Article

The Application of Hierarchical Clustering to Power Quality Measurements in an Electrical Power Network with Distributed Generation

Michał Jasiński [1],*, **Tomasz Sikorski [1]**, **Zbigniew Leonowicz [1]**, **Klaudiusz Borkowski [2]** and **Elżbieta Jasińska [3]**

1   Department of Electrical Engineering Fundamentals, Faculty of Electrical Engineering, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland
2   KGHM Polska Miedź S.A., 50-301 Lubin, Poland
3   Faculty of Law, Administration and Economics, University of Wroclaw, 50-145 Wroclaw, Poland
*   Correspondence: michal.jasinski@pwr.edu.pl; Tel.: +48-713-202-022

check for updates

Abstract: This article presents the application of data mining (DM) to long-term power quality (PQ) measurements. The Ward algorithm was selected as the cluster analysis (CA) technique to achieve an automatic division of the PQ measurement data. The measurements were conducted in an electrical power network (EPN) of the mining industry with distributed generation (DG). The obtained results indicate that the application of the Ward algorithm to PQ data assures the division with regards to the work of the distributed generation, and also to other important working conditions (e.g., reconfiguration or high harmonic pollution). The presented analysis is conducted for the area-related approach—all measurement point data are connected at an initial stage. The importance rate was proposed in order to indicate the parameters that have a high impact on the classification of the data. Another element of the article was the reduction of the size of the input database. The reduction of input data by 57% assured the classification with a 95% agreement when compared to the complete database classification.

Keywords: data mining; power quality; cluster analysis; ward algorithm; different working conditions; distributed generation

1. Introduction

A smart grid can be seen as the future of electrical power systems [1–3]. A smart grid requires the monitoring and cooperation of more and more elements, devices, and systems. Thus, it introduces the need for analyzing an increasing amount of data. Single parameter analysis, conducted by humans, has become a thing of the past in terms of the functioning of an electrical power system (EPS). Thus, a need for tools to support the long-term assessment has become very necessary [4–7].

This research is a continuation of previous work [8], which involves a method for analyzing long-term power quality (PQ) data using non-hierarchical clustering and its assessment using global indices in [9]. The presented results in Jasiński et al. [8] were based on 72 cases of clustering, which differ in terms of both the number of clusters (2/25), and also the distance definition of the items in the database (Euclidean, Chebyshev) for the K-mean algorithm. The different constructions of the database were discussed. The direct impact of the distributed generation (DG) on the PQ conditions was obtained when clustering using the K-mean algorithm and the Euclidean distance for non-standardized data that are extended by power consumption, using database C: frequency (f), voltage variations (U), short term flicker severity (P_{st}), asymmetry (k_{u2}), total harmonic distortion in voltage (THD_U), active power (P). Thus, in this article, the same input database was selected. However, the Ward

algorithm is presented in this research, which represents the hierarchical approach. Additionally, this work contains an analysis of the importance rate in order to indicate which parameters have an impact on the final classification. The comparison of clusters, which represent different working conditions of the electrical power network (EPN), obtained automatically, was only conducted for the indicated parameters with a high importance rate but not using a global index, which includes all the parameters as in [9]. Additionally, the next novelty of this work is the proposition of reducing the input database without losing data features. The proposed reduction to one value, instead of three phase-to-phase parameters, assured the classification with a 95% agreement when compared to the complete database classification.

The article is organized into four sections. Section 2 presents the state of the art of literature review. Section 3 describes the definitions and techniques of cluster analysis (CA), with special consideration for the Ward algorithm. Also, Section 3 contains the description of the research object—the EPN of the mining industry with gas-steam units and conducted long-term PQ measurements. Additionally, Section 3 contains the application of the Ward algorithm to PQ data and the results of the analysis with regards to the different working conditions of the EPN. The final element of Section 3 presents a discussion of the obtained results. Section 4 highlights the conclusions.

2. Literate Review

One solution to the problem of big data analysis is the application of data mining (DM) techniques. The literature contains many examples of the possible applications of DM for electric power systems, e.g.,

- the detection and classification of voltage events [10–15]
- the calculation and prediction of power losses [16–18]
- the diagnosis of faults in power transformers [19–23]
- load forecasting [24–29]
- load pattern segmentation [30–33]
- fault detection [34–38]
- fault prediction [39–42]
- the defining of energy consumption [43–48]
- the forecasting of energy gaining from renewable energy sources [49–52]
- the reliability assessment of renewable sources of energy [53–56]
- energy management in a household [57–60]
- the improvement of intrusion detection systems in smart grids [61–63]
- the detection of electricity theft in smart grids [64–67]

As observed above, the application of data mining is wide. This article presents the application of data mining for achieving an automatic classification of long-term power quality (PQ) data from an electrical power network (EPN) of the mining industry with distributed generation (DG). The selected technique is cluster analysis (CA).

3. Methods and Results

3.1. Cluster Analysis—Ward Algorithm

Generally, the definition of data mining in the literature concerns the achievement of knowledge from big databases. Possible algorithms and techniques are well-known and described in the literature. Examples of data mining techniques are [68–72]:

- decision trees
- neural networks
- clustering

- regression
- mining association rules
- the multilayered perceptron network—MLP network
- genetic algorithms
- fuzzy interference systems
- high-performance computing
- inductive logic programming
- memory-based reasoning methods
- fuzzy sets

One of the described techniques is cluster analysis, also known as clustering [73]. The main aim of cluster analysis is to achieve homogeneous groups (clusters) of data as defined by Witten et al. and Wu et al. in [74,75]. The homogeneous aspect of the group is defined by the similarity or dissimilarity level of the data in the same cluster. There are a lot of data similarity/dissimilarity conditions that can be selected. However, due to the grouping process approach, two basic methods of dividing are known:

- hierarchical
- non-hierarchical

In this article, the hierarchical method is presented. Hierarchical approaches are agglomeration or divisive techniques. This article presents the agglomerative approach. Agglomerative techniques represent a set of observations in which each piece of data is treated as a separate cluster at the beginning. Then, the data are aggregated into a smaller number of clusters until one single cluster is established, which represents all the data [73]. The possible methods for connecting data into clusters are [73,76]:

- the single linkage method
- the complete linkage method
- the average linkage method
- the weighted pair-group average linkage method
- the unweighted pair-group centroid linkage method
- the unweighted pair-group centroid linkage method
- the Ward method of minimum variance

The hierarchical method is selected because the agglomerative sequence is presented on a dendrogram. It is, therefore, possible to analyze if the connection is better realized by single data or by a group of similar data (achieved in the previous agglomeration) to get a final classification. The authors selected the Ward algorithm due to its features. Clustering is carried out in order to connect data concentrated in an average value until the data has a similar value (range). The hierarchical cluster analysis algorithm using the Ward method of minimal variance is presented in Figure 1.

In this paper, the hierarchical Ward method and non-hierarchical method based on the K-mean algorithm are proposed for the power quality data analysis. The indicated "finding pair of clusters which have the smallest sum of squares distance between the object and the cluster center to which this object belongs", is calculated as presented in Equation (1) [77].

$$D_{pr} = \frac{n_p + n_r}{n_p + n_q + n_r} * d_{pr} + \frac{n_q + n_r}{n_p + n_q + n_r} * d_{qr} + \frac{-n_r}{n_p + n_q + n_r} * d_{pq} \tag{1}$$

where:

D_{pr}—distance of the new cluster to cluster of number "r",
r—proceed numbers of cluster from "p" to "q",
d_{pr}—distance of primary cluster "p" from cluster "r",

d_{qr}—distance of primary cluster "q" from cluster "r",

d_{pq}—common distance of primary clusters "p" and "q",
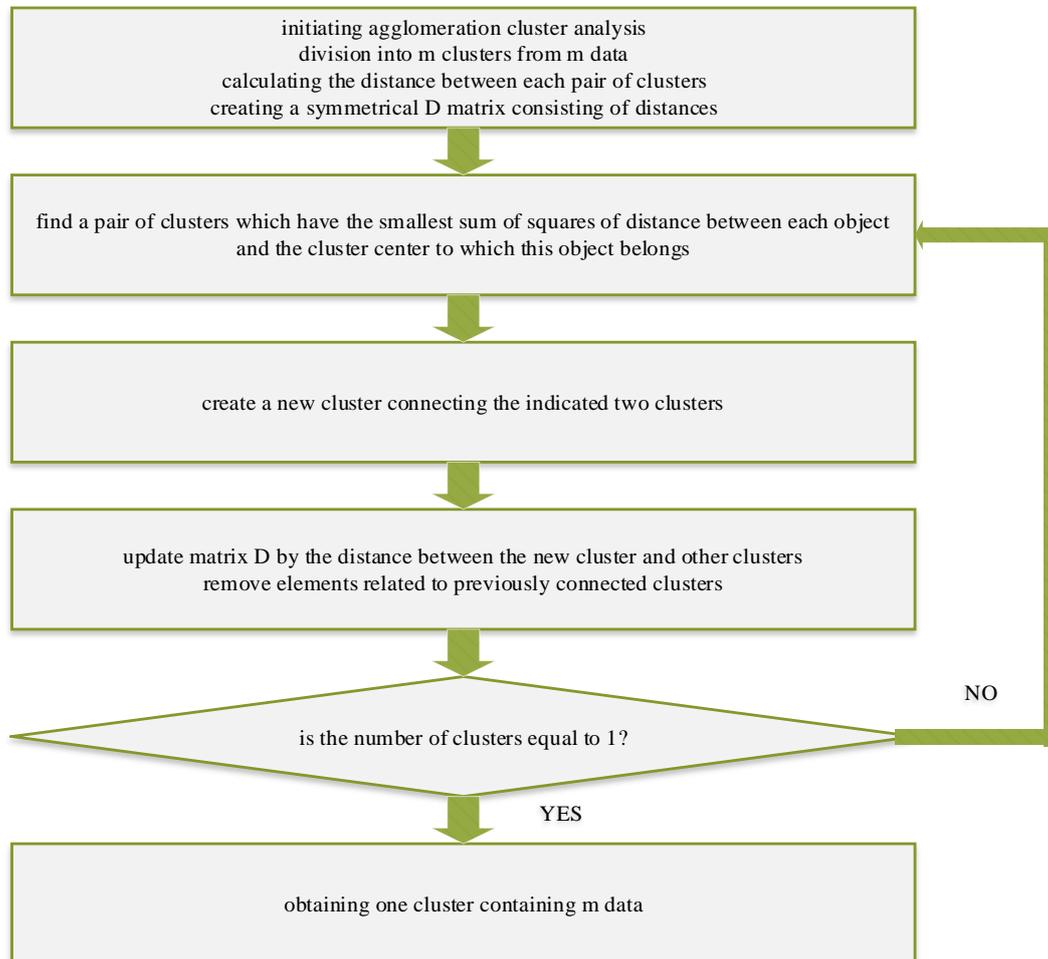
n—number of single objects in each object.

Figure 1. Cluster analysis using the Ward method of minimum variance [73,78].

Additionally, the advantage of the Ward algorithm is that it can be stopped at any moment; it can also achieve a classification represented by the excepted number of clusters. Thus, the final number of clusters should be selected in accordance with the aim of the classification. In order to support the final number of clusters, a lot of approaches have been conducted in literature. The most known are [79]:

- a dendrogram is analyzed in terms of the difference in distance between successive clusters. A big value of difference means that the data in the cluster are various. Thus, the division ends when the difference in the distance is maximal
- if a clear flattening (log vertical line) can be observed on the dendrogram, it means that in this point the clusters are distant and it is the best point for division
- an approach based on the root-mean-square standard deviation

3.2. An Electrical Power Network of the Mining Industry and the Source of the PQ Data

The PQ data used in the investigation concerns real measurements made in substations of the copper industry's electrical power network. The 110-kV substation of the mining industry works in a four-section system in cooperation with the four transformers (T1, T2, T3, T4). Normally, all the

transformers are supplied from a different 110 kV section. However, during the measurements, the T4 transformer was not loaded. Thus:

- substations R-1 work independently
- substations R-2 work independently
- substations R-3 and R-4 are coupled

The presented PQ data concerns four weeks of measurements from 27th of April to 25th of May. The measurements were conducted synchronously with class A PQ recorders [80]. This is more than the classical one week of observation time, and therefore, the PQ data may consist of different working conditions of the analyzed electrical power network of the mining industry [77]. Thus, the different working conditions may be connected to:

MAIN LOADS:

- welding machines
- conveyor belts
- drainage pumps

DISTRIBUTED GENERATION:

- combined heat and power (CHP)
- gas-steam units

Thus, the PQ measurements include the analysis of the PQ level, which concerns the impact of the DG and main load (welding machine) on the medium voltage (MV) network. The simplified scheme of the copper industry network, showing the localization of power quality recorders installed in selected bays and the localization of DG, is presented in Figure 2. The PQ recorders involve the measurements of transformers at 6 kV side (T1, T2, T3) and an outcoming feeder to a welding machine (WM).
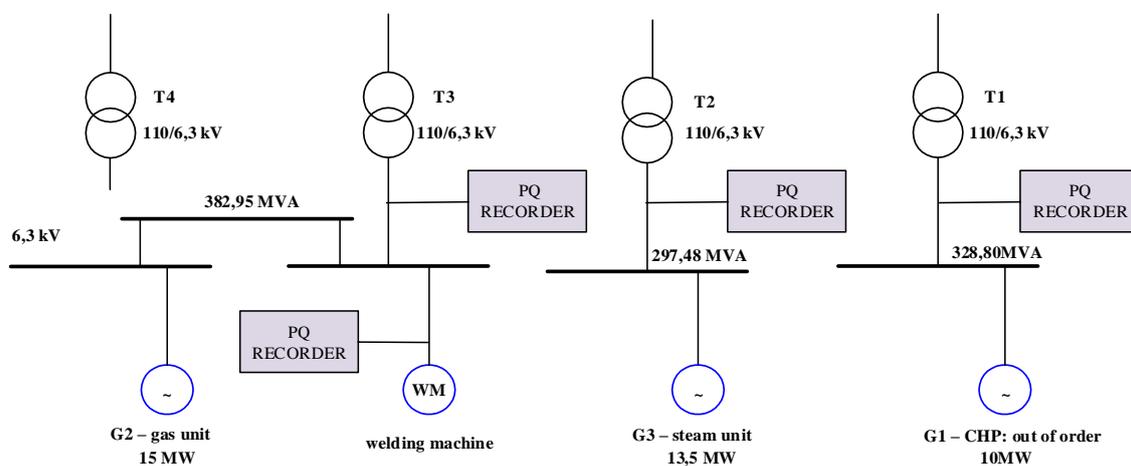

Figure 2. The simplified scheme of the electrical power network of the mining industry containing the placement of the PQ recorder and distributed generation.

It is important to note that the local generation is connected at the 6 kV level and that it consists of heat, a powerplant (G1–10 MW CHP), and steam-gas generation units (G3–15 MW gas unit and G2–13,5 MW steam unit). During the measurements, G1 was out of order and the level of generation of G2 and G3 was changing. The level of DG power (G1, G2, G3) and active power transformers (T1, T2, T3, WM) at the MV level are presented in Figure 3.
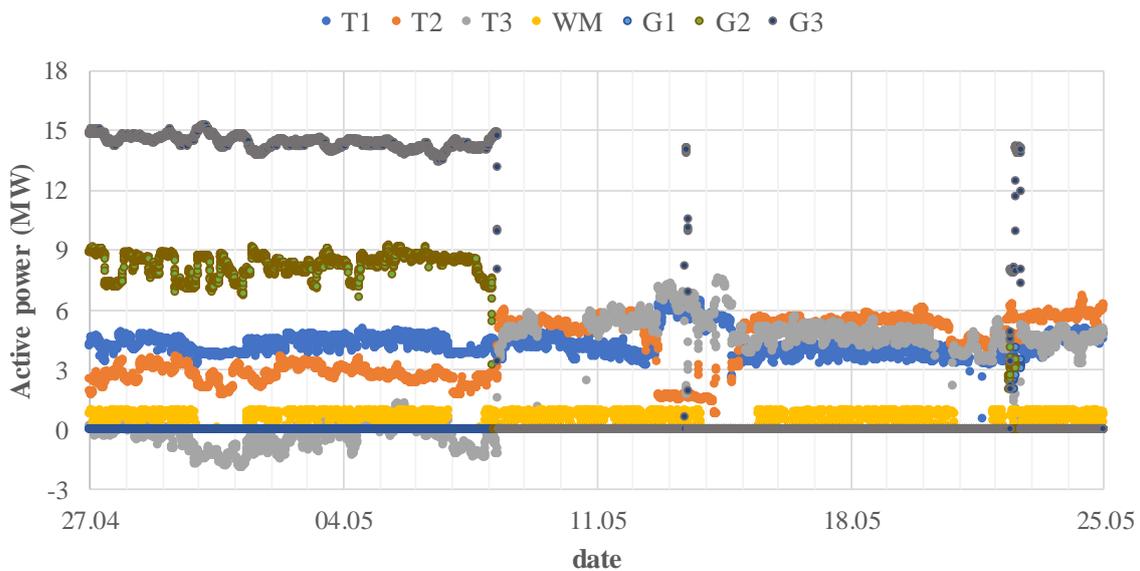
Figure 3. Active power of the high/medium (HV/MV) transformers T1, T2, and T3, welding machine (WM), and distributed generations G1, G2, G3 using the MV side measurements.

3.3. Cluster Analysis Results

3.3.1. Parameters Included to the Input Database

For the implementation of hierarchical cluster analysis, the Ward algorithm was used. The reason for this is due to the fact that the data assigned to clusters are characterized by the smallest variation of results (minimum variance of data in clusters). A data set for clustering consisted of the following PQ parameters:

- frequency variation (f)
- voltage variation (U)
- short-term flicker severity (P_{st})
- asymmetry (k_{u2})
- total harmonic distortion in voltage (THDu)
- active power level (P)

The indicated database consists of parameters, which are considered in the classical PQ assessment in accordance with the standard EN 50160 [81] but were extended to the active power in the measuring points. The noticeable change was the use of short-term flicker severity in place of long-term flicker severity. This change is connected with the time aggregations of the parameters; the long-term severity has 2 h, and the short-term one has 10 min [82,83]. Thus, the application of short term flicker severity enables a database consisting of parameters that are aggregated with 10 min intervals to be built, as is demanded in the standard of International Electrotechnical Commission (IEC) 61000-4-30 [80]. The analyzed measurement data were divided into flagged and unflagged data in accordance with the flagging concept of the standard IEC 61000-4-30 [80]. The data that was input to the CA were free of voltage events.

Additionally, due to the feature of the Ward algorithm that involves the fact that clustering is conducted in order to connect data concentrated in an average value until the data has similar values (range), the standardization process was proposed. The standardization of the parameters aims to obtain unified values by dividing the current value of a particular element of the time series by their maximum values. The decision concerning standardizing data to the average value reduces the problem with regards to different ranges and units of the PQ parameters. The standardize division 0–1 assures the possibility of comparing the changeability of the parameters.

3.3.2. Clustering to Indicate Different Working Conditions of the EPN

For the defined input database, the clustering with the Ward algorithm was carried out using the Statistica 13 program (StatSoft Polska, Kraków Polska). Figure 4 presents the CA dendrogram. The time results of clustering are presented in Figures 5–9, which show a defined final number of clusters equal to 2, 3, 4, 5, and 6. This selection of the number of clusters was realized using the dendrogram (Figure 4). The authors decided to indicate the cluster that has a connection distance greater than 100. Thus, no clusters equal or less than 6 were investigated. In the figures, the "virtual" cluster 0 was defined, which represents the data that was flagged in the initial stage. Using knowledge about the object, different working conditions, which may affect the data classification, were defined:

- working or non-working of distributed generation (G2, G3)—the knowledge was obtained from a monitoring system of gas-steam units:

  - working of DG: from 27.04, hour 00:00, to 08.05, hour 06:00; day 13.05, hours 11.00–12.00; day 22.05, hours 13.20–16.50

- reconfiguration of the network, the supply of main loads was relocated between substations—the knowledge was obtained from the Supervisory Control And Data Acquisition (SCADA) system:

  - from 12.05.2017, hour 16.10 to 14.05.2017, hour 22:20

- maintenance breaks that are connected to the mining industry's working schedule—checking the technical conditions of machines, a shift timetable, working on weekends:

  - each Monday–Saturday, hours 6.00–10.00: maintenance break during the first shift
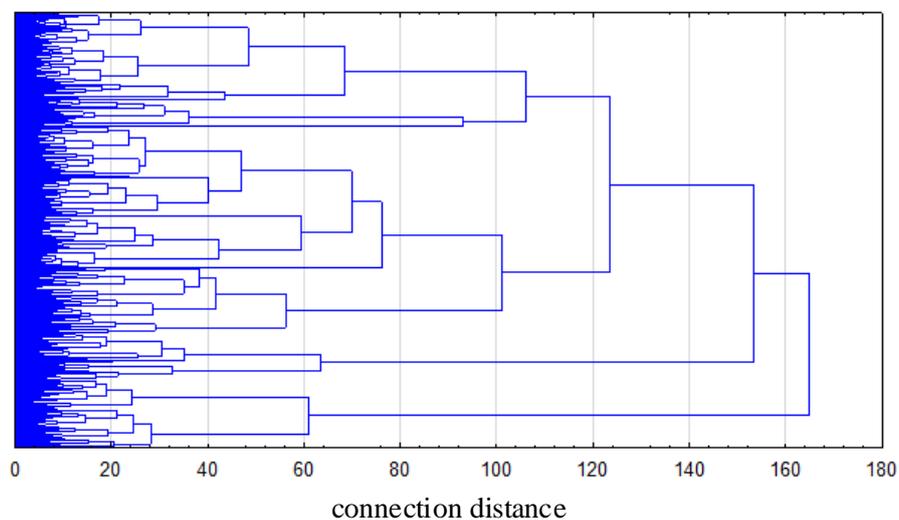 - each Saturday, hour 22.00 to Monday, hour 6.00: the weekend character of working

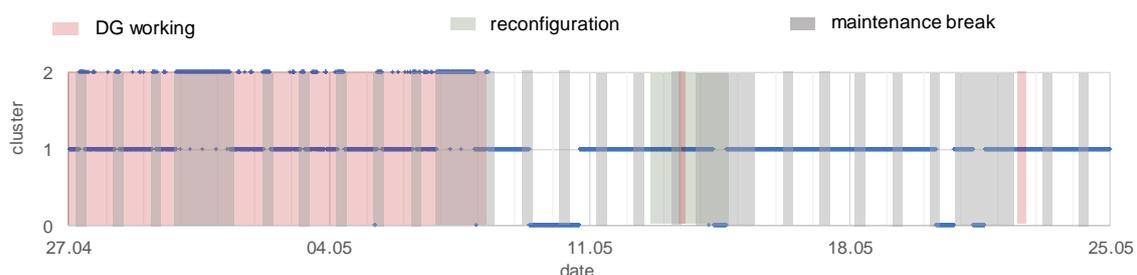**Figure 4.** Dendrogram of the CA using the Ward algorithm.

Figure 5. Cluster analysis results for the final number of clusters equal to 2.

Figure 6. Cluster analysis results for the final number of clusters equal to 3.

Figure 7. Cluster analysis results for the final number of clusters equal to 4.
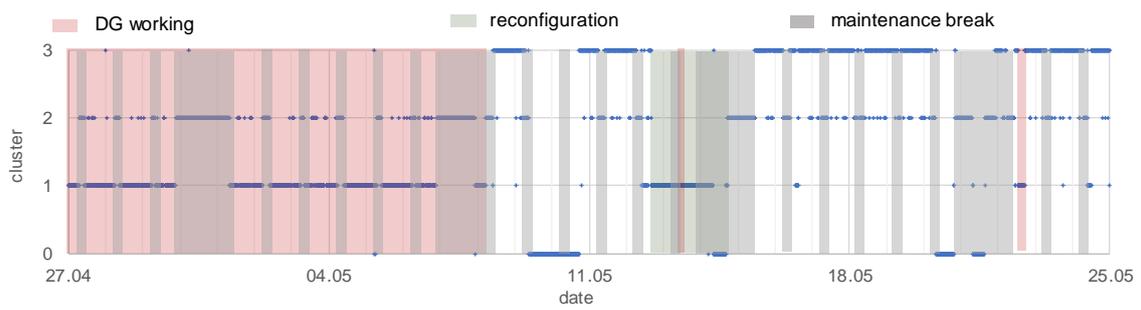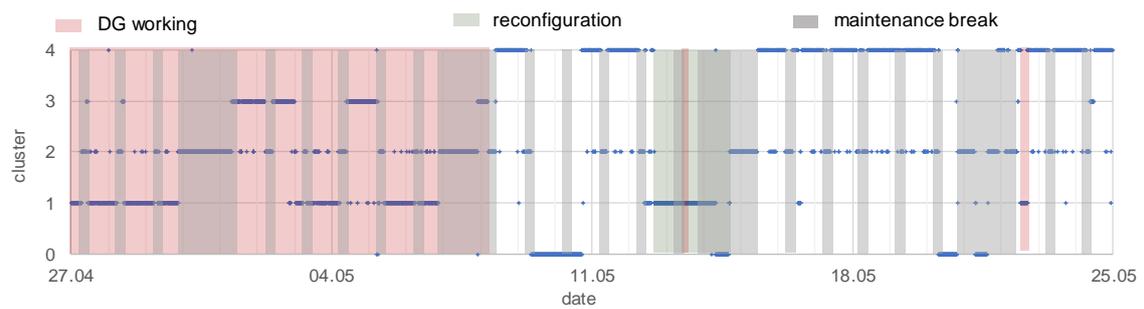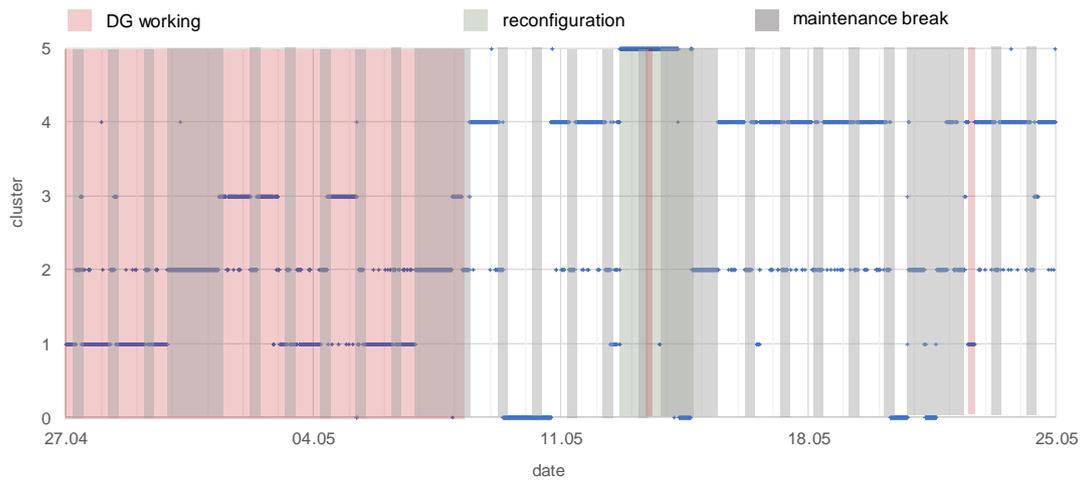

Figure 8. Cluster analysis results for the final number of clusters equal to 5.
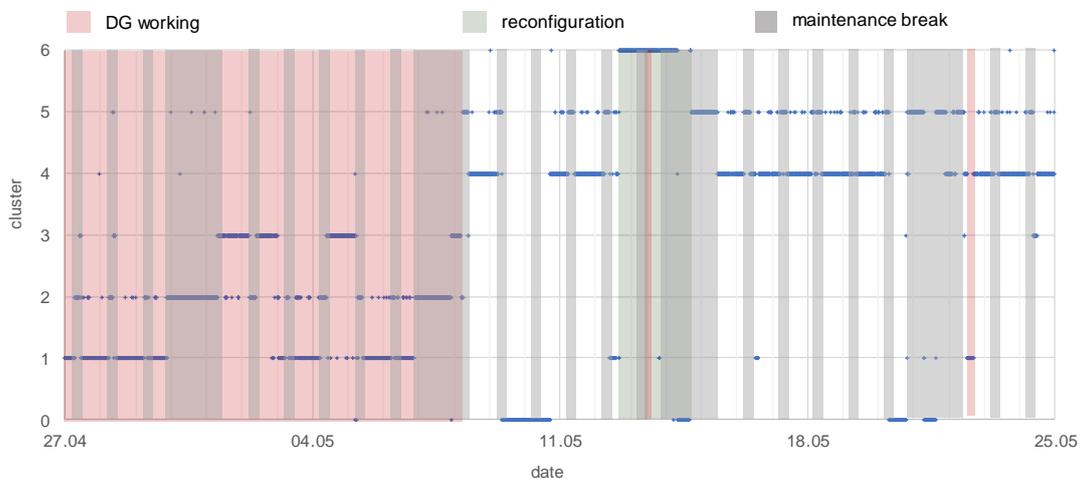

Figure 9. Cluster analysis results for the final number of clusters equal to 6.

Table 1 shows a summary of the analyzed working conditions and the assignment of clusters. The three conditions previously mentioned (DG working, reconfiguration, maintenance breaks) were indicated, and one unknown condition was observed. This unknown condition was indicated for the final number of clusters equal to at least 4. The reconfiguration of the EPN connection was indicated for the final number of clusters equal to 5. The impact of the DG and maintenance breaks was observed for all the presented classifications.

Table 1. The connection between the number of clusters and the working conditions of the electrical power network.

Condition	Final Number of Cluster				
	2	3	4	5	6
DG working	x	x	x	x	x
reconfiguration				x	x
maintenance breaks	x	x	x	x	x
other unknown condition			x	x	x

A further investigation is carried out for the final number of clusters equal to 6, although it may be realized for the other numbers too.

The analysis of the cluster assignment to the working conditions indicated that:

- c1: DG is working, exploitation time
- c2: DG is working, maintenance breaks time
- c3: DG is working, unknown working condition
- c4: DG is non-working, exploitation time
- c5: DG is non-working, maintenance breaks time
- c6: DG is non-working, reconfiguration of the network

There is an obvious question concerning which of the input parameters was important with regards to the obtained final classification. Thus, the predictor importance analysis using the Statistica 13 software (in accordance with the guidelines of a StatSoft Polska [78] and Breiman et. al. [84]) was realized for the classification of the 6 clusters. The results are presented in Figure 10. The results show that the highest impact (importance rate > 0.7) is for:

- the active power level for the transformers T1, T2, and T3
- the total harmonic distortion in the voltage for transformer T3 and the welding machine—WM
- the short-term flicker severity for the transformers T2, T3, and the welding machine—WM
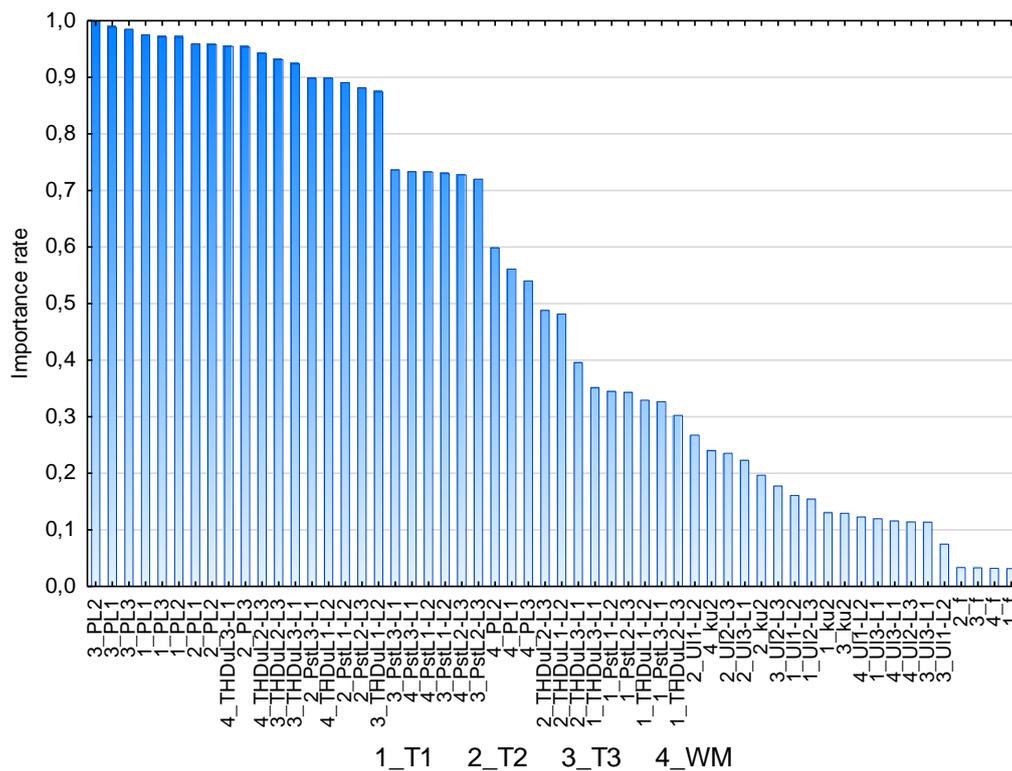
Figure 10. Importance rate of the factors to the output of the cluster analysis results for the final number of clusters equal to 6.

3.3.3. Qualitative Assessment of Clusters

A comparison of all the measurement points for each parameter in the database would lead to the analysis of the changeability of 48 parameters for each of the six clusters. Therefore, the authors suggest only analyzing those PQ parameters that were indicated as important with regards to the obtained classification (according to the predictor importance rate).

Table 2 contains the comparison of the selected PQ parameters for each cluster in terms of the mean, minimal, maximal, and standard deviation values.

Table 2. Comparison of the PQ level for different clusters.

Measurement Point	Parameter	Value	c1	c2	c3	c4	c5	c6
T2	P_{st} L1-L2	minimal	0.16	0.13	0.12	0.14	0.09	0.08
		maximal	0.50	0.30	0.54	1.42	0.74	0.57
		mean	0.22	0.21	0.18	0.24	0.26	0.11
		standard deviation	0.03	0.03	0.04	0.05	0.07	0.05
	P_{st} L2-L3	minimal	0.16	0.12	0.12	0.14	0.09	0.08
		maximal	0.50	0.29	0.53	2.07	0.73	0.38
		mean	0.22	0.21	0.18	0.24	0.26	0.11
		standard deviation	0.03	0.03	0.04	0.08	0.08	0.05
	P_{st} L3-L1	minimal	0.17	0.13	0.13	0.14	0.10	0.08
		maximal	0.50	0.30	0.52	3.54	1.06	0.53
		mean	0.23	0.21	0.19	0.25	0.27	0.11
		standard deviation	0.03	0.03	0.04	0.11	0.08	0.05
T3	P_{st} L1-L2	minimal	0.13	0.13	0.15	0.14	0.13	0.18
		maximal	0.47	0.52	0.74	0.86	0.60	0.53
		mean	0.30	0.18	0.30	0.39	0.29	0.33
		standard deviation	0.03	0.05	0.05	0.06	0.08	0.06
	P_{st} L2-L3	minimal	0.14	0.14	0.16	0.15	0.14	0.19
		maximal	0.45	0.49	0.80	2.01	0.74	0.56
		mean	0.31	0.19	0.31	0.43	0.32	0.35
		standard deviation	0.03	0.05	0.05	0.09	0.09	0.07
	P_{st} L3-L1	minimal	0.13	0.13	0.16	0.15	0.14	0.20
		maximal	0.46	0.48	0.81	1.93	1.10	0.70
		mean	0.32	0.19	0.33	0.44	0.32	0.37
		standard deviation	0.04	0.06	0.05	0.08	0.09	0.07
	THDu L1-L2	minimal	0.48	0.39	0.87	0.47	0.41	0.57
		maximal	1.20	0.99	4.99	1.50	1.11	1.38
		mean	0.67	0.56	1.53	0.83	0.64	0.80
		standard deviation	0.07	0.08	0.27	0.08	0.12	0.09
	THDu L2-L3	minimal	0.49	0.39	0.89	0.48	0.45	0.62
		maximal	1.23	0.99	5.23	1.56	1.13	1.44
		mean	0.68	0.55	1.57	0.86	0.68	0.84
		standard deviation	0.07	0.08	0.29	0.08	0.12	0.09
	THDu L3-L1	minimal	0.49	0.38	0.91	0.49	0.41	0.58
		maximal	1.28	1.02	4.87	1.63	1.18	1.50
		mean	0.70	0.55	1.63	0.89	0.67	0.87
		standard deviation	0.08	0.08	0.29	0.09	0.14	0.10
WM	P_{st} L1-L2	minimal	0.14	0.14	0.16	0.15	0.14	0.19
		maximal	0.47	0.54	0.78	6.84	0.64	0.56
		mean	0.31	0.19	0.32	0.43	0.31	0.35
		standard deviation	0.03	0.05	0.05	0.20	0.08	0.07
	P_{st} L2-L3	minimal	0.13	0.13	0.16	0.15	0.14	0.19
		maximal	0.45	0.49	0.79	6.89	0.72	0.57
		mean	0.31	0.19	0.32	0.43	0.31	0.35
		standard deviation	0.03	0.06	0.05	0.20	0.08	0.07
	P_{st} L3-L1	minimal	0.14	0.13	0.16	0.15	0.14	0.19
		maximal	0.46	0.46	0.78	6.84	1.12	0.66
		mean	0.31	0.18	0.31	0.43	0.31	0.35
		standard deviation	0.03	0.05	0.05	0.20	0.09	0.07

Table 2. *Cont.*

	THDu L1-L2	minimal	0.46	0.36	0.55	0.48	0.40	0.56
		maximal	1.23	0.99	2.40	1.54	1.16	1.42
		mean	0.67	0.53	1.56	0.85	0.65	0.81
		standard deviation	0.08	0.08	0.22	0.08	0.13	0.10
	THDu L2-L3	minimal	0.45	0.36	0.59	0.49	0.43	0.59
		maximal	1.23	0.96	2.40	1.55	1.13	1.44
		mean	0.65	0.52	1.54	0.85	0.67	0.84
		standard deviation	0.08	0.08	0.23	0.08	0.13	0.10
	THDu L3-L1	minimal	0.45	0.36	0.58	0.49	0.42	0.59
		maximal	1.22	0.94	2.37	1.53	1.12	1.42
		mean	0.65	0.52	1.50	0.86	0.67	0.85
		standard deviation	0.08	0.08	0.22	0.08	0.13	0.10

where:

minimal—the minimal value of the parameter that may be found for the observed cluster
maximal—the maximal value of the parameter that may be found for the observed cluster
mean—the mean value calculated from all the data for the observed cluster
standard deviation—the standard deviation calculated from all the data for the observed cluster.

A comparison of the level of the PQ parameters for different clusters is equivalent to the comparison of the different working conditions of an electrical power network. The examples of such a comparison may be as follows:

- (c1 with c2) and (c4 with c5)—> comparison of time with the different characters of the company that iss working (exploitation vs. maintenance break). It could be observed that the mean value of P_{st} for T3 and WM is lower during the maintenance break. Therefore, in terms of flicker severity, the time of maintenance is better.
- (c1 with c2) and (c4 with c5)—>comparison of time with the different characters of the company that is working (exploitation vs. maintenance break). It could be observed that the mean value of THDu for T3 and WM are lower during the maintenance break. Therefore, in terms of the harmonic content, the time of maintenance is better.
- (c1 with c4) and (c2 with c5)—> comparison of time with the different characters of the working DG. It could be observed that P_{st} for T3 and WM is lower for the time when the DG is working (c1, c2) compared to when the DG is switched off (c4, c5). Therefore, in terms of flicker severity, the time when the DG is working is better.
- c3 with all other clusters—> this unknown working condition represents the time when the THDu level for T3 and WM is higher than for the other clusters.
- c6 with all other clusters—> the reconfiguration that represents the time when P_{st} for T2 is very low. This is in agreement with the fact that T2 was underloaded, and therefore, the flicker is small

The presented examples about the comparison of the level of the PQ parameters for different clusters assure simplified information concerning the differences between working conditions. However, the working condition for defining the cluster c3 is unknown, but due to the indicated analysis, it is possible to define that during this time there was a higher than normal level of harmonics for T3 and WM. Thanks to this, attention could be paid to this time in order to find the reason for such high harmonic content and to reduce it in the future. Additionally, after automatic classification of the data, it is possible to show the impact of DG on the level of power quality in the electrical power network of the mining industry.

3.3.4. Reduction of the Input Database Size—Case Study

The natural question is, "is it possible to reduce or change the structure of the input database without losing the most important information". The first idea is just to exclude some parameters. However, the proposed, complete database includes, all-important points of the classical PQ parameters. Thus, excluding any of them would not seem to be adequate from the technical point of view.

In this research, the objects are represented by similar phase-to-phase values. Thus, the analysis of only one "new-multiphase" value was conducted. Moreover, the way of conducting this may be different. The minimal, maximal, mean, or median value from three phase-to-phase values may be selected. However, in this research, the authors decided to use the mean value. Thus, for each 10 min data of:

- voltage
- short-term flicker severity
- total harmonic distortion in voltage
- active power

the mean value from all three phase-to-phase values was calculated.

After such a reduction—from 16 input parameters (complete database) for each measurement point to six input parameters (reduced database)—clustering was conducted. The result of the obtained cluster using the six-parameter database, in comparison to 14-parameter clustering, is presented in Table 3. Generally, the results of this reduction in terms of indicating the same working condition for more than two clusters are positive. The obtained classification has the same result for at least 94.9% of data. The only negative classification was obtained for two clusters. The averaged data during the division to two clusters was not sensitive for DG impact.

Table 3. Comparison of clustering results for the completed database to the reduced one.

Final Number of Clusters	Do Results Indicate the Same Working Conditions?	Percent of the Data Assigned to the Same Cluster
2	no *	–
3	yes	95.7
4	yes	95.1
5	yes	95.0
6	yes	94.9

* no impact of DG is observable, only the maintenance is noticeable.

Additionally, the predictor importance for six clusters was defined. Figure 11 presents the importance rate for both classifications—(a) reduced input database, (b) complete input database. Generally, regarding the 0.7 importance rate level (noticeable importance rate), the same parameters were indicated:

- transformer T1—active power
- transformer T2—active power
- transformer T3—active power, total harmonic distortion in voltage, short-term flicker severity
- welding machine WM—active power, total harmonic distortion in voltage, short-term flicker severity

The only excluded parameter is the short-term flicker severity for transformer T2. However, the importance rate is close to 0.7.

To summarize, the size of the database has been reduced from 14 parameters to six parameters, and the obtained results are generally similar.
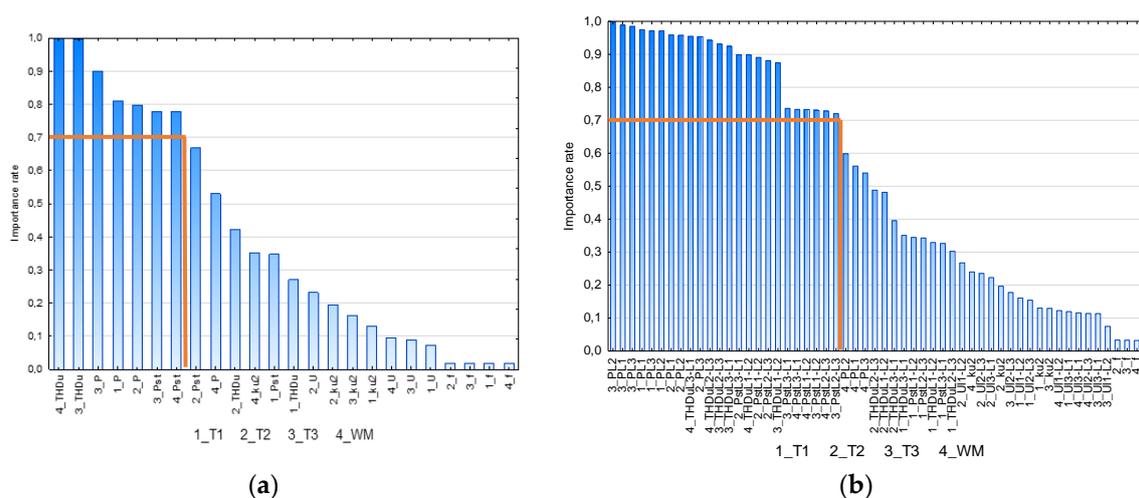
Figure 11. Importance rate for six clusters for (**a**) reduced input database; (**b**) complete input database.

3.4. Discussion

The data mining technique presented in the article is the cluster analysis. The Ward algorithm was selected as an example of the hierarchical approach. During the data preparation stage, it was necessary to uniform the data aggregation time (selection of P_{lt} to P_{st}), as well as to standardize the parameter values. For the prepared data set containing both the PQ parameters and the active power level, cluster analysis was conducted.

As a result of the cluster analysis, a dendrogram was obtained, which was illegible for the initial stages of agglomeration due to a large amount of input data. This is an unquestionable disadvantage of the hierarchical approach, but it is worth noting that it provides a division of data regardless of the final number of the obtained clusters. Additionally, on the dendrogram, there is a simple possibility of selecting the final number of clusters using methods indicated in the literature, e.g., Aggarwal [79].

Another important element of the article was to indicate the conditions that influenced the data division. On the basis of knowledge about the object, the conditions of distributed generation working, reconfiguration, and maintenance breaks were known. However, the obtained classification indicated that, in terms of the PQ level, the relevant condition was not known. It is worth highlighting the fact that the Ward algorithm is sensitive to the impact of the distributed generation on the technical conditions of the electrical power network, which confirms that the research aim was specified correctly.

The next element of the article was the analyses of the parameters that have a higher impact on the data classification. The obtained results indicated the importance of an active power level, as well as the harmonic level and flicker. The voltage variations, voltage, and frequency levels had a small impact on the classification.

Then, after obtaining the importance ranking, a comparison of the clusters in terms of the selected PQ parameters was carried out. The obtained results presented the impact of DG on the EPN. The impact of DG was indicated as positive regarding PQ. The unknown working condition was described as a time with high total harmonic distortion at the voltage level. Thus, the analysis of only this selected period of time may help to decrease the problem with harmonic pollution.

The last part of the research concerned the possibility of reducing the input database without losing the information obtained from the clustering. The authors proposed reducing the three phase-to-phase values to one mean value. Then, the comparison of the reduced input database to the completed one was conducted. The obtained classifications were similar. Around 95% of data was connected to the same clusters for both input databases and classification to more than two groups. The presented approach decreased the size of the input database by 57% (from fourteen to six parameters) without losing any data features.

The presented in-article object represents a symmetrical network, although, the method may be realized successfully for highly asymmetrical grids. Thus, if any of the phase-to-phase value was changed, the mean value of all parameters also changed. The CA is sensitive for the differences so this situation would also be indicated. The only disadvantage of this method is that there would be no information on which phase caused this situation, thus the analysis of raw data, but for the indicated period of time, is desirable.

4. Conclusions

The article presents the application of cluster analysis to long-term power quality measurements obtained in an electrical power network of the mining industry with distributed generation. The selected algorithm, due to its sensitivity to data dissimilarity, was the Ward algorithm. The article contains a discussion of the pros and cons of the hierarchical approach.

The article also contains the analysis of the sensitivity of different (known) working conditions of an electrical power network of the mining industry to the obtained classification. Conditions such as the impact of distributed generation, reconfiguration appearance, or the character of the object schedule (exploitation or maintenance breaks) are indicated. Additionally, the ranking of the impact of the parameter on the classification was conducted using predictor analysis. This analysis indicated that the level of active power, harmonic pollution, and flicker are important with regards to the obtained classification.

The obtained classification indicated the unknown working condition. After the comparison with other groups, the unknown condition was indicated as a high harmonic pollution period of time. Thanks to this, it is possible to analyze a short period of time to find the problem with harmonic pollution in an electrical power network of the mining industry.

The article contains the proposition of reducing a database concerning the calculation of one value that represents three phase-to-phase values. The results were similar (close to 95%), and the calculations were reduced by over 57%.

The presented approach of obtaining automatic data classification with regards to different working conditions (especially distributed generation or the harmonic pollution problem) is an important element of a smart grid. It is worth noting that the presented approach is conducted for area-related analysis—four different measuring points that are considered as common input data.

References

1. Oncioiu, I.; Căpușneanu, S.; Türkeș, M.; Topor, D.; Constantin, D.-M.; Marin-Pantelescu, A.; Ștefan Hint, M. The Sustainability of Romanian SMEs and Their Involvement in the Circular Economy. *Sustainability* **2018**, *10*, 2761. [CrossRef]

2. Türkeș, M.; Oncioiu, I.; Aslam, H.; Marin-Pantelescu, A.; Topor, D.; Căpușneanu, S. Drivers and Barriers in Using Industry 4.0: A Perspective of SMEs in Romania. *Processes* **2019**, *7*, 153. [CrossRef]

3. Oncioiu, I.; Bunget, O.C.; Türkeș, M.C.; Căpușneanu, S.; Topor, D.I.; Tamaș, A.S.; Rakos, I.-S.; Hint, M.Ș. The Impact of Big Data Analytics on Company Performance in Supply Chain Management. *Sustainability* **2019**, *11*, 4864. [CrossRef]

4. Salkuti, S.R. A survey of big data and machine learning. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 575. [CrossRef]

5. Ghorbanian, M.; Dolatabadi, S.H.; Siano, P. Big Data Issues in Smart Grids: A Survey. *IEEE Syst. J.* **2019**, *13*, 4158–4168. [CrossRef]

6. Dhupia, B.; Usha Rani, M.; Alameen, A. The role of big data analytics in smart grid management. In Proceedings of the 2nd International Conference on Computing, Communications Data Engineering CCODE 2019, Tirupati, India, 1–2 February 2019; Volume 1054, pp. 403–412.

7. Ding, Y. Analysis of Operation and Maintenance of Power Distribution Network Management Technology Under the Background of Big Data Era. In *International Conference on Big Data Analytics for Cyber-Physical-Systems*; Springer: Singapore, 2020; pp. 610–615.

8. Jasiński, M.; Sikorski, T.; Borkowski, K. Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry. *Electr. Power Syst. Res.* **2019**, *166*, 52–60. [CrossRef]

9. Jasiński, M.; Sikorski, T.; Kostyła, P.; Leonowicz, Z.; Borkowski, K. Combined Cluster Analysis and Global Power Quality Indices for the Qualitative Assessment of the Time-Varying Condition of Power Quality in an Electrical Power Network with Distributed Generation. *Energies* **2020**, *13*, 2050. [CrossRef]

10. Strack, J.L.; Carugati, I.; Orallo, C.M.; Maestri, S.O.; Donato, P.G.; Funes, M.A. Three-phase voltage events classification algorithm based on an adaptive threshold. *Electr. Power Syst. Res.* **2019**, *172*, 167–176. [CrossRef]

11. Shikhin, V.A.; Kochengin, A.E.; Pavliuk, G.P. Significant Events Detection and Identification through Electrical Grid Load Profile. In Proceedings of the 2018 IEEE Renewable Energies, Power Systems & Green Inclusive Economy (REPS-GIE), Casablanca, Morocco, 23–24 April 2018; pp. 1–5.

12. Ucar, F.; Alcin, O.F.; Dandil, B.; Ata, F. Power quality event detection using a fast extreme learning machine. *Energies* **2018**, *11*, 145. [CrossRef]

13. Biswal, B.; Biswal, M.; Mishra, S.; Jalaja, R. Automatic classification of power quality events using balanced neural tree. *IEEE Trans. Ind. Electron.* **2014**, *61*, 521–530. [CrossRef]

14. Jasiński, M.; Sikorski, T.; Borkowski, K. Application of cluster analysis to identification flagged power quality measurements in area-related approach. Zastosowanie eksploracji danych do identyfikacji oznaczonych wyników pomiaru jakosci energii elektrycznej w ujeciu obszarowym. *Prz. Elektrotechniczny* **2020**, *3*, 9–12.

15. Balouji, E.; Salor, O. Classification of power quality events using deep learning on event images. In Proceedings of the 3rd International Conference on Pattern Analysis Image Analysis IPRIA 2017, Shahrekord, Iran, 19–20 April 2017.

16. Dangar, B.; Josh, S.K. Interpretation of Urban Power Consumers Behaviors to Predict Power Loss in Summer. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 563–565.

17. Yun, Z.; Mengting, Y.; Junjie, L.; Ji, C.; Penghui, H. Line loss calculation of low-voltage districts based on improved K-Means. In Proceedings of the 2018 IEEE International Conference on Power System Technology (POWERCON), Beijing, China, 24–26 October 2018; pp. 4578–4583.

18. Yao, M.; Zhu, Y.; Li, J.; Wei, H.; He, P. Research on Predicting Line Loss Rate in Low Voltage Distribution Network Based on Gradient Boosting Decision Tree. *Energies* **2019**, *12*, 2522. [CrossRef]

19. Menezes, A.G.C.; Almeida, O.M.; Barbosa, F.R. Use of decision tree algorithms to diagnose incipient faults in power transformers. In Proceedings of the 2018 IEEE Simposio Brasileiro de Sistemas Eletricos (SBSE), Niteroi, Brazil, 12–16 May 2018; pp. 1–6.

20. Liu, C.H.; Chen, T.L.; Yao, L.T.; Wang, S.Y. Using data mining to dissolved gas analysis for power transformer fault diagnosis. In Proceedings of the 2012 IEEE International Conference on Machine Learning and Cybernetics, Xian, China, 15–17 July 2012; pp. 1952–1957.

21. Basuki, A. Suwarno Online Dissolved Gas Analysis of Power Transformers Based on Decision Tree Model. In Proceedings of the 2018 IEEE Conference on Power Engineering and Renewable Energy (ICPERE), Solo, Indonesia, 29–31 October 2018; pp. 1–6.

22. Ren, F.; Si, S.; Cai, Z.; Zhang, S. Transformer fault analysis based on Bayesian networks and importance measures. *J. Shanghai Jiaotong Univ.* **2015**, *20*, 353–357. [CrossRef]

23. Cheng, L.; Yu, T. Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey. *Energies* **2018**, *11*, 913. [CrossRef]

24. Almeida, V.A.; Pessanha, J.F.M.; Caloba, L.P. Load data cleaning with data mining techniques. In Proceedings of the 2018 IEEE Simposio Brasileiro de Sistemas Eletricos (SBSE), Niteroi, Brazil, 12–16 May 2018; pp. 1–6.

25. Kotriwala, A.M.; Hernandez-Leal, P.; Kaisers, M. Load Classification and Forecasting for Temporary Power Installations. In Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sarajevo, Bosnia, 21–25 October 2018; pp. 1–6.

26. Cerne, G.; Dovzan, D.; Skrjanc, I. Short-Term Load Forecasting by Separating Daily Profiles and Using a Single Fuzzy Model Across the Entire Domain. *IEEE Trans. Ind. Electron.* **2018**, *65*, 7406–7415. [CrossRef]

27. Lei, J.; Jin, T.; Hao, J.; Li, F. Short-term load forecasting with clustering–regression model in distributed cluster. *Clust. Comput.* **2019**, *22*, 10163–10173. [CrossRef]

28. Fahiman, F.; Erfani, S.M.; Leckie, C. Robust and Accurate Short-Term Load Forecasting: A Cluster Oriented Ensemble Learning Approach. In Proceedings of the 2019 IEEE International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.

29. Arun Jees, S.; Gomathi, V. Load forecasting for smart grid using non-linear model in Hadoop distributed file system. *Clust. Comput.* **2019**, *22*, 13533–13545. [CrossRef]

30. Rajabi, A.; Eskandari, M.; Ghadi, M.J.; Li, L.; Zhang, J.; Siano, P. A comparative study of clustering techniques for electrical load pattern segmentation. *Renew. Sustain. Energy Rev.* **2019**, *120*, 109628. [CrossRef]

31. Verdu, S.V.; Garcia, M.O.; Senabre, C.; Marin, A.G.; Franco, F.J.G. Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps. *IEEE Trans. Power Syst.* **2006**, *21*, 1672–1682. [CrossRef]

32. Le Ray, G.; Pinson, P. Online adaptive clustering algorithm for load profiling. *Sustain. Energy Grids Netw.* **2019**, *17*, 100181. [CrossRef]

33. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *42*, 68–80. [CrossRef]

34. Ramdasi, A.P.; Mehata, K.M. Improved Text Mining Algorithm for Fault Detection using Combined D-Matrix. *Int. J. Recent Technol. Eng.* **2019**, *8*, 1376–1379.

35. Gao, T.; Boguslawski, B.; Marié, S.; Béguery, P.; Thebault, S.; Lecoeuche, S. Data mining and data-driven modelling for Air Handling Unit fault detection. In *E3S Web of Conferences*; EDP Sciences: Jules, France, 2019; Volume 111.

36. Chen, L.; Xu, G.; Zhang, Q.; Zhang, X. Learning deep representation of imbalanced SCADA data for fault detection of wind turbines. *Measurement* **2019**, *139*, 370–379. [CrossRef]

37. Ranjbar, S.; Jamali, S. Fault detection in microgrids using combined classification algorithms and feature selection methods. In Proceedings of the 13th International Conference on Protection and Automation of Power System, IPAPS 2019, Tehran, Iran, 31 December 2019–1 January 2020; Institute of Electrical and Electronics Engineers Inc., School of Electrical Engineering, Iran University of Science and Technology (IUST): Tehran, Iran, 2019; pp. 17–21.

38. Silva, S.; Costa, P.; Gouvea, M.; Lacerda, A.; Alves, F.; Leite, D. High impedance fault detection in power distribution systems using wavelet transform and evolving neural network. *Electr. Power Syst. Res.* **2018**, *154*, 474–483. [CrossRef]

39. Sun, C.; Wang, X.; Zheng, Y.; Zhang, F. A framework for dynamic prediction of reliability weaknesses in power transmission systems based on imbalanced data. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105718. [CrossRef]

40. Sun, C.; Wang, X.; Zheng, Y. Data-driven approach for spatiotemporal distribution prediction of fault events in power transmission systems. *Int. J. Electr. Power Energy Syst.* **2019**, *113*, 726–738. [CrossRef]

41. Pal, A.; Kumar, M. DLME: Distributed Log Mining Using Ensemble Learning for Fault Prediction. *IEEE Syst. J.* **2019**, *13*, 3639–3650. [CrossRef]

42. Cynthia, S.T.; Ripon, S.H. Predicting and Classifying Software Faults. In Proceedings of the 2019 7th International Conference on Computer and Communications Management—ICCCM 2019, Bangkok, Thailand, 27–29 July 2019; ACM Press: New York, NY, USA, 2019; pp. 143–147.

43. Rathod, R.R.; Garg, R.D. Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. *Int. J. Electr. Power Energy Syst.* **2016**, *78*, 368–374. [CrossRef]

44. Benítez, I.; Quijano, A.; Díez, J.-L.; Delgado, I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int. J. Electr. Power Energy Syst.* **2014**, *55*, 437–448. [CrossRef]

45. Cil, I. Consumption universes based supermarket layout through association rule mining and multidimensional scaling. *Expert Syst. Appl.* **2012**, *39*, 8611–8625. [CrossRef]

46. Zhang, G.; Wang, G.G.; Farhangi, H.; Palizban, A. Data mining of smart meters for load category based disaggregation of residential power consumption. *Sustain. Energy Grids Netw.* **2017**, *10*, 92–103. [CrossRef]

47. Jain, P.K.; Quamer, W.; Pamula, R. *Electricity Consumption Forecasting Using Time Series Analysis BT—Advances in Computing and Data Sciences*; Singh, M., Gupta, P.K., Tyagi, V., Flusser, J., Ören, T., Eds.; Springer: Singapore, 2018; pp. 327–335.

48. Yildiz, B.; Bilbao, J.I.; Dore, J.; Sproul, A.B. Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Appl. Energy* **2017**, *208*, 402–427. [CrossRef]

49. Sheng, H.; Xiao, J.; Cheng, Y.; Ni, Q.; Wang, S. Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression. *IEEE Trans. Ind. Electron.* **2018**, *65*, 300–308. [CrossRef]

50. Anderson, W.W.; Yakimenko, O.A. Using neural networks to model and forecast solar PV power generation at Isle of Eigg. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, Qatar, 10–12 April 2018; pp. 1–8.

51. Yao, S.; Pan, L.; Yu, Z.; Kang, Q.; Zhou, M. Hierarchically Non-continuous Regression Prediction for Short-Term Photovoltaic Power Output. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 9–11 May 2019; pp. 379–384.

52. Monfared, M.; Fazeli, M.; Lewis, R.; Searle, J. Fuzzy Predictor with Additive Learning for Very Short-Term PV Power Generation. *IEEE Access* **2019**, *7*, 91183–91192. [CrossRef]

53. Su, C.; Hu, Z. Reliability assessment for Chinese domestic wind turbines based on data mining techniques. *Wind Energy* **2018**, *21*, 198–209. [CrossRef]

54. Aikhuele, D.O. Intuitionistic fuzzy model for reliability management in wind turbine system. *Appl. Comput. Inform.* **2018**. [CrossRef]

55. Uma, J.; Muniraj, C.; Sathya, N. Diagnosis of Photovoltaic (PV) Panel Defects Based on Testing and Evaluation of Thermal Image. *J. Test. Eval.* **2019**, *47*, 4249–4262. [CrossRef]

56. Harrou, F.; Dairi, A.; Taghezouit, B.; Sun, Y. An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine. *Sol. Energy* **2019**, *179*, 48–58. [CrossRef]

57. Du, S.; Li, M.; Han, S.; Shi, J.; Li, H. Multi-Pattern Data Mining and Recognition of Primary Electric Appliances from Single Non-Intrusive Load Monitoring Data. *Energies* **2019**, *12*, 992. [CrossRef]

58. Parvizimosaed, M.; Farmani, F.; Rahimi-Kian, A.; Monsef, H. A multi-objective optimization for energy management in a renewable micro-grid system: A data mining approach. *J. Renew. Sustain. Energy* **2014**, *6*. [CrossRef]

59. Ai, S.; Chakravorty, A.; Rong, C. Household Power Demand Prediction Using Evolutionary Ensemble Neural Network Pool with Multiple Network Structures. *Sensors* **2019**, *19*, 721. [CrossRef] [PubMed]

60. Singh, S.; Yassine, A. Mining Energy Consumption Behavior Patterns for Households in Smart Grid. *IEEE Trans. Emerg. Top. Comput.* **2019**, *7*, 404–419. [CrossRef]

61. El Mrabet, Z.; El Ghazi, H.; Kaabouch, N. A Performance Comparison of Data Mining Algorithms Based Intrusion Detection System for Smart Grid. In Proceedings of the 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 20–22 May 2019; pp. 298–303.

62. Gupta, S.; Sabitha, A.S.; Punhani, R. Cyber Security Threat Intelligence using Data Mining Techniques and Artificial Intelligence. *Int. J. Recent Technol. Eng.* **2019**, *8*, 6133–6140.

63. Zuo, X.; Chen, Z.; Dong, L.; Chang, J.; Hou, B. Power information network intrusion detection based on data mining algorithm. *J. Supercomput.* **2019**. [CrossRef]

64. Ahmad, T.; Chen, H.; Wang, J.; Guo, Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2916–2933. [CrossRef]

65. Razavi, R.; Gharipour, A.; Fleury, M.; Akpan, I.J. A practical feature-engineering framework for electricity theft detection in smart grids. *Appl. Energy* **2019**, *238*, 481–494. [CrossRef]

66. Maamar, A.; Benahmed, K. Machine learning Techniques for Energy Theft Detection in AMI. In Proceedings of the 2018 International Conference on Software Engineering and Information Management—ICSIM2018, Casablanca, Morocco, 4–6 January 2018; ACM Press: New York, NY, USA, 2018; pp. 57–62.

67. Jindal, A.; Dua, A.; Kaur, K.; Singh, M.; Kumar, N.; Mishra, S. Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1005–1016. [CrossRef]

68. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 12, ISBN 978-3-642-19720-8.

69. Larose, D. *Discovering Knowledge in Data. An Introduction to Data Mining*; John Wiley & Sons: Hoboken, NJ, USA, 2005; pp. 1–35. ISBN 9786468600.

70. Kantardzic, M. *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 9780470890455.

71. CIGRE. *Broshure 292: Data Mining Techniques and Applications in the Power Transmission Field*; CIGRE: Paris, France, 2006.

72. Olson, D.L.; Delen, D. *Advanced Data Mining Techniques*; Springer: Berlin/Heidelberg, Germany, 2008; ISBN 978-3-540-76916-3.

73. Wierzchoń, S.; Kłopotek, M. *Algorithms of Cluster Analysis*; Institute of Computer Science Polish Academy of Sciences: Warszaw, Poland, 2015; Volume 3, ISBN 9789638759627.

74. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 2011; ISBN 0080890369.

75. Wu, X.; Kumar, V.; Ross, Q.J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]

76. Sneath, P.H.; Sokal, R.R. *Numerical Texonomy*; Freeman: Lanzhou, China, 1973; ISBN 9780716706977.

77. Jasiński, M.; Borkowski, K.; Sikorski, T.; Kostyla, P. Cluster Analysis for Long-Term Power Quality Data in Mining Electrical Power Network. In Proceedings of the 2018 IEEE Progress in Applied Electrical Engineering (PAEE), Koscielisko, Poland, 18–22 June 2018; pp. 1–5.

78. Statsoft Polska StatSoft Electronic Statistic Textbook. Available online: http.://www.statsoft.pl/textbook/stathome.html (accessed on 15 February 2020).

79. Aggarwal, C.C. *Data Mining*; Springer: Cham, Switzerland, 2015; ISBN 978-3-319-14141-1.

80. *International Electrotechnical Commission, IEC 61000 4-30 Electromagnetic Compatibility (EMC)—Part 4-30: Testing and Measurement Techniques—Power Quality Measurement Methods*; International Electrotechnical Commission: Geneva, Switzerland, 2015.

81. *British Standards Institution, EN 50160: Voltage Characteristics of Electricity Supplied by Public Distribution Network*; British Standards Institution: UK, 2010.

82. Jasiński, M.; Sikorski, T.; Kostyła, P.; Kaczorowska, D.; Leonowicz, Z.; Rezmer, J.; Szymańda, J.; Janik, P.; Bejmert, D.; Rybiański, M.; et al. Influence of Measurement Aggregation Algorithms on Power Quality Assessment and Correlation Analysis in Electrical Power Network with PV Power Plant. *Energies* **2019**, *12*, 3547. [CrossRef]

83. Jasiński, M.; Rezmer, J.; Sikorski, T.; Szymańda, J. Integration Monitoring of On-grid Photovoltaic System: Case Study. *Period. Polytech. Electr. Eng. Comput. Sci.* **2019**, *63*, 99–105. [CrossRef]

84. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.