Geospatial Clustering of Wind/Solar Potential: Identifying High-yield Areas in Ceará, Northeast Brazil

Igor Hammom Rodrigues da Costa[1], Dalton de Araújo Honório[1], Paulo Honório Filho[1]
Ernande Eugênio Campelo Morais[1], Maria Ilana Dos Santos Guedes[2]

[1] Federal University of Ceará, Fortaleza (Brazil)

[2] Imperial College London, United Kingdom (London)

Abstract

This article addresses the identification of municipalities in the state of Ceará with the greatest potential for wind and photovoltaic power generation, using a data clustering method applied to climatic and geographic variables. For this purpose, variables such as altitude, wind speed, solar irradiation, and average temperature were considered. Based on these criteria, the cities were distributed into distinct clusters using the K-means algorithm. The results revealed the existence of distinct groups of municipalities with specific characteristics for each type of power generation. To classify the groups most suitable for power generation, a scoring methodology based on technical criteria was employed. Each group received specific scores according to its influence on the type of generation, in order to identify the favorable groups for wind and photovoltaic generation. The total score for each cluster was obtained by summing the points of the variables for each generation type, allowing for a classification of the most promising groups. Based on the results, it was found that Group 3 proved to be favorable for wind power generation due to its high temperatures and wind speed. On the other hand, Group 1 was identified as more suitable for photovoltaic generation due to its higher altitude and milder temperatures.

Keywords: Renewable Energy, Wind Energy, Photovoltaic Energy, K-means algorithm, Clusterization

1. Introduction

The state of Ceará, located in northeastern Brazil, stands out for its high potential for generating energy from renewable sources, particularly wind and photovoltaic, due to its favorable geographic and climatic characteristics. This capacity stems from favorable natural conditions, such as high solar irradiation throughout the year, constant and intense winds, high altitudes in certain regions, and suitable average temperatures (ADECE, 2024). Therefore, to fully harness this potential, it is essential to identify the municipalities with the greatest potential for the installation of renewable energy systems.

Thus, the characteristics that impact power generation, such as solar irradiation, wind speed, altitude, and temperature, will be used in this study. Based on these, it will be possible to sort the cities according to their similar characteristics. These climatic and geographic parameters play a fundamental role in the analysis and assessment of the feasibility of renewable energy generation (Marimuthu, Kirubakaran, 2014).

Furthermore, the clustering method chosen for this study is the K-means algorithm, widely used for its simplicity, computational efficiency, and ability to handle large volumes of data (BERKHIN, 2006). This algorithm promotes a significant reduction in the volume of information to be analyzed, allowing for the classification and prioritization of the set of cities.

2. Methodology

2.1. Clusterization

Clustering is a grouping method belonging to the area of unsupervised machine learning. It is an unsupervised machine learning technique, meaning it is an artificial intelligence approach where algorithms are used to discover latent patterns or structures in unlabeled datasets. Unlike supervised learning, which uses pre-labeled examples, the unsupervised method requires algorithms to autonomously identify the underlying structure in the data without pre-defined labels or categories (HONDA, et al. 2017). In general, clustering aims to find a

10.18086/swc.2025.05.06 Available at http://proceedings.ises.org

more informative representation of a large amount of available data, making it possible to analyze group patterns through techniques that measure the similarity between the information. The groups generated by this type of classification are called clusters (BERKHIN, 2002).

2.2. K-means algorithm

The K-means algorithm gets its name because its main function is to find K distinct groups within a dataset. For each identified grouping, a centroid is assigned, which represents the center of the cluster. This centroid is calculated as the mean of the values of the elements belonging to that cluster. Consequently, the K-means approach allows for an efficient organization of data into coherent groups, in addition to facilitating the analysis and interpretation of patterns present in the datasets (HONDA, 2017). This methodology commonly uses the Sum of Squared Errors (SSE) metric to minimize the distance between data points and their centroids in the K-means algorithm. After the model is created, the means for each cluster are assigned, and the nearest centroid is identified for each new data point. These centroids represent the center of each cluster, determining the organization and categorization of the data (PIRES GUEDES, 2019), as can be seen in Figure 1. In a simplified manner, the K-means algorithm can be described by the following steps, according to Fontana and Naldi (2009):

(1) **Initialization:** Assignment of initial values for the cluster centroids. These can be chosen randomly within the domain limits of each attribute or by using specific methods.
(2) **Object Assignment to Clusters:** Each object is assigned to the group whose centroid has the highest similarity. This similarity can be measured using distance metrics, such as the Euclidean distance, which was used in the present work.
(3) **Centroid Update:** The centroid value for each cluster is recalculated as the mean of the objects belonging to that cluster.
(4) **Iteration:** Steps 2 and 3 are repeated until the groups stabilize, meaning when there are no more changes in the assignment of objects to clusters or in the centroid values.

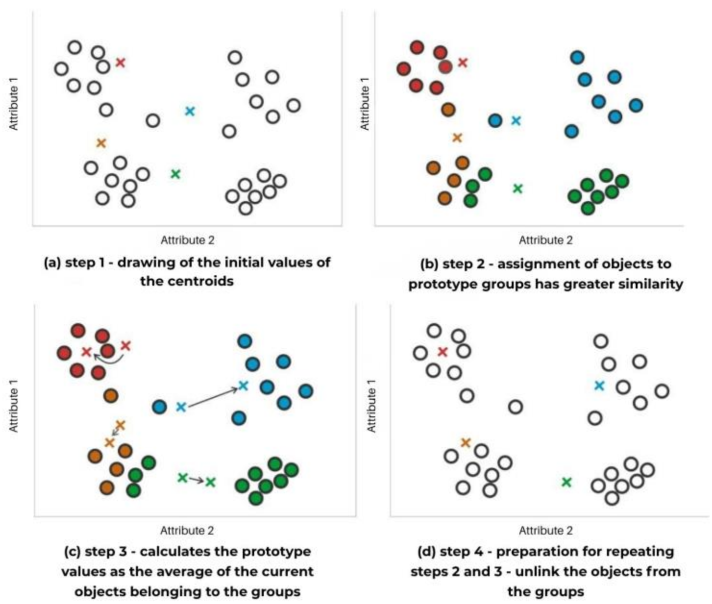

Figure 1: K-means algorithm step by step. Adapted from: Fontana, A., Naldi, M. C. 2009

Another important factor for determining the optimal number of clusters when running an algorithm is the use of the Elbow Curve method. This is a widely used technique for determining the ideal number of groups (K) in a dataset. This approach evaluates the variance of the data in relation to the number of clusters, identifying that the optimal value of K is the one that results in a lower Within Sum of Squares (WSS), while simultaneously aiming to minimize the number of clusters used. The name "Elbow Curve" is attributed to this method because the optimal value is found at the inflection point (the "elbow") of the curve, beyond which additional increases in the number of clusters do not result in significant gains in variance reduction (PIRES GUEDES, 2019), as observed in Figure 2.

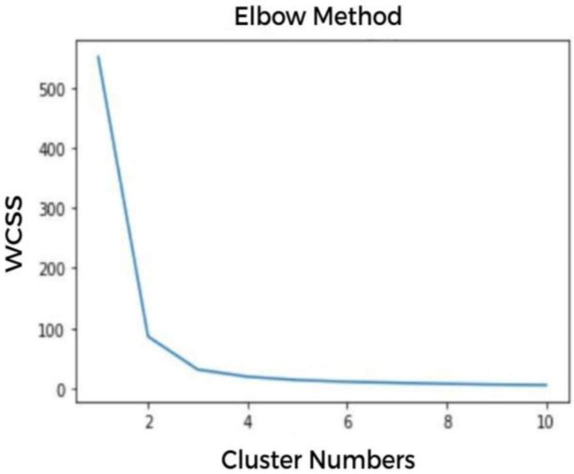

Figure 2: Representation of the elbow curve. Adapted from PIRES GUEDES, 2019.

```
Print (Irradiation_mun)
```

	Municipality	Irradiation_avg
0	Abaiara	5910.500000
1	Acarape	5508.000000
2	Acaraú	5640.800000
3	Acopiara	5746.761905
4	Aiuaba	5693.150000
..	...	...
176	Uruburetama	5353.000000
177	Uruoca	5545.000000
178	Varjota	5600.666667
179	Viçosa do ceará	5579.272727
180	Várzea Alegre	5888.714286

Figure 3: Preview of the processed annual irradiation dataframe.

2.3. Data evaluation set

To obtain the information required for this analysis, we considered key variables with potential influence on wind and photovoltaic energy generation, as well as the availability of public data. Additionally, additional data was incorporated as needed during the data transformation stage.

2.3.1 Solar Irradiation and Municipal Limits of Ceará

For this variable, we used a dataset containing information on solar irradiation at the annual and monthly average levels, both representing the total daily Global Horizontal Irradiation in the state of Ceará. Furthermore, the data were collected in shapefile format, which is an Esri vector data store widely used to store information on the position, shape, and attributes of geographic features.

2.3.2 Municipal Limits of Ceará

The dataset, available in shapefile format, contains a vector representation of the municipal boundaries of the state of Ceará. The Ceará Institute for Economic Research and Strategy (IPECE) released this information in 2021, which is essential for geographically delimiting the state's municipalities. The file was used to associate data of interest to their respective municipalities with the solar irradiation dataset, enabling georeferenced spatial and statistical analyses.

2.3.3 Temperature and Wind Speed

Historical data for these variables, aggregated by municipality, were obtained from the National Institute of Meteorology (INMET), covering the period between January and June 2023 and containing hourly information. Therefore, this data is relevant for cluster analysis because it contains information about temperature (°C) and wind speed ($m.s^{-1}$). However, for the state of Ceará, this dataset only contains information for a list of 16 municipalities: Fortaleza, Sobral, Guaramiranga, Barbalha, Iguatu, Tauá, Quixeramobim, Morada Nova, Jaguaruana, Crateús, Campos Sales, Jaguaribe, Itapipoca, Acaraú, Tianguá, and Quixadá. Then, values must be assigned to cities with missing information.

2.3.4 Altitudes of the Municipalities of Ceará

Through Google's Application Programming Interface (API) called Geocoding and using a file obtained from the "Municípios-Brasileiros" repository maintained by the user called "kelvins" on GitHub, containing the geographic coordinates (latitudes and longitudes) of all municipalities in Brazil, it was possible to extract elevation information for each municipality in the state of Ceará, which is the altitude data expressed in meters (m) above sea level, being a very impactful variable as it directly influences local conditions, affecting the generations highlighted in this work.

2.4. Data Transformation

Python (version 3.10.12) was used to process the data and implement the clustering method, with support from several libraries for data manipulation, calculation, and analysis (Pandas, NumPy), graph plotting (Matplotlib, Seaborn), and map visualization and manipulation (Folium). In particular, the sklearn.cluster library, which imports K-Means, was used to perform the final clustering of the data. Power BI was also used to generate additional maps and visualizations. Given the assumption that the clustering would be implemented at the city level in Ceará, it was first necessary to standardize all files to this format. Therefore, municipalities were initially associated with the irradiation shapefile, as it contained only latitude and longitude information. For this association, another dataset provided by IPECE was used, also in shapefile format, containing the state boundaries, to assign city names to each coordinate present in the initial database. Then, using the pandas library, the average annual irradiation per municipality was calculated, resulting in a data frame with 181 municipalities, as shown in Figure 3.

Considering that the state has 184 municipalities (IPECE, 2021), three cities with missing data were identified and designated using the haversine method in Python, which calculates the geodesic distance between two points based on their latitude and longitude. This procedure determined the three closest cities to each of these municipalities with missing data and then designated the arithmetic means of solar irradiation from three adjacent cities as the estimate to fill in the missing values for these cities. Furthermore, the INMET meteorological reference database only contains information for 16 municipalities in Ceará, which are distributed across all regions of the state, as shown in Figure 4.

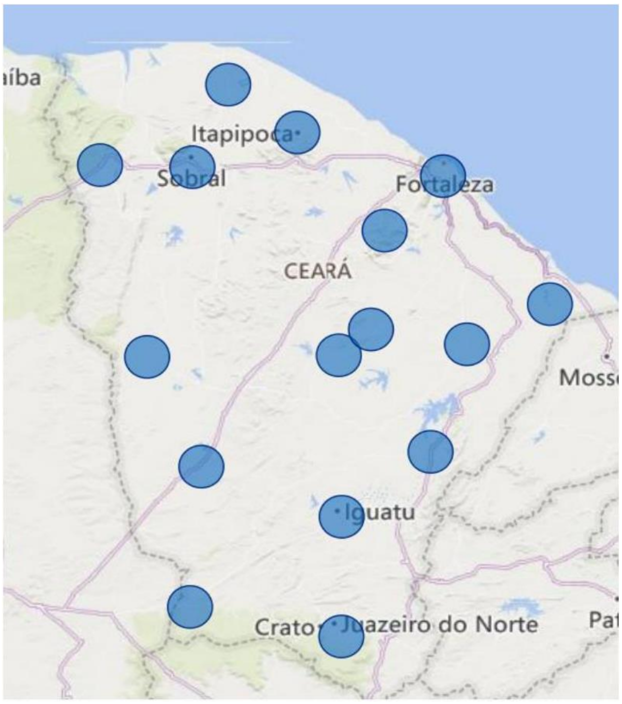

Figure 4: Map of municipalities with historical information at INMET.

Due to the information gaps, it was also necessary to apply the haversine method to estimate the average wind speed and average temperature values for municipalities without recorded data. To this end, the average value of the three cities closest to all the coordinates of the missing municipalities was calculated. Therefore, to obtain the final table required for the first stage of the analysis, the two processed datasets were merged. This merger involved integration using the linking key formed by the lowercase, unaccented names of the 184 municipalities, combining information on average annual irradiation and altitude from the first prepared database, along with the average wind speed and average temperature data obtained from the INMET database. Figure 5 represents the processing of the analyzed data.

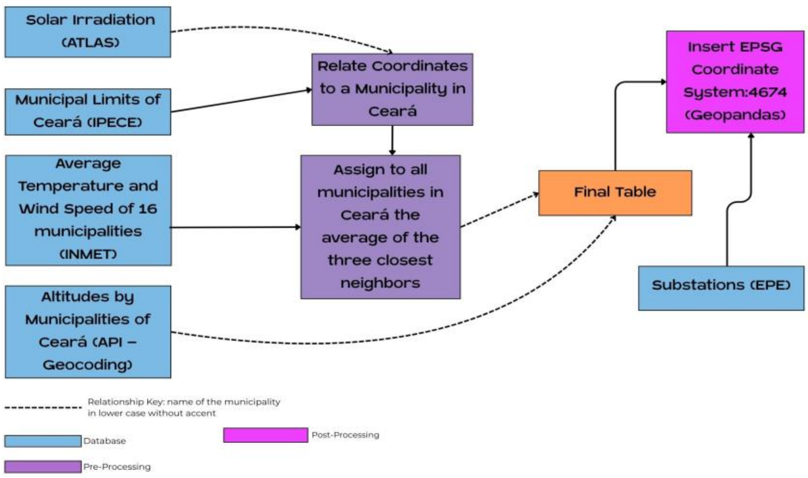

Figure 5: Data processing schematic.

2.5. Application of K-means

Using the unified and processed database, the K-means algorithm was executed, considering the four selected variables. The implementation was performed using the sklearn.cluster library in Python, following the steps described below:

(1) Data normalization: Initially, the data was normalized to adjust the attribute values to the same scale. This process is essential to ensure that all attributes have equal importance when calculating the distances between points, preventing attributes with larger scales from dominating the clustering process. Therefore, a standardization technique was adopted that converts the mean to zero and the standard deviation to one for all variables. This approach ensures that all variables have equal weight in calculating the Euclidean distances used by the K-means algorithm.

(2) Determining the optimal number of clusters: Using the elbow curve method, it was possible to identify the optimal number of clusters. This method involves analyzing the graph shown in Figure 6, which shows the relationship between the number of clusters and the variance explained by the clusters. The inflection points in the graph, known as the "elbow", represent the optimal number of clusters.

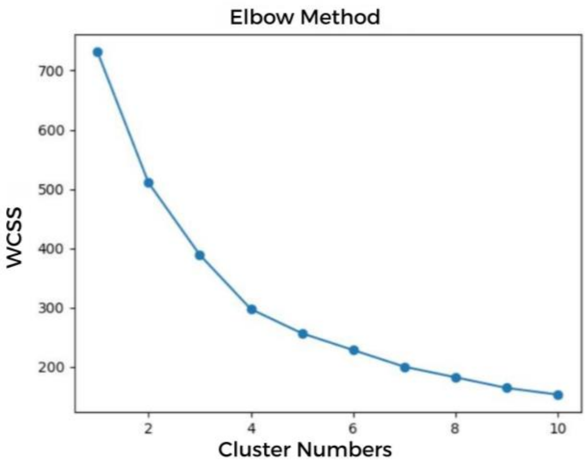

Figure 6: Elbow graph for clustering municipalities.

When analyzing the graph, it was observed that this inflection point, known as the "elbow", was identified at k = 4, indicating that dividing the data into four groups represents the most appropriate configuration for the set analyzed.

(3) Execution of the clustering algorithm: based on the number of clusters determined in the previous steps, the K-means algorithm was executed to group the municipalities according to the similarity of their attributes, taking into account the four variables analyzed.

(4) Cluster label assignment: At the end of the clustering process, the cluster_label label was added to the initial database, indicating which cluster each city belongs to. This information allows for subsequent analysis and interpretation of the results.

3. Results

3.1. Cluster analysis

By applying the clustering algorithm to the georeferenced data from Ceará, it was possible to identify the formation of four distinct groups of municipalities, characterized by their similarities in relation to the variables analyzed. Each group was constituted as follows: Cluster 0: 65 municipalities; Cluster 1: 35 municipalities; Cluster 2: 55 municipalities; Cluster 3: 29 municipalities.

According to the procedures described above, it was possible to create distinct groups of municipalities, characterized by similar patterns based on the variables analyzed, as shown in Figure 7, facilitating understanding and contributing to the understanding of the specificities of the locations and the formulation of more efficient strategies for the generation of renewable energy. In fact, Figure 8 provides a clearer understanding of the distribution of clusters across Ceará, allowing us to identify a dispersion of some clusters within the state. Cluster 3 displays remarkable uniformity compared to the others. This characteristic can be attributed to the influence of altitude, as the clusters exhibit more significant variation in this topological factor than in the other climatic factors mentioned above. This hypothesis is corroborated by the data presented in Table 1, which summarizes the averages of the variables for each cluster. It can be seen that the most divergent values between the clusters are related to altitude, reinforcing that this variable plays a significant and discriminating role in differentiating the clusters.

```
[66] base_final.head()
```

	Municipality	State	Latitude	Longitude	Altitude	Average_Wind_Speed	Average_Temperature	Key	Average_Irradiation	Cluster_Label
0	Abaiara	CE	-7.34588	-39.0416	386.262207	1.605062	25.078958	abaiara	5910.500000	2
1	Acaraú	CE	-2.88769	-40.1183	19.753691	1.783420	26.391786	acarau	5508.000000	0
2	Acarape	CE	-4.22083	-38.7055	95.597076	1.618619	25.025069	acarape	5640.800000	0
3	Acopiara	CE	-6.08911	-39.4480	313.874878	1.871038	26.854507	acopiara	5746.761905	3
4	Aiuaba	CE	-6.57122	-40.1178	467.503235	1.882526	25.511495	aiuaba	5693.150000	1

Figure 7: Clustered Database Preview.

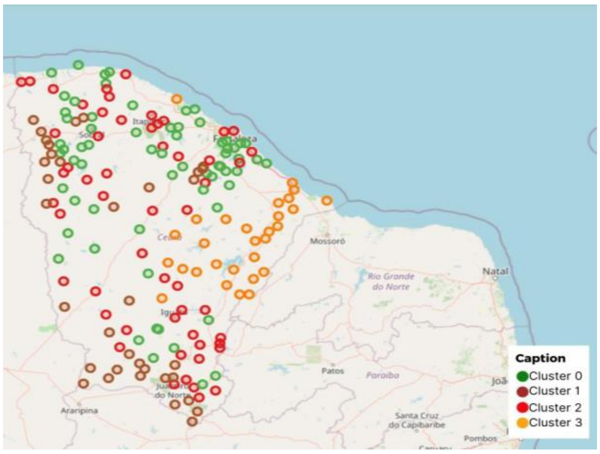

Figure 8: Map of clustered municipalities.

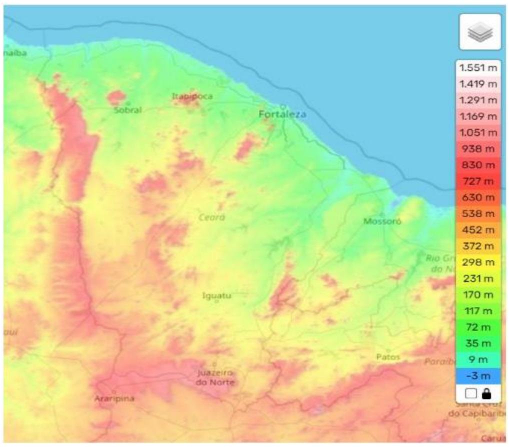

Figure 9: Topographic map of Ceará.

Furthermore, the elevation map available on the topographic-map website, shown in Figure 9, illustrates the variation in altitude throughout the state of Ceará. It can be seen that the area corresponding to the group 3 region presents a low variation in altitude, characterized by a relatively uniform topography.

Table 1: Average of variables by group.

Cluster	Avg. Altitude (m)	Avg. Irradiation (Whm^{-2})	T Avg. (°C)	Avg. Wind speed (ms^{-1})
0	136.52	5564.69	25.40	1.70
1	653.08	5675.31	25.14	1.65
2	214.39	5822.27	25.71	1.72
3	119.60	5686.03	27.23	2.10

With the data grouped and segmented according to the variables analyzed, Tables 2, 3, 4, and 5 present the descriptive statistics of each attribution used in the grouping. This information complements the other analyses and allows for a more precise and detailed characterization of the groups that were formed, complementing the understanding of the structural differences between the clusters and the factors that distinguish them.

Table 2: Altitude statistics by cluster.

Cluster	Average	Standard Deviation	Minimum	Median	Maximum
0	136.52	103.52	8.49	103.45	367.85
1	653.08	153.03	404.93	671.16	921.45
2	214.39	135.35	7.02	237.22	506.29
3	119.60	118.66	8.45	117.38	560.88

Table 3: Irradiation statistics by cluster.

Cluster	Average	Standard Deviation	Minimum	Median	Maximum
0	5564.69	86.97	5353.00	5562.00	5715.80
1	5675.31	135.40	5433.00	5648.73	5935.33
2	5822.27	78.98	5662.00	5822.50	5995.86
3	5686.03	147.84	5408.00	5649.69	5980.25

Table 4: Temperature statistics by cluster.

Cluster	Average	Standard Deviation	Minimum	Median	Maximum
0	25.40	0.58	24.88	25.09	26.59
1	25.14	0.29	24.93	25.08	26.59
2	25.71	0.70	24.79	25.51	27.05
3	27.23	0.26	26.61	27.18	27.58

3.2. Cluster analysis

3.2.1 Wind Generation

Based on the analysis of the variables during the clustering process, it was identified that certain aspects have a positive impact on wind generation. These characteristics include high temperatures, higher wind speeds,

Table 5: Average wind speed statistics by cluster.

Cluster	Average	Standard Deviation	Minimum	Median	Maximum
0	1.70	0.13	1.19	1.73	1.90
1	1.65	0.09	1.55	1.61	1.80
2	1.72	0.16	1.19	1.75	2.10
3	2.10	0.12	1.78	2.12	2.24

and higher altitudes (Marimuthu, Kirubakaran, 2014). Given the above, a comparative scoring methodology was developed, using these three factors as indicators to determine the most favorable group for wind generation. Therefore, the following procedures were adopted:

(1) Assigning scores by variable: Using the means calculated for each variable in each group, a score from 1 to 4 was assigned to each variable (mean altitude, mean wind speed, and mean temperature). A score of 1 represents the minimum value, and a score of 4 represents the maximum value. For example, as shown in Table 6, group 3 obtained the highest mean altitude value, resulting in a score of 4, indicating its positive influence on wind generation.

(2) Calculation of scores: Finally, the three scores of the variables assigned to each cluster were added, resulting in a final score, called "Total score" in Table 6. This index compares the potential of each cluster for wind generation, making it possible to make a hierarchical classification of the groups for the conditions analyzed.

Table 6: Groups classified for wind generation.

Cluster	Avg. altitude. (m)	Avg. wind speed ($m\ s^{-1}$)	Temp. (°C)	Altitude grade	Avg. wind speed grade	Temp. grade	Total grade
2	214.39	1.72	25.71	3	3	3	9
3	119.60	2.10	27.23	4	4	1	9
0	136.52	1.70	25.40	2	2	2	6
1	653.08	1.65	25.14	1	1	4	6

Based on Table 6, we can understand that there was a tie in the total scores between groups 2 and 3. However, considering the degree of impact on wind generation efficiency, it is preferable to prioritize group 3, as it stands out for its average wind speed and temperature values, with all group members presenting above-average values for these variables. Regarding altitude, group 3 has a lower average, with 93% of its members having altitudes below the overall average of 255.39 meters. Therefore, for these reasons and based on the analysis of Table 6, cluster 3 was chosen as the group most conducive to wind energy generation. The cities that make up cluster 3 are: Acopiara, Alto Santo, Aracati, Banabuiú, Fortim, Ibicuitinga, Icapuí, Iracema, Itaiçaba, Jaguaretama, Jaguaribara, Jaguaribe, Jaguaruana, Limoeiro do Norte, Milhã, Morada Nova, Palhano, Pereiro, Potiretama, Quixadá, Quixeramobim, Quixeré, Russas, São João do Jaguaribe, Senador Pompeu, Solonópole, Tabuleiro do Norte, Trairi, and Ererê.

In addition to the classification obtained, Figure 10 also shows the EPE WebMap of existing wind farm projects in Ceará, made available by the Energy Research Company (EPE). Currently, it was found that there is a greater concentration of these projects on the Ceará coast, in contrast to the area encompassing cluster three, identified in this study as more favorable to wind generation based on the variables analyzed. This indicates the need to redirect efforts to relocate or install more wind farms to regions with more favorable conditions.

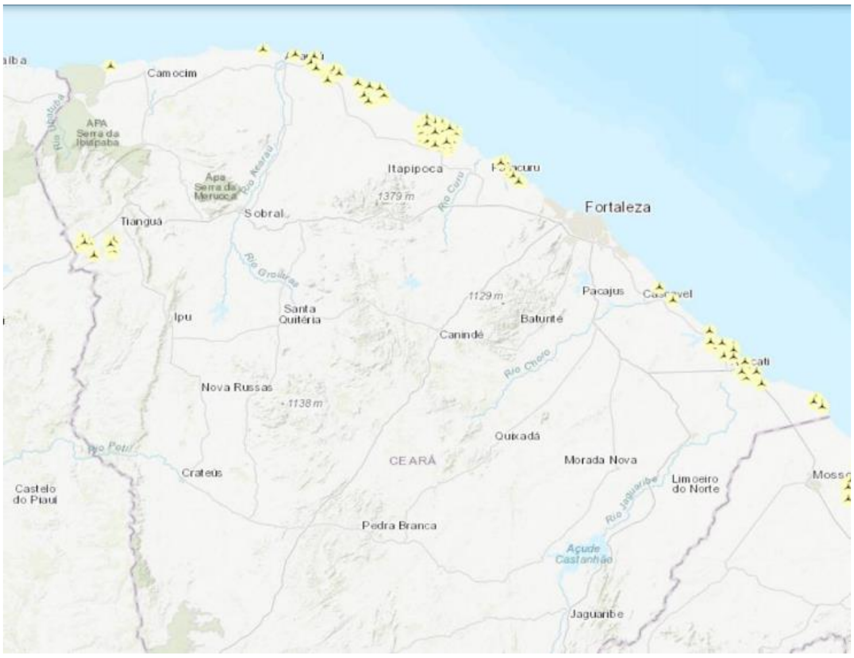

Figure 10: Wind power generation projects in Ceará.

3.2.2 Photovoltaic Generation

Based on the analysis of the variables during the clustering process, it was identified that certain aspects have a positive impact on solar generation. These characteristics include low temperatures, high levels of solar radiation, and higher altitudes (Marimuthu, Kirubakaran, 2014). Given the above, a comparative scoring methodology was developed, using these three factors as indicators to determine the most favorable group for solar generation. Therefore, the following procedures were adopted:

(1) Assigning scores by variable: Using the means calculated for each variable in each group, a score from 1 to 4 was assigned to each variable (mean altitude, mean irradiation, and mean temperature). A value of 1 represents the minimum value and 4 the maximum value. For example, as shown in Table 7, group 1 had the lowest mean temperature value, resulting in a score of 4, indicating its positive influence on solar generation.

(2) Calculation of scores: Finally, the three scores of the variables assigned to each cluster were added, resulting in a final score, called "Total grade" in Table 7. This index compares the potential of each cluster for solar generation, enabling a hierarchical classification of the groups for the conditions analyzed.

Table 7: Groups classified for photovoltaic generation.

Cluster	Avg. altitude. (m)	Avg. Irradiation (Wh/m²)	Temp. (°C)	Altitude grade	Avg. Irradiation grade	Temp. grade	Total grade
1	653.08	5675.31	25.14	4	2	4	10
2	214.39	5822.27	25.71	3	4	2	9
0	136.52	5564.69	25.40	2	1	3	6
3	119.60	5686.03	27.23	1	3	1	5

Analyzing Table 4, it is noted that cluster 1 stands out for its higher altitudes compared to the other groups, presenting a minimum altitude of 404.93 m, while in the other groups, it is lower than 8.5 m. Regarding temperature, the average temperature is 25.14 degrees Celsius (°C). Approximately 43% of the municipalities in this group have solar irradiation greater than the general average of 5681.85 Wh m^{-2}. Therefore, for these reasons and the analysis of Table 7, cluster 1 was chosen as the group most conducive to solar energy generation. The cities that make up cluster 1 are: Aiuaba, Alcântaras, Altaneira, Araripe, Assaré, Aratuba, Campos Sales, Caririaçu, Carnaubal, Catarina, Crato, Croatá, Guaraciaba do Norte, Ibiapina, Itatira, Guaramiranga, Jardim, Jati, Juazeiro do Norte, Meruoca, Monsenhor Tabosa, Nova Olinda, Parambu, Penaforte, Poranga, Porteiras, Potengi, Salitre, Mulungu, Santana do Cariri, São Benedito, Pacoti, Tianguá, Ubajara, and Viçosa do Ceará.

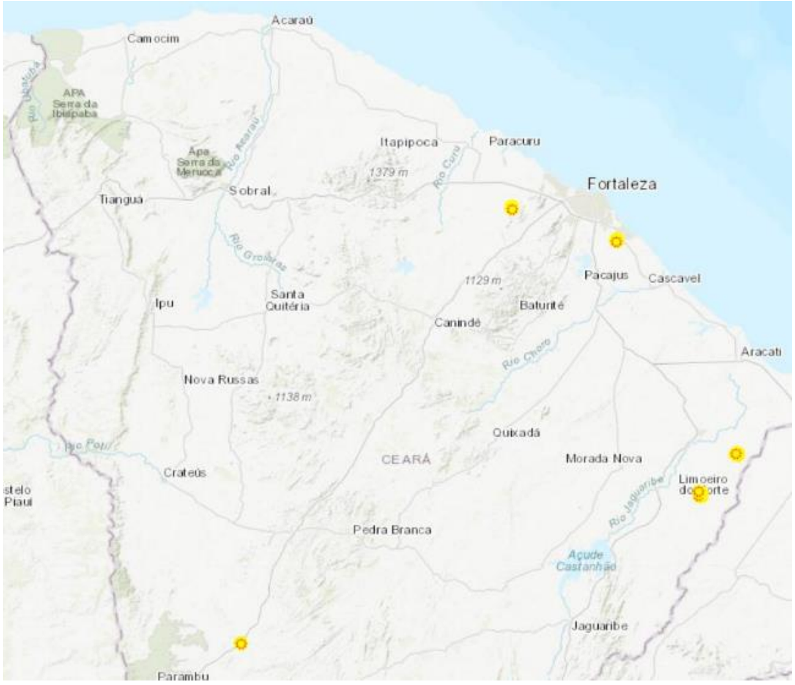

Figure 11: Photovoltaic generation projects in Ceará.

In addition to the classification obtained, Figure 11 also shows the EPE WebMap of existing photovoltaic projects in Ceará, made available by the Energy Research Company (EPE). Currently, it was found that there is a greater concentration of these projects, in contrast to the area encompassing cluster one, identified in this study as more favorable to solar generation based on the variables analyzed. This indicates the need to redirect efforts toward relocating or installing more photovoltaic parks to regions with more favorable conditions for better use of this resource.

3.2. Analysis of the Geolocation of Electrical Substations in Ceará in Relation to Areas Suitable for Generation

At this stage of the study, distances were observed to assess the proximity of the selected municipalities to the nearest substations. This will allow us to obtain a comprehensive view of the energy flow structure in these areas and assess the efficiency and potential for expansion of renewable generation in the state of Ceará. Analyzing Figure 12, it is clear that the substations are concentrated mainly in the coastal region of the state of Ceará, totaling 42 substations in operation in the state. Thus, this geographic distribution of substations suggests a greater energy flow infrastructure along the Ceará coast, providing more direct connectivity with renewable energy generation areas, such as wind and photovoltaic power, which are often located in regions close to the coast. This concentration of substations in this region may have implications for the efficiency of the transmission system and the potential for integrating energy generated in these generation areas into the distribution grid.

Figure 12: Map of Ceará substations.

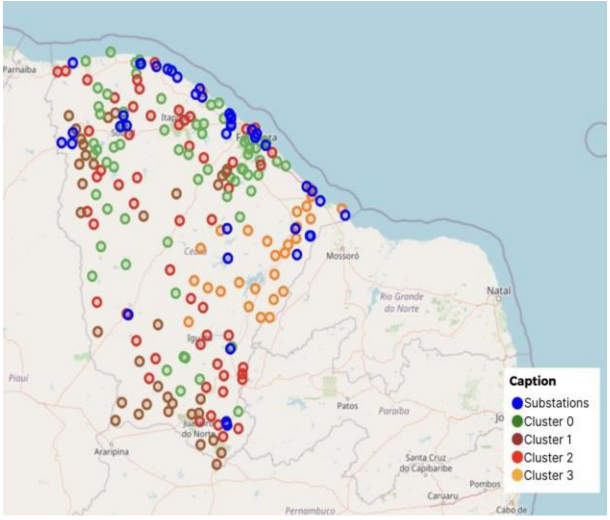

Figure 13: Map of Ceará substations.

According to Figure 13, it can be observed that the clusters with the largest number of municipalities located near the substations correspond to groups 0 and 2, that is, considering that the clusters selected based on the analyses carried out previously are group 1 and group 3 for solar and wind energy respectively, it can be observed that the locations of the substations are not in accordance with the municipalities that have the greatest potential for the generation of solar and wind energy, which highlights not only a logistical disadvantage for the flow of the energy produced, but also a loss of strategic potential for the state's energy planning.

4. Conclusion

Therefore, it can be concluded that the cluster analysis of Ceará's municipalities identified four distinct groups with specific characteristics of altitude, temperature, solar irradiation, and wind speed. Group 3 showed greater potential for wind generation, while Group 1 showed greater potential for photovoltaic generation, although existing projects are not always located in these more favorable areas. Regarding the proximity of substations, it was observed that the groups with the greatest renewable potential are not necessarily close to them, highlighting the need to expand the transmission and distribution system. Despite this, Group 3, being more compact, demonstrated greater ease in integrating cities near substations, which may be a strategic advantage. The selection of the best groups is not definitive, as it may vary depending on the variables analyzed, making the study dynamic. In summary, the study achieved its objective of characterizing Ceará's municipalities in detail, providing support for energy planning and directing public policies aimed at expanding renewable sources. It should also be noted that the selection of the most suitable groups should not be definitive, especially given the tie observed for wind generation, since different variables can alter prioritization. Future work recommends integrating climatic, geographic, socioeconomic, and population data to make the analyses more comprehensive and contribute to better utilization of the state's energy resources.

5. Acknowledgments

The authors would like to thank the Federal University of Ceará (UFC) for its institutional and scientific support, which made this research possible. The infrastructure provided, as well as the academic and collaborative environment, were essential for the execution of the activities and for the consolidation of the results presented here.

6. References

ADECE – Agência de Desenvolvimento do Estado do Ceará, 2024. https://www.adece.ce.gov.br/2024/04/10/ceara-apresenta-potencialidades-para-a-geracao-de-energias-renovaveis-no-intersolar-summit-brasil-nordeste/. Accessed on: August 18, 2023.

BERKHIN, Pavel. A survey of clustering data mining techniques. In: Grouping multidimensional data: Recent advances in clustering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. 25-71.

Empresa de Pesquisa Energética. WebMap EPE. EPE, [s.d]. available at: https://gisepeprd2.epe.gov.br/WebMapEPE/. Accessed on: August 18, 2023.

Fontana, A., Naldi, M. C. Estudo de Comparação de Métodos para Estimação de Números de Grupos em Problemas de Agrupamento de Dados. 2009. Universidade de São Paulo. ISSN - 0103-2569

GOOGLE, 2025. Geocoding API. Google Developers. available at: https://developers.google.com/maps/documentation/geocoding/overview. Accessed on: August 18, 2023.

HONDA, Hugo; FACURE, Mateus; YAOHAO, Peng, 2017. Os Três Tipos de Aprendizado de Máquina. available at: https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/. Accessed on: August 15, 2023.

HONDA, Hugo. Introdução básica à clusterização. LAMFO – Laboratório de Aprendizado de Máquina em Finanças e Organizações, Brasília, 5 out. 2017. available at: https://lamfo-unb.github.io/2017/10/05/Introducao_basica_a_clusterizacao/. Accessed on: August 10, 2023.

INMET – Instituto Nacional de Meteorologia. BDMEP – Banco de Dados Meteorológicos para Ensino e Pesquisa. available at: https://bdmep.inmet.gov.br/. Accessed on: August 17, 2023.

IPECE - Instituto de Pesquisa e Estratégia Econômica do Ceará. Ceará ocupa o 13º lugar no ranking energético nacional e tem como principais matrizes de energia a termelétrica e a eólica. 2018. available at: https://www.ipece.ce.gov.br/2018/12/12/ceara-ocupa-o-13o-lugar-no-ranking-energetico-nacional-e-tem-como-principais-matrizes-de-energia-a-termeletrica-e-a-eolica/. Accessed on: August 17, 2023.

KELVINS, 2023. Municípios-Brasileiros. GitHub. available at: https://github.com/kelvins/municipios-brasileiros. Accessed on: August 10, 2023.

Marimuthu, C., Kirubakaran, V., 2014. Wind and Solar Energy Applications. Springer.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, v. 12, p. 2825-2830, 2011.

PIRES GUEDES, Erivelton. Clusterização (k-means), 2019. available at: http://www.kaggle.com/code/eriveltonguedes/7-clusteriza-o-k-means-erivelton. Accessed on: August 14, 2023.

Topographic-Map, 2025. Mapa topográfico Ceará, altitude, relevo. available at: https://pt-br.topographic-map.com/map-tgxm2/Cear%C3%A1/. Accessed on: August 13, 2023.