energies

*Article*

# Clustering Methods for Power Quality Measurements in Virtual Power Plant

**Fachrizal Aksan, Michał Jasiński *, Tomasz Sikorski, Dominika Kaczorowska, Jacek Rezmer, Vishnu Suresh, Zbigniew Leonowicz, Paweł Kostyła, Jarosław Szymańda and Przemysław Janik**

Faculty of Electrical Engineering, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland; 254212@student.pwr.edu.pl (F.A.); tomasz.sikorski@pwr.edu.pl (T.S.); dominika.kaczorowska@pwr.edu.pl (D.K.); jacek.rezmer@pwr.edu.pl (J.R.); vishnu.suresh@pwr.edu.pl (V.S.); zbigniew.leonowicz@pwr.edu.pl (Z.L.); pawel.kostyla@pwr.edu.pl (P.K.); jaroslaw.szymanda@pwr.edu.pl (J.S.); przemyslaw.janik@pwr.edu.pl (P.J.)

* Correspondence: michal.jasinski@pwr.edu.pl; Tel.: +48-713202022

**Citation:** Aksan, F.; Jasiński, M.; Sikorski, T.; Kaczorowska, D.; Rezmer, J.; Suresh, V.; Leonowicz, Z.; Kostyła, P.; Szymańda, J.; Janik, P. Clustering Methods for Power Quality Measurements in Virtual Power Plant. *Energies* **2021**, *14*, 5902. https://doi.org/10.3390/en14185902

Academic Editor: Abu-Siada Ahmed

Received: 25 June 2021
Accepted: 23 July 2021
Published: 17 September 2021

**Abstract:** In this article, a case study is presented on applying cluster analysis techniques to evaluate the level of power quality (PQ) parameters of a virtual power plant. The conducted research concerns the application of the K-means algorithm in comparison with the agglomerative algorithm for PQ data, which have different sizes of features. The object of the study deals with the standardized datasets containing classical PQ parameters from two sub-studies. Moreover, the optimal number of clusters for both algorithms is discussed using the elbow method and a dendrogram. The experimental results show that the dendrogram method requires a long processing time but gives a consistent result of the optimal number of clusters when there are additional parameters. In comparison, the elbow method is easy to compute but gives inconsistent results. According to the Calinski–Harabasz index and silhouette coefficient, the K-means algorithm performs better than the agglomerative algorithm in clustering the data points when there are no additional features of PQ data. Finally, based on the standard EN 50160, the result of the cluster analysis from both algorithms shows that all PQ parameters for each cluster in the two study objects are still below the limit level and work under normal operating conditions.

**Keywords:** power quality; cluster analysis; K-means; agglomerative; virtual power plant

## 1. Introduction

The electrical power system aims to generate electrical power and deliver it through the transmission and distribution system to customers' devices in a stable, secure, reliable, and sustainable manner [1]. However, nowadays, various electronic devices such as AC/DC converters, switching power supplies, and industrial non-linear load are becoming the factors responsible for the increasing PQ disturbances{XE "PQ"\t "*Power Quality*"} [2,3]. These devices tend to significantly distort the waveform of the supply and voltage [4,5]. The term power quality refers to a wide range of electromagnetic phenomena that characterize the voltage and current quality at a given time and location in the power system [6]. Generally, PQ disturbances are defined into two types based on the characteristics: voltage variations [7,8] and voltage events [9,10]. The analysis of PQ can be used to monitor the characteristic disturbances to capture PQ events that potentially detect faults associated with power quality problems in electrical power systems [11–13].

Power quality assessment has become a critical issue since the increase in sensitive electronic equipment in industrial and household applications [14,15]. The original lack of standardization of the measurement method has led to significant differences in each of the main parameters calculated by different devices. Thus, the consequences of power quality disturbances can lead to equipment malfunctions and process shutdowns [16,17]. The IEC 61000-4-30 standard [18] satisfies the standardization problem related to power quality by specifying a precise procedure, mathematical relationships, and required measurement accuracy for power quality analyzers [19,20]. During power quality monitoring, a short averaging time may be sufficient to evaluate the performance and disturbances related to PQ problems [21,22]. The ideal analysis time for PQ surveying is usually over one week, which is typically an integration period set at 10 min [23,24]. A network survey over a long period means that a large amount of data needs to be collected, which can be difficult and time-consuming to process. It is necessary to use appropriate tools or techniques to analyze the power quality measurement data to obtain valuable information or patterns within the huge amount of data. The data mining technique is one of the proposed solutions to process the huge amount of power quality measurement data to discover useful information related to power quality disturbances [25,26]. However, when working with data mining techniques, a basic understanding of the workflow and processing steps is required to achieve optimal results [27,28].

This research concerns the application of data mining techniques, especially on the clustering analysis method with the non-hierarchical and hierarchical approach for power quality problems in a virtual power plant (VPP). In general, a VPP consists of three effective components. The components used are the conventional dispatchable power plants, energy storage systems, and responsive or flexible loads [29]. Since cluster analysis is a part of unsupervised learning that has no labeling [30], it is used for exploratory data analysis of a VPP to group a collection of data items that are similar to each other and dissimilar from data items in other clusters [31].

In terms of a non-hierarchical approach, reference [32] introduces cluster analysis with the K-means algorithm on the long-term measurement of power quality data in the real virtual power plant that operates in Poland. The research aimed to identify the different working conditions of the VPP based on data features. Then, the different input power quality databases that consist of global index values as PQ parameters were used to identify how the algorithm works with different input features. In contrast, identifying the optimal number of clusters was defined by using v-fold cross-validation and a cost sequence chart. Reference [33] proposed the K-means algorithm to identify the energy features of the prosumers. The algorithm allows one to obtain the categories of prosumers from two specific indicators, which can assist the distribution network operators in optimizing networks' operation.

On the other hand, with the hierarchical approach, reference [34] proposed the application of the Ward algorithm to detect short-term working conditions of a VPP by analyzing the classical PQ parameters as the input data, and the qualitative assessment of the clustering process was realized by using the cubic clustering criterion. In comparison, reference [35] involves the same approach but using the dendrogram to select the final number of clusters, which was the unquestionable disadvantage of the hierarchical approach. Reference [36] uses the Ward algorithm for a prosumer fair load sharing and surplus trading approach based on the micro-grid concept of the transactive energy concept. This article presented a dendrogram to separate the daily energy consumption and maximum power from each micro-grid bus.

The exploration of PQ data in mining electrical power networks has been introduced in reference [37]. The proposed solution is using both the K-means algorithm and the Ward algorithm for cluster analysis. The K-means algorithm was used to classify the non-flagged data and flagged data, while the Ward algorithm was proposed to obtain a dendrogram for determining the optimal number of clusters in non-flagged data to identify the different working conditions of the electrical network. Since PQ monitoring collects a huge amount of measurement data, it is necessary to explore the pattern of data to obtain useful information that can inform upgrades or changes. Reference [26] proposed the methodology known as mixture modeling or intrinsic classification to recognize network problems at medium voltage (MV) electrical distribution systems, while reference [38] presented the use of the K-means algorithm and a hierarchical method to identify the reference nodes in distribution generation (DG).

Due to working with clustering analysis, it is very important to define the optimal number of clusters of the dataset for further analysis [39] because the main problem in applying many of the existing clustering techniques is that the optimal number of clusters needs to be pre-specified before the clustering is carried out [40]. Various methods have been proposed in the literature for determining the optimal number of clusters. Previous literature presents the use of a dendrogram and v-fold cross-validation, while reference [41] proposed the elbow method by calculating a sum of squares at each number of clusters and graphing. The optimal number of clusters is the one that looks like an arm in the graph. Reference [11] also proposed applying the elbow method to determine the optimal number of clusters on the K-means method. The result shows that the algorithm is sensitive to selecting the initial centroid of the cluster, and the elbow method allows the achievement of the same number of clusters on different data.

On the other hand, reference [42] reveals that the K-means algorithm is widely used for processing quantitative data with numeric attributes. However, this algorithm has drawbacks because it needs to determine the number of clusters before being applied. This research proposed applying the elbow method to assume the optimal number of clusters for the initial selection of centroids to fill the basic requirement of the K-means algorithm. The result shows that the elbow method can reduce 25% of iterations of processing using the K-means algorithm.

Based on the literature review, cluster analysis is suitable to analyze the power quality disturbance in a VPP. This research focuses on comparing two cluster analysis algorithms on a hierarchical approach with a non-hierarchical approach to the power quality measurement data of a VPP. The analysis concerns the medium voltage distribution network in a VPP, consisting of a 1.25 MW hydropower plant (HPP) and 0.5 MW energy storage system (ESS). The research is based on PQ measurements, which were realized synchronically in measurement points on the HPP, ESS, and associated with medium voltage lines. The measurement values consist of classic PQ parameters and an application of the global index, which is taken from the duration 1 May to 29 October 2020. Due to the huge amount of data collection varying on different feature units and scales, the dataset should be standardized to achieve a compatible input for the cluster analysis algorithms. The K-means algorithm has been selected from a non-hierarchical approach, while the agglomerative algorithm is a representative of a hierarchical approach, and both of the algorithms were performed by using a Euclidean distance metric. Additionally, the method for determining the optimal number of clusters for the K-means algorithm using the elbow method and determining the optimal number of clusters of the agglomerative algorithm was conducted by observing the dendrogram. Finally, the qualitative assessment of clustering results from both algorithms was presented to identify the power quality characteristic of a VPP based on the standard EN 50160 [43]. The analysis research and calculation method was performed using a scikit-learn machine learning library [44] for python [45] users. The contribution of this article is described below:

- The power quality dataset based on the long-term measurement in a VPP was standardized to cluster analysis.
- The proposed K-means algorithm and agglomerative algorithm were performed to compare the hierarchical and non-hierarchical approach for PQ data, which have different sizes of input data.
- The elbow method and dendrogram were performed to obtain the optimal number of clusters for PQ data.
- The global index was used for the comparative assessment of PQ parameters between clustering results of the K-means algorithm and agglomerative algorithm.
- The cluster algorithm evaluation and comparison are used to determine which algorithm is suitable for the investigation object.

This article is organized into five sections to achieve these contributions. Section 2 contains the description of the VPP, power quality data parameters, and the cluster analysis methodology. Section 3 presents the optimal number of clusters for PQ data of the VPP and the qualitative assessment of PQ for each clustering result. Section 4 reveals the discussion of the results. Finally, Section 5 covers the conclusion.

## 2. Research Object Description and Methodology

### *2.1. Object Investigated*

The research object of this article is to study the power quality problem in a virtual power plant operating in Lower Silesia, Poland. The studied subject refers to reference research [32], which concerns a case study of a VPP that analyzed medium voltage (MV) distribution networks. The real VPP consists of a 1.25 MW hydroelectric power plant (HPP) connected to a 0.5 MW battery energy storage system (ESS), which relies on a 20 kV distribution network connected to the HV/MV substation via an MV line (1-MV). The study was conducted by analyzing PQ measurement data from power quality recorders marked "R" in Figure 1. For further analysis, the research object in this article is divided into two subsections: the first investigation object and the second investigation object.

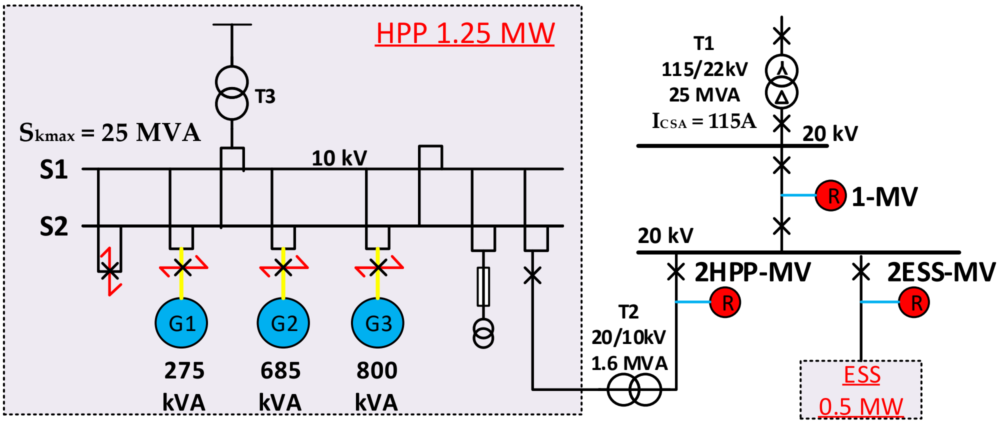


**Figure 1.** Investigated object of VPP with the location of PQ recorders, where 2HPP-MV: Hydro power plant with medium voltage, 2ESS-MV: Medium voltage energy storage system, and 1-MV: medium voltage line.

#### 2.1.1. First Investigation Object

The first object of study includes the power quality measurement data from PQ recorders at the measurement points 2HPP-MV and 2EES-MV, presented in Figure 1. They are treated as one point but reported as two for PQ issues because the hydropower plant and the energy storage system are connected to one node, and their PQ recorders are connected to the same voltage transformer [32]. However, the active power level of 2ESS-MV and 2HPP-MV is measured separately.

#### 2.1.2. Second Investigation Object

The second object of study has a wider scope than the previous object. It includes the PQ data of the first investigation object and additional PQ data from the PQ recorders of 1-MV, shown in Figure 1. For further analysis, the extension of the dataset into two sub-datasets is due to observation on how the different sizes and features of the input dataset will influence the performance of cluster analysis algorithms and on how the optimal number of clusters will be determined.

### *2.2. Parameter of Dataset Description*

The most common power quality standards used nowadays [46] are IEC 61000-4-30 [18] and EN 50160 [43]. IEC 61000-4-30 [18] is an IEC standard that specifies procedures for measuring and interpreting power quality parameters in 50/60 Hz AC networks. This standard also describes measurement procedures to obtain reliable, repeatable, and comparable results with any compliant measuring instrument.

In this section, the PQ measurement parameters from all PQ recorders of the VPP were used as the input dataset corresponding to the demand of the standard IEC 61000-4-30 [18]. The global power quality index consists of a classical power quality measurement, and the active power level was included in the proposed dataset. The measurement from PQ recorders was taken for a 26-week period, which already went through the preprocessing stage to exclude the voltage events. Thus, the dataset of the VPP that was used has parameters from a 10-min interval of about 24,612 data points. Since there are two study objects in this article, the first study object and the second study object, the dataset is independently divided into two different datasets, as shown in Figure 2.

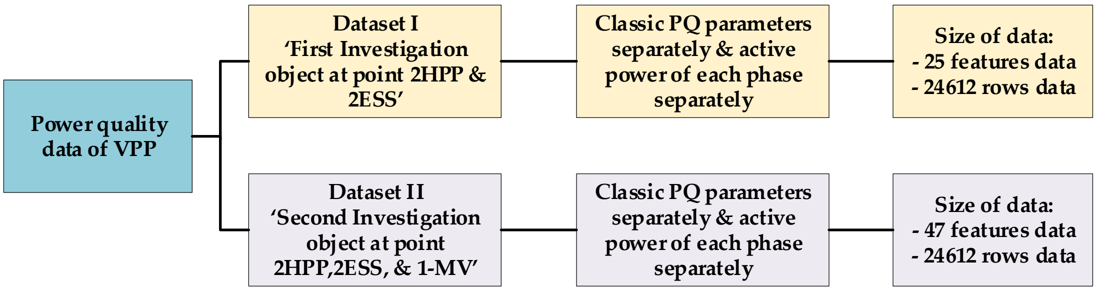

Power quality data of VPP
Dataset I 'First Investigation object at point 2HPP & 2ESS'
Classic PQ parameters separately & active power of each phase separately
Size of data: - 25 features data - 24612 rows data
Dataset II 'Second Investigation object at point 2HPP,2ESS, & 1-MV'
Classic PQ parameters separately & active power of each phase separately
Size of data: - 47 features data - 24612 rows data

**Figure 2.** Proposed dataset schema: description of dataset separation, parameters, and size of data.

Dataset I was used for the first study object and contains classical 10-min PQ parameters and active power levels of PQ recorders at points 2HPP-MV and 2ESS-MV. The dataset has 25 features as input parameters and 24,612-row data, and it contains about 615,300 single cells. On the other hand, dataset II is larger than dataset I, and it contains classical 10-min PQ parameters and active power levels of PQ recorders at points 2HPP-MV, 2ESS-MV and 1-MV. This dataset is in the form of an array with 47 features as input parameters and 24,612-row data, and the size of the second dataset is 1,156,764 individual cells. Both the first and second datasets consist of classical power quality parameters, as shown below:

- Three phases of voltage;
- Three phases of 200 ms minimal voltage;
- Three phases of 200 ms maximal voltage;
- Voltage unbalance;
- Three phases of active power;
- Three phases of total harmonic distortion in voltage;
- Three phases of 200 ms maximum of total harmonic distortion in voltage.

### 2.3. *Proposed Methodology*

This section proposes cluster analysis as a data mining technique to group the data points based on their similarity [47]. The different algorithms were used to compare the grouping process and clustering result between hierarchical and non-hierarchical methods in this research approach. K-means is one of the algorithms from the non-hierarchical family, while the agglomerative technique is representative of the hierarchical method. The summary of processing steps to explore the dataset by using clustering techniques can be seen in the following schema in Figure 3.

The proposed methodology schema consists of five major steps: load dataset, feature engineering, applying clustering algorithm, qualitative assessment of PQ parameters, and the last is cluster algorithm evaluation and comparison. These five stages are described below:

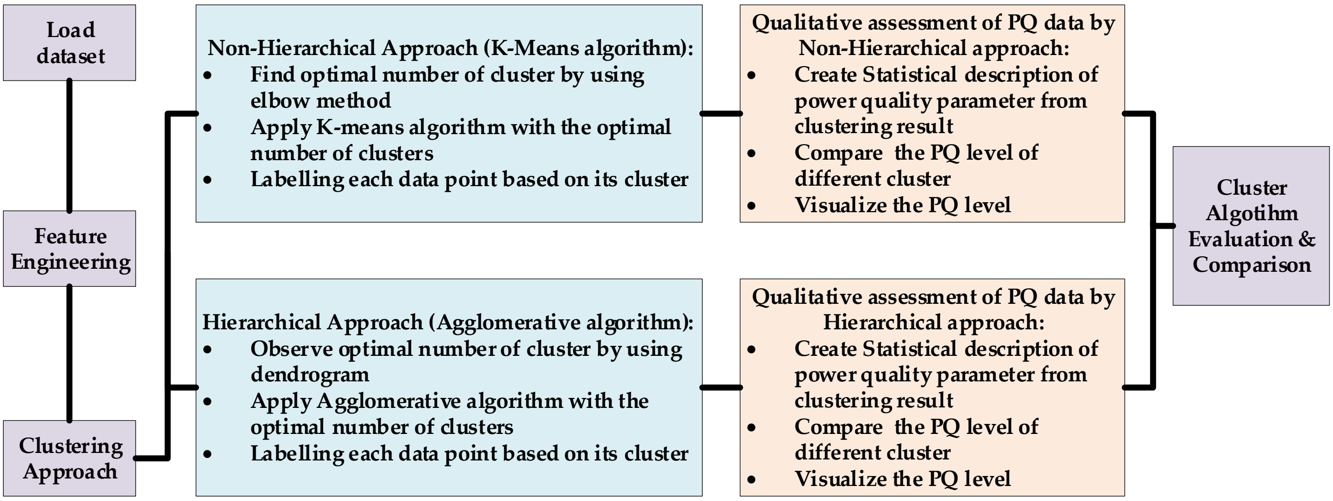

Load dataset
Feature Engineering
Clustering Approach
Non-Hierarchical Approach (K-Means algorithm):
• Find optimal number of cluster by using elbow method
• Apply K-means algorithm with the optimal number of clusters
• Labelling each data point based on its cluster
Hierarchical Approach (Agglomerative algorithm):
• Observe optimal number of cluster by using dendrogram
• Apply Agglomerative algorithm with the optimal number of clusters
• Labelling each data point based on its cluster
Qualitative assessment of PQ data by Non-Hierarchical approach:
• Create Statistical description of power quality parameter from clustering result
• Compare the PQ level of different cluster
• Visualize the PQ level
Qualitative assessment of PQ data by Hierarchical approach:
• Create Statistical description of power quality parameter from clustering result
• Compare the PQ level of different cluster
• Visualize the PQ level
Cluster Algotihm Evaluation & Comparison

**Figure 3.** Proposed methodology schema.

### 2.3.1. Load Dataset

The data are obtained from reliable single or multiple sources. This step is crucial for clustering since the quantity and quality of the dataset affect the accuracy of the output [41]. The dataset used in this article contains power quality parameters without flagged data. In this work, two different datasets are loaded to conduct the research: dataset I was used for the first research object, and dataset II was used for the second research object.

### 2.3.2. Feature Engineering

Feature engineering is a process of preparing the right input dataset that is compatible with the requirements of the clustering analysis algorithm with the distance calculation [48], and it can improve the performance of the clustering analysis model [49]. In this stage, standardization is needed because the dataset contains features with different measurement unit scales, where these differences in the ranges of initial features cause trouble for the clustering algorithm. To prevent this problem, the standardization scaling technique by the Z-score method was used to transform the features to comparable scales. The authors used the Standard Scaler library from scikit-learn [50] and Pandas library [51] as tools to standardize the features of the dataset.

### 2.3.3. Clustering Approach

In this step, the K-means [52] and the agglomeration [53] methods are used to determine the cluster of the data points. To meet this requirement, the elbow method is proposed to obtain the optimal number of clusters in the dataset by measuring the value of the sum of squares within the clusters when using the K-means algorithm. When using the agglomeration algorithm, the optimal number of clusters does not need to be defined specifically, but the dendrogram graph was proposed to observe the optimal number of clusters for this method. After finding the optimal number of clusters for each algorithm, the authors applied the clustering algorithms to the dataset to assign each data point to belong to their clusters. The summary of the cluster analysis stage can be seen in Figure 3.

### 2.3.4. Qualitative Assessment

Qualitative assessment was required to analyze the PQ problems that belong to the cluster. To perform the power quality assessment, the data points were grouped based on their cluster, and the statistical value of each cluster was measured to observe the qualitative data analysis in each cluster and compare the average value of PQ parameters from each cluster. In this article, the standard limit based on the European standard EN

50160 [43] was used as an indicator to observe the visualization of the PQ level. The acceptable values based on the standard used for the global index are presented in Table 1.

**Table 1.** Voltage characteristic of standard EN 50160 [43] used for the global index.

| **Parameter** | **Voltage Characteristic** |
|---|---|
| Voltage | 10% of declared voltage |
| Short-term flicker severity | 1.0 |
| Total harmonic distortion in voltage | 8% |
| Voltage unbalance | 2% |

### 2.3.5. Cluster Algorithm Evaluation and Comparison

Cluster evaluation determines how well the clustering algorithm can separate the dataset with the optimal number of clusters [54]. Since cluster analysis is not part of supervised learning and is performed as a part of exploratory data analysis, it is challenging to assess and evaluate the algorithm. However, there are several methods to define the validation of the clustering result. In this research, the Calinski–Harabasz index and silhouette score were used to evaluate the clustering algorithms. In this article, the metric was performed by using the scikit-learn library [44]. The silhouette function computes the mean silhouette coefficient of all samples based on the mean intra-cluster distance and the mean nearest-cluster distance for each sample. The formula is described mathematically [55,56]:

$$s = \frac{b - a}{max\ (a, b)} \tag{1}$$

where:

- $s$: the silhouette coefficient score;
- $a$: the mean distance between all other points and a sample in the same class;
- $b$: the mean distance between all other points and a sample in the next closest cluster.

The Calinski–Harabasz index, also known as the Variance Ratio Criterion, is the ratio of the sum of the dispersion between clusters and the dispersion between clusters for all clusters—the higher the value, the better the performance [57,58]. The math formula for this method is shown below [59]:

$$CH = \frac{SS_B}{SS_w}\ x\ \frac{N - k}{k - 1} \tag{2}$$

where:

- $CH$: the Calinski–Harabasz score;
- $k$: the number of clusters;
- $N$: the total number of observations (data points);
- $SS_w$: the overall within-cluster variance;
- $SS_B$: the overall between-cluster variance.

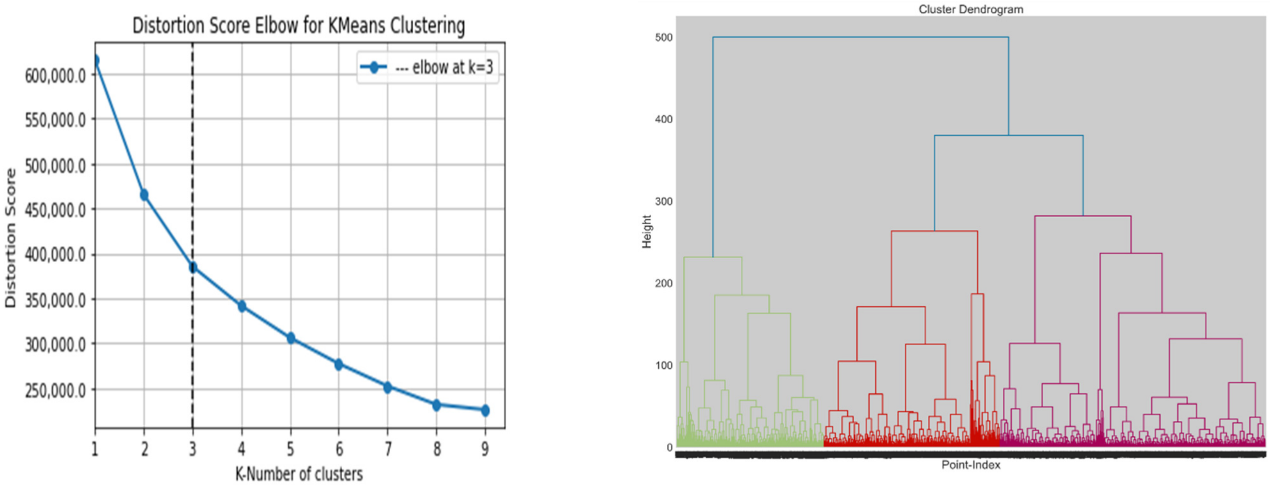
## 3. Result

This section presents the observation of the selected optimal number of clusters using the elbow method and a dendrogram for the first and second investigated object. Then, after applying the selected optimal number of clusters, the qualitative assessment was performed to compare the PQ level for different clusters. Finally, the clustering algorithm evaluation and comparison were calculated to determine which algorithm is suitable as a cluster analysis algorithm approach for the investigation objects.

### *3.1. Optimal Number of Clusters*

The optimal number of clusters is a fundamental problem in partitioning clustering such as K-means clustering or hierarchical clustering [60]. It is necessary to know the number of clusters to achieve an optimal result. In this section, the optimal number of clusters of the first investigation object and second investigation object was performed using the elbow method provided by the Yellowbrick library [61] for machine learning visualization and the dendrogram provided by using SciPy library [62] for scientific computing.

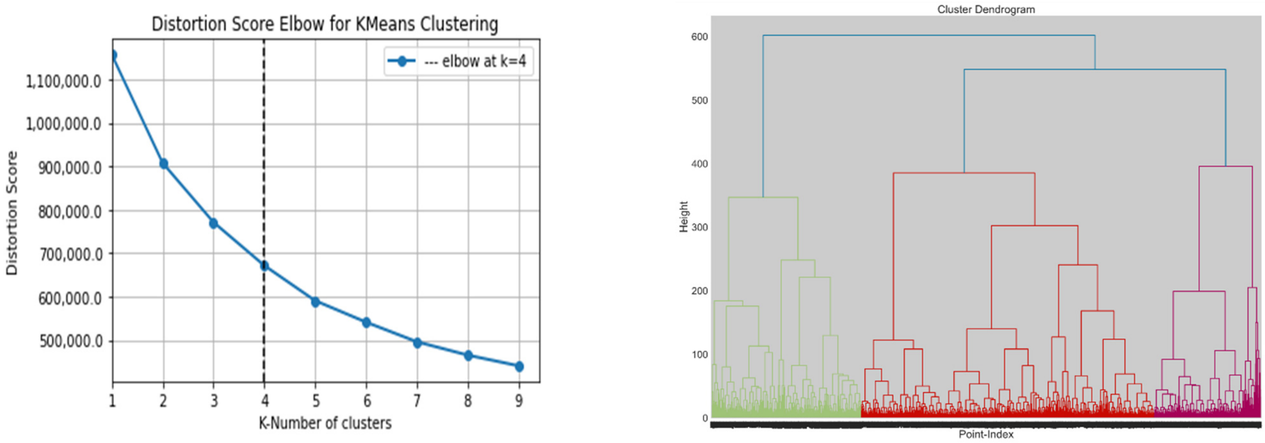
The first investigation object, the elbow method associated with the K-means algorithm, was performed for the minimal number of clusters equal to 1, and the maximal is equal to 9. However, the analysis presents that the optimal number of clusters is equal to 3, shown in the elbow curve in Figure 4a, while on a hierarchical approach by using the agglomerative algorithm with a dendrogram, the optimal number of clusters is determined equal to 3, as presented in Figure 4b, after truncating the threshold distance at height 300 in the dendrogram.



(**a**) (**b**)

**Figure 4.** The optimal number of clusters for the first investigation object: the elbow method (**a**) shows that the optimal number of clusters is equal to 3, while on the dendrogram (**b**), the possibility of the optimal number of clusters is also equal to 3.

On the other hand, dataset II belongs to the second investigation object with a bigger input size than the previous object. By using the elbow method associated with the K-means algorithm as a non-hierarchical approach, the minimal number of clusters is equal to 1, and the maximal is equal to 9. However, the analysis shows that the optimal number of clusters is equal to 4, which can be seen in the elbow curve presented in Figure 5a. On the contrary, with the hierarchical approach using a dendrogram, which is shown in Figure 5b, the optimal number of clusters is equal to 3 after determining to truncate the threshold distance at height 400 in the dendrogram.



(**a**) (**b**)

**Figure 5.** The optimal number of clusters for the second investigation object: the elbow method (**a**) shows that the optimal number of clusters is equal to 4, while on the dendrogram (**b**), the possibility of the optimal number of clusters is also equal to 3.

## 3.2. *Qualitative Assessment of Clusters*

As defined in the previous subsection, the optimal number of clusters has been determined. Thus, the qualitative assessment from the clustering result of the K-means algorithm approach and agglomerative algorithm approach is performed in this section. The authors deal with power quality parameters that are only considered in the elements of the power quality report [63]: short-term flicker magnitude and total harmonic distortion in voltage.

The assessment leads to comparing the average value of the selected parameters for different clusters based on the K-means clustering result and agglomerative clustering result to achieve knowledge about the PQ parameter level of the investigation object. The visualization of the PQ parameter was graphed on each cluster to observe the limit level of the PQ parameter through the standard EN 50160 [43,64].

### 3.2.1. Qualitative Assessment of the First Investigation Object

The qualitative assessment of the first investigation object, Table 2, shows the comparison of the statistical value of the PQ parameter level for different clusters by the K-means approach and agglomerative approach at measurement points 2HPP-MV and 2ESS-MV.

The clustering result by the K-means algorithm shows that the average value of cluster "1" is characterized by a medium level of Pst and a high level of THDu compared to other clusters. Cluster "2" is characterized by a high level of Pst and a low level of THDu compared to the other clusters. However, cluster "3" has a low level of Pst and a medium level of THDu.

On the other approach, by using an agglomerative algorithm, the comparison of the average value of the PQ level for different clusters presented in Table 2 indicates that cluster "1" represents a medium level of Pst and a high percentage level of THDu compared to others. Cluster "2" represents a medium level of THDu with a low level of Pst. Only cluster "3" has a high value for Pst and a low level of THDu.

**Table 2.** Comparison of PQ level for different clusters on the first investigation object.

| The PQ Parameter Level at Measurement Points 2HPP-MV and 2ESS-MV | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Clustering Algorithm** | **Cluster** | **Value** | **Pst** | | | **THDu (%)** | | |
| | | | **L1** | **L2** | **L3** | **L1** | **L2** | **L3** |
| K-means Approach | Cluster 1 | Mean | 0.111 | 0.115 | 0.125 | 1.3 | 1.3 | 1.3 |
| | | Min | 0.032 | 0.028 | 0.03 | 0.6 | 0.5 | 0.6 |
| | | Max | 1.358 | 1.704 | 1.584 | 2.2 | 2.4 | 2.3 |
| | Cluster 2 | Mean | 0.138 | 0.143 | 0.147 | 0.8 | 0.8 | 0.8 |
| | | Min | 0.034 | 0.032 | 0.04 | 0.4 | 0.3 | 0.4 |
| | | Max | 1.586 | 1.63 | 2.788 | 1.5 | 1.6 | 1.6 |
| | Cluster 3 | Mean | 0.101 | 0.107 | 0.111 | 1 | 1 | 1 |
| | | Min | 0.028 | 0.024 | 0.03 | 0.3 | 0.3 | 0.4 |
| | | Max | 2.386 | 2.18 | 2.864 | 1.7 | 1.8 | 1.8 |
| Agglomerative Approach | Cluster 1 | Mean | 0.104 | 0.111 | 0.119 | 1.158 | 1.108 | 1.142 |
| | | Min | 0.028 | 0.028 | 0.03 | 0.36 | 0.375 | 0.403 |
| | | Max | 0.36 | 0.352 | 0.412 | 2.234 | 2.384 | 2.319 |
| | Cluster 2 | Mean | 0.096 | 0.1 | 0.104 | 0.994 | 1.007 | 1.031 |
| | | Min | 0.028 | 0.024 | 0.03 | 0.317 | 0.308 | 0.378 |
| | | Max | 0.814 | 0.658 | 0.82 | 1.907 | 1.938 | 1.99 |
| | Cluster 3 | Mean | 0.159 | 0.163 | 0.167 | 0.905 | 0.886 | 0.894 |
| | | Min | 0.04 | 0.038 | 0.036 | 0.366 | 0.345 | 0.351 |
| | | Max | 2.386 | 2.18 | 2.864 | 1.846 | 2.011 | 1.914 |

The level of PQ parameters by the K-means algorithm approach. Cluster "2" represents the average of short-term flicker severity (Pst) in each phase at a higher level than the other clusters, while cluster "1" represents the highest average value of THDu in each phase compared to the other clusters. This can be seen clearly in Figure 6.

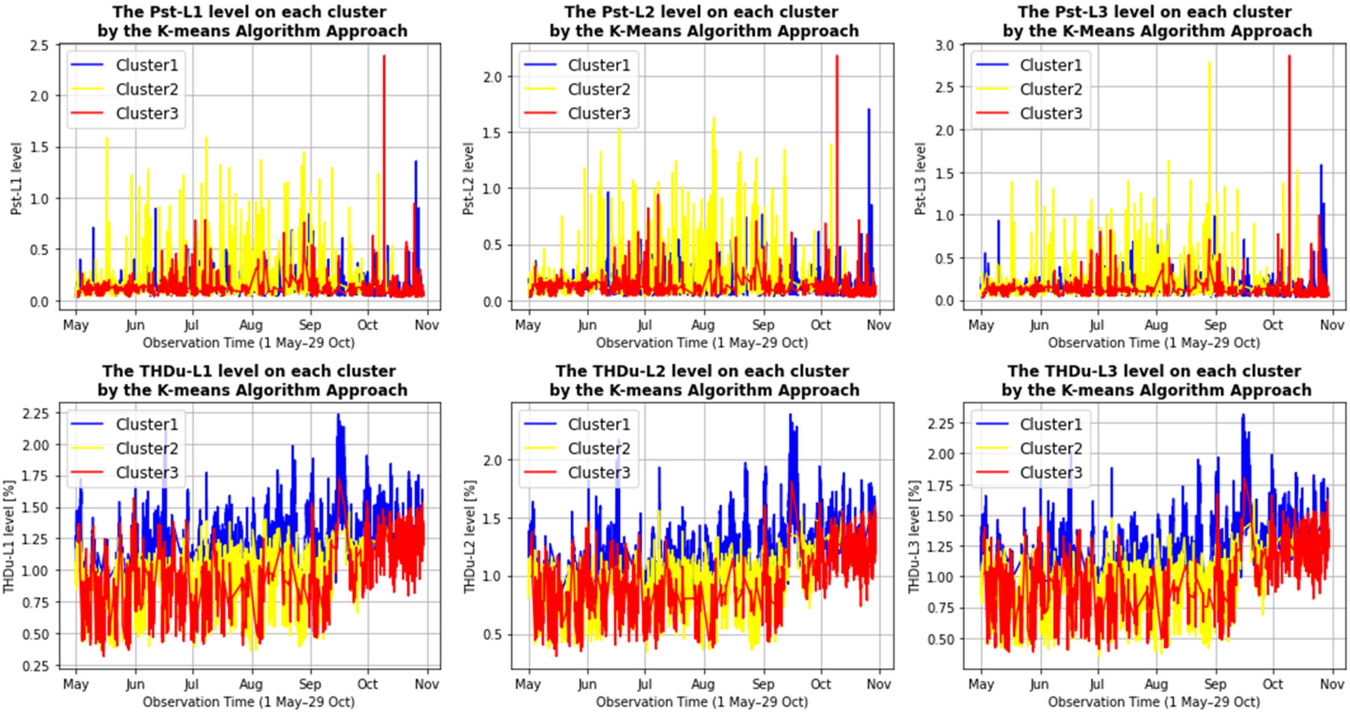


**Figure 6.** Visualization of PQ parameters' level of the first investigation object at measurement points 2HPP-MV and 2ESS-MV by the K-means algorithm approach.

Based on the standard EN 50160 [43], the visualization of the recorded set of PQ parameters of the first investigation object that is graphed in Figure 6 indicates that short-term flicker severity and total harmonic distortion in voltage on each cluster are under normal operating conditions due to the fact that all the levels of PQ parameters are below compatibility for 95% of the time.

The agglomerative algorithm approach shows that cluster “3” is indicated as the group with the highest average value of Pst in each phase compared to the other clusters, but in terms of THDu, cluster “1” has the highest average value over 1.1% among other clusters. This can be proven by the visualization of PQ parameters’ level in Figure 7. According to the standard EN 50160 [43], the recorded set of THDu and Pst, which are shown in Figure 7, is below the limit level and working under normal operating conditions during any period.

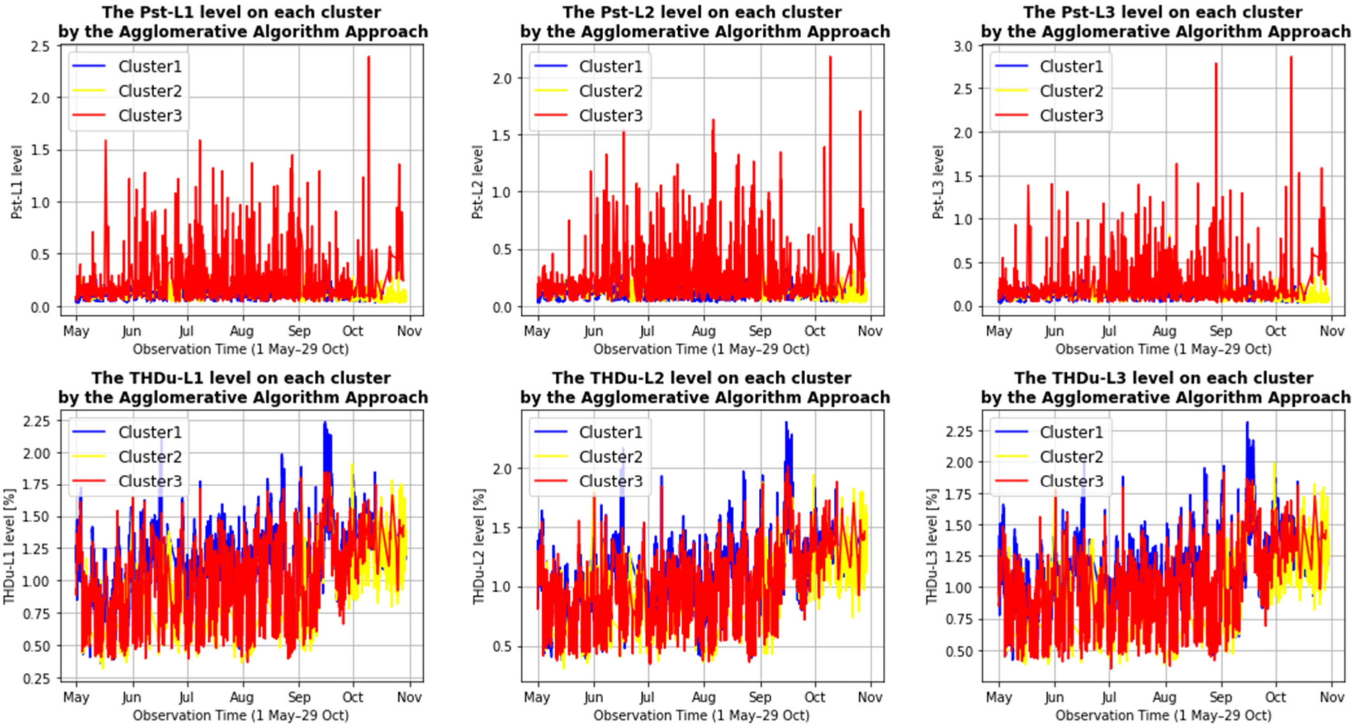


**Figure 7.** Visualization of PQ parameters’ level of the first investigation object at measurement points 2HPP-MV and 2ESS-MV by the agglomerative algorithm approach.

### 3.2.2. Qualitative Assessment of the Second Investigation Object

Table 3 shows the comparison of the statistical value of the PQ level for different clusters by the K-means algorithm approach and agglomerative algorithm approach at the measurement points 2HPP-MV and 2ESS-MV and additionally at the measurement point 1-MV for the second investigation object.

With a non-hierarchical approach using the K-means algorithm, when examining the average values of the PQ level in certain clusters at the measurement point 1-MV, cluster “1” represents a low level of Pst but an intermediate level of THDu compared to other clusters. Cluster “2” represents a high level of Pst with a low level of THDu. Cluster “3” represents a low level of Pst with a high level of THDu. Cluster “4” indicates a medium level of Pst and THDu compared to other clusters. While the average comparison value of the PQ level at measurement points 2HPP-MV and 2ESS-MV shows that cluster “1” has a medium level of Pst and low level of THDu among other clusters, cluster “2” has a high level of Pst but a low level of THDu, and cluster “3” has a high level of THDu but a medium level of Pst. Cluster “4” has a low level of Pst and a high level of THDu compared to the other clusters.

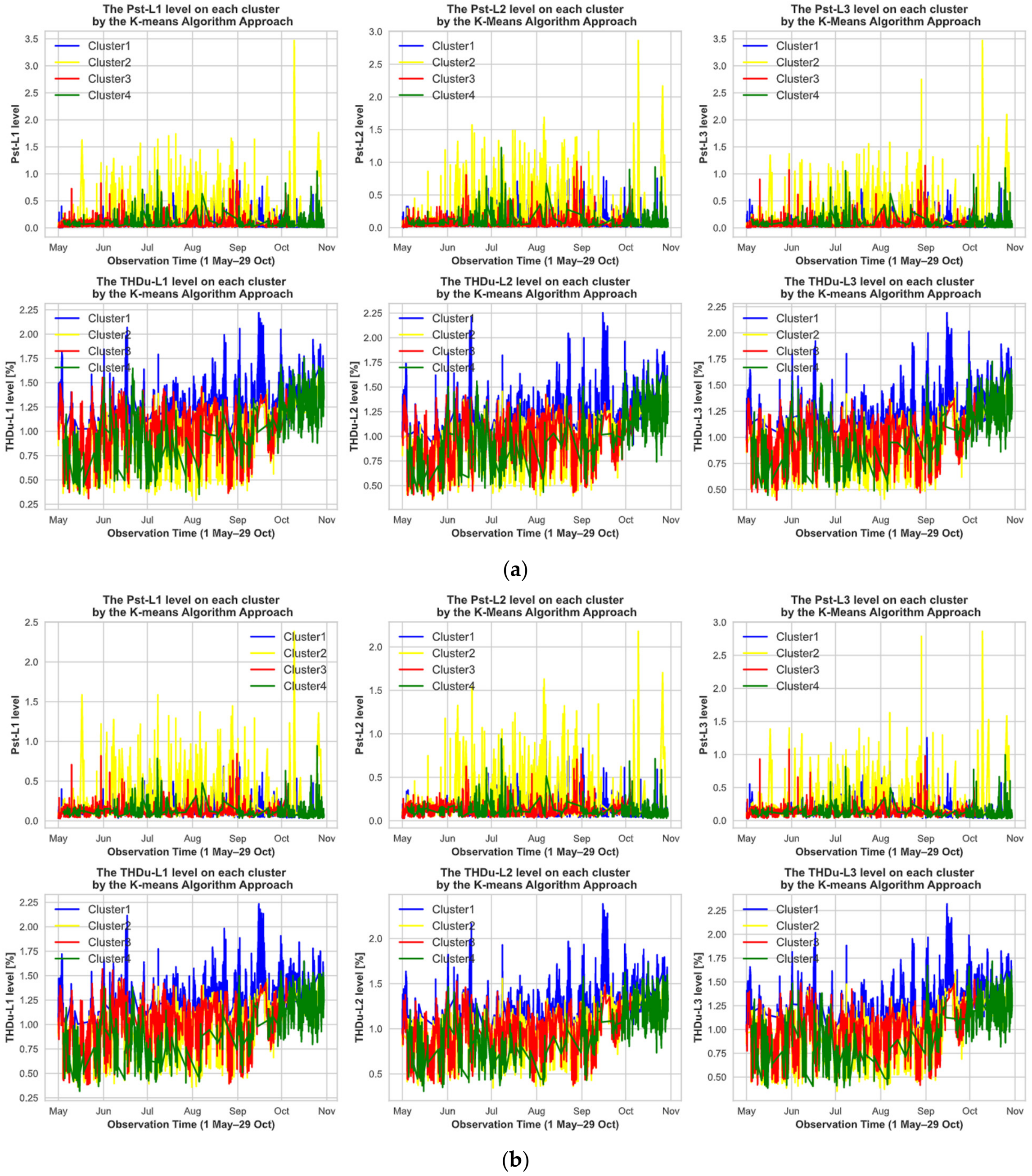
Through a hierarchical approach, the agglomerative algorithm represents the comparative evaluation of the average value of the PQ level in specific clusters. At the measurement point 1-MV, it reveals that cluster “1” has a high value of Pst and a low value of THDu compared to the others. Cluster “2” shows a medium level of both the THDu value and

Pst value. Cluster "3" shows a low level for Pst and a high level for THDu compared to the other clusters. Regarding the comparative assessment of the average value of the PQ level in specific clusters at the measurement points 2HPP-MV and 2ESS-MV, it is found that only cluster "1" has a high Pst value with a low THDu value compared to the others, while cluster "2" is characterized by a medium value at the THDu level but still has a low Pst value. Cluster "3" is the opposite of cluster "1" as it has a high THDu value with a medium Pst value compared to the other clusters.

**Table 3.** Comparison of PQ level for different clusters on the second investigation object.

| **The PQ Parameter Level at Measurement Point 1-MV** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Clustering Algorithm** | **Cluster** | **Value** | **Pst** | | | **THDu (%)** | | |
| | | | **L1** | **L2** | **L3** | **L1** | **L2** | **L3** |
| K-means Approach | Cluster 1 | Mean | 0.064 | 0.06 | 0.058 | 0.933 | 0.878 | 0.913 |
| | | Min | 0.006 | 0.008 | 0.006 | 0.308 | 0.357 | 0.4 |
| | | Max | 1.074 | 1.012 | 1.154 | 1.553 | 1.498 | 1.422 |
| | Cluster 2 | Mean | 0.121 | 0.113 | 0.114 | 0.778 | 0.785 | 0.803 |
| | | Min | 0.012 | 0.014 | 0.014 | 0.293 | 0.354 | 0.409 |
| | | Max | 3.47 | 2.864 | 3.47 | 1.587 | 1.584 | 1.56 |
| | Cluster 3 | Mean | 0.065 | 0.064 | 0.061 | 1.382 | 1.338 | 1.33 |
| | | Min | 0.004 | 0 | 0.006 | 0.61 | 0.519 | 0.671 |
| | | Max | 0.87 | 0.968 | 0.844 | 2.218 | 2.252 | 2.191 |
| | Cluster 4 | Mean | 0.07 | 0.068 | 0.066 | 1.121 | 1.082 | 1.112 |
| | | Min | 0.006 | 0.002 | 0.004 | 0.351 | 0.397 | 0.446 |
| | | Max | 1.07 | 1.226 | 1.11 | 1.755 | 1.743 | 1.727 |
| Agglomerative Approach | Cluster 1 | Mean | 0.148 | 0.138 | 0.14 | 0.758 | 0.764 | 0.784 |
| | | Min | 0.014 | 0.014 | 0.008 | 0.293 | 0.354 | 0.409 |
| | | Max | 3.47 | 2.864 | 3.47 | 1.953 | 1.929 | 1.944 |
| | Cluster 2 | Mean | 0.067 | 0.064 | 0.063 | 1.068 | 1.04 | 1.071 |
| | | Min | 0.006 | 0.002 | 0.004 | 0.335 | 0.378 | 0.433 |
| | | Max | 0.506 | 0.444 | 0.654 | 2.057 | 1.999 | 2.014 |
| | Cluster 3 | Mean | 0.061 | 0.06 | 0.057 | 1.147 | 1.096 | 1.108 |
| | | Min | 0.004 | 0 | 0.006 | 0.308 | 0.357 | 0.4 |
| | | Max | 0.57 | 0.744 | 0.528 | 2.218 | 2.252 | 2.191 |
| The PQ parameter level at measurement points 2HPP-MV and 2ESS-MV | | | | | | | | |
| K-means Approach | Cluster 1 | Mean | 0.111 | 0.119 | 0.124 | 0.934 | 0.874 | 0.919 |
| | | Min | 0.028 | 0.028 | 0.032 | 0.384 | 0.378 | 0.415 |
| | | Max | 0.846 | 0.766 | 1.078 | 1.569 | 1.526 | 1.526 |
| | Cluster 2 | Mean | 0.152 | 0.156 | 0.16 | 0.81 | 0.802 | 0.807 |
| | | Min | 0.04 | 0.038 | 0.04 | 0.36 | 0.345 | 0.351 |
| | | Max | 2.386 | 2.18 | 2.864 | 1.471 | 1.556 | 1.608 |
| | Cluster 3 | Mean | 0.111 | 0.116 | 0.125 | 1.392 | 1.366 | 1.383 |
| | | Min | 0.032 | 0.028 | 0.03 | 0.552 | 0.537 | 0.613 |
| | | Max | 0.736 | 0.982 | 1.254 | 2.234 | 2.384 | 2.319 |
| | Cluster 4 | Mean | 0.097 | 0.1 | 0.105 | 1.027 | 1.033 | 1.057 |
| | | Min | 0.028 | 0.024 | 0.03 | 0.317 | 0.308 | 0.385 |
| | | Max | 0.944 | 0.942 | 0.996 | 1.648 | 1.727 | 1.718 |
| Agglomerative Approach | Cluster 1 | Mean | 0.171 | 0.174 | 0.179 | 0.792 | 0.779 | 0.786 |
| | | Min | 0.04 | 0.038 | 0.04 | 0.366 | 0.345 | 0.351 |
| | | Max | 2.386 | 2.18 | 2.864 | 1.801 | 1.865 | 1.914 |
| | Cluster 2 | Mean | 0.097 | 0.101 | 0.105 | 0.991 | 1.004 | 1.026 |
| | | Min | 0.028 | 0.024 | 0.03 | 0.317 | 0.308 | 0.378 |
| | | Max | 0.452 | 0.398 | 0.602 | 1.907 | 1.938 | 1.99 |
| | Cluster 3 | Mean | 0.11 | 0.117 | 0.124 | 1.159 | 1.11 | 1.14 |
| | | Min | 0.028 | 0.028 | 0.03 | 0.388 | 0.391 | 0.418 |
| | | Max | 0.54 | 0.742 | 0.552 | 2.234 | 2.384 | 2.319 |

Table 3 contains two sets of the statistical value of the PQ parameters at different measurement point locations. Therefore, the qualitative analysis is performed separately. Using the K-means algorithm approach, the group of average values at the measurement point 1-MV shows that cluster "2" has the highest average value of three-phase short-term flicker severity with a value above 0.11 for each phase, and cluster “3” is a group with a high value of THDu (these parameters are visualized in Figure 8a), while the analysis at the measurement points 2HPP-MV and 2ESS-MV shows that cluster “2” has the highest average value at the Pst level and cluster “3” has a high average value of THDu, as presented in Figure 8b. According to the standard EN 50160 [43], the visualization of the recorded set of PQ parameters on the second investigation object at measurement points 2HPP-MV and 2ESS-MV (a) and 1-MV (b) by the K-means approach indicates that short-term flicker severity and total harmonic distortion in voltage on each cluster are under normal operating conditions, due to the fact that all the levels of PQ parameters are below compatibility for 95% of the time.

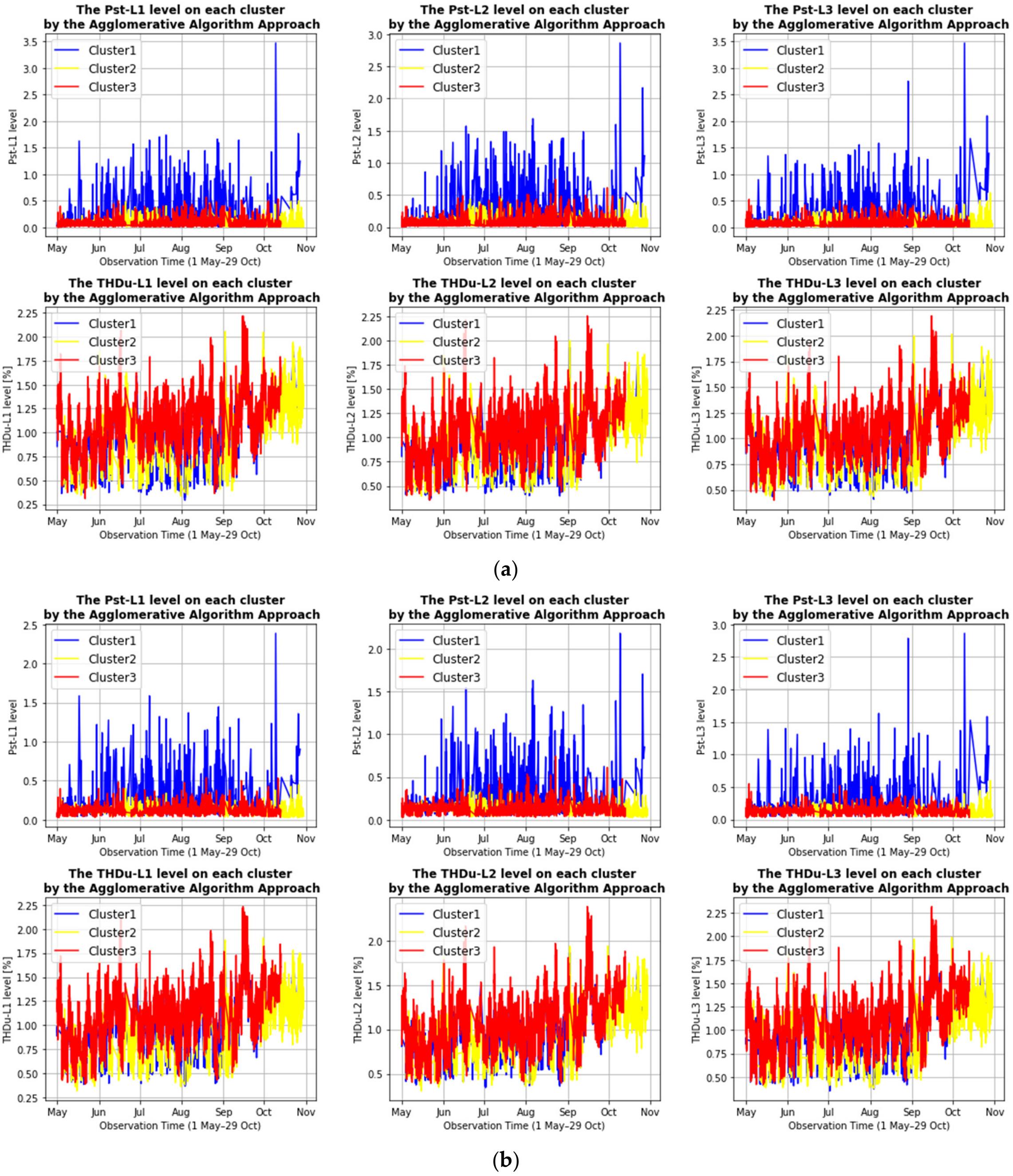


(**a**)

(**b**)

**Figure 8.** Visualization of PQ parameters’ level of the second investigation object by the K-means algorithm approach: Part (**a**) is a PQ level’s graph at measurement point 1-MV and part (**b**) is a PQ level’s graph at measurement points 2HPP-MV and 2ESS-MV.

In terms of the agglomerative algorithm approach, the average value of PQ phenomena at the measurement point 1-MV shows that cluster "1" has the highest mean value of three-phase short-term flicker severity with an average value on each phase around 0.14, followed by cluster "2" and cluster "3". For THDu, cluster "3" represents the highest value among the other clusters. The PQ parameters' level at the measurement point 1-MV (a) can be seen in Figure 9. Therefore, the qualitative evaluation of the PQ data of the measurement points 2HPP-MV and 2ESS-MV shows that cluster "1" has the highest value for Pst compared to the other clusters, and cluster "3" is the group with a high value for THDu. This can be observed in the graph of measurement points 2HPP-MV and 2ESS-MV (b), as presented in Figure 9.


The Pst-L1 level on each cluster by the Agglomerative Algorithm Approach

The Pst-L2 level on each cluster by the Agglomerative Algorithm Approach

The Pst-L3 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L1 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L2 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L3 level on each cluster by the Agglomerative Algorithm Approach

(**a**)

The Pst-L1 level on each cluster by the Agglomerative Algorithm Approach

The Pst-L2 level on each cluster by the Agglomerative Algorithm Approach

The Pst-L3 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L1 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L2 level on each cluster by the Agglomerative Algorithm Approach

The THDu-L3 level on each cluster by the Agglomerative Algorithm Approach

(**b**)

**Figure 9.** Visualization of PQ parameters' level of the second investigation object by the agglomerative algorithm approach: Part (**a**) is a PQ level's graph at measurement point 1-MV and part (**b**) is a PQ level's graph at measurement points 2HPP-MV and 2ESS-MV.

When referring to the standard EN 50160 [43], the visualization of the recorded set of PQ parameters' level at measurement point 2HPP-MV (b) and the measurement point 1-MV (a), presented in Figure 9, indicates that short-term flicker severity and total harmonic distortion in voltage on each cluster are under normal operating conditions, due to the fact that all the levels of PQ parameters are below compatibility for 95% of the time.

### *3.3. Cluster Algorithm Evaluation and Comparison*

In the first study object, dataset I was used to observe the power quality parameter issues only at measurement points 2HPP-MV and 2ESS-MV. The proposed cluster analysis algorithms, K-means and the agglomerative method, are considered to conduct qualitative data analysis on power quality measurement data. In this section, the evaluation of the algorithms is performed to investigate how well the algorithms group the data points to belong to their clusters. The silhouette coefficient and Calinski–Harabasz index were used as the clustering performance evaluation metric.

The silhouette coefficient validates clustering performance by computing the pairwise difference between and within clusters. Here, the silhouette has a value that expresses how similar an object is to its cluster compared to other clusters. The coefficient varies between $-1$ and 1, where a low value or close to $-1$ indicates that the object is assigned to the wrong cluster. A high value or close to 1 means that the object is close to its cluster and matches well with its cluster [54,65]. At the same time, the Calinski–Harabasz index is an evaluation index based on the degree of dispersion between clusters. A higher value indicates that the clusters are dense and well-separated, which is the standard concept of a cluster [66].

Table 4 contains the optimal number of clusters determined by the elbow method and the dendrogram and the evaluation of clustering performance scores by the K-means and agglomeration algorithms performed in dataset I and dataset II.

**Table 4.** Clustering algorithm performance evaluation on the investigation object.

| **Object Dataset** | **Cluster Algorithm** | **Find Optimal Number Method** | **Optimal Number of Clusters** | **Clustering Performance Evaluation Metric** | |
|---|---|---|---|---|---|
| | | | | **Silhouette Coefficient** | **Calinski–Harabasz Index** |
| Dataset I | K-Means | Elbow method | 3 | 0.235 | 7336.44 |
| | Agglomerative | Dendrogram | 3 | 0.201 | 5811.29 |
| Dataset II | K-Means | Elbow method | 4 | 0.213 | 5923.020 |
| | Agglomerative | Dendrogram | 3 | 0.219 | 4954.945 |

The silhouette coefficient score and Calinski–Harabasz index score indicate that the K-means algorithm performs better than the agglomerative algorithm to cluster the data points of dataset I, where the optimal number of clusters is equal to 3.

In the second object of study, dataset II was used to investigate the PQ disturbances at measurement points 2HPP-MV and 2ESS-MV with additional measurement point 1-MV. By the elbow method associated with the K-means algorithm, the optimal number of clusters is equal to 4, while the agglomerative algorithm determines that the optimal number of clusters is equal to 3, based on the consideration of the dendrogram. Due to the different values for the optimal number of clusters, it is challenging to compare these algorithms as they lead to a different view. However, according to the silhouette coefficient, the agglomerative algorithm performs well in separating the data points belonging to their cluster compared to the K-means algorithm. However, for the Calinski–Harabasz index, the K-means algorithm performs better than the agglomerative algorithm.

## 4. Discussion

This research aimed to apply clustering analysis as a data mining technique to study the qualitative assessment of the power quality level from a virtual power plant operating at medium voltage levels of the distribution network in Lower Silesia in Poland. The object of this study focuses on the analysis of the measurement data of three PQ recorders of the VPP. The measurement dataset consists of 10 min of classical power quality parameters with a measurement period of 26 weeks, recorded from 1 May to 29 October 2020. In previous research [32], the application of the global index was proposed to reduce the size of the input dataset from all PQ parameters to maintain the data features for cluster analysis. However, this article is a further development of the reference study [32]. This study only deals with analyzing a part of the measured data from PQ recorders to obtain comparative results between two clustering algorithms.

The object of study is divided into two sub-studies for further analysis. The first study object includes only the measurement data of PQ recorders at the measurement points 2HPP-MV and 2ESS-MV, which have 25 parameters, while the second study object for further analysis includes the first study object's data with the additional dataset of measurement point 1-MV. Hence, it has 45 features. The reason to extend the dataset into two sub-datasets is to observe how different sizes and features of the input of the dataset will influence the algorithm of cluster analysis to separate the data points and how the optimal number of clusters will be determined.

The measurement data coming from PQ recorders are not standardized. In the first attempt to perform cluster analysis without scaling the dataset, the optimal number of clusters from both algorithms was inconsistent. For this reason, the authors are concerned with the scaling of measurement data parameters to maintain the consistency and regularity of the value between features, and the standardization of the dataset can also improve the performance of the clustering analysis model.

Determining the optimal number of clusters is difficult because there is no specific answer to this question [67,68]. Prior domain knowledge can help choose the optimal number of clusters [40], but when working with an unknown dataset, it is necessary to study the dataset to determine the number of clusters [69]. In this article, the elbow method and the dendrogram are proposed to observe the relationship between the data. The elbow method is related to the K-means algorithm, and the dendrogram refers to the application of the agglomerative hierarchical algorithm. All clustering algorithms were built on Python [45] using Visual Studio Code IDE [70]. The machine learning library was used to run the K-means algorithm and the agglomerative algorithm is from scikit-learn [44]. The computer used for this experiment has a processor with a specification of intel® Core™ i5-4300U CPU@1.90 GHz~2.49 GHz, with 8 GB RAM and a total memory 2160 MB graphics card from intel® HD Graphics Family installed. The operating system of the computer is Windows 10 Pro 64-bit.

The experimental results of this study show that the execution of the dendrogram method provided by the SciPy library [62] required a long processing time of about 1350.28 s for dataset I, which belongs to the first study object, and for dataset II, which belongs to the second study object, the execution time was about 1421.64 s. This method is difficult to compute, and a crowded graph also makes it challenging to obtain the specific number of clusters.

On the other hand, the elbow method provided by Yellowbrick [61] requires the fast computation of the algorithm to calculate the sum of the squared distance between each point and the centroid in a cluster (WCSS). The execution time of the elbow method for dataset I, which belongs to the first study object, was about 13.33 s, and for dataset II, which belongs to the second study object, it was 18.08 s.

A comparison cluster analysis between the two algorithms reveals that in the first object of study, using the K-means algorithm associated with the elbow method, the optimal number of clusters is determined to equal 3, while the hierarchical relationship between the objects in the dendrogram shows that the optimal number of clusters is also equal to

3, which is suitable for the agglomerative algorithm. On the other hand, for the second study item, there was a shift in the value of the optimal number of clusters by the elbow method from 3 to 4. The current study shows that the additional features of the dataset can change the interpretation of the K-means algorithm for grouping data points based on their centroid. However, by using the dendrogram, the authors were able to decide on the optimal number of clusters, which was still set equal to 3 for the dataset of the second study object.

There are some advantages and disadvantages between the two methods presented. The elbow method with K-means provides a fast calculation, but it gives an inconsistent value of the optimal number when there are additional parameters or features, while the dendrogram for the agglomerative algorithm is hard to calculate but still produces a consistent result even if there are additional parameters or features.

## 5. Conclusions

In this article, a case study of the cluster analysis technique for analyzing the power quality level of a virtual power plant was presented. The algorithm was developed to compare the qualitative assessment of the dataset. The input dataset was standardized to have the same scale between the different feature units to achieve a compatible input for the cluster analysis algorithm. The optimal number of clusters was still difficult to choose because the additional dataset parameters may affect the performance of the clustering algorithm. The result shows that the K-means algorithm is a simple algorithm that allows fast computation, while the agglomerative algorithm is the opposite. In order to run the K-means algorithm, the specific number of clusters must be determined first, which is a disadvantage of this algorithm. On the other hand, there is no specific number of clusters required for the agglomerative algorithm, which must be determined first.

The study at the first object investigation shows that there is a possibility that the optimal number of clusters is equal to 3, which determines both the elbow method and the dendrogram. The comparison result of cluster algorithm evaluation in this object study shows that the K-means algorithm works better than the agglomerative algorithm to separate the data points of dataset I, with a higher score of silhouette coefficient and Calinski–Harabasz index of 0.235 and 7336.49, respectively.

When examining the second object, the optimal number of clusters for the two algorithms is different. The elbow method defines 4 as the optimal number of clusters for the K-means algorithm, while by the dendrogram, it is found that 3 is the optimal number of clusters for the agglomerative algorithm. Moreover, due to the different values of the optimal number, it is not fair to compare the cluster evaluation results between the two algorithms since the score is also generated differently. For the silhouette coefficient, the agglomerative algorithm performs better than the K-means algorithm. However, based on the Calinski–Harabasz index, the K-means algorithm performs better than the agglomerative algorithm.

A qualitative assessment was performed to define the comparison of statistical data analysis between clusters to determine the characteristics of the power quality parameters. It is necessary to use the standard EN 50160 [43] to consider the work of the electrical distribution network in normal or abnormal operating conditions. The cluster analysis results show that all PQ parameters on each cluster for both the first study object and the second study object indicate that all PQ levels are still below the limit value and work under normal operating conditions.

**Author Contributions:** Conceptualization, M.J. and T.S.; methodology, F.A., M.J. and T.S.; software, F.A., M.J. and T.S.; validation, D.K., J.R. and V.S.; formal analysis, F.A., M.J., T.S., D.K. and J.R.; investigation, F.A., M.J. and T.S.; resources, P.K. and P.J.; data curation, F.A., V.S. and J.S.; writing—original draft preparation, F.A. and M.J.; writing—review and editing, T.S., D.K. and J.R.; visualization, F.A., D.K. and J.R.; supervision, M.J., T.S., J.R. and Z.L.; project administration, T.S. and P.J.; funding acquisition, T.S. and P.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Center of Research and Development in Poland, the project "Developing a platform for aggregating generation and regulatory potential of dispersed renewable energy sources, power retention devices and selected categories of controllable load" supported by European Union Operational Programme Smart Growth 2014–2020, Priority Axis I: Supporting R&D carried out by enterprises, Measure 1.2: Sectoral R&D Programmes, POIR.01.02.00-00-0221/16, performed by TAURON Ekoenergia Ltd.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are not publicly available due to the policy of the associate company.

**Acknowledgments:** The authors would like to acknowledge the support of Tauron Ekoenergia Ltd.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hong, Y.-Y. Electric Power Systems Research. *Energies* **2016**, *9*, 824. [CrossRef]
2. Atputharajah, A.; Ramachandaramurthy, V.K.; Pasupuleti, J. Power Quality Problems and Solutions. In Proceedings of the IOP Conference Series Earth and Environmental Science, Putrajaya, Malaysia, 5–6 March 2013; Volume 16, p. 012153. [CrossRef]
3. Naik, C.A.; Kundu, P. Identification of Short Duration Power Quality Disturbances Employing S-Transform. In Proceedings of the 2011 International Conference on Power and Energy Systems, Chennai, India, 22–24 December 2011; pp. 1–5.
4. Mekhamer, S.F.; Abdelaziz, A.Y.; Ismael, S.M. Design Practices in Harmonic Analysis Studies Applied to Industrial Electrical Power Systems. *Eng. Technol. Appl. Sci. Res.* **2013**, *3*, 467–472. [CrossRef]
5. More, T.G.; Asabe, P.R.; Chawda, S. Power Quality Issues and It's Mitigation Techniques. *Int. J. Eng. Res. Appl.* **2014**, *4*, 8.
6. Yadav, J.R.; Vasudevan, K.; Kumar, D.; Shanmugam, P. Power Quality Assessment for Industrial Plants: A Comparative Study. In Proceedings of the 2019 IEEE 13th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), IEEE, Sonderborg, Denmark, 23–25 April 2019; pp. 1–6.
7. Jena, R. Electrical Power Quality. *Dep. Electr. Eng. CET BBSR 66.* Available online: https://www.cet.edu.in/noticefiles/227_Electrical_Power_Quality-PEEL5403-8th_Sem-Electrical.pdf (accessed on 10 June 2021).
8. Crotti, G.; Giordano, D.; D'Avanzo, G.; Femine, A.D.; Gallo, D.; Landi, C.; Luiso, M.; Letizia, P.S.; Barbieri, L.; Mazza, P.; et al. Measurement of Dynamic Voltage Variation Effect on Instrument Transformers for Power Grid Applications. In Proceedings of the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 25–28 May 2020; pp. 1–6.
9. Barros, J.; Pérez, E.; Diego, R.I. Measurement and Analysis of Voltage Events. In *Power Quality*; Moreno-Muñoz, A., Ed.; Power Systems; Springer: London, UK, 2007; pp. 73–102. ISBN 978-1-84628-771-8.
10. Tur, M.R.; Bayindir, R. Comparison of Power Quality Distortion Types and Methods Used in Classification. In Proceedings of the 2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE), Odisha, India, 29–31 July 2020; pp. 1–7.
11. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. In Proceedings of the IOP Conference Series Earth and Environmental Science, Banda Aceh, Indonesia, 26–27 September 2018; Volume 336, p. 012017. [CrossRef]
12. Lin, S.; Xie, C.; Tang, B.; Liu, R.; Pan, A. The Data Mining Application in the Power Quality Monitoring Data Analysis. In Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, China, 5–7 June 2016; pp. 338–342.
13. Sangepu, R. Effect of Power Quality Issues in Power System and Its Mitigation by Power Electronics Devices. 2015. Available online: https://www.researchgate.net/publication/325676538_Effect_of_Power_Quality_Issues_in_Power_System_and_Its_Mitigation_by_Power_Electronics_Devices (accessed on 6 June 2021).
14. Zakaria, M.F.; Ramachandaramurthy, V.K. Assessment and Mitigation of Power Quality Problems for Puspati Triga Reactor (RTP). *J. Appl. Phys.* **2017**, 020011. [CrossRef]
15. Gul, O. An Assessment of Power Quality and Electricity Consumer's Rights in Restructured Electricity Market in Turkey. *Electric. Power Qual. Utilis. J.* **2008**, *14*, 29–34.
16. Mindykowski, J. Fundamentals of Electrical Power Quality Assessment. 2003. Available online: https://www.imeko.org/publications/wc-2003/PWC-2003-TC4-027.pdf (accessed on 20 June 2021).
17. Batkiewicz-Pantula, M. The Problem of Selected Parameters of the Power Quality in the Perspective of Tightening Normative Requirements. In Proceedings of the 2019 Modern Electric Power Systems (MEPS), Wroclaw, Poland, 9 September 2019; pp. 1–4.
18. Legarreta, A.E.; Figueroa, J.H.; Bortolin, J.A. An IEC 61000-4-30 Class a-Power Quality Monitor: Development and Performance Analysis. In Proceedings of the 11th International Conference on Electrical Power Quality and Utilisation, Barcelona, Spain, 9–11 October 2011; pp. 1–6.

19. What Is the IEC61000-4-30 Standard for Power Quality Analysers? *Power Qual. Anal.* Available online: https://www.fluke.com/en-gb/learn/blog/power-quality/what-does-the-iec-61000-4-30-class-a-standard-mean-to-me (accessed on 15 June 2021).
20. Bollen, M.H.J.; Milanović, J.V.; Čukalevski, N. CIGRE/CIRED JWG C4.112-Power Quality Monitoring. *Renew. Energy Power Qual. J.* **2014**. [CrossRef]
21. Ferracci, P. Cahier Technique No. 199 Power Quality. *Power Qual.* 2000. Available online: https://eduscol.education.fr/sti/sites/eduscol.education.fr.sti/files/ressources/techniques/3361/3361-ect199.pdf (accessed on 14 July 2021).
22. Karabiber, A. Controllable AC/DC Integration for Power Quality Improvement in Microgrids. *Adv. Electr. Comput. Eng.* **2019**, *19*, 97–104. [CrossRef]
23. Vokas, G.A.; Langouranis, P.A.; Kontaxis, P.A.; Topalis, F.V. Analysis of Power Quality Field Measurements and Considerations on the Power Quality Standard 14. Available online: https://www.researchgate.net/publication/312031589_Analysis_of_power_quality_field_measurements_and_considerations_on_the_power_quality_standard (accessed on 21 June 2021).
24. Sezi, I.T.; Zimmer, I.K.; Lang, J. Power Quality Monitoring and Analysis System. In Proceedings of the 18th International Conference and Exhibition on Electricity Distribution (CIRED 2005), Turin, Italy, 15–18 June 2005; Volume 2005, p. v2-62-v2-62.
25. Nourollah, S.; Moallem, M. A data mining method for obtaining global power quality index. In Proceedings of the 2011 2nd International Conference on Electric Power and Energy Conversion Systems (EPECS), Sharjah, United Arab Emirates, 15–17 November 2011. [CrossRef]
26. Asheibi, A.; Stirling, D.; Robinson, D. Identification of Load Power Quality Characteristics Using Data Mining. In Proceedings of the 2006 Canadian Conference on Electrical and Computer Engineering, Ottawa, ON, Canada, 7–10 May 2006; pp. 157–162.
27. Larose, D.T. *Data Mining Methods and Models*; Wiley: New York, NY, USA, 2006.
28. Larose, D.T. *Discovering Knowledge in Data: An Introduction to Data Mining*; Wiley-Interscience: Hoboken, NJ, USA, 2005; ISBN 978-0-471-66657-8.
29. Ghavidel, S.; Li, L.; Aghaei, J.; Yu, T.; Zhu, J. A Review on the Virtual Power Plant: Components and Operation Systems. In Proceedings of the 2016 IEEE International Conference on Power System Technology (POWERCON), Wollongong, Australia, 28 September–1 October 2016; pp. 1–6.
30. Taylor, K. Oracle Data Mining Concepts 11g Release 2 (11.2). *Doc. E1680807 Oracle 2013.* Available online: https://docs.oracle.com/cd/E11882_01/datamine.112/e16808/title.htm (accessed on 14 July 2021).
31. Ullman, S.; Poggio, T.; Harari, D.; Zysman, D.; Seibert, D. Unsupervised Learn. Slides 2014, Fall 2014 Lecture 13. Available online: http://www.mit.edu/~{}9.54/fall14/Classes/class13.html (accessed on 24 August 2021).
32. Jasiński, M.; Sikorski, T.; Kaczorowska, D.; Rezmer, J.; Suresh, V.; Leonowicz, Z.; Kostyła, P.; Szymańda, J.; Janik, P.; Bieńkowski, J.; et al. A Case Study on Data Mining Application in a Virtual Power Plant: Cluster Analysis of Power Quality Measurements. *Energies* **2021**, *14*, 974. [CrossRef]
33. Dandea, V.; Grigoras, G.; Neagu, B.-C.; Scarlatache, F. K-Means Clustering-Based Data Mining Methodology to Discover the Prosumers' Energy Features. In Proceedings of the 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 25 March 2021; pp. 1–5.
34. Jasiński, M.; Sikorski, T.; Kaczorowska, D.; Rezmer, J.; Suresh, V.; Leonowicz, Z.; Kostyła, P.; Szymańda, J.; Janik, P.; Bieńkowski, J.; et al. A Case Study on a Hierarchical Clustering Application in a Virtual Power Plant: Detection of Specific Working Conditions from Power Quality Data. *Energies* **2021**, *14*, 907. [CrossRef]
35. Jasiński, M.; Sikorski, T.; Leonowicz, Z.; Borkowski, K.; Jasińska, E. The Application of Hierarchical Clustering to Power Quality Measurements in an Electrical Power Network with Distributed Generation. *Energies* **2020**, *13*, 2407. [CrossRef]
36. Neagu, B.-C.; Grigoras, G. A Fair Load Sharing Approach Based on Microgrid Clusters and Transactive Energy Concept. In Proceedings of the 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 25–27 June 2020; pp. 1–4.
37. Jasilski, M.; Borkowski, K.; Sikorski, T.; Kostvla, P. Cluster Analysis for Long-Term Power Quality Data in Mining Electrical Power Network. In Proceedings of the 2018 Progress in Applied Electrical Engineering (PAEE), Koscielisko, Poland, 18–22 June 2018; pp. 1–5.
38. Jureedi, N.V.V. Karunakar.; Rosalina, K.M.; Prema Kumar, N. Clustering Analysis and Its Application in Electrical Distribution System. *Int. J. Recent Adv. Eng. Technol.* **2020**, *8*, 38–43. [CrossRef]
39. Determining the Optimal Number of Clusters: 3 Must Know Methods. Available online: https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/ (accessed on 22 June 2021).
40. Patil, C.; Baidari, I. Estimating the Optimal Number of Clusters k in a Dataset Using Data Depth. *Data Sci. Eng.* **2019**, *4*, 132–140. [CrossRef]
41. Aksan, F.F.; Azizah, A.; Prihastomo, E.D. Prediction of Earthquake Magnitude Based on the Clusters in Sulawesi Island, Indonesia. *Int. J. Sci. Res.* **2021**, *7*, 7.
42. Umargono, E.; Suseno, J.E.; Vincensius Gunawan, S.K. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. In *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019), Yogyakarta, Indonesia, 25 November 2019*; Atlantis Press: Yogyakarta, Indonesia, 2020.
43. EN 50160: Voltage Characteristics of Electricity Supplied by Public Distribution Network. Available online: https://orgalim.eu/position-papers/en-50160-voltage-characteristics-electricity-supplied-public-distribution-system (accessed on 22 June 2021).

44. Scikit-Learn: Machine Learning in Python—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/ (accessed on 21 June 2021).
45. Welcome to Python.Org. Available online: https://www.python.org/ (accessed on 4 May 2021).
46. What Are the Standards for Power Quality Measurements?—Power Quality Analysers. Available online: https://powerqualityanalysers.com/knowledgebase/what-are-the-standards-for-power-quality-measurements/ (accessed on 24 August 2021).
47. Kassambara, A. *Multivariate Analysis 1: Practical Guide To Cluster Analysis in R*, 1st ed.; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2015; Volume 1, ISBN 1542462703.
48. Rençberoğlu, E. Fundamental Techniques of Feature Engineering for Machine Learning. Available online: https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114 (accessed on 19 April 2021).
49. Aggarwal, C.C. *Data Mining*; Springer International Publishing: Cham, Germany, 2015; ISBN 978-3-319-14141-1.
50. Sklearn Preprocessing StandardScaler—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed on 16 June 2021).
51. Pandas-Python Data Analysis Library. Available online: https://pandas.pydata.org/ (accessed on 16 June 2021).
52. Sklearn Cluster KMeans—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html (accessed on 30 April 2021).
53. Sklearn Cluster AgglomerativeClustering—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html (accessed on 16 June 2021).
54. Manimaran Clustering Evaluation Strategies. Available online: https://towardsdatascience.com/clustering-evaluation-strategies-98a4006fcfc (accessed on 16 June 2021).
55. Sklearn Metrics Silhouette_score—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html (accessed on 14 June 2021).
56. Nanjundan, S.; Sankaran, S.; Arjun, C.R.; Anand, G.P. Identifying the Number of Clusters for K-Means: A Hypersphere Density Based Approach. Available online: https://arxiv.org/abs/1912.00643 (accessed on 22 June 2021).
57. Wei, H. How to Measure Clustering Performances When There Are No Ground Truth? Available online: https://medium.com/@haataa/how-to-measure-clustering-performances-when-there-are-no-ground-truth-db027e9a871c (accessed on 14 June 2021).
58. Milligan, G.W.; Cooper, M.C. An Examination of Procedures for Determining the Number of Clusters in a Data Set. Available online: https://link.springer.com/article/10.1007/BF02294245 (accessed on 22 June 2021).
59. Calinski-Harabasz Index and Boostrap Evaluation with Clustering Methods. Available online: https://ethen8181.github.io/machine-learning/clustering_old/clustering/clustering.html (accessed on 23 June 2021).
60. Kutbay, U. Partitional Clustering. In *Recent Applications in Data Clustering*; Pirim, H., Ed.; InTech: Nappanee, IN, USA, 2018; ISBN 978-1-78923-526-5.
61. Yellowbrick: Machine Learning Visualization—Yellowbrick v1.3.Post1 Documentation. Available online: https://www.scikit-yb.org/en/latest/ (accessed on 16 June 2021).
62. SciPy.Org. Available online: https://www.scipy.org/ (accessed on 16 June 2021).
63. Janik, P.; Sikorski, T. *Control in Electrical Power Engineering*; Wiley: New York, NY, USA, 2009; Volume 168, pp. 1–65.
64. Markiewicz, H. 5.4.2 Standard EN 50160 Voltage Characteristics in Public Distribution Systems. Available online: http://copperalliance.org.uk/uploads/2018/03/542-standard-en-50160-voltage-characteristics-in.pdf (accessed on 22 June 2021).
65. Prabhu, P. *Method for Determining Optimum Number of Clusters*; Social Science Research Network: Rochester, NY, USA, 2012.
66. Wang, X.; Xu, Y. An Improved Index for Clustering Validation Based on Silhouette Index and Calinski-Harabasz Index. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *569*, 052024. [CrossRef]
67. Subbalakshmi, C.; Krishna, G.R.; Rao, S.K.M.; Rao, P.V. A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *Procedia Comput. Sci.* **2015**, *46*, 346–353. [CrossRef]
68. Zhou, S.; Xu, Z.; Liu, F. Method for Determining the Optimal Number of Clusters Based on Agglomerative Hierarchical Clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 3007–3017. [CrossRef] [PubMed]
69. Rahman, M.M.; Masud, M.d.A.; Mazumder, B. Estimation of the Number of Clusters Based on Simplical Depth. In Proceedings of the 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 19 December 2020; pp. 1–5.
70. Visual Studio Code-Code Editing. Redefined. Available online: https://code.visualstudio.com/ (accessed on 15 July 2021).