UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE ELECTRICIDAD

ANÁLISIS DE TÉCNICAS DE CLUSTERIZACIÓN EN SISTEMAS ELÉCTRICOS DE DISTRIBUCIÓN

Trabajo de titulación previo a la obtención del

título de Ingeniero Eléctrico

AUTOR: ALFREDO VINICIO ZÚÑIGA PEÑALOZA
TUTOR: ING. PABLO DANIEL ROBLES LOVATO

Cuenca - Ecuador

2023

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Alfredo Vinicio Zuñiga Peñaloza con documento de identificación N° 0107599508 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 18 de Mayo del 2023

Atentamente,

Alfredo Vinicio Zuñiga Peñaloza
0107599508

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Alfredo Vinicio Zuñiga Peñaloza con documento de identificación Nº 0107599508, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del artículo científico : “Análisis de técnicas de clusterización en sistemas eléctricos de distribución ”, el cual ha sido desarrollado para optar por el título de Ingeniero Eléctrico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 18 de Mayo del 2023

Atentamente,

Alfredo

Alfredo Vinicio Zúñiga Peñaloza
0107599508

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Pablo Daniel Robles Lovato con documento de identificación N° 0101663342, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ANÁLISIS DE TÉCNICAS DE CLUSTERIZACIÓN EN REDES ELÉCTRICAS DE DISTRIBUCIÓN, realizado por Alfredo Vinicio Zúñiga Peñaloza con documento de identificación N° 0107599508, obteniendo como resultado final el trabajo de titulación bajo la opción de artículo científico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 18 de Mayo del 2023

Atentamente,

PABLO DANIEL ROBLES LOVATO

Firmado digitalmente por PABLO DANIEL ROBLES LOVATO
Fecha: 2023.05.22 23:22:56 -05'00'

Ing. Pablo Daniel Robles Lovato
0101663342

RESUMEN

El diseño de nuevas líneas eléctricas de distribución es una tarea a la que se enfrentan con frecuencia los ingenieros; este trabajo propone diseñar redes eléctricas de media y baja tensión con la ayuda del Clustering. Clustering se han utilizado desde la antigüedad para el análisis de datos por su facilidad de aplicación y la calidad de sus resultados. y la calidad de sus resultados. Existen varios métodos de clustering, pero se hará una comparación entre los métodos de clustering más utilizados, K-Means y K-Medoids. Uno de los grandes problemas de de los métodos de clustering es que es imposible determinar con precisión el número de K clusters que se deben por caso. Aun así, varios principios nos dan una idea de cuántos clusters pueden generarse; analizaremos varios de estos principios para que nuestro sistema esté lo más optimizado posible. Con la ayuda MSP, obtendremos la ruta más corta que el conductor tiene que recorrer a lo largo de toda la red eléctrica proyectada, ahorrando costes y reduciendo costes. toda la red eléctrica proyectada, ahorrando costes y reduciendo los tiempos de análisis. Finalmente, con los resultados obtenidos, concluiremos qué método da resultados más óptimos, analizando costes de implementación y pérdidas a lo largo de la red.

ABSTRACT

The design of new distribution power lines is a task that engineers frequently encounter; this paper proposes to design medium and low-voltage power networks with the help of Clustering. Clustering methods have been used since ancient times for data analysis because of their ease of implementation and quality in delivering results. There are several clustering methods, but a comparison will be made between the most used clustering methods, K-Means and K-Medoids. One of the big problems of clustering methods is that it is impossible to precisely determine the number of K clusters that must be generated per case study. Still, several principles give us an idea of how many clusters can be generated; we will analyze several of these principles to have our system as optimized as possible. With the help of minimun spanning tree MSP, we will obtain the shortest route the driver has to travel along the entire projected electrical network, saving costs and reducing analysis times. Finally, with the results obtained, we will conclude which method gives more optimal results, analyzing implementation costs and losses along the network.

PALABRAS CLAVES TEMÁTICAS

K-Means

K-Medoids

Power flow

Cluster comparison

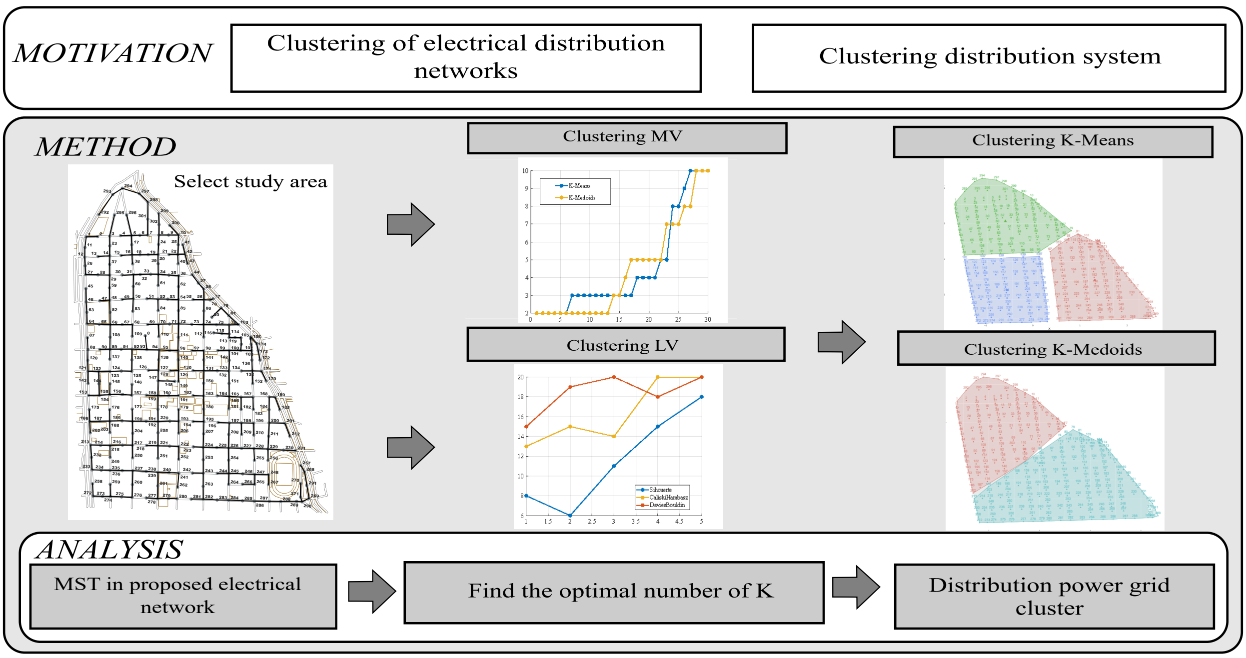
Minimum spanning tree

Electrical distribution system

Graphical Abstract

Analysis of clustering techniques in electrical distribution system

Pablo Robles,Alfredo Zuñiga



Highlights

Analysis of clustering techniques in electrical distribution system

Pablo Robles,Alfredo Zuñiga

- Custering of electric distribution networks in MV and LV.
- Change of clusters based on the increase in demand over time.
- Analysis of how the cluster acts in case a feeder goes out of service due to failures.

Analysis of clustering techniques in electrical distribution system

Pablo Robles[a,b,**], Alfredo Zuñiga[b,**]

[a]Smart Grid Research Group, Quito, 010205, Ecuador
[b]Salesian Polytechnic University, Cuenca, 010105, Ecuador

ARTICLE INFO

Keywords:
K-Means
K-Medoids
Power flow
Cluster comparison
Minimum spanning tree
Electrical distribution system

ABSTRACT

The design of new distribution power lines is a task that engineers frequently encounter; this paper proposes to design medium and low-voltage power networks with the help of Clustering. Clustering methods have been used since ancient times for data analysis because of their ease of implementation and quality in delivering results. There are several clustering methods, but a comparison will be made between the most used clustering methods, K-Means and K-Medoids. One of the big problems of clustering methods is that it is impossible to precisely determine the number of K clusters that must be generated per case study. Still, several principles give us an idea of how many clusters can be generated; we will analyze several of these principles to have our system as optimized as possible. With the help of minimun spanning tree MSP, we will obtain the shortest route the driver has to travel along the entire projected electrical network, saving costs and reducing analysis times. Finally, with the results obtained, we will conclude which method gives more optimal results, analyzing implementation costs and losses along the network.

1. Introduction

The electrical distribution networks systems have an essential role since they are responsible for transporting energy from generation to the end user; several mathematical methods allow modelling these networks faster [1], saving time and resources; we will focus on the clustering methods applied to electrical distribution networks. Clustering is one of the most widely used methodologies for data analysis; even though it was proposed more than 50 years ago [2], it is still valid when analyzing databases due to its simplicity and effectiveness. In electrical engineering, these mathematical methods can be applied to observe their behaviour when planning the construction of an electric distribution network [3] [4] (in MV as well as in LV) as applied in the case study of Figure 1. There are several ways to cluster data; this article mainly focuses on partition clustering, which requires the user to define the number of clusters K [5]. One of the main problems of clustering methods is to choose the optimal number of clusters [6] since there is no fixed mathematical method. To solve this, we propose to analyze several criteria and compare them to choose the optimal number of clusters. A cost-benefit comparison [7] is proposed to analyze which clustering method is more efficient when grouping distribution networks; with this article, the researchers will be able to consider which clustering method is more suitable for their study. The proposed cost-benefit comparison will analyze the following clustering criteria: accuracy, speed of execution, complexity, scalability, and robustness. To evaluate these criteria, we will use several techniques, such as the Davies-Bouldin index for accuracy, the time complexity analysis for speed of execution, the number of parameters for complexity, the ability to scale for scalability, and the ability to cope with outliers for robustness. The results of our study show that partition clustering is the most efficient method for clustering electrical distribution networks, as it offers the highest accuracy and speed of execution while being simple and scalable. Moreover, it is robust enough to cope with outliers. Therefore, it is the best option for clustering electrical distribution networks. We suggest the following criteria to consider when choosing the optimal number of clusters:

- Elbow method: This method determines the optimal number of clusters by looking at each cluster's squared errors (SSE). The elbow point is the point at which the rate of decrease of SSE starts to decrease [8].
- Silhouette analysis: This method measures how close the data points in a cluster are to other points in the same and other clusters. The optimal number of clusters is determined by the highest average silhouette score [9].
- Dendrogram: This method builds a hierarchical structure of clusters by looking at the similarities between the data points. The optimal number of clusters is determined by the height at which the dendrogram is cut [10].
- Gap statistic: This method compares the observed SSE of a KMN clustering model with a null reference distribution of SSEs. The optimal number of clusters is determined by the point at which the difference between the observed and the null distribution is the greatest [11].

[⋆] This document is the result of the research project of the Faculty of Electrical Engineering of the Salesian Polytechnic University

[*]Corresponding author
[**]Principal corresponding author
probles@ups.edu.ec (P. Robles); azunigap1@est.ups.edu.ec (A. Zuñiga)
www.ups.edu.ec (P. Robles); www.ups.edu.ec (A. Zuñiga)
ORCID(s): 0000-0001-8673-3974 (P. Robles); 0000-0001-6400-7103 (A. Zuñiga)

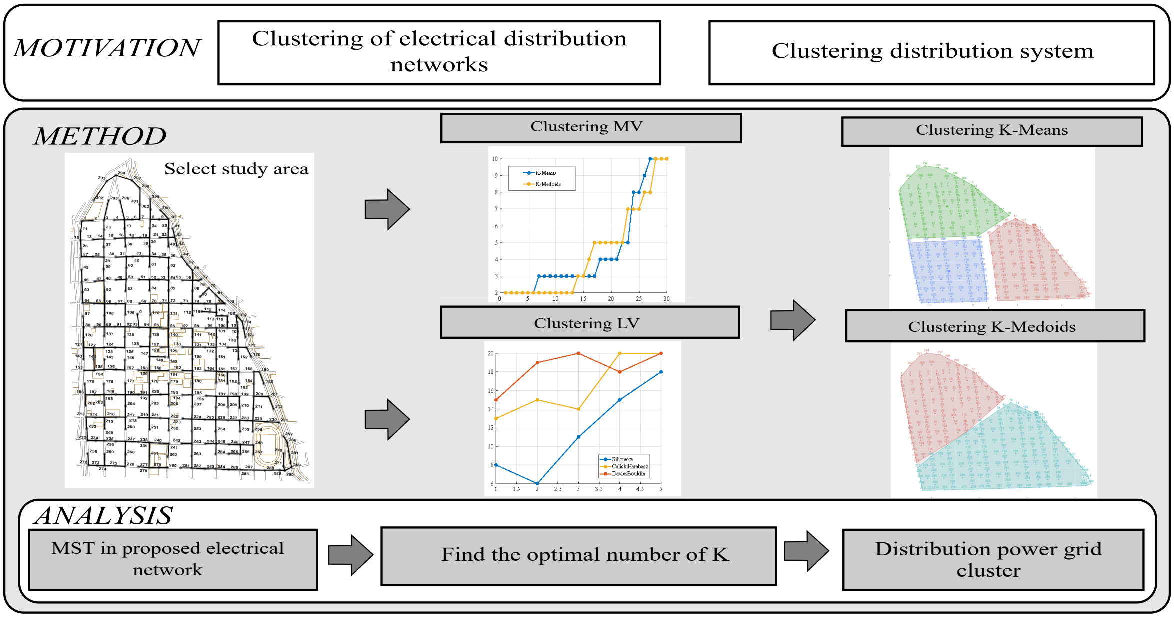


Figure 1: Figure Conceptual.

- Calinski-Harabasz Index: This method measures the ratio between-cluster to within-cluster variance. The optimal number [12].

2. Related works

Proper planning of the power system and the correct allocation of transformer stations mainly focus on the growing demand [13], which is why continuous network optimisation is necessary. When discussing data analysis, it is known about the clear distinction between learning problems. This is a widely used method for data analysis, as it can analyze large databases quickly, reliably and without user supervision [14], and can be used in various branches of research to have a clear and orderly view of the problem to be analyzed. Several researchers have applied them to optimize the location of electrical distribution transformers with the help of heuristic methods in a georeferenced way; with the help of these methods, it has been possible to locate: users, candidate sites, possible routes for the conductor along the network and other parameters that will help us to optimize the network, minimizing investment and maintenance costs. To obtain the database of users [15].

Obtaining the user database is a tedious and complicated task due to the confidentiality that the distribution company maintains with the user, so to perform all these analyses, it was decided to create an electrical network in simulation programs such as CYMDIST and, in this way be able to perform the respective analyses in an electrical network as close to reality as possible. There are many variables for analysing voltage drops along the network, such as the type of network (rural or urban), the nominal voltage, the transformation capacity, etc. Researchers have focused on reducing the voltage drops along the network, with clustering methods [21] to particular group characteristics of the network and to be able to sample the losses along the circuit. One of the most important elements in the electrical distribution networks is the electrical transformers, so it is important to determine the useful life of these transformers. To perform preventive and predictive maintenance, it has been proposed to analyze the useful life of the distribution transformers (TDE) with clustering methods by analyzing the daily load profiles with the KMN method [22]. The purpose of clustering is to reduce the amount of wire required between S/E and homes, thereby reducing installation time, costs, and risk of service interruption [23]. It is tough to perform individual analysis of each feeder due to its size and complexity, so sampling can be applied to analyze the losses [24] in such networks and classify them by clusters [19]. The smart grid is a growing field that integrates electric grids and other electronic systems. This integration aims to improve the reliability and performance of these systems. To achieve this, many smart meters and smart appliances will be connected to the grid to provide accurate data about the state of the grid in real time. However, when many devices are connected to the grid, it can be difficult to monitor them all. This is where clustering techniques can be used [25]. Clustering algorithms divide a given dataset into different clusters. Each cluster represents a grouping of data that has similar characteristics. These algorithms identify behaviour patterns and monitor the devices that make up the grid. Electric distribution systems are innovating new technologies, facilitating data exchange between consumers and the distribution company. In the case of smart grids, smart meters (AMI) can be located quickly and accurately by applying clustering methods [17], streamlining the smart grid planning process. It is interesting to see how these mathematical methods are increasingly used in research and field applications. In a

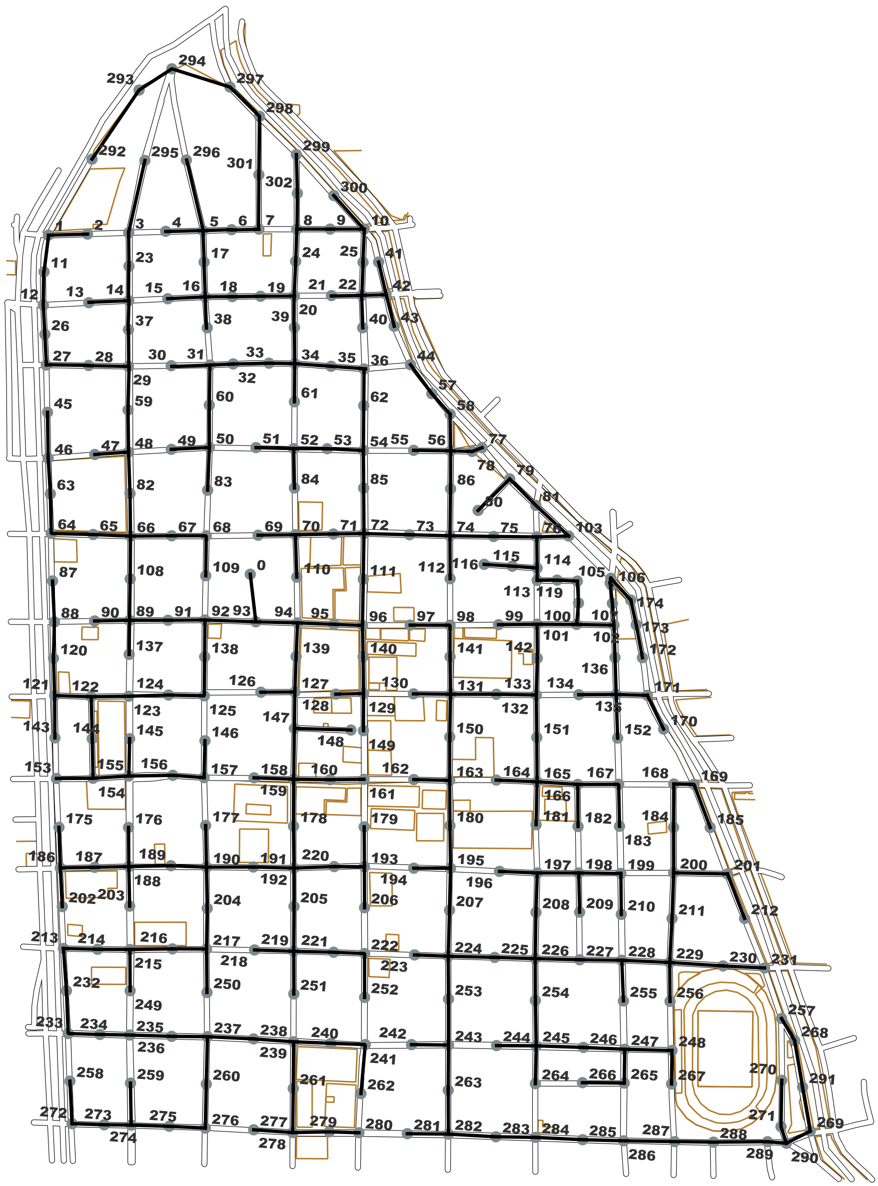


Figure 2: Scenario.

distributed generation, these methods can be used to cluster microgrids and have a network as balanced as possible, opening the way to reconfigure existing power grids [26]; the purpose behind clustering in power grid reconfiguration is to group similar nodes in a network into clusters and then reconfigure the network using these clusters instead of individual nodes. This can improve the efficiency and stability of the network since nodes within a cluster tend to have similar characteristics [27]. The results showed a significant reduction in energy losses and improved network stability. In addition, it was found that the use of clustering methods resulted in an optimal solution in terms of cost and efficiency compared to other reconfiguration techniques [28]. Renewable energies are the future, and clustering methods are present to guarantee the accuracy of operating data by improving the data quality, and detecting anomalies in the network for respective repairs and maintenance with the help of a characteristic curve created by KMN [18]. Distributed generation is an approach to producing electricity locally rather than relying on centralized power plants. This is achieved by using small renewable energy sources, such as solar panels or wind turbines, instead of large power plants. Clustering can be an effective tool for improving the efficiency of distributed generation. By grouping small power sources into clusters, energy management and control can be optimized, improving efficiency and reducing costs. In addition, clustering can help maximize energy efficiency by enabling better integration of renewable energy sources and the electricity system. Distributed generation with clustering offers numerous advantages, including reduced costs,

Table 1
Summary of articles related to flaw detection equipment

		Parameters considered				Thematic				
Author, year	Objectives	K-Mean	K-Medoid	Resizing	Geo-referencedlt Distance	Clustering	Mircrogrid	Optimization	Graph theor	MST
Nazru, 2020 [16]	Clustering in electrical distribution system	✠	-	✠	✠	✠	-	✠	✠	✠
JL. Gallardo, 2021 [17]	Clustering in electrical distribution system	-	✠	✠	-	✠	✠	✠	✠	✠
S Pinzón, 2020 [15]	Power system reconfiguration	-		✠	✠	✠	-	✠	✠	✠
L Liu, 2020 [18]	Power system reconfiguration	✠	-	✠	-	✠	-	✠	-	-
GC Cabrera,2017 [3]	Clustering in electrical distribution system	✠	✠	✠	-	✠	✠	✠	-	-
G Cartina, 2009 [19]	Power system losses	✠	-	-	-	✠	-	✠	-	-
GC Pamuji, 2020 [20]	Comparation of cluster methods	✠	-	-	-	✠	-	✠	✠	-
Present work	Clustering in electrical distribution system	✠	✠	✠	✠	✠	✠	✠	✠	✠

Table 2
Nomenclature & Description

Nomenclature	Description
MST	Minimun Spanning tree
AMI	Advanced Metering Infrastructure
TDE	Electrical distribution transformer
K	Number of clusters
S/E	Substation
MV	Medium voltage
LV	Low voltage
KMN	K-Mean
KMD	K-Medoid
SSE	Sum of Square Error
μ_i	Cluster Centorid
S	data set
x_j	Real objects
P	vertices belonging to the MST
S	Edges belonging to the MST
V	Vertices in the graph
A	Edges in the graph.
R	Auxiliary variable
x_1	Start node
x_2	End node
Kwh	kilowatt hour

improved energy efficiency, and greater resilience in power system failures. In addition, distributed generation with clustering enables better energy management and control, improving energy security and reducing reliance on centralized power plants [25].

3. Problem Formulation and Methodology

Clustering is grouping similar items, each with advantages and disadvantages. The two most popular types of clustering are KMN and KMD clustering; we will discuss the differences and some of their pros and cons. A cluster is a group of objects closely connected and separated from other objects. In other words, all of the objects in a cluster are similar in some way but different from other objects in the cluster. To perform clustering, one needs to input data into an algorithm and then use it to determine similarities between the objects in the set and organize them, the steps to follow for this grouping are shown in figure 3. The first data obtained from the electrical network are the coordinates of

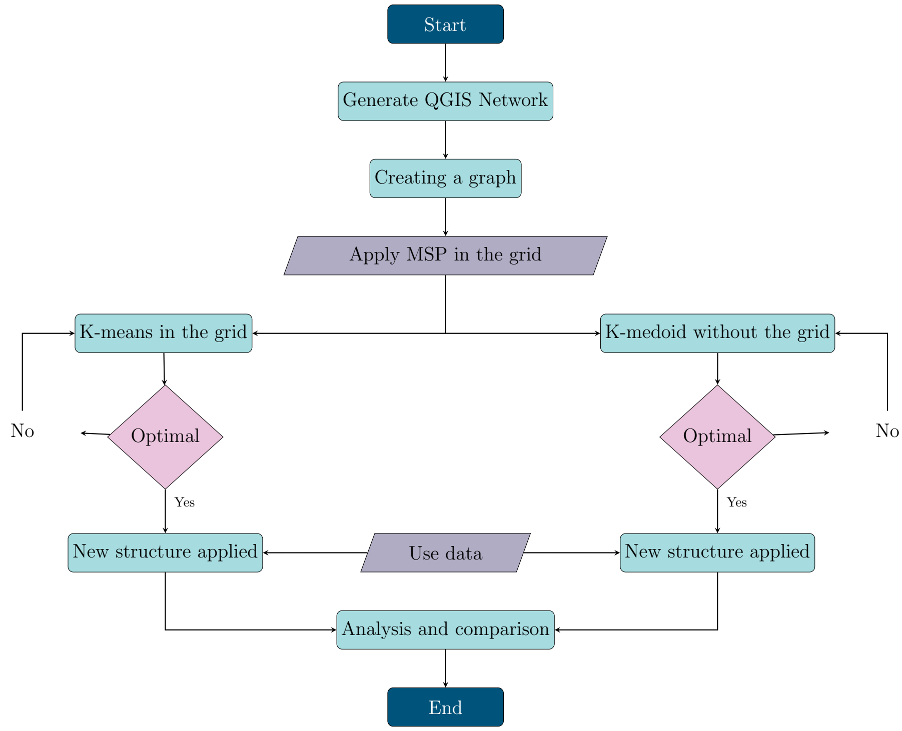


Figure 3: Flow Diagram.

the poles; with the Manhattan algorithm, we obtain the distance between them; with the georeferenced system, we can calculate the distance between nodes used in the electrical network, see equation 1.

$$f(M) = |(x_1 - x_2)| + |(y_1 - y_2)| \tag{1}$$

After determining the similarities, the algorithm uses these values to assign each object to its respective cluster. Sometimes, there may be more than one grouping method based on similarity. The clustering algorithm needs to determine which method will best organize the data. The case study is designed as a georeferenced electrical network which allows you to create a graph; the topology is defined by the minimum spanning tree algorithm. The basic idea is to create a path between two network nodes and keep that path as short as possible. This is achieved by adding each edge of the network to the tree one at a time, connecting each node of the network to the farthest node in the tree and adding each new edge to the existing tree. This process continues until the network has no more edges to add, thus having our optimized electrical network.

- **Step 1):** Create a tree structure by connecting all nodes to the furthest node, using the edges from the source to each destination; for example: if the source is labelled two and the destinations are labelled 4, 5 and 6, then the edge from 2 to 4 will be added first, followed by edges 2 to 5 and 2 to 6.
- **Step 2):** Add a new edge to each existing tree by adding the two edge nodes and connecting the new edge to the appropriate parent at its new location.

The KMN algorithm uses the distance of each observation from the average of the centroids of other observations to determine the best cluster. It is simple to implement and a good choice for smaller datasets; however, it may not be the best choice for large datasets because finding an optimal solution can take a long time. It is also possible that the clusters the algorithm finds will not be well separated, which can lead to wrong results. These are the steps that are followed to apply the K-means algorithm:

1. **Initialization:** K data points are randomly selected as initial centroids.
2. **Assignment of points to clusters:** Each data point is assigned to its nearest centroid using the Euclidean distance.
3. **Centroid update:** The centroid of each cluster is calculated as the average of the data points assigned to that cluster.
4. **Reassignment of points:** If the centroids have changed, each data point is reassigned to its nearest centroid.
5. **Iteration:** Steps 3 and 4 are repeated until the centroids no longer change or a maximum number of iterations is reached.

Algorithm information MST:

P: set of vertices belonging to the MST;

S: set of edges belonging to the MST;

V: set of vertices in the graph;

A: set of edges in the graph;

R: auxiliary variable.

Steps of the algorithm:
Step 0: Initialization of variables
Step 1: Construction of the minimal spanning tree

- Choose a vertex $i \in V$
- $S \leftarrow \phi$
- $P \leftarrow \{i\}$
- $R \leftarrow V - \{i\}$
- While $|P| \neq |V|, \forall j \in P$ do:
    - Find the smallest edge $(j,k) \in A$ such that $j \in P, k \in R$
    - $P \leftarrow P \cup \{k\}$
    - $R \leftarrow R - \{k\}$
    - $S \leftarrow S \cup \{(j,k)\}$
    - End (while)

Step 2: END.

The objects to be clustered can be represented by real vectors $x_1, x_2, ..., x_n$, which are then placed within each group $S = (S_1, S_2, ..., S_K)$, referred to as a centroid, formulated as follows. Where S is the data set consisting of the base of x_j, having K groups based on the centroids μ_i, see equation 1.

$$minE(\mu_i) = min \sum_{i=1}^{K} \sum_{x_j \in S_i} ||x_j - \mu_i||^2 \tag{2}$$

K-medoids is a clustering algorithm that divides objects into k groups based on similarity. This algorithm is similar to K-means, but instead of using the mean data points as centroids, it uses specific data points as medoids. The basic pseudocode of the K-medoids algorithm is as follows:

1. Select k random medoids from the data set.
2. Assign each data point to the nearest medoid.
3. Calculate the sum of the distances of the data points to their assigned medoids.
4. For each medoid, swap its position with a data point and calculate the sum of the distances to its assigned data points. If the sum decreases, keep the swap. If not, undo the swap.
5. Repeat steps 3 and 4 until no more change in the total sum of distances.
6. Return the k medoids and the assignment of each data point to a medoid.

The algorithm continues to iterate until the total sum of the distances of the data points to their assigned medoids no longer changes or until a specified number of iterations is reached. The final result is k medoids and the assignment of each data point to a particular medoid. But being unsupervised clustering methods, the user is in charge of imposing the number of clusters K in which the database will be divided, as mentioned above; several criteria help the end user to choose this constant K. One of the most used criteria is the elbow method, is based on the idea that the optimal number of clusters is one in which increasing the number of clusters no longer provides a significant reduction in the sum of the intra-cluster distances. It is important to note that the elbow method does not guarantee finding the optimal number of clusters in all situations. In some cases, it may be necessary to try different techniques and consider other factors, such as the interpretability of the results, to determine the optimal number of clusters. In this case study, the weight of the cluster will be changed to perform the analysis of how many low-voltage transformers have to be installed per feeder; for this analysis, it is considered that the installed load is between $61 < Kwh <= 110$, with this data we will have a better result of how many transformers will go per feeder, depending on the load. Clustering in distribution power networks refers to grouping similar elements; this approach improves network operation, maintenance efficiency, and effectiveness. There are different applications of clustering in distribution power networks, including grouping of transformers in terms of characteristics, such as load, geographical location and power consumption; improving power supply planning and control; Segmentation of transmission lines into different groups based on criteria such as transmission capacity, distance and cost; Grouping of similar consumers in terms of power consumption, consumption schedule and geographical location, to improve efficiency in power delivery and demand management. This paper, 3 case studies are performed to see how the clustering groups a projected electrical network. In the first case study, the cluster is applied twice; the first clustering results in the 22KVA medium voltage network, and the second results in the 220V low voltage network, KMN and KMD will be used to observe how the distribution networks are clustered under normal conditions. For the second case study, the load is modified over 10 years, to observe how the number of LV transformers varies as the demand increases over time, considering an annual growth in demand of 4.6 % as considered by [29]. For the third case study, it is assumed that one of the feeders goes out of service due to a failure in one of its components. This causes the residential area fed by that feeder to be without a power supply. In this situation, the decision is made to reconnect the load to another feeder of the same substation to restore the power supply in the

Table 3
Optimal number of K-cluster MV.

	K-Means		K-Medoids	
Item	Method	Clusters	Method	Clusters
1	Elbow method	3	Elbow method	2
2	Silhouette method	3	silhouette method	3
3	Gap stat Method	1	Gap stat Method	2
4	kl Method	4	kl Method	8
5	ch Method	9	ch Method	7
6	Hartigan Method	3	Hartigan Method	4
7	Scott Method	3	Scott Method	5
8	Trcovw Method	4	Trcovw Method	3
9	Hubert Method	3	Hubert Method	5
10	Dindex Method	4	Dindex Method	2
	Optimum number of 30 methods	**3**	**Optimum number of 30 methods**	**2**

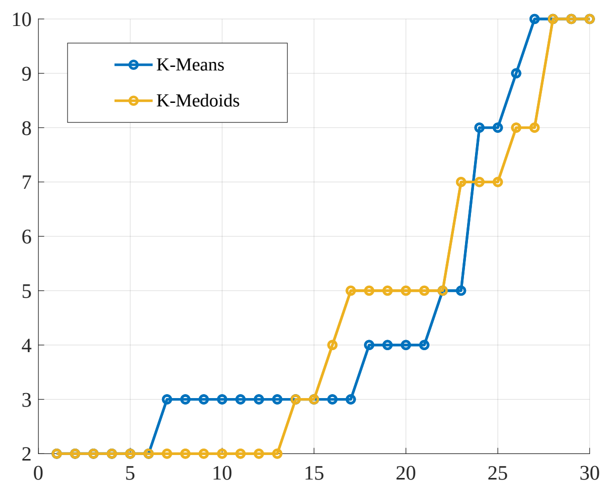
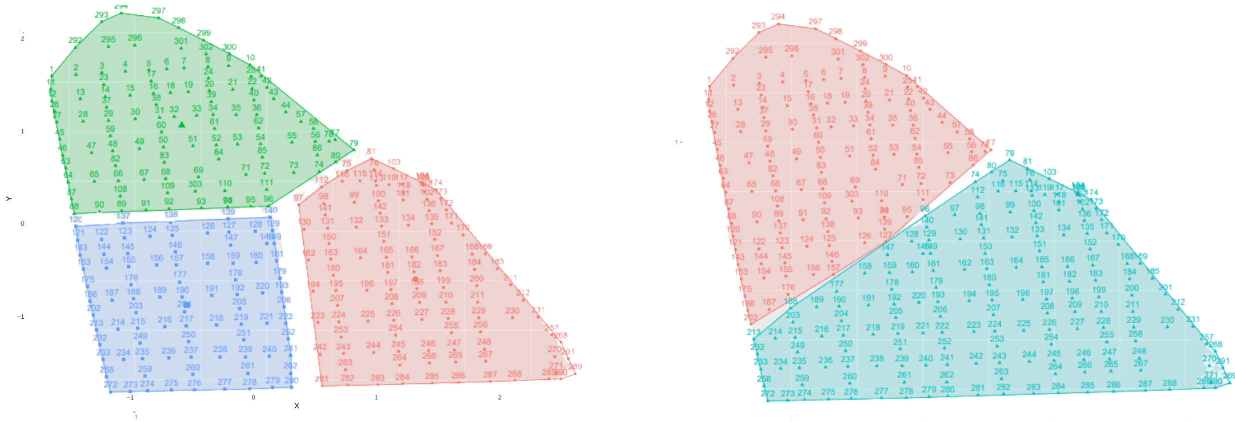


Figure 4: Optimal number of K-cluster MV.

affected area. With this failure, it will be analyzed how the power flow changes at the substation level. The intention is to see how the system reacts in case of failures, interconnecting one feeder with another and see if the voltage drops are optimal throughout the reconnected substation.

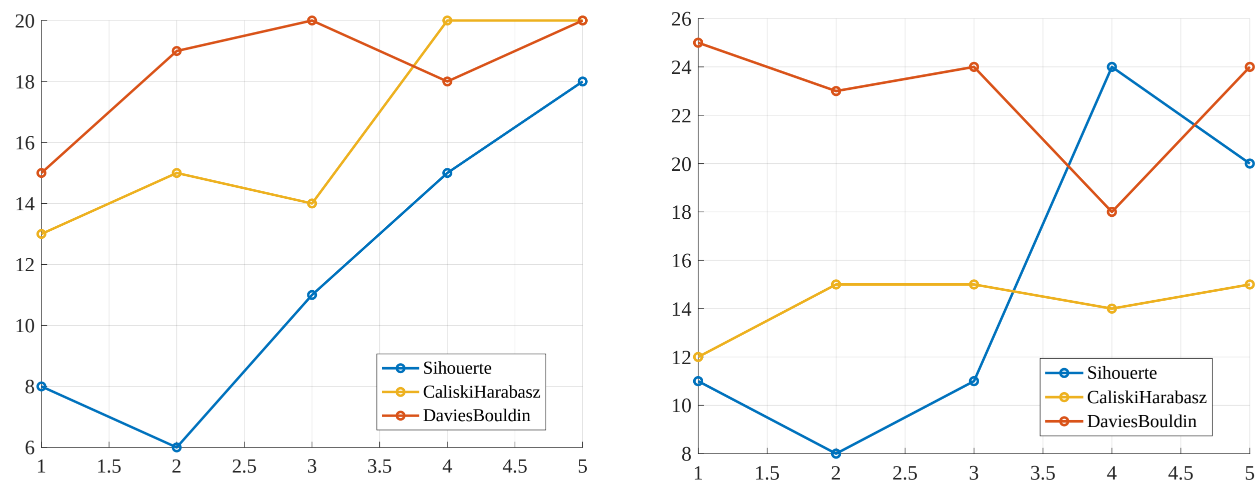
4. Analysis of results

To cluster an electrical distribution network, there is an unsupervised algorithm, and there is no specific method to find K, but there are several criteria that give us an idea of how many groups can cluster in the network; in Table 3, it can see some of the criteria applied for the analysis, However, to obtain more precise data, we compare between 30 criteria and calculate the mode to find the optimal number of K, in Figure 4 it is observed that on the abscissa axis is the number of criteria, while on the ordinate axis the number of K proposed by each criterion. On the one hand, K-Means proposes that the electrical system divides into 2 clusters for the medium voltage network, figure 5a. In contrast, K-medoids proposes dividing the system into 3 clusters, figure 5b. Each of these clusters represents an electrical feeder; see table 5, defining the clusters necessary to perform an MST to obtain the optimal topology from each of the proposed feeders to the farthest node of the same. For the low voltage network, a similar methodology is used, with the main difference that in the clustering methods, the weight variable change; instead of calculating the distance between nodes by Manhattan and finding the optimal number of K, the demand of each node use considering an area with residential load, with this change expect to have a better distribution of transformers, see figure 6a, calculating the average of the data obtained between the 3 proposed criteria: Silhouette method, Caliski Harabasz method and Davies Boulding method, for each feeder by both K-Means and K-medoids, this procedure also apply with constant demand growth of 4.8% in 10 years, see figure 6b. It is essential to consider the energy expected to be consumed by the low-voltage clusters to find the appropriate transformer power based on the low-voltage cluster load. In the proposed case studies, the power consumed is stable by analyzing the demand of the area where the transformers are to install; it is necessary to determine the cluster's power factor (PF). The PF is a measure of efficiency that indicates how much of



(a) (b)

Figure 5: a) k-Means Cluster MV. (b) K-medoids cluster MV.



(a) (b)

Figure 6: a) Optimal number of K LV. (b) Optimal number of K LV after 10 years.

the supplied energy is used to perform practical work; a PF of 1 means that all the supplied energy is used to perform practical work, while a PF of 0.8 indicates that only 80% of the supplied energy use to perform practical work, and 20% is lost as heat and other forms of non-useful energy. Finally, the transformer input and output voltage level and efficiency must be considered to determine the transformer power rating needed to power the cluster. CYMDIST allows simulation of the electrical distribution system's behaviour; the feeders' topology is maintained with clustering through K-means and K-medoids, Table 4, with the main difference in the number of transformers. The third case study considers a fault in one of the distribution feeders; common causes of electrical feeder failures include short circuits, overheating, connection problems, and component failures. Each of these causes can have different effects. Still, in general, all of them can cause interruptions in the power supply to the end user, as well as damage to electrical and electronic equipment connected to the network. To ensure reliability, the network reconnects the load through its protections, shunting it from the feeder that fails to another feeder in the same substation see figure 7. In the network proposed by K-Means, figure 7a observes that one of the three feeders fails, so it goes out of service. The load is transferred to the nearest feeder, resulting in two feeders in this new network; on the other hand, the network proposed by K-Medoids, figure 7b, feeder 2 fails, which means that only one feeder remains in service in this substation, being this the one that feeds the entire load affected by the failure. The results of the losses obtained by this reconfiguration of feeders are in table 4; it is possible to observe that there is a slight variation in the voltage drop in the feeders when the load reconnect when the load derive from one feeder to another, a temporary overload can occur in the system, which can cause a voltage drop in the affected feeder. This voltage drop may be small, but it can still affect the operation of equipment connected to the network. It is essential to carry out preventive measures and monitor the electrical system to prevent this situation. These include

Table 4
Summary of electrical network

	Total energy production			Total loads			Total losses		
	kW	kVAR	KVA	kW	kVAR	KVA	kW	kVAR	KVA
KMN-SE	1.416,28	1.043,36	1.759,10	1.392,47	1.044,22	1.740,51	23,80	45,70	51,53
KMDI-SE	1.164,51	1.083,03	1.590,29	1.139,99	1.082,97	1.572,38	24,52	47,02	53,03
KMNS-SE Reconnected	1.416,31	1.043,16	1.759,79	1.392,47	1.044,22	1.740,51	23,84	45,98	51,80
KMDI-SE Reconnected	1.164,70	1.083,67	1.590,87	1.139,99	1.082,97	1.572,38	24,71	47,75	53,76

(a) (b)

Figure 7: a) K-Means reconnected by feeder failure 3. (b) K-Medoids reconnected by feeder failure 2.

the installation of voltage and current measuring equipment on feeders, the implementation of protection and control devices, and the early identification of system problems.

5. Discussion

Figure 8a shows how K-Means divides the medium and low-voltage electrical network. Once the optimal number of K is analyzed independently, K-Means is faster and more scalable for data analysis. Still, it is sensitive to the initialization of the centroids and outliers in the database; on the other hand, K-Medoids is more robust and tolerant to outliers, being slower in its convergence; this does not prevent it from optimally clustering the electrical distribution networks. It is essential to analyze the importance of the weight at the time of clustering a database; for the medium voltage network, the cluster performs considering as weight of the distance between nodes; being the network less prone to large voltage drops, it is possible to calculate the collections without significant difficulty, for the low voltage network, it is necessary to change the weight to consumption demand per node using a residential load, this ensures an improvement in the algorithm at the time of clustering, giving a consistent number of transformers to distribute the electrical energy to the loads. With the growth of the 10-year demand, the installed power increases, requiring transformers with a higher power or, in turn, more transformer units since the existing power grid would become saturated at a certain point; this behaviour can to seen in Figure 6b, where it sees that the number of transformers increases according to the criteria for finding K. In the case in which one of the feeders fails for both K-Means and K-Medoids, it observed how the voltage drops increase due to the load increase, which makes its technical losses for energy transport also increase, as can be seen in Table

Table 5
System characteristics in normal state

S/E Capacity	5 MVA
Total Load	
Real Power	3.686 MW
Reactive Power	2.325 MVAR
Apparent Power	4.358 MVA
Load Used	
Real Power	3.614 MW
Reactive Power	1,534 MVAR
Apparent Power	3,926 MVA
Total losses	0,823 MVA
Max. ΔV	2,85 %
Length	15899 m

(a) (b)

Figure 8: a) k-Means Cluster LV. (b) K-medoids cluster LV.

4; this situation ensures that the system remains reliable to supply electricity continuously and with the minimum possible interruptions.

6. Conclusions

This article mainly aims to analyse cauterisation methods in electrical distribution networks. To cover this purpose, the analysis of three case studies with different types of problems that involve a degree of difficulty to analyse the behaviour of the clustering algorithm so that it can have a reliable, safe and balanced power grid; it observes that clustering methods have proven to be highly efficient to create projected power grids, achieving favourable results in meeting standards. It is essential to note that this type of algorithm frequently use to model distribution power grids [4] quickly and efficiently; the choice of one method or another depends on the specific characteristics of each data set, and the objectives of the analysis, cauterisation of distribution power grids can be helpful to identify patterns of consumption, power generation and loss assessment to segment customers or network users into homogeneous groups [19]. This can help researchers make informed decisions about grid management and maintenance and analyse fair and efficient pricing and tariff policies. Finally, it is essential to note that clustering is one tool within the broad field of data analysis; other techniques and methods can be helpful to analyse and better understand energy consumption and generation patterns. Therefore, combining data analysis tools and methods are recommended to obtain a more complete and accurate network view. In conclusion, the comparative study of K-Means and K-Medoids clustering methods in electrical distribution networks is essential to data analysis in electrical distribution. The results indicate that both methods are helpful and have specific strengths and weaknesses that should be considered depending on the data set's characteristics and the analysis's objectives.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Pablo Robles: Conceptualization of this study, Methodology, Software. **Alfredo Zuñiga:** Methodology, Software, Formal analysis.

References

[1] E. B. Herrera-Cisneros and E. M. Inga-Ortega, "Planeación óptima de redes aéreas de distribución basada en teoría de grafos," *Iteckne*, vol. 15, no. 2, pp. 122–130, 2018.

[2] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010, award winning papers from the 19th International Conference on Pattern Recognition (ICPR). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865509002323

[3] G. C. Cabrera-Celi, E. G. Novoa-Guaman, and P. F. Vasquez-Miranda, "Design of secondary circuits of distribution networks using clustering and shortest path algorithms," in *2017 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*. IEEE, 2017, pp. 1–6.

[4] J. D. Watson, N. R. Watson, D. Santos-Martin, S. Lemon, A. Wood, and A. Miller, "Low voltage network modelling," 2014.

[5] K. Lin, X. Jing, Z. Hengjun, and W. Beibei, "Planning method for distribution networks based on load growth characteristics of the industry," in *2016 China International Conference on Electricity Distribution (CICED)*. IEEE, 2016, pp. 1–5.

[6] D. Aggarwal and D. Sharma, "Application of clustering for student result analysis," *Int J Recent Technol Eng*, vol. 7, no. 6, pp. 50–53, 2019.

[7] G. A. Argudo Guaillas and J. P. Samaniego Placencia, "Optimización de los alimentadores de media tensión y transformadores de distribución de la s/e 17 los cerezos proyectada por la centrosur," B.S. thesis, 2019.

[8] C. C. Aggarwal and C. R. D. Clustering, "Algorithms and applications," 2014.

[9] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, p. 759, 2021.

[10] M. Forina, C. Armanino, and V. Raggio, "Clustering with dendrograms on interpretation variables," *Analytica Chimica Acta*, vol. 454, no. 1, pp. 13–19, 2002.

[11] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[12] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and calinski-harabasz index," in *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5. IOP Publishing, 2019, p. 052024.

[13] L. Chao, Z. Lei, and L. Yuhang, "Topology checking method for low voltage distribution network based on fuzzy c-means clustering algorithm," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE, 2020, pp. 1077–1080.

[14] J. A. Rodrigo, "Clustering y heatmaps: aprendizaje no supervisado," 2017. [Online]. Available: https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps

[15] S. Pinzón, S. Yánez, and M. Ruiz, "Optimal location of transformers in electrical distribution networks using geographic information systems," *Enfoque Ute*, vol. 11, no. 1, pp. 84–95, 2020.

[16] F. N. Shimim, H. Nehrir, M. Bahramipanah, and Z. Shahooei, "A graph-theory-based clustering method for improving resiliency of distribution systems," in *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2020, pp. 1–6.

[17] J. L. Gallardo, M. A. Ahmed, and N. Jara, "Clustering algorithm-based network planning for advanced metering infrastructure in smart grid," *IEEE Access*, vol. 9, pp. 48 992–49 006, 2021.

[18] Y. Li, W. Tongxun, T. Meng, L. Yaqiong, and Z. Pi, "Cluster feature based multivariate data analysis and recovery method for renewable energy operation and control," in *2020 IEEE International Conference on Energy Internet (ICEI)*. IEEE, 2020, pp. 1–5.

[19] G. Cartina, G. Grigoras, and E. Bobric, "Clustering techniques for energy losses evaluation in distribution networks," in *2009 IEEE Bucharest PowerTech*. IEEE, 2009, pp. 1–5.

[20] G. C. Pamuji and H. Rongtao, "A comparison study of DBScan and k-means clustering in jakarta rainfall based on the tropical rainfall measuring mission (TRMM) 1998-2007," *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, p. 012057, jul 2020.

[21] G. Grigoras, G. Cartina, M. Istrate, and F. Rotaru, "The efficiency of the clustering techniques in the energy losses evaluation from distribution networks," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 5, no. 1, pp. 133–141, 2011.

[22] E. Mulliez, M. A. Sanz-Bobi, A. Sanchez, P. Mazidi, A. Gonzalez, and R. Bachiller, "Life characterization of power distribution transformers using clustering techniques," in *European Conference of the Prognostics and Health Management Society, PHM Society, Utrecht*, 2018.

[23] E. Chelaru and G. Grigoraş, "Decision support system to determine the replacement ranking of the aged transformers in electric distribution networks," in *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE, 2020, pp. 1–6.

[24] H. Fathabadi, "Power distribution network reconfiguration for power loss minimization using novel dynamic fuzzy c-means (dfcm) clustering based ann approach," *International Journal of Electrical Power &*

Energy Systems, vol. 78, pp. 96–107, 2016.

[25] E. Bullich-Massague, F. Díaz-González, M. Aragüés-Peñalba, F. Girbau-Llistuella, P. Olivella-Rosell, and A. Sumper, "Microgrid clustering architectures," *Applied energy*, vol. 212, pp. 340–361, 2018.

[26] A. Valenzuela, I. Montalvo, and E. Inga, "A decision-making tool for electric distribution network planning based on heuristics and georeferenced data," *Energies*, vol. 12, no. 21, p. 4065, 2019.

[27] Y. Zhang, X. Ji, J. Xu, Z. Yin, X. Han, and C. Zhang, "Dynamic reconfiguration of distribution network based on temporal constrained hierarchical clustering and fireworks algorithm," in *2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*. IEEE, 2020, pp. 1702–1708.

[28] K. Gholami, S. Karimi, and E. Dehnavi, "Optimal unified power quality conditioner placement and sizing in distribution systems considering network reconfiguration," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 32, no. 1, p. e2467, 2019.

[29] C. A. Espinosa and J. L. Hernández, "Electricity demand and supply in ecuador: Current status, future projections, and policy implications," *Energy Policy*, vol. 134, p. 110995, 2019.