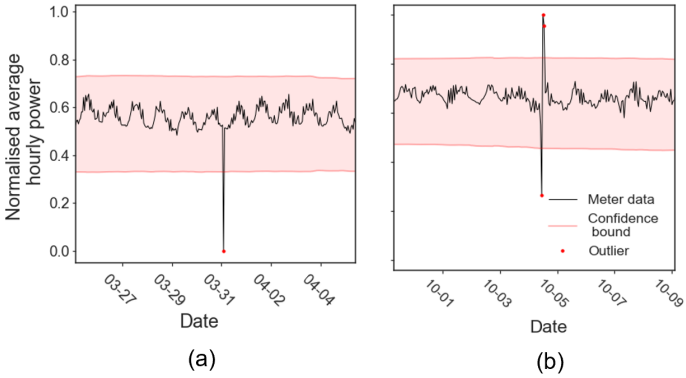

Figure 2: Examples of typical outliers identified through the *Filtering* task of the *Data Preprocessing* module.

For the sake of simplicity, the following analyses focuses on only of those clusters, which consists of 6 COs in the area of Turin. Fig. 3 shows carpet plots of the normalized hourly load profiles of the selected building group. These plots provide a concise representation of the entire dataset, and highlight typical seasonal trends, such as the effect of cooling loads on summer days, which increase the load consumption patterns from about 12 to 19. Looking at the consumption graphs of these buildings as a whole, it can be seen that they show the same behaviour in the months from June to September. In other words, it can be observed that the peak energy demand for the six buildings occurs on the same days in June and August. It can be seen, when comparing the buildings, that the consumption patterns instead vary more during the colder months. Buildings 2 and 5, unlike the other buildings in the cluster, do not show a low level of consumption during the months of January to April, and instead maintain an intermediate level of consumption. This may be a first sign of the possibility of an improvement in the management of the two buildings.

The *Data Reduction* module involves SAX optimization, which is achieved by testing different time and energy intervals to identify the optimal combination (n_t, n_e) of the SAX encoding. As detailed in Section 3, the MIA index is calculated for each combination. Therefore, the resulting matrix of MIA indices, which is depicted in Fig. 4 as a three-dimensional histogram, is thus obtained. The optimal combination is obtained from the elbow point identification of the MIA matrix, which results in $(4, 4)$ the best time and energy interval parameter configuration.

For the sake of clarity, the following two paragraphs only report the procedure used to obtain the SAX encoding of the identified optimal combination $(4, 4)$:

- The *Time Interval Identification* task identifies four-time intervals, i.e., [00-09], [10-11], [12-19], [20-23] considering the 24-hour clock system, by exploiting a CART algorithm implementation. At the end of the task, the load profiles are reduced in size by calculating the average power consumption for each identified time window. In this way, each daily observation is no longer composed of 24 variables but only of 4 mean energy consumption values. The resulting segmentation effectively characterizes the con-

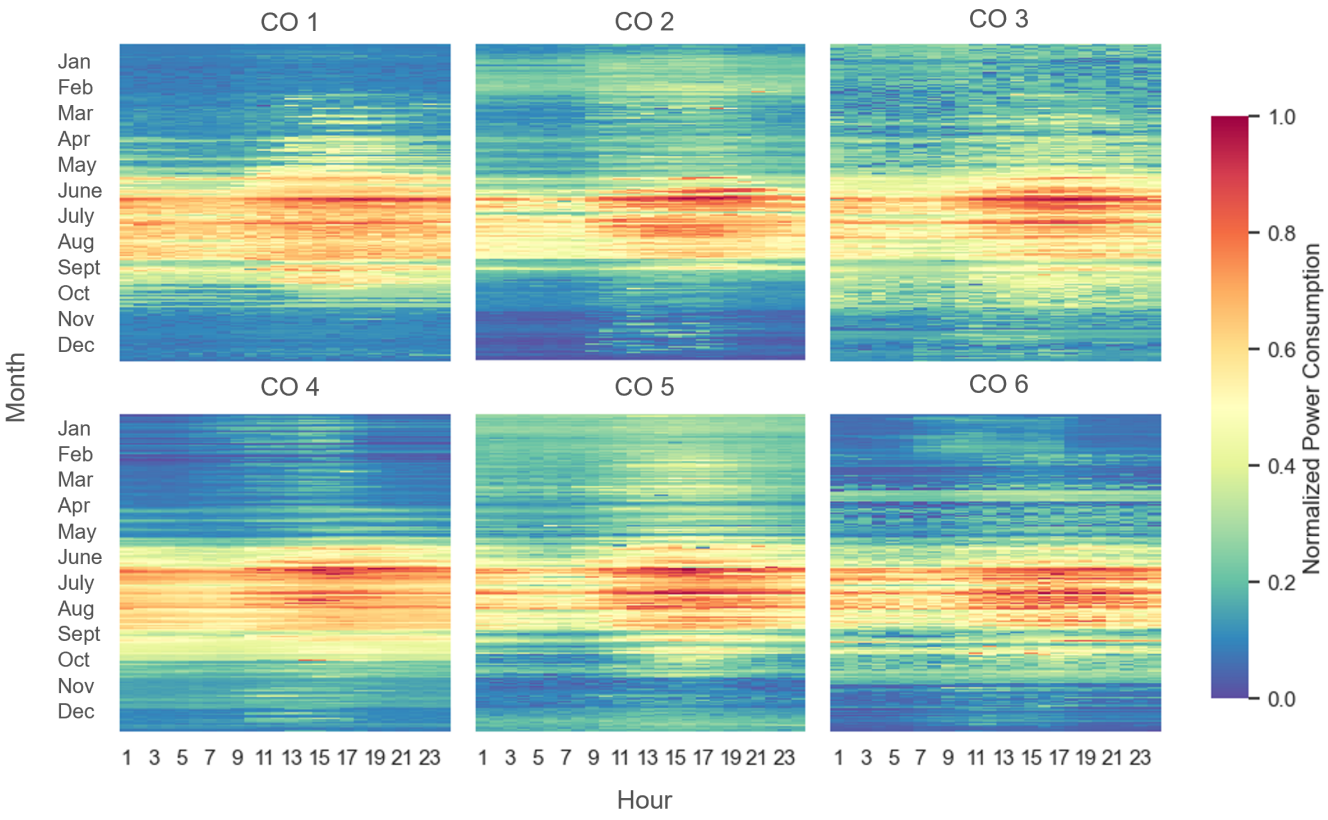

Figure 3: Carpet plot representation of the selected homogeneous group of six COs. The color scale shows the power consumption of each CO.

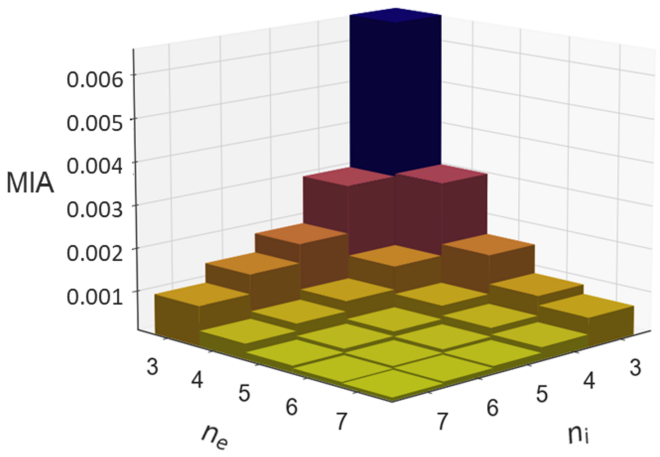

Figure 4: Graphical representation of the MIA matrix considering the different combinations (n_t, n_e).

sumption pattern of a CO. The first time interval corresponds to a low consumption range. The next window [10-11] identifies a period of sudden load variations, which is commonly called system rump-up. The central time interval [12-19] represents the period of maximum consumption of the CO. Finally, the [20-23] window represents the period in which the consumption settles at low levels.

- The *Energy Interval Identification* task identifies four equiprobable power intervals, as reported in Fig. 5*(a)*. It can be noted that the distribution of the power sampling is markedly unbalanced toward low values. Therefore, the energy intervals are denser for low values and sparser for high consumption values. This aspect represents a weak output to be used for SAX transformation, as it could lead to inadequately detailed the high consumption intervals. A k-means division of the domain is adopted to avoid this type of coding weakness. The new subdivision of the domain, as depicted in Fig. 5*(b)*, shows more balanced intervals. It can be observed that the high consumption interval has been reduced by almost a third, while the low and medium-high consumption intervals have been increased, accordingly.

A comprehensive graphical representation of the information that can be extracted from the selected SAX encoding is illustrated in Fig. 6. The figure depicts a Sankey chart that show the sequential generation of the load profiles obtained through concatenation of the symbols of each time window. Each block represents a time window and reports two types of information: the consumption level and the number of days described by the same code. The consumption level of the considered window can be deduced either through color code of the block or by considering the last letter of the code placed next to each block. The number of load profiles belonging to each block can insted be deduced by considering the thickness of each block or path connecting two blocks. Moreover, Fig. 6 shows the carpet graph of the load profiles for each of the 57 letter sequences identified by the SAX codes of the homogeneous building cluster. By

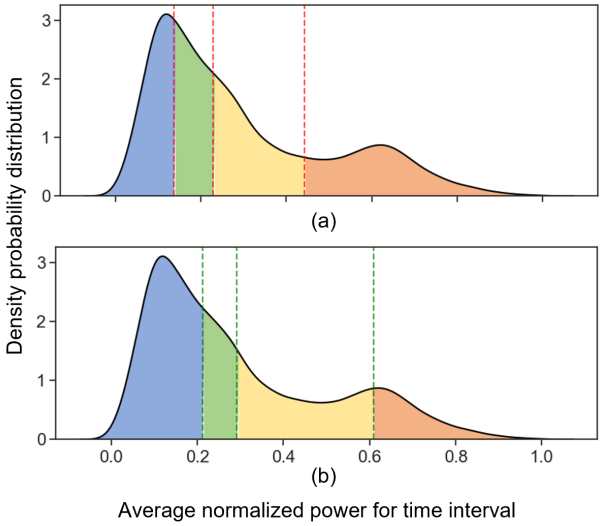

Figure 5: Graphical representation of the Probability Density Function (PDF) of the population of power demand values averaged over the time windows. *(a)* shows the equiprobable subdivision while *(b)* shows the domain subdivision obtained through the application of k-means.

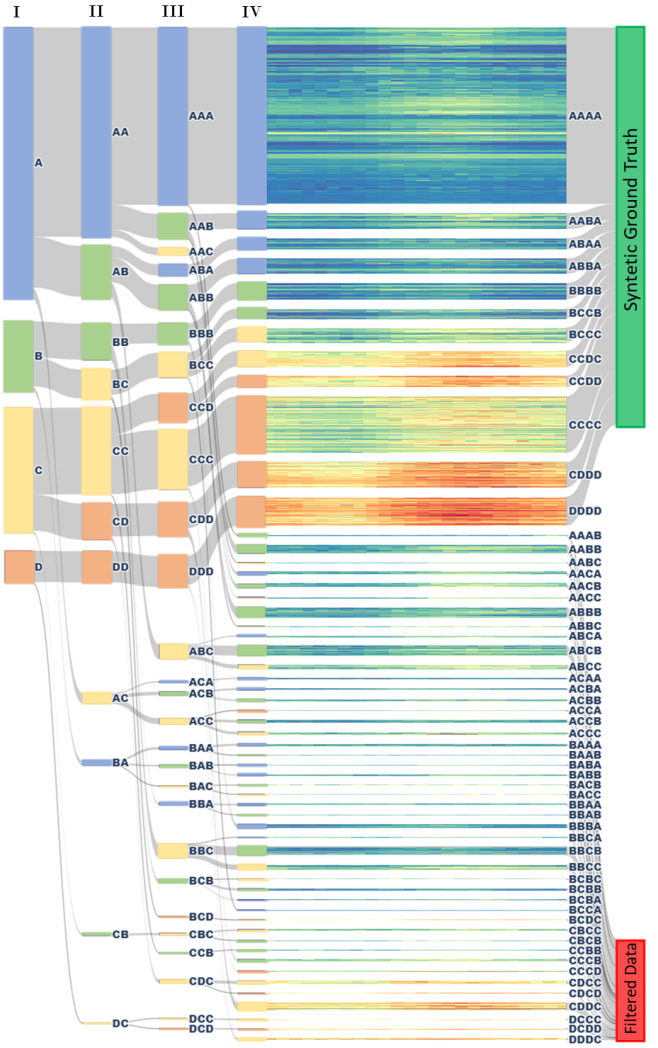

Figure 6: Sankey chart of the SAX encoding. Each day is represented in the graph by the sequence of symbols assigned to it for each time window. The thickness of each path is related to the number of days belonging to such a pat. The figure also reports, for each path, the carpet graph of the consumption profile. The SGT is obtained by following the Pareto principle.

analyzing the carpet plot, it is possible to see that the most significant cluster, i.e., the densely populated ones, report consistent information. This shows that the SAX encoding process is able to preserve the most relevant information of the analyzed load profiles. Moreover, it is possible to see that each cluster presents consistent load profiles. Furthermore, it is also possible to notice differences between the clusters. This is particular evident when looking at the clusters related to high consumption, e.g., the days coded as 'CCDD' differ considerably from those described by the 'CDDD' term, although only the second letter between the two codings varies. It is also worth noting that the first cluster, 'AAAA', shows some behaviors that could be presented in more detail, thus representing a constraint of the proposed approach. This aspect is related to a trade-off in the representation of the SAX. However, it is necessary to accept this inaccuracy in order to achieve a codification that generalizes building behavior and which is not overly specific. However,cataloguing the different consumption patterns of buildings in a reduced set would in fact be unattainable.

Once the SAX codingof the data set has been obtained, the *Anomaly Detection* analysis is conducted. First, the SGT from the historical consumption series is obtained. The results of this process can already be observed in Fig. 6, where it can in fact be seen that all the paths are separated into two groups: the SGT and the filtered data. The Pareto diagram is represented in Fig. 7 to explicate the details of the procedure.

The Pareto diagram shows the daily encodings for each of the 57 resulting words on the x-axis, ordered according to the recurrence of each code within the dataset. The recurrence of each word is reported quantitatively on the left axis. The right axis shows the cumulative percentage of the selected words. The 13 most recurrent symbols, which account for around 20% of the 57 identified words, represent approximately 80% of the dataset. The remaining part of the dataset (20%) is instead represented by 45 different symbols. Consequently, the dispersion of this part of the dataset poses a learning challenge to to the ML algorithms, which may lead to a deterioration of their performance.

Taking advantage of the results obtained so far, it is possible to extend the applications of the SAX coding from the definition of SGT to the detection of anomalies, as described in Section 3, through the application of semi-supervised ML algorithms. The training results of the two SAX-CART and SAX-MLP algorithms are presented in the following paragraphs. The outputs of the two anomaly detection algorithms are compared using the carpet plots outlined in Section 4.1.

Since the optimal number of time windows in the SAX-CART application is four, an equal number of decision trees was initialized, and a grid search with cross-validation was performed for each of them. The results of the grid search are

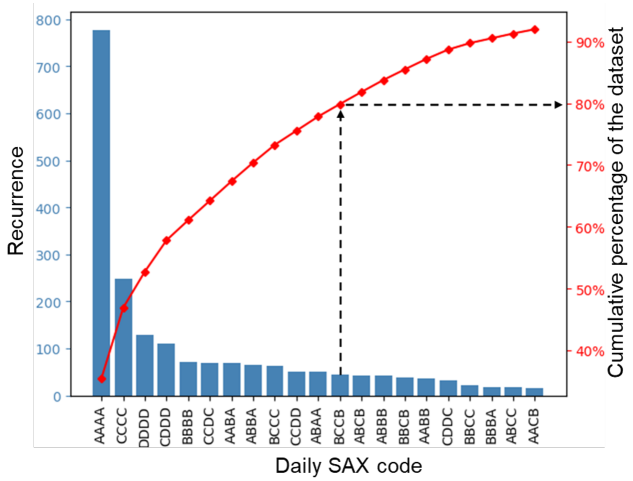

Figure 7: Pareto diagram of the resulting (4, 4) Clustering step.

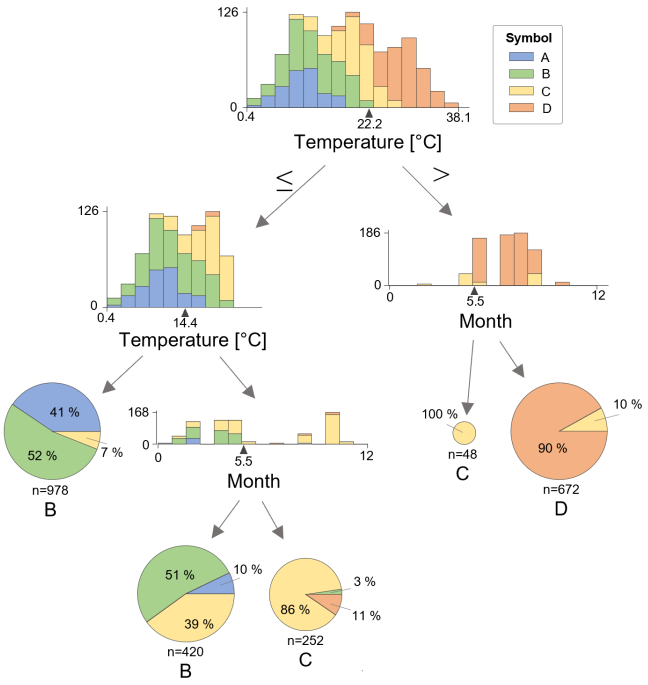

Figure 8: Graphical representation of the CART obtained for the [10-11] time window.

reported in Table 4, where the optimal hyperparameters identified through grid search and the Key Performance Indicators (KPIs) achieved for each time window are shown. It is possible to compare the accuracy values achieved for the three different split tests, as described in Section 3.3.2, and also the mean and standard deviation values calculated for them.

The average prediction accuracy over the four time windows is 80.9 %. Any anomalous periods can be identified on the basis of this result. For example, the CART obtained for the second time window [10-11], is depicted in Fig. 8. In this case, five leaf nodes can be identified, which are analyzed from left to right. The first leaf node belongs to days characterized by an average temperature that is less than or equal to 14.4 °C. According to the results obtained from CART, consumption levels A and B within this category can be considered acceptable as they exceed the support threshold of 20 %. These levels are therefore deemed normal under these specific circumstances. On the other hand, consumption level C is only expected with a probability of 7 %, and can therefore be considered anomalous in these particular circumstances, as can the consumption level D. The second leaf node is defined by a higher external temperature than 14.4 °C and for the days of the months between January and May. Again, two consumption levels (B and C) are identified as normal. What was done for the two previously described leaf nodes can now be extended to all the leaf nodes and all the trees, and this allows all the different scenarios to be mapped. Therefore, by exploiting the identified rules, it is possible to distinguish any abnormal consumption from normal consumption.

As far as the SAX-MLP application is concerned, the NN was trained following a similar methodology to what had previously been used for CART. The optimal parameters for the network were defined through a grid search, the domain of which is explicitly stated in Section 3. Table 5 reports the hyperparameters and KPIs that were tested for the different configurations. The optimal configuration is the one with the lowest mean MAPE, which is equal to 14.49 %. Finally, the anomaly score can be calculated from the optimal configuration using Equation (4).

4.1. Anomaly Detection analysis

If only the performance indices of the two models are compared, it could be argued that the performance of the SAX-MLP exceeds that of the SAX-CART approach. However, the score obtained when predicting consumption is not completely indicative of the performance of the same model in defining anomalous behavior. In fact, it should be pointed out that the ultimate goal of the analysis was not to obtain an accurate consumption estimation but was rather the dual task of identifying anomalies and supporting energy managers in verifying the authenticity of an alarm by combining both approaches. For this reason, the first characteristic that should be evaluated for the two models is the type of output they provide. The NN-based method only provides the distance between the prediction and the actual consumption value, whereas the SAX-CART approach not only provides a warning about possible anomalies and the most likely consumption level for a specific condition, but also the variables that appear for the drivers that best characterize a given context. This latter aspect is undoubtedly helpful for those stakeholders who want to identify the incidents that have triggered a specific outlier.

The analysis carried out for CO 2 is reported as an example of the application of the developed ADF. The anomaly detection analysis was conducted for the year following the one used for model training. Fig. 9 depicts four carpet plots that that depict, from left to right: *i)* the external temperature, which represent the primary consumption driver, *ii)* the normalized consumption demand, *iii)* the anomaly score defined by means of SAX-MLP, and *iv)* the identified anomalies extracted from

Table 4: Result of the top 5 combinations of the hyperparameters tested through a grid search for the CART models applied in the SAX-CART algorithm

CART	Hyperparameters					KPIs				
	CCP-α	Class weight	Criterion	Max depth	Min impurity decrease	Split0 accuracy	Split1 accuracy	Split2 accuracy	Mean accuracy ▽	Std accuracy
CART 1 [00 - 09]	**0**	**balanced**	**entropy**	**4**	**0.05**	**0.871**	**0.859**	**0.918**	**0.882**	**0.025**
	0	None	entropy	4	0.05	0.871	0.882	0.871	0.875	0.006
	0.35	None	entropy	3	0.15	0.871	0.871	0.859	0.867	0.006
	0.25	None	entropy	3	0.1	0.871	0.871	0.847	0.863	0.011
	0.05	None	gini	5	0.05	0.859	0.871	0.847	0.859	0.01
CART 2 [10 - 11]	**0**	**None**	**entropy**	**5**	**0.05**	**0.718**	**0.765**	**0.776**	**0.753**	**0.025**
	0.1	balanced	entropy	4	0.05	0.753	0.765	0.729	0.749	0.015
	0.05	balanced	entropy	4	0.1	0.753	0.765	0.718	0.745	0.02
	0.05	None	gini	2	0.05	0.682	0.753	0.788	0.741	0.044
	0	balanced	entropy	2	0.05	0.718	0.753	0.718	0.729	0.017
CART 3 [12 - 19]	**0**	**None**	**gini**	**2**	**0.05**	**0.847**	**0.847**	**0.8**	**0.831**	**0.022**
	0.05	None	entropy	5	0.05	0.824	0.824	0.776	0.808	0.022
	0.05	balanced	entropy	4	0.15	0.859	0.776	0.765	0.8	0.042
	0.05	balanced	gini	2	0.05	0.859	0.741	0.765	0.788	0.051
	0	balanced	gini	3	0.05	0.835	0.741	0.776	0.784	0.039
CART 4 [20 - 23]	**0.2**	**balanced**	**entropy**	**4**	**0.15**	**0.812**	**0.776**	**0.729**	**0.773**	**0.034**
	0.05	balanced	gini	2	0.1	0.788	0.776	0.729	0.765	0.025
	0.2	None	entropy	4	0.25	0.776	0.741	0.765	0.761	0.015
	0	None	gini	2	0.05	0.765	0.741	0.765	0.757	0.011
	0	None	entropy	5	0.05	0.788	0.718	0.753	0.753	0.029

Table 5: Result of the top 5 combinations of the hyperparameters tested through a grid search for the MLP models.

Hyperparameters						KPIs					
Decay	Dropout rate	Learning rate	Ns HL1	Ns HL2	Batch size	Split0 MAPE	Split1 MAPE	Split2 MAPE	Mean MAPE △	Std MAPE	Mean fit time
0.001	**0**	**0.3**	**150**	**45**	**24**	**14.99**	**14.19**	**14.29**	**14.49**	**0.4**	**114**
0	0	0.001	150	45	24	14.17	14.94	14.45	14.52	0.3	140
1.00E-06	0	0.1	150	45	24	14.83	13.91	14.84	14.52	0.4	138
1.00E-05	0	0.0001	150	45	24	15	14.32	14.4	14.57	0.3	131
0.001	0	0.001	150	45	24	14.02	15.61	14.19	14.61	0.7	114

SAX-CART. When the two methods used to define anomalies are analyzed, the presence of more noise in SAX-MLP is immediately evident, which indicates thatseveral individual hours are anomalous. Instead, the output derived from SAX-CART appears more defined. However, even in this case, there are some warnings that do not correspond to actual anomalies but instead refer to biases related to the encoding of the leaf nodes for which conditions very close to existence domain of a leaf node can lead to a misclassification. However, these types of errors are easily verifiable and bypassable thanks to the interpretability of the tree rules.

The period with the largest number of anomalies identified by both algorithms covers the period from the last days of March to the first half of April, as depicted in Fig. 9. In this period, it can be noted that the consumption does not follow the load profiles that are typical of those specific temperatures. This type of information can be deduced through the SAX-CART approach, which allows the most significant variables to be clearly highlighted in order to define a specific context. For example, by analyzing one of the days marked as abnormal by both models, i.e., 03/29, it is possible to interpret the causes that produced the alarms in the [00-09] and [20-23] time windows. Table 6 can be used to support analysts as it shows the main outputs that can be obtained from the application of SAX-CART. According to Table 6, the SAX reports a high average consumption level, which has the symbol C, for both of the indicated windows, i.e. with an anomaly score equal to one. If we consider the [00-09] time window, the rules obtained from SAX-CART suggest that, the most probable consumption level for temperature levels below 10 °C is the lowest one, i.e., symbol A. This estimate is supported by 80 % of the trained dataset. Therefore, the actual consumption level recorded (symbol C) may not be related to the external temperature, and the algorithm targets it as anomalous. On the other hand, only SAX-MLP identifies anomalous

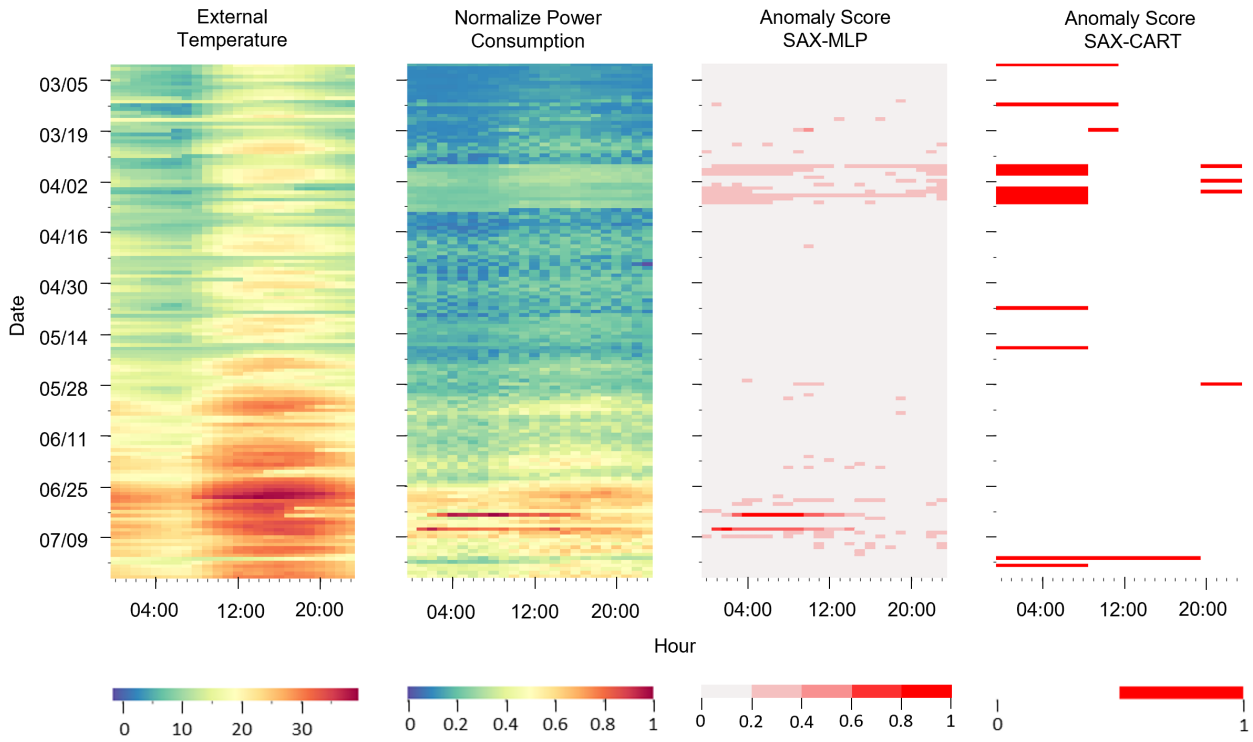

Figure 9: Carpet plot representation of the External Temperature, Normalized Power Consumption, and the Anomaly Scores for the SAX-MLP and SAX-CART models regarding the building CO 2.

behaviors on the days 07/03 and 07/07, which SAX-CART does not identify. This inaccuracy is not due to an error of the CART process, but is instead a feature related to the implementation of theSAX encoding, for which the maximum power interval is still too wide, even after applying k-means, as almost 40% of the existence domain of the normalized power variable is included within this interval. In conclusion, the combined use of these approaches allows the limitations of SAX-CART, pertaining to the accuracy of anomaly detection, to be overcome, and vice versa, it extends the information obtained from the SAX-MLP approach. Indeed, even in cases where the SAX-CART approach fails to identify anomalous behavior, it can still provide information about which drivers are the most significant for the particular case under examination. In summary, the conducted study has successfully identified anomalous behaviors within the historical consumption series of a CO, highlighting instances of excessive consumption compared to climatic and temporal conditions. While the level of detail attained may not precisely pinpoint the origin of the anomaly, it does allow for the formulation of initial hypotheses on the operating state of the air-conditioning system.

Table 6: SAX-CART output related to day 03/29 to support the analysis of the detected anomalies. Abbreviation: DT (Day Type), M (Month), S (Season), T (Temperature), RH (Relative Humidity)

Cart Window	Anomaly Score	SAX	SAX-CART Inputs					CART Symbol Prob. [%]				CART Rules
			DT	*M*	*S*	*T* [°C]	*RH* [%]	A	B	C	D	
[00 - 09]	**1**	**C**	0	3	1	**7.9**	63	80	20	**0**	0	$T < 10$
[10 - 11]	0	C	0	3	1	15.5	39	7	52	41	0	$14.4 < T \leq 22.2$ $M \leq 5$
[12 - 19]	0	C	0	3	1	16.7	35	6	55	39	0	$13.6 < T \leq 18.7$ $M \leq 6$
[20 - 23]	**1**	**C**	0	**3**	1	**13.5**	44	72	28	**0**	0	$5 < T \leq 16.8$ $M \leq 5$

5. Conclusion

This work proposes an Anomaly Detection Framework that is based on two complementary semi-supervised machine-learning approaches which can be used for building electricity consumption data. The aim of both approaches is to detect contextual anomalies and they attain this goal through the adoption of non-autoregressive algorithms. In this study, emphasis is placed on the type of output that the algorithms return to the end user, as the interpretability of anomaly detection is fundamental. These types of algorithms are often used in information systems, and the intelligibility of the outputs is therefore crucial to fully exploit them. The application of the framework defined in Section 3 allowed for the identification of anomalous consumption occurrences in the historical series of a CO, validating the effectiveness of the method for the specific case study. However, this same application also revealed the main limitations of the methodology. The approach defined as SAX-CART can lead to both false alarms and missed alarms, which is why it was complemented with the SAX-MLP method, that provides higher accuracy. Nonetheless, excluding the SAX-CART method from the monitoring system is not a viable choice, as it is crucial for verifying the typical behavior of the power plant under certain conditions. Therefore, it

is believed that the greatest benefits can be achieved through a combined approach of both methodologies. By comparing the results from both methods, one can quickly determines which alarms are unequivocally false positives and which ones require further investigation. The integration of both methods allows for a more efficient and effective monitoring system. In conclusion, it is deemed that the best approach involves the use of both methodologies together, as it provides a more comprehensive understanding of the system's behavior. Furthermore it helps to improve the accuracy of anomaly detection while retaining the ability to assess the power plant's typical behavior easily and effectively.

Future work will be directed toward attempt to validate the identified techniques by using them in the field and evaluating the alarms provided by the algorithms, in order to be able to detect the presence of false alarms and consequently identify and test possible improvements to the tested algorithms.

Credit Authorship Contribution Statement

L. Mascali: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Visualization **D.S. Schiera:** Conceptualization, Software, Investigation, Data Curation, Writing - Review & Editing, Visualization, Supervision **S. Eiraudo:** Conceptualization, Software, Investigation, Data Curation **L. Barbierato:** Conceptualization, Writing - Review & Editing, Supervision **R. Giannantonio:** Resources, Supervision, Project administration **E. Patti:** Writing - Review & Editing **L. Bottaccioli:** Writing - Review & Editing, Supervision **A. Lanzini:** Conceptualization, Writing - Review & Editing, Supervision, Project administration

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data has been provided by TIM S.p.A., and it is confidential due to the presence of Non-Disclosure Agreements.

References

[1] M. González-Torres, L. Pérez-Lombard, J. F. Coronel, I. R. Maestre, D. Yan, A review on buildings energy information: Trends, end-uses, fuels and drivers, Energy Reports 8 (2022) 626–637.

[2] Z. Zhongming, Z. Wangqiang, L. Wei, et al., World energy balances overview (2020 edition) (2020).

[3] S. Katipamula, R. M. Underhill, N. Fernandez, W. Kim, R. G. Lutes, D. Taasevigen, Prevalence of typical operational problems and energy savings opportunities in us commercial buildings, Energy and Buildings 253 (2021) 111544.

[4] C. Fan, D. Yan, F. Xiao, A. Li, J. An, X. Kang, Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches, in: Building Simulation, Vol. 14, Springer, 2021, pp. 3–24.

[5] L. Mascali, S. Eiraudo, L. Barbierato, D. S. Schiera, R. Giannantonio, E. Patti, L. Bottaccioli, A. Lanzini, Synthetic ground truth generation of an electricity consumption dataset, in: 2022 International Conference on Smart Energy Systems and Technologies (SEST), 2022, pp. 1–6. doi:10.1109/SEST53650.2022.9898444.

[6] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM computing surveys (CSUR) 41 (3) (2009) 1–58.

[7] G. Pang, C. Shen, L. Cao, A. V. D. Hengel, Deep learning for anomaly detection: A review, ACM computing surveys (CSUR) 54 (2) (2021) 1–38.

[8] M. E. Villa-Pérez, M. A. Alvarez-Carmona, O. Loyola-Gonzalez, M. A. Medina-Pérez, J. C. Velazco-Rossell, K.-K. R. Choo, Semi-supervised anomaly detection algorithms: A comparative summary and future research directions, Knowledge-Based Systems 218 (2021) 106878.

[9] S. Maleki, S. Maleki, N. R. Jennings, Unsupervised anomaly detection with lstm autoencoders using statistical data-filtering, Applied Soft Computing 108 (2021) 107443.

[10] S.-V. Oprea, A. Bâra, F. C. Puican, I. C. Radu, Anomaly detection with machine learning algorithms and big data in electricity consumption, Sustainability 13 (19) (2021) 10963.

[11] G. Chicco, R. Napoli, F. Piglione, Load pattern clustering for short-term load forecasting of anomalous days, in: 2001 IEEE Porto Power Tech Proceedings (Cat. No. 01EX502), Vol. 2, IEEE, 2001, pp. 6–pp.

[12] M. Fahim, A. Sillitti, An anomaly detection model for enhancing energy management in smart buildings, in: 2018 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm), IEEE, 2018, pp. 1–6.

[13] J.-S. Chou, A. S. Telaga, Real-time detection of anomalous power consumption, Renewable and Sustainable Energy Reviews 33 (2014) 400–411.

[14] C. Xu, H. Chen, A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data, Energy and Buildings 215 (2020) 109864.

[15] C. Chahla, H. Snoussi, L. Merghem, M. Esseghir, A deep learning approach for anomaly detection and prediction in power consumption data, Energy Efficiency 13 (2020) 1633–1651.

[16] L. Lei, B. Wu, X. Fang, L. Chen, H. Wu, W. Liu, A dynamic anomaly detection method of building energy consumption based on data mining technology, Energy 263 (2023) 125575.

[17] H. Rashid, P. Singh, Monitor: An abnormality detection approach in buildings energy consumption, in: 2018 IEEE 4th international conference on collaboration and internet computing (CIC), IEEE, 2018, pp. 16–25.

[18] M. Jesmeen, J. Hossen, A. B. A. Aziz, Unsupervised anomaly detection for energy consumption in time series using clustering approach, Emerging Science Journal 5 (6) (2021) 840–854.

[19] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, A. Amira, Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives, Applied Energy 287 (2021) 116601.

[20] C. Miller, Z. Nagy, A. Schlueter, A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings, Renewable and Sustainable Energy Reviews 81 (2018) 1365–1377.

[21] W. Cui, H. Wang, Anomaly detection and visualization of school electricity consumption data, in: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, 2017, pp. 606–611.

[22] H. Janetzko, F. Stoffel, S. Mittelstädt, D. A. Keim, Anomaly detection for visual analytics of power consumption data, Computers & Graphics 38 (2014) 27–37.

[23] Â. P. Alves, A. M. Milani, I. H. Manssour, Visual analytics system for energy data in smart cities and buildings, in: 2020 IEEE International Smart Cities Conference (ISC2), IEEE, 2020, pp. 1–8.

[24] C. Zhang, Y. Zhao, T. Li, X. Zhang, M. Adnouni, Generic visual data mining-based framework for revealing abnormal operation patterns in building energy systems, Automation in Construction 125 (2021) 103624.

[25] C. Fan, F. Xiao, M. Song, J. Wang, A graph mining-based methodology for discovering and visualizing high-level knowledge for building energy management, Applied Energy 251 (2019) 113395.

[26] D. Streeb, Y. Metz, U. Schlegel, B. Schneider, M. El-Assady, H. Neth, M. Chen, D. A. Keim, Task-based visual interactive modeling: decision

trees and rule-based classifiers, IEEE Transactions on Visualization and Computer Graphics 28 (9) (2021) 3307–3323.

[27] X. Liu, Y. Ding, H. Tang, F. Xiao, A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data, Energy and Buildings 231 (2021) 110601.

[28] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, G. Chicco, Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings, Energy 157 (2018) 336–352.

[29] C. Miller, Z. Nagy, A. Schlueter, Automated daily pattern filtering of measured building performance data, Automation in Construction 49 (2015) 1–17.

[30] P. Patel, E. Keogh, J. Lin, S. Lonardi, Mining motifs in massive time series databases, in: 2002 IEEE International Conference on Data Mining, 2002. Proceedings., 2002, pp. 370–377. doi:10.1109/ICDM.2002.1183925.

[31] E. Keogh, J. Lin, A. Fu, Hot sax: Efficiently finding the most unusual time series subsequence, in: Fifth IEEE International Conference on Data Mining (ICDM'05), Ieee, 2005, pp. 8–pp.

[32] S. Yang, Y. Wang, J. Zhang, A similarity measure for time series based on symbolic aggregate approximation and trend feature, in: 2020 39th Chinese Control Conference (CCC), 2020, pp. 6386–6390. doi:10.23919/CCC50068.2020.9189060.

[33] K. Zhang, Y. Li, Y. Chai, L. Huang, Trend-based symbolic aggregate approximation for time series representation, in: 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 2234–2240. doi:10.1109/CCDC.2018.8407498.

[34] Y. Zhang, L. Duan, M. Duan, A new feature extraction approach using improved symbolic aggregate approximation for machinery intelligent diagnosis, Measurement 133 (2019) 468–478.

[35] N. D. Pham, Q. L. Le, T. K. Dang, Two novel adaptive symbolic representations for similarity search in time series databases, in: 2010 12th International Asia-Pacific Web Conference, 2010, pp. 181–187. doi:10.1109/APWeb.2010.23.

[36] M. S. Gallimore, C. M. Bingham, M. J. W. Riley, Self-organising symbolic aggregate approximation for real-time fault detection and diagnosis in transient dynamic systems, in: 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 2017, pp. 000043–000048. doi:10.1109/SAMI.2017.7880350.

[37] S. Eiraudo, L. Barbierato, R. Giannantonio, A. Porta, A. Lanzini, R. Borchiellini, E. Macii, E. Patti, L. Bottaccioli, A machine learning based methodology for load profiles clustering and non-residential buildings benchmarking, IEEE Transactions on Industry Applications (2023).

[38] S. A. Notari, G. Chicco, F. Piglione, Data size reduction with symbolic aggregate approximation for electrical load pattern grouping, IET Generation, Transmission & Distribution 7 (2) (2013) 108–117.

[39] R. Sanders, The pareto principle: its use and abuse, Journal of Services Marketing (1987).

[40] L. Wilkinson, Revising the pareto chart, The American Statistician 60 (4) (2006) 332–334.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.