



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA - DEE
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WESLEY BEZERRA NOGUEIRA

**AVALIAÇÃO DE DESEMPENHO DE UMA USINA SOLAR FOTOVOLTAICA COM
BASE NO AGRUPAMENTO K-MEANS**

FORTALEZA

2022

WESLEY BEZERRA NOGUEIRA

AVALIAÇÃO DE DESEMPENHO DE UMA USINA SOLAR FOTOVOLTAICA COM BASE
NO AGRUPAMENTO K-MEANS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Orientadora: Prof^ª. Ph.D. Ruth P.S. Leão

Coorientadora: Eng. Danielle Baltazar
Cavalcante

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

N716a Nogueira, Wesley Bezerra.
Avaliação de desempenho de uma usina solar fotovoltaica com base no agrupamento k-means / Wesley Bezerra Nogueira. – 2022.
57 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Elétrica, Fortaleza, 2022.

Orientação: Profa. Dra. Ruth Pastôra Saraiva Leão.

Coorientação: Profa. Danielle Baltazar Cavalcante.

1. Desempenho de Usinas Fotovoltaicas. 2. Clusterização de Inversores. 3. K-means. 4. Sistemas Fotovoltaicos. I. Título.

CDD 621.3

WESLEY BEZERRA NOGUEIRA

AVALIAÇÃO DE DESEMPENHO DE UMA USINA SOLAR FOTOVOLTAICA COM BASE
NO AGRUPAMENTO K-MEANS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Aprovada em:

BANCA EXAMINADORA

Prof^ª. Ph.D. Ruth P.S. Leão (Orientadora)
Universidade Federal do Ceará (UFC)

Eng. Danielle Baltazar Cavalcante (Coorientadora)
Universidade Federal do Ceará (UFC)

Eng. Mestrando André Wagner de Barros Silva
Universidade Federal do Ceará (UFC)

Eng. Jefferson Neri de Melo
Universidade de Fortaleza (UNIFOR)

A todos aqueles que, pela Graça de Deus, fizeram, fazem e farão parte da minha vida.

AGRADECIMENTOS

Primeiramente a Deus, pelo dom da minha vida e por me permitir ultrapassar as dificuldades encontradas ao longo da realização deste trabalho, mas que, em todo e qualquer momento é o maior amigo que alguém pode ter.

Aos meus familiares (pais, avós, irmão, tios, primos...), por todo o apoio e sustento que têm em minha vida, transmitindo-me amor e palavras de sabedoria.

À professora Ruth Pastôra Saraiva Leão, por toda a orientação, pois desempenhou tal função com dedicação, zelo e amizade.

Aos engenheiros André Wagner e Danielle Baltazar, por todo o apoio dado durante o desenvolvimento deste trabalho.

Aos meus colegas (graduação, PET, faculdade), com quem convivi, intensamente, durante vários momentos, pelo companheirismo nas situações de felicidades e de adversidades.

Em especial, aos meus amigos/irmãos Bruno Faustino, Nathanael Vasconcelos, Marcos Felipe, Milena Kevely e Saynarah Cruz, por se fazerem verdadeiros pilares durante esta caminhada.

Aos meus amigos Cássio Lopes, Eric Saunders, Felipe Tabosa, Giovana Bruna, Giovanni Eduardo, João Manuel, Jonathan Fernandes, Nicolas Matos, Pedro Neto e Vinicius Passos por todos os momentos vividos fora do ambiente institucional, ajudando a tornar os meus dias cada vez melhores.

A todos os professores, pelos conhecimentos técnicos, mas, sobretudo, pelo auxílio na formação do meu caráter.

À esta instituição de ensino, essencial no meu processo de formação, pela dedicação, e por tudo o que aprendi ao longo dos anos do curso.

“Comece fazendo o que é necessário, depois o que é possível, e de repente você estará fazendo o impossível.”

(São Francisco de Assis)

RESUMO

Em 2021, o Brasil alcançou, segundo a Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR), a marca de 13 GW de potência operacional em usinas solares fotovoltaicas de grande porte, bem como em pequenos e médios sistemas, correspondendo a um valor acima de 70% da potência da usina hidrelétrica de Itaipu. Com isso, o país é o único latino-americano no Top 15 da classificação mundial de capacidade instalada de energia solar elaborada pela Agência Internacional para as Energias Renováveis (IRENA). Por esse motivo, faz-se importante analisar o uso dos recursos energéticos, buscando entender o comportamento e o desempenho de sistemas fotovoltaicos, identificando possíveis irregularidades na laboração. Portanto, neste trabalho, será implementada um método de data mining, aplicado em clusterização de dados, conhecido como K-means. Tal método objetiva agrupar 36 estações de transformadores, pertencentes a 4 usinas fotovoltaicas, em diferentes conjuntos, de tal forma que os componentes de um mesmo grupo apresentem características semelhantes. O algoritmo de clusterização K-means foi desenvolvido em linguagem de programação Python, usando as variáveis de potência dos inversores e de irradiação solar como entrada de dados. Os dados de entrada foram extraídos do Sistema de Supervisão e Aquisição de Dados (SCADA) e discretizados em intervalos de 60 minutos. Tal metodologia permite analisar o desempenho dos sistemas por meio de grupos, em detrimento de análises individuais, amenizando o esforço para acompanhar a operação de usinas fotovoltaicas com grande quantidade de inversores. Os resultados obtidos demonstram que inversores com especificações semelhantes entre si, dispostos em um mesmo sistema fotovoltaico, podem apresentar comportamentos divergentes. Em grupos onde o desempenho dos inversores é inferior aos demais, é necessária a realização de um estudo para compreender os possíveis motivos que resultaram em tais deficiências.

Palavras-chave: Desempenho de Usinas Fotovoltaicas. Clusterização de Inversores. K-means. Sistemas Fotovoltaicos.

ABSTRACT

In 2021, Brazil reached, according to the Brazilian Association of Photovoltaic Solar Energy (ABSOLAR), the mark of 10 GW of operational power in large-scale solar photovoltaic plants, as well as in small and medium-sized systems, corresponding to a value above 70% of the power of the Itaipu hydroelectric plant. With that, the country is the only Latin American in the Top 15 of the world ranking of installed solar energy capacity prepared by the International Renewable Energy Agency (IRENA). For this reason, it is important to analyze the use of energy resources, seeking to understand the behavior and performance of photovoltaic systems, identifying possible irregularities in the work. Therefore, in this work, a data mining method will be implemented, applied in data clustering, known as K-means. This method aims to group 36 transformer stations, belonging to 4 photovoltaic plants, in different sets, in such a way that the components of the same group have similar characteristics. The K-means clustering algorithm was developed in Python programming language, using the inverter power and solar irradiance variables as data input. The input data were extracted from the Supervisory Control and Data Acquisition (SCADA) and discretized at 60-minute intervals. This methodology allows analyzing the performance of systems through groups, rather than individual analyses, easing the effort to monitor the operation of photovoltaic plants with a large number of inverters. The results obtained demonstrate that inverters with similar specifications, arranged in the same photovoltaic system, can present divergent behaviors. In groups where the performance of inverters is inferior to the others, it is necessary to carry out a study to understand the possible reasons that resulted in such deficiencies.

Keywords: Performance of Photovoltaic Plants. Cluster of Inverters. K-means. Photovoltaic Systems.

LISTA DE FIGURAS

Figura 1 – Efeito Fotoelétrico	18
Figura 2 – Níveis de energia em isolantes (a), em condutores (b) e em semicondutores (c)	20
Figura 3 – Dopagem da rede cristalina de átomos de silício (Si) com fósforo (P) e com boro (B)	21
Figura 4 – Esquemático de uma junção P-N	21
Figura 5 – Esquemático das camadas de uma célula fotovoltaica	23
Figura 6 – Modelo do circuito elétrico equivalente de uma célula fotovoltaica	24
Figura 7 – Curvas características de corrente versus tensão (a) e de potência versus tensão (b) para um painel fotovoltaico em condições padrões de temperatura e irradiância	25
Figura 8 – Comportamento da curva de corrente versus tensão de um painel fotovoltaico diante da variação de temperatura	26
Figura 9 – Comportamento da curva de corrente versus tensão de um painel fotovoltaico diante da variação dos níveis de irradiância	26
Figura 10 – Esquemático dos fluxos de calor sobre um módulo fotovoltaico	29
Figura 11 – Esquemático do diagrama unifilar de um sistema fotovoltaico conectado à rede elétrica	30
Figura 12 – Esquemático do diagrama unifilar de um sistema fotovoltaico isolado da rede elétrica	31
Figura 13 – Estrutura de organização dos circuitos nas usinas fotovoltaicas	33
Figura 14 – Esquemático da vista superior do complexo fotovoltaico	33
Figura 15 – Esquemático para diferentes técnicas de clusterização	35
Figura 16 – Exemplo ilustrativo de um dendograma	35
Figura 17 – Determinação do valor ótimo de clusters em um dendograma	36
Figura 18 – Esquemático da clusterização de um conjunto de dados brutos em: dois <i>clusters</i> , três <i>clusters</i> e cinco <i>clusters</i>	36
Figura 19 – Fluxograma representativo do algoritmo K-means	38
Figura 20 – Processo de implementação do algoritmo K-means para um conjunto de dados de entrada (a), com determinação do número de clusters e centroides iniciais (b), realizando-se a segunda iteração (c), a terceira iteração (d) e a última iteração (e)	39

Figura 21 – Curva do método Elbow	40
Figura 22 – Escolha do valor ótimo de K pelo método Elbow	40
Figura 23 – Fluxograma simplificado do código python	41
Figura 24 – Atributos da base de dados	42
Figura 25 – Método Elbow: fevereiro 2020	43
Figura 26 – Método Elbow: maio 2020	44
Figura 27 – Método Elbow: agosto 2020	44
Figura 28 – Método Elbow: novembro 2020	44
Figura 29 – Dados irradiância x potência: fevereiro 2020	45
Figura 30 – <i>Clusters</i> irradiância x potência: fevereiro 2020	45
Figura 31 – Dados irradiância x potência: maio 2020	46
Figura 32 – <i>Clusters</i> irradiância x potência: maio 2020	46
Figura 33 – Dados irradiância x potência: agosto 2020	46
Figura 34 – <i>Clusters</i> irradiância x potência: agosto 2020	47
Figura 35 – Dados irradiância x potência: novembro 2020	47
Figura 36 – <i>Clusters</i> irradiância x potência: novembro 2020	47
Figura 37 – Potência e irradiância ao longo do dia para o <i>cluster</i> 3 do mês de fevereiro	50
Figura 38 – Potência e irradiância ao longo do dia para o <i>cluster</i> 0 do mês de maio	50
Figura 39 – Potência e irradiância ao longo do dia para o <i>cluster</i> 0 do mês de agosto	51
Figura 40 – Potência e irradiância ao longo do dia para o <i>cluster</i> 3 do mês de novembro	51

LISTA DE TABELAS

Tabela 1 – Funções trabalho de alguns materiais	18
Tabela 2 – Distribuição das TS nas usinas fotovoltaicas	32
Tabela 3 – Potência total instalada do complexo fotovoltaico	33
Tabela 4 – Localização das estações meteorológicas (WS)	34
Tabela 5 – Atribuição das estações de transformadores (TS) nas estações meteorológicas (WS)	34
Tabela 6 – Centroides: fevereiro 2020	48
Tabela 7 – Centroides: maio 2020	48
Tabela 8 – Centroides: agosto 2020	48
Tabela 9 – Centroides: novembro 2020	48
Tabela 10 – <i>Clusters</i> : fevereiro 2020	49
Tabela 11 – <i>Clusters</i> : maio 2020	49
Tabela 12 – <i>Clusters</i> : agosto 2020	49
Tabela 13 – <i>Clusters</i> : novembro 2020	49
Tabela 14 – Dados de referência para curvas de comportamento dos <i>clusters</i>	50

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	16
1.2	Estrutura do Trabalho	16
2	FUNDAMENTOS DA GERAÇÃO FOTOVOLTAICA	17
2.1	Princípios Energéticos	17
<i>2.1.1</i>	<i>Natureza da Luz</i>	17
<i>2.1.2</i>	<i>Teoria dos Semicondutores</i>	19
<i>2.1.2.1</i>	<i>Materiais Intrínsecos</i>	19
<i>2.1.2.2</i>	<i>Níveis de Energia</i>	19
<i>2.1.2.3</i>	<i>Materiais Extrínsecos</i>	20
<i>2.1.2.4</i>	<i>Junção P-N</i>	21
2.2	Sistemas Fotovoltaicos	22
<i>2.2.1</i>	<i>Inversores de Frequência</i>	22
<i>2.2.2</i>	<i>Células Fotovoltaicas</i>	23
<i>2.2.2.1</i>	<i>Modelagem</i>	23
<i>2.2.2.2</i>	<i>Curvas Características</i>	24
<i>2.2.2.2.1</i>	<i>Fator de Forma</i>	26
<i>2.2.2.2.2</i>	<i>Rendimento</i>	27
<i>2.2.2.3</i>	<i>Perdas</i>	27
<i>2.2.2.3.1</i>	<i>Balanco Energético em uma Célula Fotovoltaica</i>	28
<i>2.2.3</i>	<i>Tipos de Sistemas Fotovoltaicos</i>	30
3	METODOLOGIA	32
3.1	Características do Complexo Fotovoltaico	32
3.2	Fundamentos sobre Clusterização	35
<i>3.2.1</i>	<i>Metodologia Hierárquica</i>	35
<i>3.2.2</i>	<i>Metodologia Particional</i>	36
3.3	Técnica K-means	37
<i>3.3.1</i>	<i>Método Elbow</i>	39
<i>3.3.2</i>	<i>Algoritmo Python</i>	40
3.4	Base de Dados	41

4	ANÁLISE DOS RESULTADOS	43
4.1	Determinação do Valor Ótimo de K	43
4.2	Clusterização: Irradiância x Potência	45
4.3	Análise de Performance	49
5	CONCLUSÕES E TRABALHOS FUTUROS	52
5.1	Conclusão	52
5.2	Trabalhos Futuros	53
	REFERÊNCIAS	54
	APÊNDICES	56
	APÊNDICE A – Código Python K-Means	56
	APÊNDICE B – Código Python Elbow	58

1 INTRODUÇÃO

É fato que as características da matriz elétrica brasileira estão mudando, principalmente, pelo aumento da participação de energias renováveis, com destaque para a geração eólioeleétrica e a geração fotovoltaica. Além disso, o avanço das tecnologias, relacionadas à era da informação, possibilita análises de dados mais lapidadas e, desta forma, torna-se possível elaborar estratégias que otimizem a operação e os processos de manutenção dos sistemas. Nesse contexto, a organização destas estratégias é uma ferramenta que pode auxiliar no desenvolvimento de medidas que visam melhorar o desempenho de usinas fotovoltaicas.

Para a obtenção dos dados a serem analisados, normalmente, utiliza-se o Sistema de Supervisão e Aquisição de Dados (SCADA - *Supervisory Control and Data Acquisition*). Tais informações serão necessárias para o estabelecimento dos parâmetros de entrada dos algoritmos de análise de dados. Ao analisar um conjunto de dados, agrupá-los é uma excelente estratégia para amenizar a complexidade do estudo, possibilitando um melhor entendimento do comportamento dinâmico e da relação existente entre tais dados. Por isso, a análise de desempenho de usinas fotovoltaicas pode ser feita por intermédio de técnicas de clusterização, sendo possível adotar diversas metodologias existentes.

Em (PARK; PARK, 2018), o autor utiliza métodos de clusterização para estudar sistemas fotovoltaicos sem a necessidade de grandes simulações e análises, com a aplicação de algoritmos hierárquicos, visando perceber a relação existente entre os agrupamentos. Tais semelhanças podem ser percebidas por meio dos dendogramas formados. Como resultado do estudo, foi possível classificar os agrupamentos como semelhantes e defeituosos. Além disso, o autor cita a necessidade de que os elementos estejam sob as mesmas condições, destacando que diferentes condições meteorológicas, a exemplo de temperatura e pluviosidade, podem alterar os resultados.

Nota-se que, em (CAVALCANTE, 2020), o autor realiza o estudo de um complexo de aerogeradores por meio de clusterização. Aplicando-se a estratégia conhecida como K-means, foi possível analisar o desempenho de 65 turbinas eólicas, onde o comportamento dos centroides de cada *cluster* foi comparado com a performance esperada diante da curva de potência dada pelos fabricantes. Com isso, percebeu-se que aerogeradores pertencentes a um mesmo parque eólico podem apresentar desempenhos diferentes.

A complexidade dos sistemas elétricos exige, com o avanço das novas tecnologias, um gerenciamento eficaz e eficiente da qualidade de geração e distribuição de energia elétrica.

É nesse contexto que, em (FENSTERSEIFER, 2008), o autor utiliza métodos de clusterização, por meio dos algoritmos K-Means e DBS-CAN, para analisar agrupamentos de consumidores. Tal estudo permite uma avaliação concisa, identificando consumidores que podem afetar, negativamente, a qualidade de energia.

O estudo do padrão de consumo de energia elétrica é uma ação de extrema importância que visa atender, da melhor forma possível, às necessidades dos consumidores. É nesse contexto que, em (CLAUDIO, 2018), o autor utiliza da estratégia de clusterização, implementada por intermédio do algoritmo K-means, para realizar uma análise dos perfis de consumidores, caracterizados, principalmente, pelos dados de consumo ao longo das diferentes estações do ano. Desta forma, foi possível dimensionar sistemas fotovoltaicos, comparando-se as características dos diferentes grupos de consumidores formados.

Em (CHRISTOFORIDIS *et al.*, 2013), o autor faz uso de técnicas de clusterização em uma instalação fotovoltaica de 20 kWp localizada na Grécia. Agrupando-se os dados de produção, foi gerada uma curva representativa de potência, denominada centroide. Por meio da utilização de vários métodos de clusterização, o autor fez a comparação dos centroides obtidos em cada um deles e, com isso, foi possível analisar o comportamento da instalação, objetivando o estabelecimento de estratégias que otimizem o armazenamento de energia em sistemas fotovoltaicos.

Neste trabalho, aplicou-se, por meio de um algoritmo em linguagem Python, a técnica de clusterização K-means para o estudo de uma planta fotovoltaica. Como parâmetros de entrada, foram utilizados a potência de 136 inversores e a irradiância solar. Com o objetivo de otimizar a escolha do número de agrupamentos a serem formados, implementou-se o método Elbow. Desta forma, foi possível analisar o desempenho dos sistemas fotovoltaicos que constituem a planta.

1.1 Objetivos

O objetivo global deste trabalho é analisar o desempenho de sistemas fotovoltaicos, integrantes de uma planta formada por diferentes usinas fotovoltaicas, aplicando-se o algoritmo K-means como técnica de clusterização.

Os objetivos específicos deste trabalho são:

- a) Analisar o histórico de dados de operação do sistema SCADA de 136 inversores de complexos fotovoltaicos para desenvolver uma base de dados concisa;
- b) Agrupar inversores de comportamentos similares, por meio da aplicação, em linguagem Python, de um algoritmo de clusterização;
- c) Implementar o método Elbow para otimizar a escolha do número de *clusters* a serem formados;
- d) Analisar e comparar o padrão de comportamento dos diferentes grupos de inversores ao longo dos meses de fevereiro, maio, agosto e novembro do ano de 2020.

1.2 Estrutura do Trabalho

A estrutura deste trabalho foi dividida em 5 capítulos:

- O Capítulo 1 apresenta uma ideia geral do tema, a justificativa, os objetivos propostos e a estrutura de organização do trabalho.
- O Capítulo 2 apresenta uma fundamentação teórica acerca da geração fotovoltaica, abordando conceitos sobre princípios físicos, semicondutores e fluxos de calor. Além disso, contempla a modelagem da célula fotovoltaica, bem como suas curvas características e os tipos de sistemas fotovoltaicos.
- O Capítulo 3 descreve a metodologia do trabalho, bem como os conceitos e definições necessários para seu entendimento. Neste capítulo, também, encontram-se a modelagem matemática da técnica de clusterização implementada e a descrição do algoritmo desenvolvido em Python.
- O Capítulo 4 apresenta os resultados obtidos na realização deste trabalho, bem como discussões e análises sobre as simulações feitas.
- O Capítulo 5 apresenta as conclusões do estudo e as recomendações para trabalhos futuros que possam agregar na linha de pesquisa.

2 FUNDAMENTOS DA GERAÇÃO FOTOVOLTAICA

Faz-se necessário entender que o aproveitamento da energia proveniente do Sol para a geração de eletricidade é pautado no princípio de conservação de energia, pois, segundo Lavoisier: “Na natureza, nada se cria, nada se perde, tudo se transforma”.

Portanto, as teorias, equações e princípios da conversão de energia solar em energia elétrica, mais especificamente a conversão fotovoltaica, serão apresentados nesse capítulo, bem como as principais partes integrantes de um sistema fotovoltaico.

2.1 Princípios Energéticos

Sabe-se, do estudo da física, que onda é um fenômeno físico de transporte de energia e de quantidade de movimento. Nesse contexto, pode-se discutir a possibilidade de aproveitar a energia transportada pelas emissões de ondas eletromagnéticas solares para a geração de energia elétrica.

No entanto, para que isso seja possível, faz-se necessário compreender a natureza dessa energia solar, bem como a elaboração de um sistema que seja capaz de realizar tal conversão.

2.1.1 Natureza da Luz

Por meio de vários estudos e pesquisas, em 1905, Einstein elaborou um postulado intitulado como Teoria do Fóton. Nessa teoria, foi possível, utilizando-se das ideias propostas por Max Planck para explicar as propriedades de energia do corpo negro, comprovar que um feixe de luz pode ser entendido como uma composição de pequenos pacotes de energia chamados de fótons. Diante disso, a energia transportada por um fóton é dada pela seguinte relação (YOUNG; FREEDMAN, 2016):

$$E = hf = \frac{h \cdot c}{\lambda}. \quad (2.1)$$

Onde:

E : Energia de um Fóton.

h : Constante de Planck ($6,6261 \times 10^{-34}$ J.s).

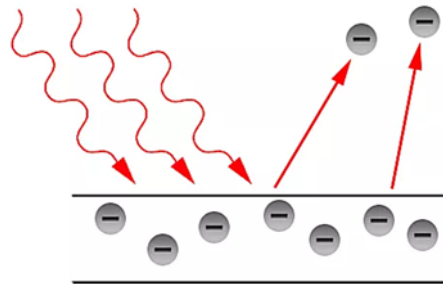
f : Frequência da Luz Emitida.

c : Velocidade da Luz no Vácuo (3×10^8 m/s)

λ : Comprimento de Onda da Luz Emitida.

Nesse contexto, percebeu-se que o efeito fotoelétrico pode ser entendido como um fenômeno de extremos, ou seja, quando uma superfície é “bombardeada” por uma certa quantidade de fótons, tem-se 2 possibilidades: um elétron absorve toda a energia dos fótons e se desprende da superfície ou não absorve nenhuma quantidade de energia e permanece na superfície.

Figura 1 – Efeito Fotoelétrico



Fonte: (JONES, 2020)

Ou seja, existe uma quantidade mínima de energia que um elétron deve receber para que o efeito fotoelétrico ocorra. Tal quantidade de energia mínima é conhecida como função trabalho e depende do material de composição da superfície (YOUNG; FREEDMAN, 2016).

Tabela 1 – Funções trabalho de alguns materiais

Elemento	Função Trabalho [eV]
Alumínio	4,3
Cobre	4,7
Ouro	5,1
Silício	4,8
Sódio	2,7

Fonte: (YOUNG; FREEDMAN, 2016)

Com isso, pode-se equacionar o efeito fotoelétrico da seguinte forma:

$$K_{\text{máx}} = \frac{1}{2} \cdot m \cdot v_{\text{máx}}^2 = h \cdot f - \phi. \quad (2.2)$$

Onde:

$K_{\text{máx}}$: Energia Cinética Máxima do Elétron Desprendido.

m : Massa do Elétron ($9,1094 \times 10^{-31}$ Kg).

$v_{\text{máx}}$: Velocidade Máxima do Elétron Desprendido.

h : Constante de Planck ($6,6261 \times 10^{-34}$ J.s).

f : Frequência da Luz Emitida.

ϕ : Função Trabalho.

Portanto, para que a conversão fotovoltaica possa acontecer, o efeito fotoelétrico, modelado pela Equação 2.2, deve estar presente. É, justamente, nesse contexto que surge a ideia da utilização de semicondutores na confecção dos conversores fotovoltaicos.

2.1.2 Teoria dos Semicondutores

Os semicondutores são materiais que possuem comportamento intermediário entre os materiais condutores e os materiais isolantes. No entanto, para que a Teoria dos Semicondutores seja entendida, faz-se necessário, o entendimento sobre os materiais puros (intrínsecos), sobre os níveis de energia e sobre os materiais dopados (extrínsecos).

2.1.2.1 Materiais Intrínsecos

Basicamente, é um cristal com alta pureza formado por 4 elétrons na camada de valência, a exemplo do Silício (Si) e Germânio (Ge). Nessas redes de cristais, todos os 4 átomos são mantidos ligados aos átomos vizinhos por ligações covalentes (BOYLESTAD; NASHELSKY, 2013).

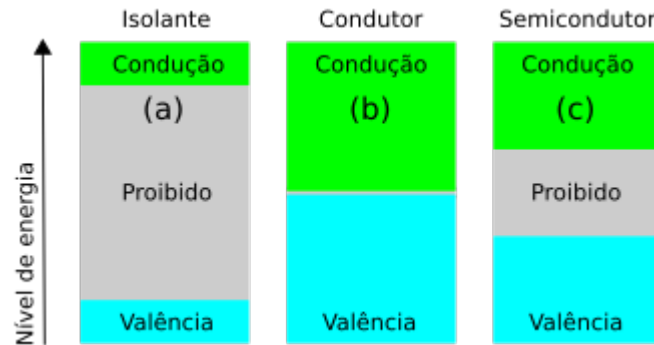
2.1.2.2 Níveis de Energia

Com o intuito de facilitar o entendimento do comportamento dos elétrons nas redes cristalinas dos átomos de um elemento, pode-se analisar os níveis de energia.

Para que haja a condução de corrente elétrica em um material, é necessário que ocorra uma transferência de elétrons da banda de valência para a banda de condução. Tal situação é possibilitada por meio do recebimento de energia. Nos materiais condutores, existe, basicamente, uma sobreposição entre as bandas de valência e de condução, fazendo com que todos os elétrons estejam disponíveis pra conduzirem corrente, ou seja, existe uma condutividade elevada. No entanto, é possível perceber a presença de uma zona de separação entre as bandas de valência e de condução nos materiais isolantes e semicondutores. Tal região é denominada de banda proibida, pois os elétrons não podem situar-se nessa zona. Nos materiais isolantes, a banda proibida

possui uma extensão considerável e, por esse motivo, a quantidade de energia envolvida nessa transferência de elétrons é muito elevada e, conseqüentemente, dificulta a propriedade condutiva do material. Os semicondutores, entretanto, possuem uma banda proibida de extensão mais branda, resultando numa menor necessidade de energia para que mais elétrons sejam transferidos para a banda de condução (BOYLESTAD; NASHELSKY, 2013).

Figura 2 – Níveis de energia em isolantes (a), em condutores (b) e em semicondutores (c)



Fonte: (SEMICONDUCTORES, 2018)

Nesse contexto, quando os elétrons recebem energia suficiente para romper as ligações covalentes da estrutura cristalina, tornam-se elétrons livres na banda de condução, ou seja, são portadores de carga negativa, deixando um “espaço vazio”. Este espaço deixado pelo elétron livre, comporta-se como um portador de carga positiva e passa a ser chamado de lacuna. A lacuna pode ser preenchida por outro elétron de uma ligação covalente próxima ou por algum outro elétron livre e, neste último caso, os dois portadores de carga desaparecem, processo conhecido como recombinação. A recombinação é oriunda do movimento aleatório dos portadores de carga. Esse movimento tem o intuito de reestabelecer o equilíbrio por meio de uma corrente de distribuição dos portadores de carga. Tal corrente é denominada corrente de difusão (BOYLESTAD; NASHELSKY, 2013).

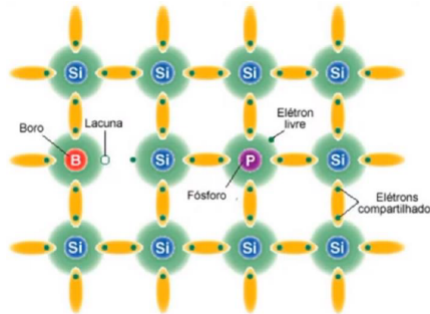
2.1.2.3 *Materiais Extrínsecos*

Os materiais extrínsecos surgem a partir da dopagem, ou seja, da introdução de impurezas nos materiais intrínsecos.

Quando as impurezas possuem 5 elétrons na camada de valência, a exemplo do Fósforo (P) e do Arsênio (As), os elétrons excedentes tornam-se elétrons livres na banda de condução, ou seja, as lacunas são portadoras minoritárias de carga e os elétrons livres são portadores majoritários de carga. Por esta razão, tais materiais são ditos do tipo N. Todavia, quando as

impurezas possuem 3 elétrons na camada de valência, a exemplo do Boro (B) e do Gálio (Ga), as lacunas são portadoras majoritárias de carga e os elétrons livres são portadores minoritários de carga. Por esta razão, tais materiais são ditos do tipo P (BOYLESTAD; NASHELSKY, 2013).

Figura 3 – Dopagem da rede cristalina de átomos de silício (Si) com fósforo (P) e com boro (B)

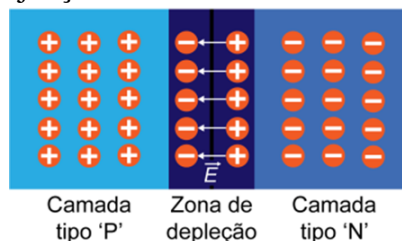


Fonte: (SEMICONDUCTORES, 2007)

2.1.2.4 Junção P-N

Ao ocorrer a junção de materiais do tipo P e do tipo N, os portadores majoritários da parte N são difundidos para a parte P. Essa movimentação resulta no aumento de elétrons do lado P, deixando-o com um potencial negativo, e na diminuição de elétrons do lado N, deixando-o com um potencial positivo. Dessa forma, ocorre a formação de um campo elétrico nessa junção, ou seja, uma espécie de barreira potencial. A movimentação dos portadores de carga cessa quando o efeito do campo elétrico formado é capaz de impedir o transporte dos elétrons livres restantes do lado N para o lado P. Essa região de “bloqueio” é denominada de zona de depleção (SEDRA; SMITH, 2007).

Figura 4 – Esquemático de uma junção P-N



Fonte: (MORAES, 2020)

Quando essa junção P-N recebe uma polarização externa inversa (polo positivo no lado N), ocorre o aumento no efeito do campo elétrico da zona de depleção, dificultando ainda mais o fluxo de elétrons. No entanto, quando é aplicada uma tensão elétrica direta (polo positivo no lado P), capaz de vencer o efeito da barreira potencial, os elétrons conseguirão passar do

lado N para o lado P, ou seja, irá surgir uma corrente elétrica. Por meio da iluminação de uma junção P-N, os elétrons que ali se encontram absorvem energia, já que a luz é entendida como um fluxo de fótons. Ao absorverem os fótons, os elétrons conseguem saltar da zona de valência para a zona de condução, vencendo a camada de depleção. Dessa forma, o fluxo de elétrons será responsável pelo surgimento de uma corrente elétrica que, por sua vez, é capaz de gerar energia elétrica. As características gerais dessa corrente podem ser entendidas por meio da Equação de Shockley: (SEDRA; SMITH, 2007).

$$I_D = I_S \cdot \left(e^{\frac{q \cdot V_D}{n \cdot K \cdot T}} - 1 \right). \quad (2.3)$$

Onde:

I_D : Corrente na Junção P-N.

I_S : Corrente de Saturação (referente à polarização inversa).

e : Número de Neper.

q : Carga do Elétron ($1,6 \times 10^{-19}$ C).

V_D : Tensão Direta aplicada na Junção P-N.

n : Fator de Idealidade (varia de 1 a 2 e depende do material).

K : Constante de Boltzmann ($1,38 \times 10^{-23}$ J/K).

T : Temperatura Ambiente em Kelvin.

2.2 Sistemas Fotovoltaicos

Nesse contexto, faz-se necessário desenvolver uma estrutura capaz de realizar a conversão da radiação solar em energia elétrica. Essa estrutura é denominada de sistema fotovoltaico e constitui-se de vários componentes, sendo os inversores de frequência e os painéis fotovoltaicos os principais.

2.2.1 Inversores de Frequência

Pode-se perceber que a corrente elétrica, formada pelo efeito fotoelétrico nos semicondutores, possui caráter unidirecional, ou seja, é uma corrente elétrica contínua. No entanto, certas cargas conectadas à rede de alimentação possuem padrão de corrente alternada. Diante disso, os inversores de frequência são instalados com a finalidade de converter a energia de corrente elétrica contínua em alternada com a menor dissipação, possível, de potência e baixa

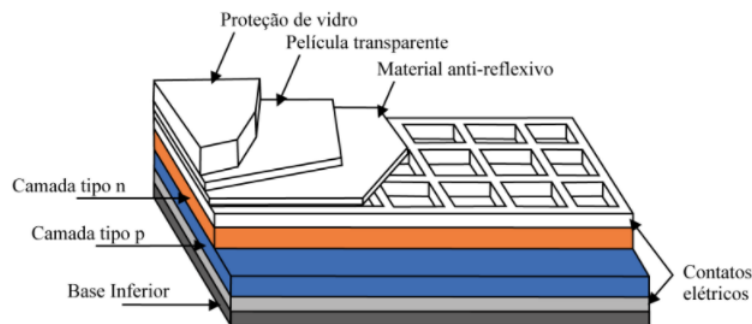
concentração de componentes harmônicas, sincronizando as características de fase da rede de alimentação e do sistema fotovoltaico (MOREIRA, 2019).

As chaves semicondutoras que compõem os inversores, são as responsáveis pelo fenômeno de inversão e podem ser comutadas pela própria rede ou por auto comutação (controle interno do próprio inversor). Os inversores de frequência mais utilizados são os que produzem ondas quadradas e ondas retangulares (quadradas modificadas).

2.2.2 Células Fotovoltaicas

Diante da análise acerca da natureza das emissões solares e dos estudos referentes à teoria dos semicondutores, foram desenvolvidas tecnologias que possibilitaram o aparecimento das células fotovoltaicas. Tais dispositivos podem ser, basicamente, entendidos como semicondutores que, quando expostos à luz, possuem a capacidade de produzir corrente elétrica (NETO; CARVALHO, 2012).

Figura 5 – Esquemático das camadas de uma célula fotovoltaica



Fonte: (COELHO, 2008)

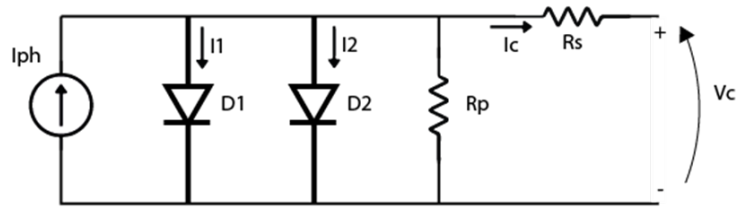
2.2.2.1 Modelagem

Por meio do entendimento da estrutura física da célula fotovoltaica, pode-se partir para a modelagem de um circuito elétrico que represente o funcionamento de tal dispositivo.

No circuito da Figura 6 (modelo de dois diodos da célula fotovoltaica), o diodo D1 indica a região de junção P-N da célula. O diodo D2 é utilizado para representar as perdas por corrente de difusão, proveniente do fenômeno de recombinação. A resistência em série R_s é responsável por descrever as perdas ôhmicas no material semicondutor, na comunicação entre o metal e o semicondutor e nos contatos metálicos. Por fim, a resistência em paralelo R_p representa as perdas por impurezas, por perturbações na estrutura e por efeito joule nas ligações elétricas

entre as partes da célula (NETO; CARVALHO, 2012).

Figura 6 – Modelo do circuito elétrico equivalente de uma célula fotovoltaica



Fonte: Adaptado de (NETO; CARVALHO, 2012)

Com isso, é possível, ao aplicar a Lei de Kirchhoff das Correntes e a 1ª Lei de Ohm, determinar a seguinte expressão para relacionar a tensão e a corrente fornecidas pelo painel fotovoltaico:

$$I_C = I_{PH} - I_S \cdot \left(e^{\frac{q \cdot V_D}{n \cdot K \cdot T}} - 1 \right) - \frac{V_C + R_S \cdot I_C}{R_p}. \quad (2.4)$$

Onde:

I_C : Corrente Fornecida pela Célula Fotovoltaica.

I_{PH} : Corrente Fotogerada.

I_S : Corrente de Saturação (referente à polarização inversa).

e : Número de Neper.

q : Carga do Elétron ($1,6 \times 10^{-19}$ C).

V_D : Tensão Direta aplicada na Junção P-N.

n : Fator de Idealidade (varia de 1 a 2 e depende do material).

K : Constante de Boltzmann ($1,38 \times 10^{-23}$ J/K).

T : Temperatura Ambiente em Kelvin.

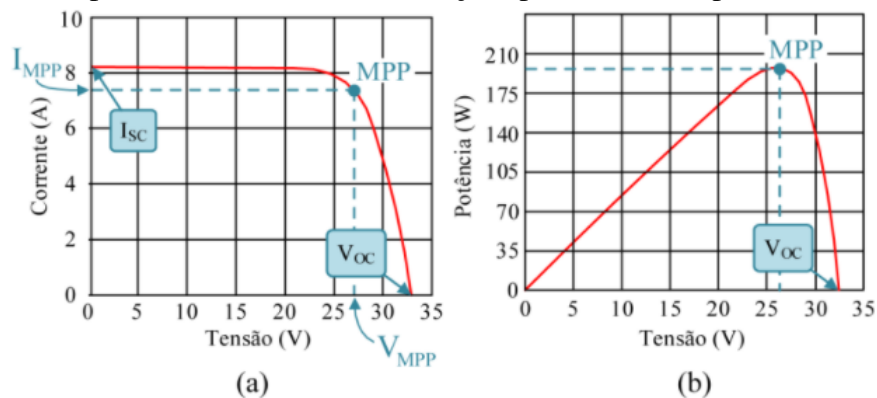
V_C : Corrente na Junção P-N.

2.2.2.2 Curvas Características

Em posse da modelagem realizada para a célula fotovoltaica, é possível realizar a análise de seu comportamento e, diante disso, descrevê-lo por meio de curvas características. A necessidade de poucas grandezas físicas, nessa caracterização, facilita o estudo de tal comportamento (NETO; CARVALHO, 2012).

Para a obtenção das curvas características, faz-se necessário adotar condições padrões, possibilitando a comparação entre diferentes tipos de células. As condições padrões de teste (STC – *Standard Test Conditions*) são definidas para irradiância de 1000 W/m^2 , para temperatura de $25 \text{ }^\circ\text{C}$ e 1,5 para a massa de ar (parâmetro calculado em posse do ângulo de altura solar e relacionado aos caminhos percorridos pelas ondas eletromagnéticas emitidas pelo Sol) (PINHO; GALDINO, 2014).

Figura 7 – Curvas características de corrente versus tensão (a) e de potência versus tensão (b) para um painel fotovoltaico em condições padrões de temperatura e irradiância

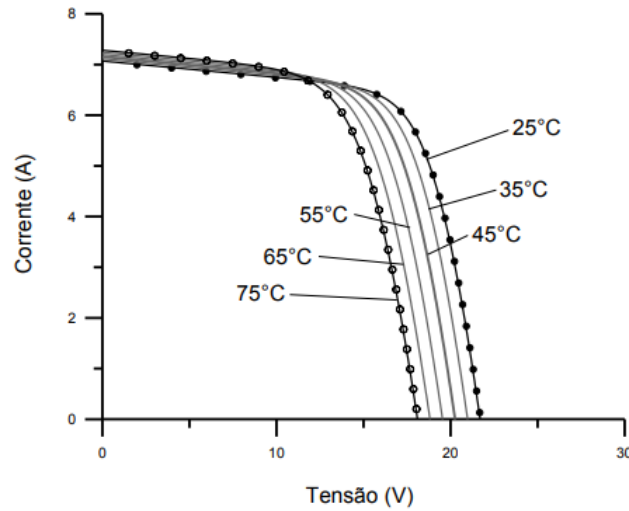


Fonte: (BRITO, 2013)

Observando-se as curvas ilustradas na Figura 7, tem-se que, quando a tensão externa na célula é nula, ou seja, quando ocorre a situação de curto-circuito, a corrente é máxima e é denominada corrente de curto-circuito (I_{sc}). Além disso, quando a corrente externa é nula, ou seja, não há surgimento de corrente por não superação da barreira de potencial da região dopada, a tensão externa recebe o nome de tensão de circuito aberto (V_{oc}). Nota-se, ainda, a existência de um ponto em que a potência fornecida, pela célula fotovoltaica, é máxima. Tal ponto de operação recebe o nome de ponto de máxima potência (MPP – *Maximum Power Point*). No MPP, a tensão é representada por V_{MPP} e a corrente por I_{MPP} (NETO; CARVALHO, 2012). No entanto, quando o sistema opera fora das condições padrões, tais curvas são modificadas.

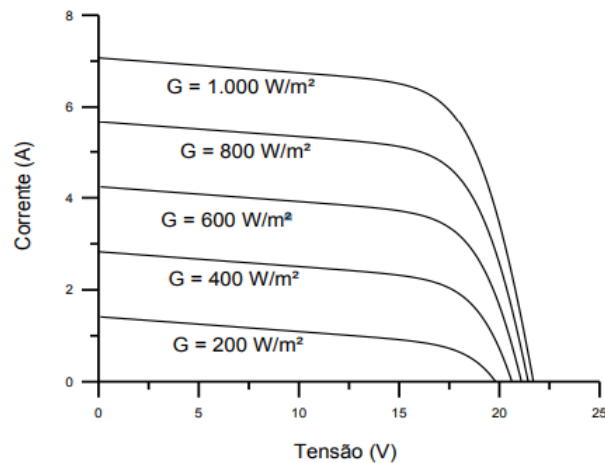
O aumento da temperatura resulta em um pequeno incremento de corrente de curto-circuito, pois gera-se um número maior de portadores de cargas. Todavia, a tensão de circuito aberto diminui em maior proporção. Desta forma, a potência máxima decai, normalmente, numa taxa de $0,5\%/^\circ\text{C}$. Analisando-se o comportamento da curva com o aumento de irradiância, é possível perceber que o efeito de elevação da corrente de curto-circuito é maior que o efeito de elevação da tensão de circuito aberto, resultando em um aumento na potência máxima fornecida pela célula (NETO; CARVALHO, 2012). As Figuras 8 e 9 ilustram tais comportamentos.

Figura 8 – Comportamento da curva de corrente versus tensão de um painel fotovoltaico diante da variação de temperatura



Fonte: (PINHO; GALDINO, 2014)

Figura 9 – Comportamento da curva de corrente versus tensão de um painel fotovoltaico diante da variação dos níveis de irradiância



Fonte: (PINHO; GALDINO, 2014)

Nesse contexto, é de grande importância a definição dos parâmetros conhecidos como fator de forma (FF) e como rendimento (η).

2.2.2.2.1 Fator de Forma

O fator de forma (FF) pode ser entendido como a razão entre a máxima potência da célula, ou seja, a potência no MPP e a potência dada pelo produto entre a tensão de circuito aberto e a corrente de curto-circuito.

$$FF = \frac{V_{MPP} \cdot I_{MPP}}{V_{OC} \cdot I_{SC}}. \quad (2.5)$$

Onde:

FF : Fator de Forma.

V_{MPP} : Tensão no MPP.

I_{MPP} : Corrente no MPP.

V_{OC} : Tensão de Circuito Aberto.

I_{SC} : Corrente de Curto-Circuito.

Pode-se salientar que, usualmente, o fator de forma varia de 70% a 80% (NETO; CARVALHO, 2012).

2.2.2.2.2 Rendimento

O rendimento (η) é definido como a relação entre a máxima potência da célula, ou seja, a potência no MPP e a potência total relacionada à incidência solar. Desta forma, o rendimento é capaz de quantificar a efetividade da conversão de energia solar em energia elétrica.

$$\eta = \frac{V_{MPP} \cdot I_{MPP}}{A \cdot G} = \frac{FF \cdot V_{OC} \cdot I_{SC}}{A \cdot G}. \quad (2.6)$$

Onde:

η : Rendimento.

V_{MPP} : Tensão no MPP.

I_{MPP} : Corrente no MPP.

A : Área da Célula Fotovoltaica.

G : Irradiância Solar Incidente.

FF : Fator de Forma.

V_{OC} : Tensão de Circuito Aberto.

I_{SC} : Corrente de Curto-Circuito.

Atualmente, os painéis solares comerciais mais comuns operam no sistema com um rendimento que varia de 13% a 15% (NETO; CARVALHO, 2012).

2.2.2.3 Perdas

Como visto anteriormente, os módulos fotovoltaicos convertem uma pequena quantidade de energia da radiação solar em energia elétrica (13% a 15%), transformando o restante, principalmente, na forma de calor. Uma parcela desse calor é responsável pelo aquecimento

das células fotovoltaicas, reduzindo a eficiência de conversão, a outra parcela é liberada para a vizinhança por convecção ou irradiação e a última parcela diz respeito à não absorção dos fótons que carregam uma quantidade de energia menor que a necessária para saltar a banda proibida e da não absorção dos fótons que estão em excesso. Apenas nesta última parcela, o calor liberado corresponde a um percentual de, aproximadamente, 51% de perdas totais das células comerciais de Silício (Si).

2.2.2.3.1 Balanço Energético em uma Célula Fotovoltaica

Uma estratégia para se compreender as perdas de um painel fotovoltaico é realizar o estudo do balanço energético, pois um sistema termodinâmico, no estado estacionário, possui uma equivalência entre os fluxos de energia de entrada e os fluxos de energia de saída. Da termodinâmica, sabe-se que a energia pode ser transferida por condução, por radiação ou por convecção.

O processo de condução é decorrente do contato entre superfícies de diferentes temperaturas, obedecendo a Lei de Fourier (HOLMAN, 2009).

$$Q_{cond} = \frac{k_{mat} \cdot A \cdot \Delta T}{e}. \quad (2.7)$$

Onde:

Q_{cond} : Taxa de Calor por Condução [W].

k_{mat} : Condutividade Térmica do Material [W/m.K].

A : Área Superficial [m²].

ΔT : Diferença de Temperatura entre os Meios [K].

e : Espessura do Material [m].

Na transmissão por radiação, a taxa de energia transmitida pode ser calculada por meio da Lei de Stefan-Boltzmann (HOLMAN, 2009).

$$Q_{rad} = \varepsilon \cdot \sigma \cdot A \cdot T^4. \quad (2.8)$$

Onde:

Q_{rad} : Taxa de Calor por Radiação [W].

ε : Emissividade do Material (0 para corpos completamente refletivos e 1 para corpos negros).

σ : Constante de Stefan-Boltzmann ($5,6697 \times 10^{-8} \text{ W/m}^2 \cdot \text{K}^4$).

A : Área Superficial [m^2].

T : Temperatura do Corpo [K].

Quando a troca de calor é realizada por meio do processo de convecção, a taxa da energia térmica trocada com o ambiente externo obedece a Lei de Resfriamento de Newton (HOLMAN, 2009).

$$Q_{convec} = h \cdot A \cdot \Delta T. \quad (2.9)$$

Onde:

Q_{convec} : Taxa de Calor por Convecção [W]

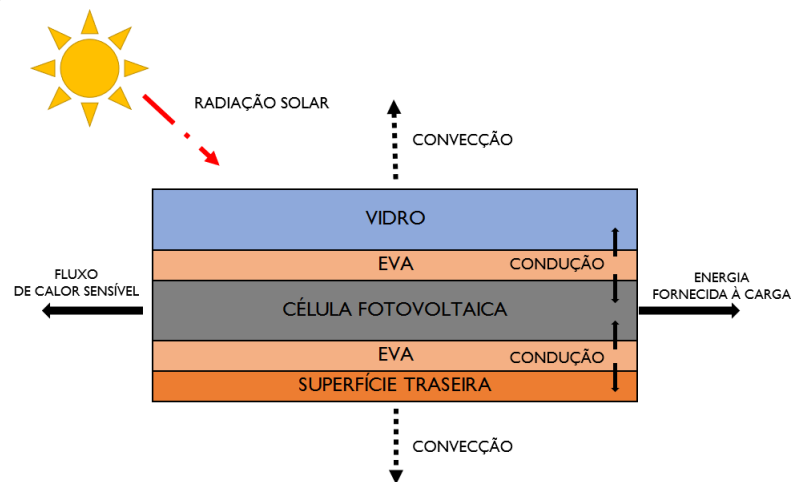
h : Coeficiente de Transferência de Calor [$\text{W/K} \cdot \text{m}^2$]

A : Área Superficial [m^2].

ΔT : Diferença de Temperatura entre os Meios [K].

Em posse de tais relações, por meio da Figura 10, torna-se possível analisar os fluxos de energia que estão relacionados com o painel fotovoltaico.

Figura 10 – Esquemático dos fluxos de calor sobre um módulo fotovoltaico



Fonte: Adaptado de (LIMA, 2019)

Com isso, o balanço de energia da superfície do módulo fotovoltaico pode ser escrito por meio da seguinte expressão (SMITH *et al.*, 2014):

$$q_{oc,S} + q_{ol,at} + q_{ol,solo} = q_{oc,fv} + q_{ol,fv-c} + q_{ol,fv-s} + H + E_{prod}. \quad (2.10)$$

Onde:

$q_{oc,S}$: Radiação de Ondas Curtas vinda do Sol (difuso ou direto).

$q_{ol,at}$: Radiação de Ondas Longas que chega da Atmosfera (difuso).

$q_{ol,solo}$: Radiação de Ondas Longas Refletidas Interceptadas pelo Pannel (albedo).

$q_{oc,fv}$: Radiação de Ondas Curtas Refletidas pelo Pannel para o Céu.

q_{ol,fv_c} : Radiação de Ondas Longas Refletidas pelo Pannel para o Céu.

q_{ol,fv_s} : Radiação de Ondas Longas Emitidas pelo Pannel para o Telhado ou para o Solo.

H : Fluxo de Calor Sensível entre Pannel e Atmosfera.

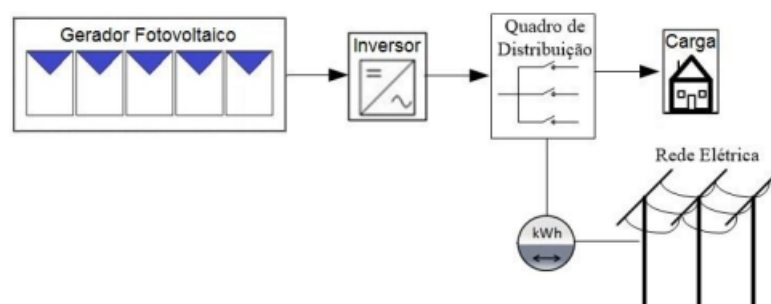
E_{prod} : Energia Elétrica Fornecida pelo Pannel.

2.2.3 Tipos de Sistemas Fotovoltaicos

É importante destacar que um sistema fotovoltaico é classificado dependendo do tipo de relação existente entre a rede de alimentação elétrica e o sistema fotovoltaico.

Quando o sistema opera conectado à rede elétrica, denomina-se um sistema *On-grid*. Tal tipo de operação possui a vantagem da não dependência exclusiva da conversão fotovoltaica para alimentar as cargas, pois a rede elétrica, também, é responsável por tal alimentação. Outra característica que merece destaque, nesse tipo de operação, é a possibilidade de fluxo bidirecional de energia. Ou seja, caso a energia convertida pelo sistema fotovoltaico seja maior que a necessária para suprir a alimentação das cargas, a quantidade excedente pode ser transmitida para a rede de alimentação (MOREIRA, 2019).

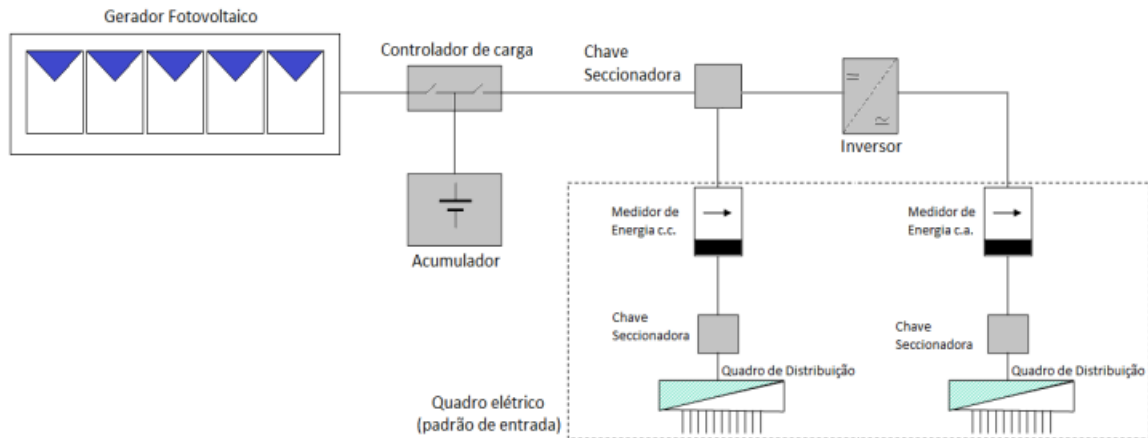
Figura 11 – Esquemático do diagrama unifilar de um sistema fotovoltaico conectado à rede elétrica



Fonte: (PINHO; GALDINO, 2014)

No entanto, quando a operação é isolada da rede (*Off-grid*), o sistema fotovoltaico é o único responsável pela alimentação das cargas. Diante disso, faz-se necessária a utilização de um banco de baterias (responsável pelo armazenamento de energia, evitando interrupções no fornecimento para as cargas) e de um controlador de carga.

Figura 12 – Esquemático do diagrama unifilar de um sistema fotovoltaico isolado da rede elétrica



Fonte: (PINHO; GALDINO, 2014)

A principal função dos controladores de carga é a proteção do banco de baterias contra cargas e descargas excessivas, visando uma maior vida útil do sistema. Tal proteção é feita baseada no carregamento das baterias, quando a energia elétrica gerada estiver em excesso, e na utilização do banco de baterias quando a energia elétrica gerada é insuficiente. Nesse caso, as baterias são utilizadas, apenas, até um nível mínimo de segurança da carga remanescente (MOREIRA, 2019).

3 METODOLOGIA

A partir de agora, dar-se-á início ao processo de descrição da metodologia utilizada no desenvolvimento deste trabalho. Em posse dos dados de um complexo fotovoltaico, fornecidos pelo sistema de supervisão e aquisição de dados (SCADA), aplicou-se o algoritmo K-means de clusterização em linguagem Python.

Neste trabalho, a clusterização dos inversores foi realizada por intermédio da análise dos parâmetros de potência e de irradiância solar. Os sensores localizados nos inversores das estações de transformadores (TS – *Transformer Stations*) são os responsáveis por fornecerem as informações de potência, enquanto que as informações de irradiância solar são fornecidas pelos sensores acoplados nas estações meteorológicas (WS – *Weather Station*).

3.1 Características do Complexo Fotovoltaico

O complexo, localizado no interior do Ceará, é dividido em 4 usinas fotovoltaicas, que são constituídas por 9 estações de transformadores cada uma, como indicado na Tabela 2.

Tabela 2 – Distribuição das TS nas usinas fotovoltaicas

Usina	TS
I	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9
II	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9
III	3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9
IV	4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9

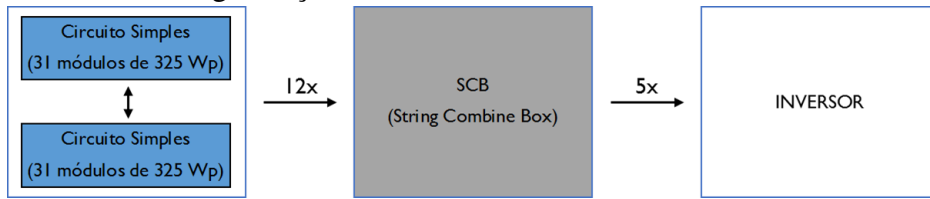
Fonte: Própria

Com exceção das estações de transformadores 1.5, 2.5, 3.5 e 4.5, que possuem apenas 2 inversores, todas as outras possuem 4 inversores. Vale salientar que cada inversor é conectado a 5 estruturas denominadas de SCB (*String Combine Box*).

Cada SCB é alimentada por 12 circuitos duplos, onde 1 circuito duplo indica a configuração estabelecida por 2 circuitos simples ligados em paralelo. Um circuito simples é composto por 31 módulos de 325 W. A Figura 13 ilustra o esquemático da estrutura descrita.

Diante disso, o complexo fotovoltaico analisado possui um total de 505.920 módulos, resultando em uma potência instalada de 164,424 MW, como indicado na Tabela 3.

Figura 13 – Estrutura de organização dos circuitos nas usinas fotovoltaicas



Fonte: Própria

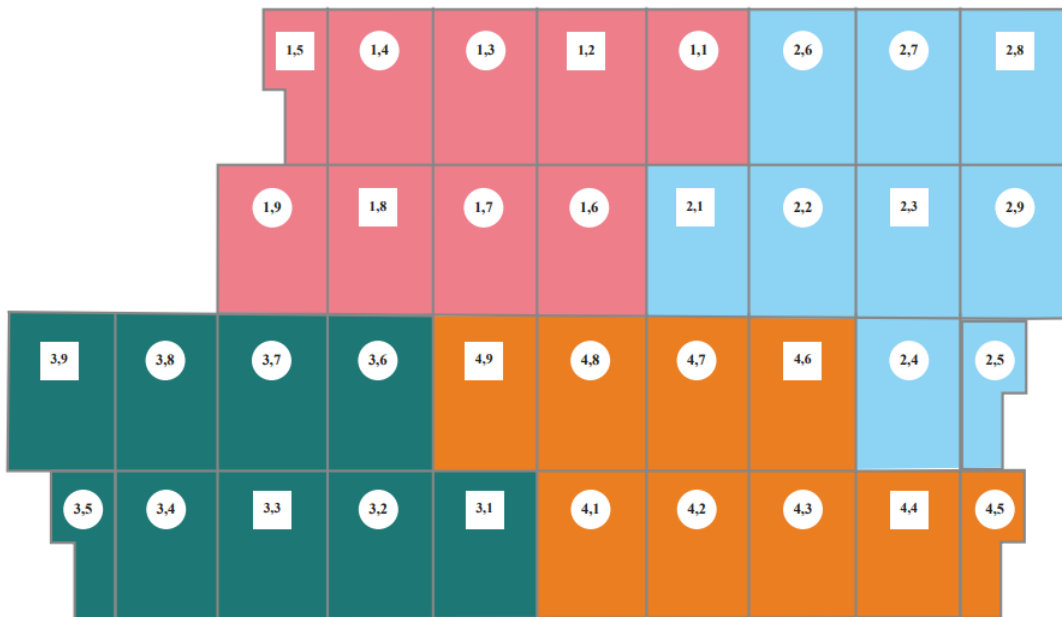
Tabela 3 – Potência total instalada do complexo fotovoltaico

Total de Módulos do Complexo	Potência Unitária dos Módulos (W)	Potência Total Instalada (MW)
505920	325	164,424

Fonte: Própria

Na Figura 14 é mostrado um esquemático da vista superior do complexo, onde as TS marcadas por símbolos retangulares abrigam as estações meteorológicas, enquanto que as TS representadas por símbolos circulares não possuem WS integradas. Além disso, a diferença na tonalidade das cores faz referência às diferentes usinas fotovoltaicas.

Figura 14 – Esquemático da vista superior do complexo fotovoltaico



Fonte: Própria

As estações meteorológicas estão situadas em apenas 12 das 36 estações de transformadores, como indicado na Tabela 4, fazendo com que as informações meteorológicas não estejam disponibilizadas para as outras 24 TS. Por esta razão, realizou-se um estudo do comportamento mensal da irradiância nas WS, concluindo-se que, pela análise de valores médios e

desvios padrões, a irradiância possui pequena variação de uma WS para outra, em um mesmo período mensal.

Com isso, de posse das informações de posição de cada TS, foi possível atribuir valores de irradiância às TS que não abrigam alguma WS, considerando-se que cada TS possui a mesma magnitude de irradiância da WS mais próxima. A Tabela 5 organiza a correlação entre as WS e TS.

Tabela 4 – Localização das estações meteorológicas (WS)

Estação Meteorológica	Localização
WS 02	TS 1.2
WS 03	TS 1.5
WS 04	TS 1.8
WS 05	TS 2.1
WS 06	TS 2.3
WS 07	TS 2.8
WS 08	TS 3.1
WS 09	TS 3.3
WS 10	TS 3.9
WS 11	TS 4.4
WS 12	TS 4.6
WS 13	TS 4.9

Fonte: Própria

Tabela 5 – Atribuição das estações de transformadores (TS) nas estações meteorológicas (WS)

Estação Meteorológica	Localização	TS Atribuídas
WS 02	TS 1.2	1.1 e 1.3
WS 03	TS 1.5	1.4
WS 04	TS 1.8	1.9 e 3.7
WS 05	TS 2.1	1.6 e 4.7
WS 06	TS 2.3	2.4, 2.5 e 2.9
WS 07	TS 2.8	2.6 e 2.7
WS 08	TS 3.1	4.1 e 4.2
WS 09	TS 3.3	3.2, 3.4 e 3.5
WS 10	TS 3.9	3.8
WS 11	TS 4.4	4.3 e 4.5
WS 12	TS 4.6	2.2
WS 13	TS 4.9	1.7, 3.6 e 4.8

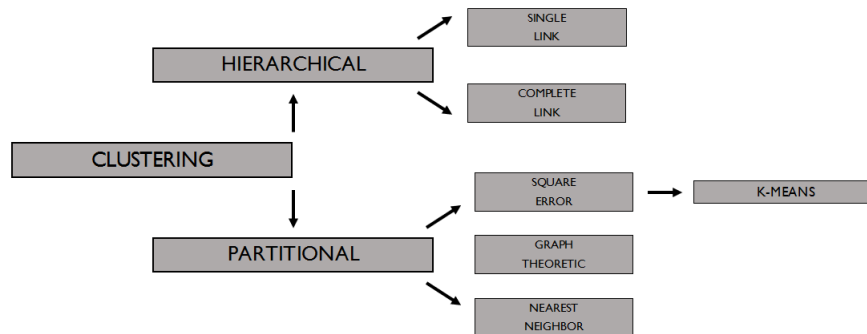
Fonte: Própria

3.2 Fundamentos sobre Clusterização

A clusterização consiste na ação de analisar dados, com o objetivo de separá-los em agrupamentos distintos denominados *clusters*, possuindo, por isso, grande importância no universo da pesquisa de dados, possibilitando reconhecimento de padrões, previsões de comportamento e identificação de problemas (ARORA *et al.*, 2016).

A divisão é baseada na observação das características dos dados analisados, visando a criação de grupos com alta homogeneidade interna e maior heterogeneidade externa possível. Em outras palavras, os dados de um mesmo *cluster* possuem uma forte similaridade, enquanto que *clusters* distintos devem possuir uma baixa similaridade entre si (OLIVEIRA, 2008).

Figura 15 – Esquemático para diferentes técnicas de clusterização

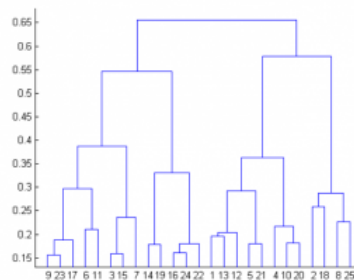


Fonte: Adaptado de (LACHI; ROCHA, 2005)

3.2.1 Metodologia Hierárquica

Os algoritmos aplicados na formação de *clusters* por metodologia hierárquica são responsáveis por organizar uma estrutura de hierarquia, formando uma espécie de fluxograma divisional. Essa estrutura é denominada de dendograma.

Figura 16 – Exemplo ilustrativo de um dendograma



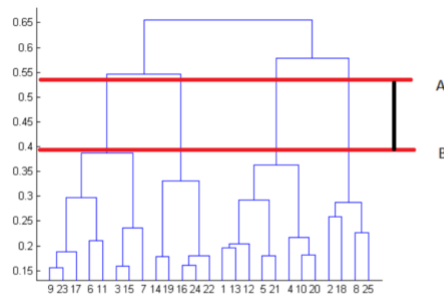
Fonte: (INTRODUCAO..., 2018)

Por meio do dendograma, é possível observar as características dos dados analisados, devido à existência de subconjuntos. Nesse contexto, percebe-se que um *cluster* pode dar origem a *clusters* menores, denominados filhos. No entanto, quando um *cluster* não possui filhos, ele é denominado de folha (LACHI; ROCHA, 2005).

Para que a escolha do número ótimo de *clusters* representativos do conjunto de dados analisados seja feita, faz-se necessário encontrar a maior distância de separação entre dois nós do dendograma. É, justamente, o número de linhas verticais que estão nessa região de separação que determina o número ótimo de *clusters* (INTRODUCAO..., 2018).

A Figura 17 mostra a aplicação, de tal estratégia, ao dendograma da Figura 16. Nesta situação, o número ótimo de agrupamentos seria 4.

Figura 17 – Determinação do valor ótimo de clusters em um dendograma



Fonte: (INTRODUCAO..., 2018)

3.2.2 Metodologia Particional

Diferente do que ocorre na metodologia hierárquica, em que os dados são analisados em uma série de agrupamentos, na metodologia particional os dados são organizados em agregados mais simples.

Figura 18 – Esquemático da clusterização de um conjunto de dados brutos em: dois *clusters*, três *clusters* e cinco *clusters*



Fonte: (OLIVEIRA, 2008)

A ideia principal é determinar o número de *clusters* do conjunto de dados e, a partir daí, realizar a distribuição dos elementos, analisando a similaridade com os chamados centros, ou centroides, daquele grupo. Os centroides podem ser entendidos como o objeto representativo de um determinado *cluster*. Tal distribuição de dados, nos *clusters*, é feita por meio da aplicação de estratégias iterativas, visando otimizar uma função objetivo. Pode-se perceber que os algoritmos particionais possuem uma maior vantagem, em relação aos algoritmos hierárquicos, na análise de conjuntos com grande quantidade de dados, pois, em situações como estas, a elaboração de um dendograma se torna, altamente, complexa (LACHI; ROCHA, 2005).

Vale salientar que a determinação do objeto representante do grupo de dados pode ser dada tanto pela escolha do centro de gravidade daquele *cluster*, quanto pela escolha do elemento que mais se aproxima do centro de gravidade. A primeira possibilidade é denominada de K-means e a segunda possibilidade é conhecida como K-medoids.

3.3 Técnica K-means

Como foi representado na Figura 15, uma das possibilidades de aplicação de metodologias particionais é, justamente, o estudo de funções de erros quadrados. Um dos tipos de algoritmo que utiliza tais funções é o K-means.

Por conta de sua fácil implementação e baixa complexidade, o algoritmo K-means é extremamente popular no universo da análise de dados (LACHI; ROCHA, 2005). Por esta razão, tal algoritmo será utilizada neste trabalho.

Proposto em 1957 por Stuart Lloyd, mas publicado em 1982, o algoritmo de Lloyd é um modelo iterativo que visa otimizar a função erro-quadrado, pois, desta forma, a distância entre os dados de um *cluster* e o seu centroide tendem a ser minimizadas. A ideia consiste em realizar a análise de um conjunto de n dados observados $(x_1, x_2, x_3, \dots, x_n)$, particionados em k conjuntos $(S_1, S_2, S_3, \dots, S_k)$, onde k é diferente de n (OLIVEIRA, 2008).

A estrutura do algoritmo de Lloyd pode ser representada pela equação abaixo:

$$\operatorname{argmin} \sum_{i=1}^k \sum_{x_j \in S_i} \min \left(\|x_j - \mu_i\|^2 \right). \quad (3.1)$$

Onde:

μ_i : centroide no conjunto S_i .

$\|x_j - \mu_i\|^2$: distância entre um ponto de um *cluster* e o centroide.

Diante da formação de um *cluster* S_i , os centroides são reestabelecidos, analisando-se as médias das distâncias de cada ponto x_j do agrupamento, conforme a seguinte relação:

$$\mu'_i = \frac{1}{|S_i|} \cdot \sum_{x_j \in S_i} x_j. \quad (3.2)$$

Onde:

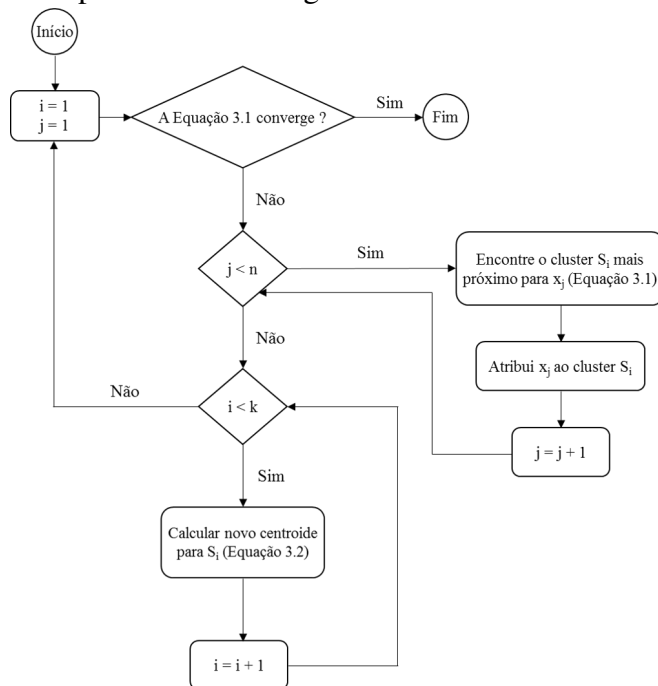
μ'_i : novo centroide do conjunto S_i .

$|S_i|$: número de elementos do *cluster* S_i .

Nesse contexto, o algoritmo K-means pode ser descrito pela seguinte ordem de processos (LACHI; ROCHA, 2005):

1. Determinação do número K de conjuntos (*clusters*) a serem formados, bem como dos respectivos centroides;
2. Atribuição de cada um dos n dados ao *cluster* com maior similaridade;
3. Atualização dos centroides de cada um dos *clusters* por intermédio da estratégia descrita pela Equação 3.2;
4. Retorna ao processo de atribuição dos dados aos *clusters* (passo 2) até que o critério de convergência seja contemplado.

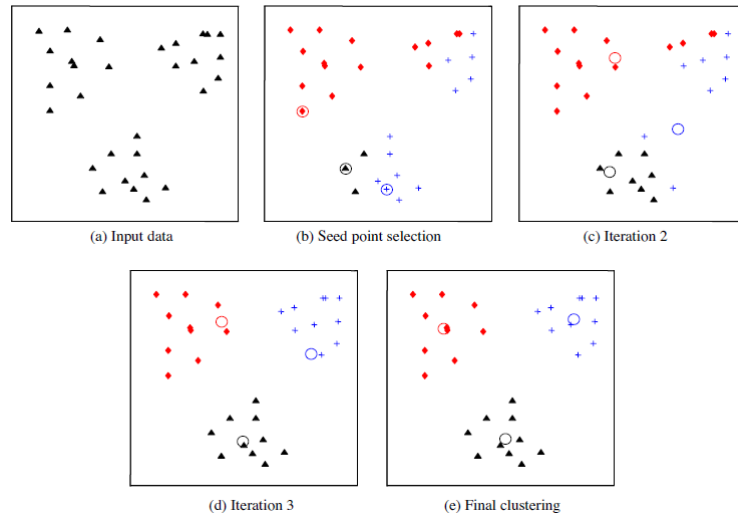
Figura 19 – Fluxograma representativo do algoritmo K-means



Fonte: Adaptado de (CAVALCANTE, 2020)

O fluxograma da Figura 19 ilustra a estrutura lógica para a execução do algoritmo K-means. Além disso, tal processo é exemplificado na Figura 20.

Figura 20 – Processo de implementação do algoritmo K-means para um conjunto de dados de entrada (a), com determinação do número de clusters e centroides iniciais (b), realizando-se a segunda iteração (c), a terceira iteração (d) e a última iteração (e)



Fonte: (JAIN, 2010)

Normalmente, os critérios de convergência considerados são a não possibilidade de otimização da função objetivo (decréscimo mínimo da função erro quadrado) ou a ausência do surgimento de um novo centroide ($\mu'_i = \mu_i$) (OLIVEIRA, 2008).

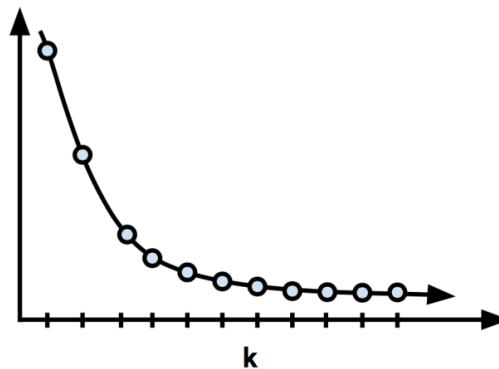
3.3.1 Método Elbow

Como visto, anteriormente, na utilização do algoritmo K-means é necessário que se determine o número K de agrupamentos que se deseja formar. No entanto, determinar a quantidade de *clusters* não é algo tão imediato. Nas situações em que o valor K é superestimado, o esforço computacional será muito elevado para que seja feita a divisão dos dados analisados, além de não possibilitar o reconhecimento de padrões. Por outro lado, quando o valor de K é muito pequeno, a homogeneidade dos *clusters* não será muito grande, resultando em grupos com dados de baixa correlação interna.

Diante disso, uma das estratégias existentes, para a otimização da determinação do número de *clusters*, é o chamado método Elbow. Este método consiste na plotagem do gráfico de um fator representativo da dispersão dos dados (variância, erro médio, soma dos erros quadrados, dentre outros), no eixo y, para seus correspondentes valores de K no eixo x. Cada ponto do gráfico faz referência ao centroide daquele *cluster*.

A Figura 21 representa a aplicação do método Elbow.

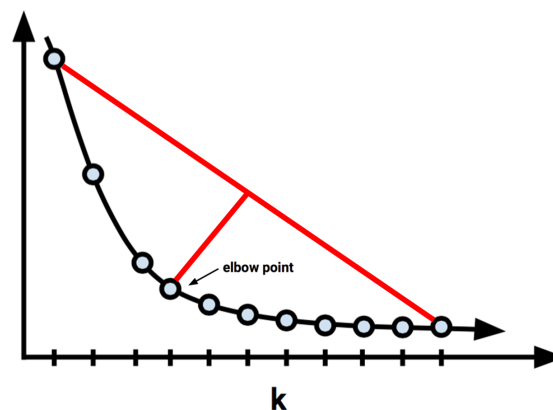
Figura 21 – Curva do método Elbow



Fonte: (SANTANA, 2017)

Com isso, pode-se perceber que, a partir de um certo momento, os valores do eixo vertical são estabilizados, ou seja, mesmo com o aumento de K , os valores correspondentes ao eixo y são, praticamente, constantes. Ao traçar uma reta que conecte os centroides extremos, é possível notar que o ponto da curva mais distante desta reta indica o momento de estabilização dos valores do eixo y . É, justamente, por esta razão que este ponto indica o valor ótimo de K .

Figura 22 – Escolha do valor ótimo de K pelo método Elbow



Fonte: Adaptado de (SANTANA, 2017)

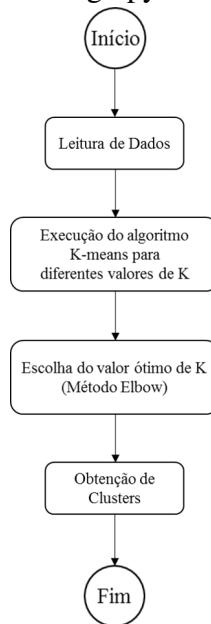
Nota-se que o gráfico é, visualmente, semelhante a um braço curvado, sendo o ponto escolhido análogo ao cotovelo. Por isso, a denominação de método Elbow.

3.3.2 Algoritmo Python

Com o objetivo de estabelecer a clusterização, por meio do algoritmo K -means, utilizou-se a linguagem de programação python. Tal linguagem, proposta por Guido van Rossum

em 1991, consiste em uma linguagem de programação de alto nível, pois a sua sintaxe é bem próxima da estrutura comunicativa das pessoas. Além disso, a existência da biblioteca pré-estabelecida *scikit-learn* possibilita um menor esforço na estruturação do programa, visto que tal biblioteca é munida de diversos algoritmos orientados a métodos de classificação, com base nos fundamentos da aprendizagem de máquina. As instruções sobre a utilização, de tal biblioteca, podem ser, encontradas em (PEDREGOSA *et al.*, 2011).

Figura 23 – Fluxograma simplificado do código python



Fonte: Adaptado de (CAVALCANTE, 2020)

A Figura 23 ilustra, simplificada, a organização da sequência de passos utilizadas no desenvolvimento do código. A estrutura completa do algoritmo encontra-se nos Apêndices A e B.

3.4 Base de Dados

Para o desenvolvimento deste trabalho, adotou-se a estratégia de realizar o agrupamento dos inversores, visando diminuir a complexidade de estudo das estações de transformadores. Foram utilizados, para isso, os dados de potência, colhidos por sensores instalados nos inversores das estações de transformadores, e os dados de irradiância solar, colhidos nas estações meteorológicas. Tal procedimento foi realizado, pois o entendimento, com relação ao processo de produção de energia elétrica, de uma planta fotovoltaica é de extrema importância para a dinâmica das energias renováveis.

Os dados fornecidos pelo SCADA, com granularidade de 1 hora, foram organizados, em arquivos de formato CSV, possibilitando a estruturação das informações de identificação dos inversores e dos parâmetros de entrada utilizados para a clusterização.

Figura 24 – Atributos da base de dados

BASE DE DADOS		
IDENTIFICAÇÃO DO INVERSOR [N° INV_TS]	IRRADIÂNCIA SOLAR [W/m ²]	POTÊNCIA DO INVERSOR [kW]

Fonte: Própria

A tabela apresentada na Figura 24 mostra 3 atributos: identificação do inversor, irradiância solar e potência do inversor. É importante destacar que os valores de potência são referentes às médias dos valores mensais dos inversores que compõe a TS e os valores de irradiância são referentes às médias dos valores mensais das WS.

Durante a análise dos dados, o processo de normalização não foi implementado, pois os parâmetros de potência e de irradiância utilizados na clusterização possuem ordens de grandeza similares. Além disso, o fato de terem sido analisados os valores médios dos atributos fez com que não houvesse a necessidade de tratamento de *outliers*.

4 ANÁLISE DOS RESULTADOS

O algoritmo utilizou os dados de entrada, descritos em 3.4, para analisar o desempenho dos inversores. As etapas de análise podem ser, resumidamente, descritas por:

1. Execução do algoritmo K-means para diferentes valores de K;
2. Determinação do valor ótimo de K pelo método Elbow;
3. Clusterização dos inversores por intermédio dos parâmetros de potência e de irradiância;
4. Análise da performance dos *clusters*.

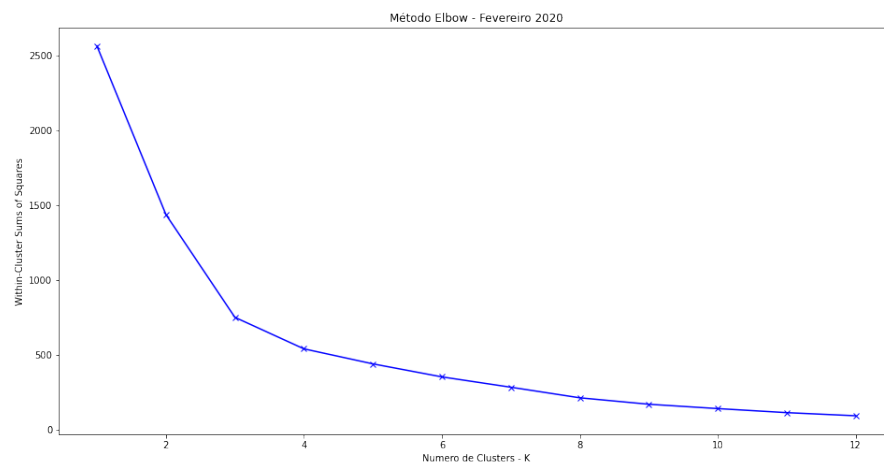
A partir de agora, as etapas descritas, acima, serão executadas e os resultados deste processo serão analisados e discutidos.

4.1 Determinação do Valor Ótimo de K

Para que o número de agrupamentos formados seja aquele que possibilite uma distribuição otimizada dos dados, utilizou-se o método Elbow. O algoritmo, em linguagem python, para a implementação deste método encontra-se descrito no Apêndice B.

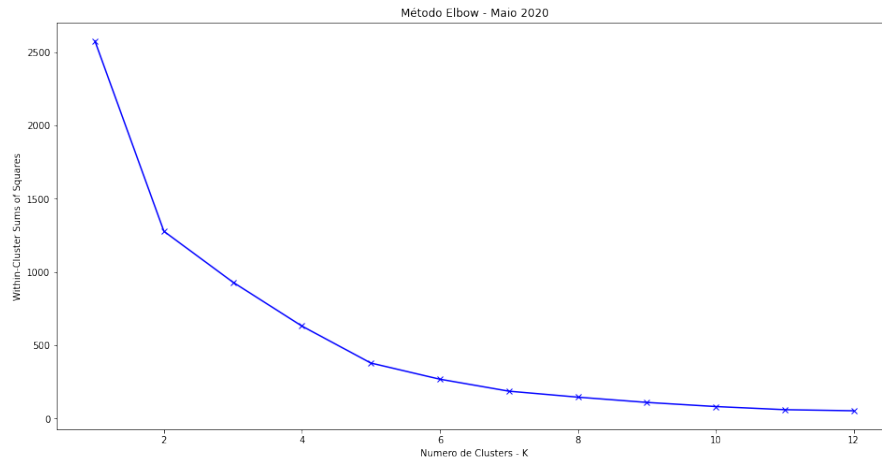
Pela visualização das Figuras 25 a 28, pode-se perceber que a soma dos erros quadrados (WCSS - *Within-Cluster Sums of Squares*) não varia, consideravelmente, para valores de $K > 4$. Portanto, o número de *clusters* escolhidos para a utilização do algoritmo K-means é $K = 4$.

Figura 25 – Método Elbow: fevereiro 2020



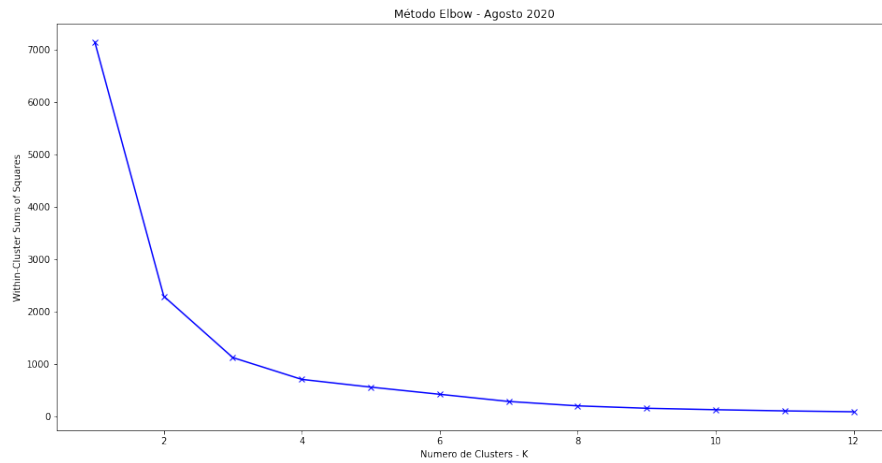
Fonte: Própria

Figura 26 – Método Elbow: maio 2020



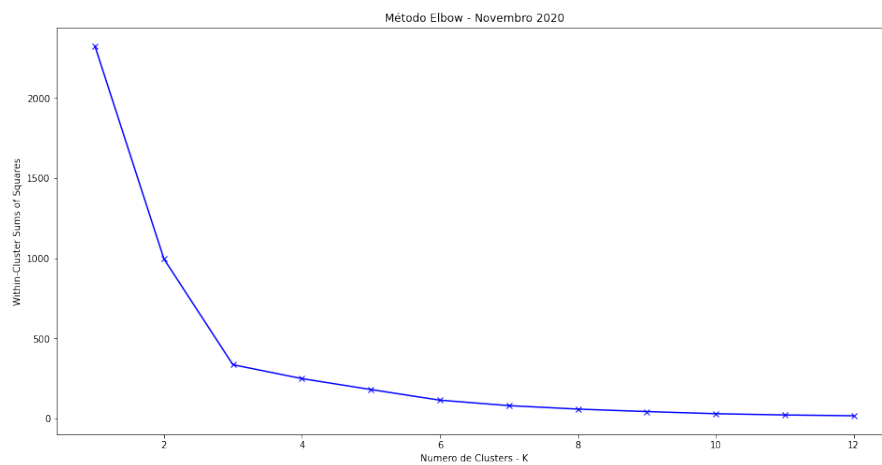
Fonte: Própria

Figura 27 – Método Elbow: agosto 2020



Fonte: Própria

Figura 28 – Método Elbow: novembro 2020



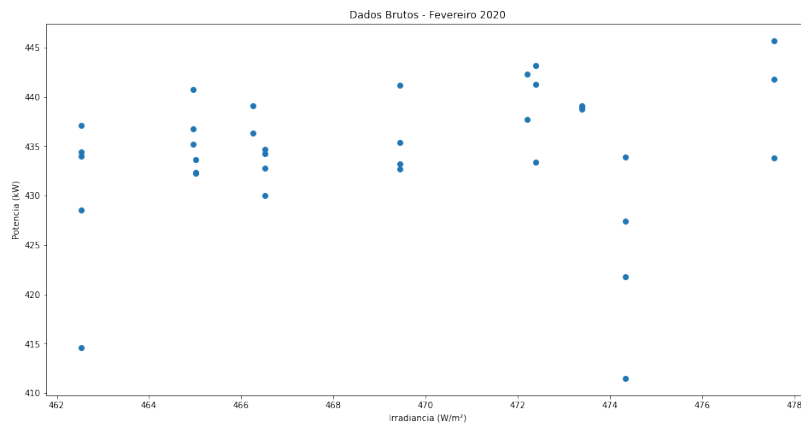
Fonte: Própria

4.2 Clusterização: Irradiância x Potência

As Figuras 29, 31, 33 e 35 permitem a visualização dos gráficos que mostram os inversores sem agrupamento, distribuídos segundo os parâmetros de irradiância e de potência. Além disso, as Figuras 30, 32, 34 e 36 representam a distribuição dos dados nos 4 *clusters* obtidos pelo método K-means, implementado segundo o algoritmo descrito no Apêndice A.

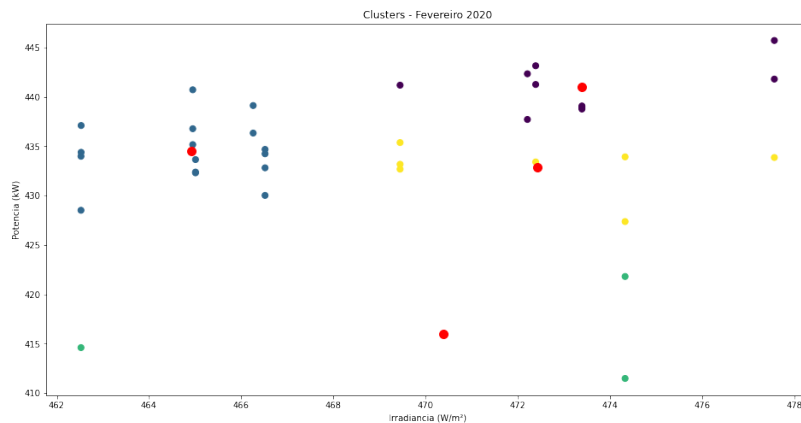
As análises foram feitas para os meses de fevereiro, maio, agosto e novembro do ano de 2020, pois são os meses centrais de cada trimestre do ano. Para o mês de maio, utilizou-se os dados de irradiância horizontal multiplicados por um fator de 1,3, pois os dados de irradiância inclinada (irradiância no plano dos painéis solares) estavam inconsistentes. O fator de 1,3 foi escolhido por ser um valor médio observado na relação entre a irradiância inclinada e a irradiância horizontal nos outros meses.

Figura 29 – Dados irradiância x potência: fevereiro 2020



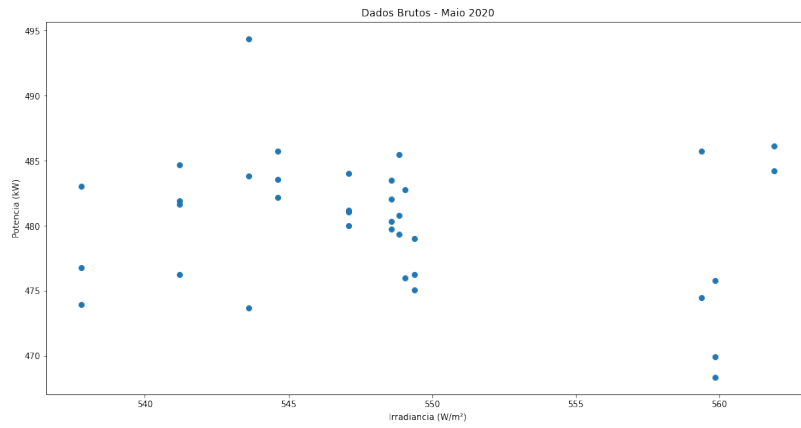
Fonte: Própria

Figura 30 – *Clusters* irradiância x potência: fevereiro 2020



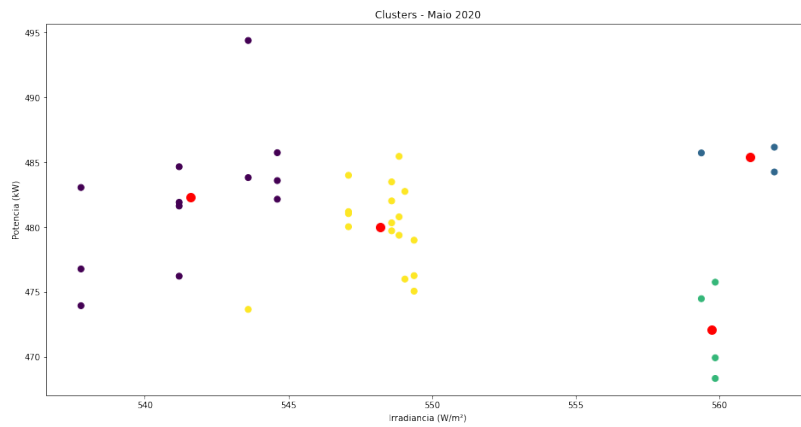
Fonte: Própria

Figura 31 – Dados irradiância x potência: maio 2020



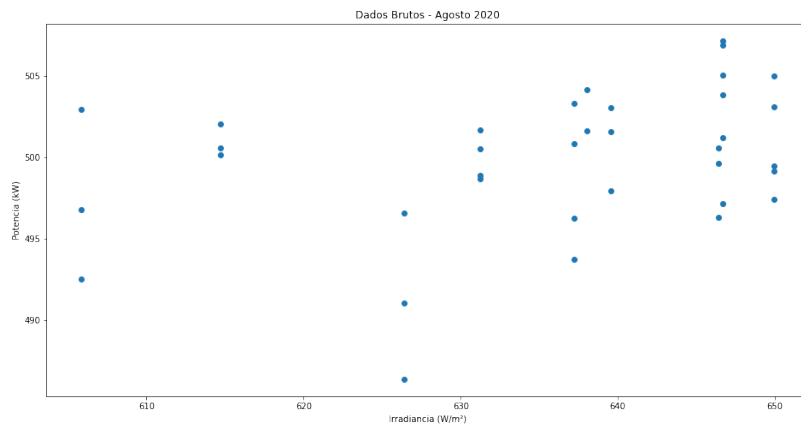
Fonte: Própria

Figura 32 – Clusters irradiância x potência: maio 2020



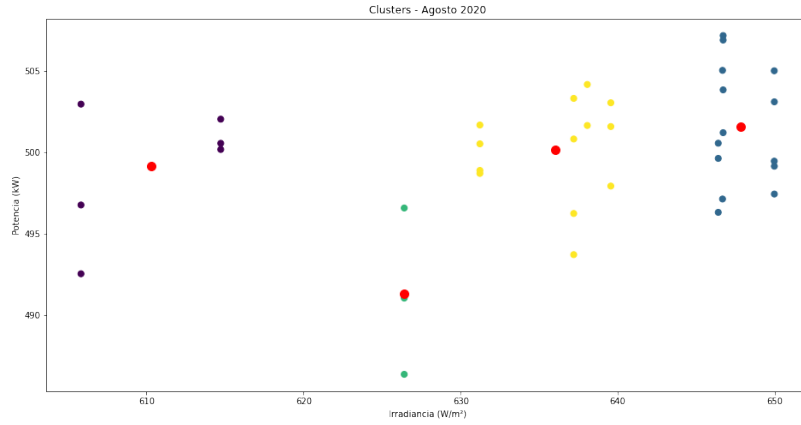
Fonte: Própria

Figura 33 – Dados irradiância x potência: agosto 2020



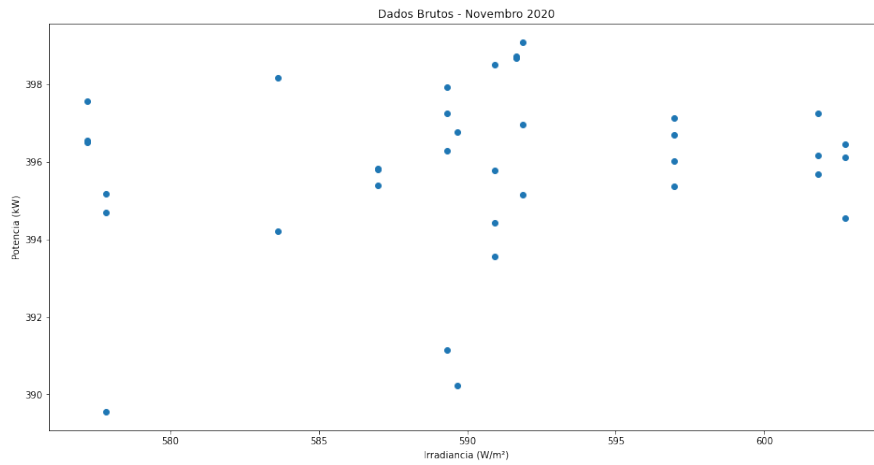
Fonte: Própria

Figura 34 – Clusters irradiância x potência: agosto 2020



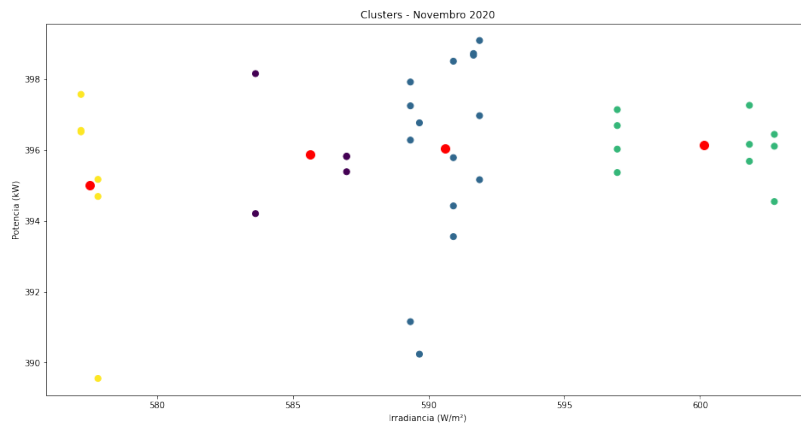
Fonte: Própria

Figura 35 – Dados irradiância x potência: novembro 2020



Fonte: Própria

Figura 36 – Clusters irradiância x potência: novembro 2020



Fonte: Própria

As Tabelas 6, 7, 8 e 9 indicam as características dos centroides para os meses de fevereiro, maio, agosto e novembro de 2020, respectivamente.

Tabela 6 – Centroides: fevereiro 2020

<i>Cluster</i>	Irradiância do Centroide [W/m ²]	Potência do Centroide [kW]
0	473,40	441,01
1	464,91	434,52
2	472,42	432,83
3	470,39	415,95

Fonte: Própria

Tabela 7 – Centroides: maio 2020

<i>Cluster</i>	Irradiância do Centroide [W/m ²]	Potência do Centroide [kW]
0	561,06	485,39
1	541,60	482,34
2	548,18	480,03
3	559,73	472,15

Fonte: Própria

Tabela 8 – Centroides: agosto 2020

<i>Cluster</i>	Irradiância do Centroide [W/m ²]	Potência do Centroide [kW]
0	647,80	501,56
1	636,04	500,16
2	610,28	499,16
3	626,41	491,31

Fonte: Própria

Tabela 9 – Centroides: novembro 2020

<i>Cluster</i>	Irradiância do Centroide [W/m ²]	Potência do Centroide [kW]
0	600,15	396,15
1	590,62	396,04
2	585,64	395,88
3	577,51	395,01

Fonte: Própria

Além disso, as Tabelas 10, 11, 12 e 13, possibilitam a visualização da distribuição dos inversores nos clusters. Para facilitar a acomodação das informações, os clusters serão organizados em ordem decrescente de potência dos centroides, ou seja, o cluster 0 será o que possui o centroide de maior potência e o cluster 3 será o que possui o centroide de menor potência.

Tabela 10 – *Clusters*: fevereiro 2020

<i>Cluster</i>	TS	Qtde.
0	1.1, 1.3, 1.4, 1.5, 1.6, 1.7, 2.1, 2.6, 2.7, 4.7	10
1	1.8, 1.9, 2.2, 3.1, 3.2, 3.3, 3.4, 3.5, 3.7, 3.8, 3.9, 4.1, 4.2, 4.3, 4.4, 4.6	16
2	1.2, 2.3, 2.4, 2.8, 3.6, 4.8, 4.9	7
3	2.5, 2.9, 4.5	3

Fonte: Própria

Tabela 11 – *Clusters*: maio 2020

<i>Cluster</i>	TS	Qtde.
0	2.2, 3.8, 3.9	3
1	1.6, 1.8, 1.9, 2.1, 2.3, 2.4, 2.5, 2.9, 3.1, 3.7, 4.1, 4.2	12
2	1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 2.6, 2.7, 2.8, 3.2, 3.3, 3.4, 3.5, 3.6, 4.7, 4.8, 4.9	17
3	4.3, 4.4, 4.5, 4.6	4

Fonte: Própria

Tabela 12 – *Clusters*: agosto 2020

<i>Cluster</i>	TS	Qtde.
0	1.4, 1.5, 1.8, 1.9, 2.2, 2.6, 2.7, 2.8, 3.2, 3.3, 3.4, 3.5, 3.7, 4.6	14
1	1.7, 2.3, 2.4, 2.5, 2.9, 3.1, 3.6, 3.8, 3.9, 4.1, 4.2, 4.8, 4.9	13
2	1.1, 1.2, 1.3, 1.6, 2.1, 4.7	6
3	4.3, 4.4, 4.5	3

Fonte: Própria

Tabela 13 – *Clusters*: novembro 2020

<i>Cluster</i>	TS	Qtde.
0	1.1, 1.2, 1.3, 1.6, 1.7, 2.1, 3.6, 4.7, 4.8, 4.9	10
1	1.4, 1.5, 2.3, 2.4, 2.5, 2.9, 3.1, 3.2, 3.3, 3.4, 3.5, 3.8, 3.9, 4.1, 4.2	15
2	2.2, 2.6, 2.7, 2.8, 4.6	5
3	1.8, 1.9, 3.7, 4.3, 4.4, 4.5	6

Fonte: Própria

4.3 Análise de Performance

Por meio das informações observadas nas Tabelas 6 a 13, é possível perceber que algumas TS, como 3.8 e 3.9, possuem uma tendência de ocupar *clusters* com centroides de maiores potências. Por outro lado, certas TS tendem a permanecer nos *clusters* com centroides de potências menores, a exemplo da TS 4.5.

Com o objetivo de se analisar os comportamentos mensais, ao longo do dia, da potência e da irradiância de cada agrupamento, optou-se por utilizar, como referência, a curva traçada pelos valores médios das TS que compõem o *cluster*. Para os meses de fevereiro e novembro, foram analisadas as curvas dos *clusters* com centroides de menores potências, enquanto que para os meses de maio e agosto foram analisados os *clusters* com centroides de

maiores potências, conforme indicado na Tabela 14.

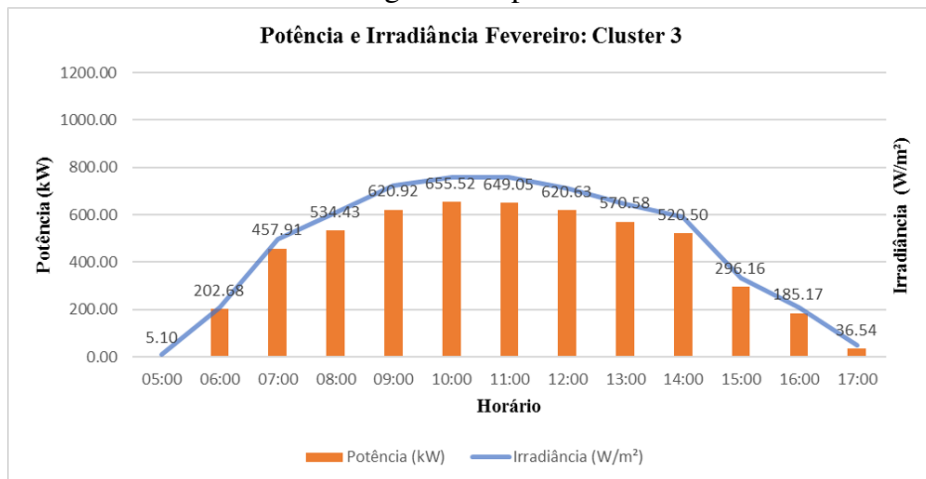
Tabela 14 – Dados de referência para curvas de comportamento dos *clusters*

Mês	Cluster	Irradiância do Centróide [W/m ²]	Potência do Centróide [kW]
Fevereiro	3	470,39	415,95
Mai	0	561,06	485,39
Agosto	0	647,80	501,56
Novembro	3	577,51	395,01

Fonte: Própria

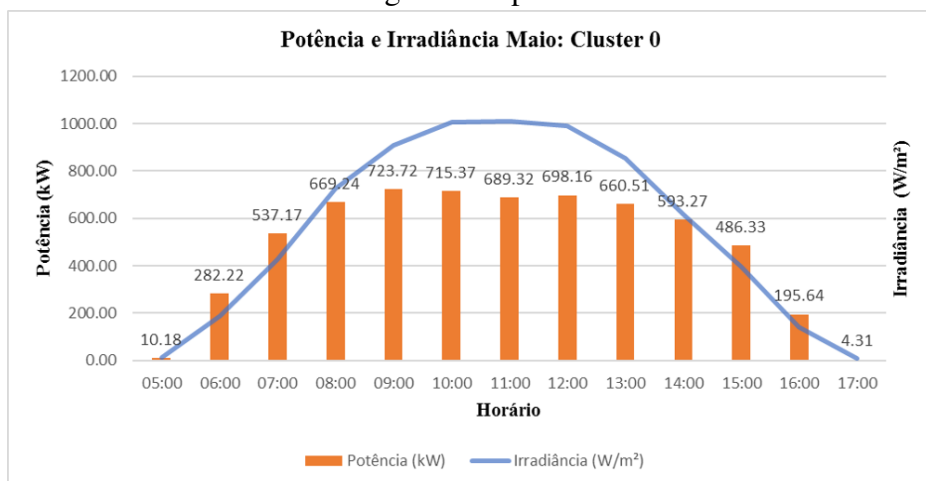
Os gráficos das Figuras 37 a 40 representam o padrão dos clusters seguindo as informações da Tabela 14.

Figura 37 – Potência e irradiância ao longo do dia para o *cluster* 3 do mês de fevereiro



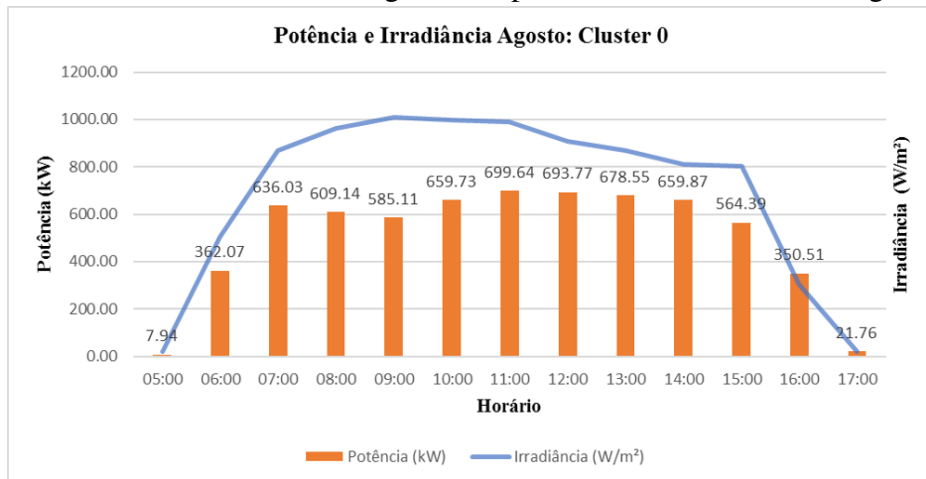
Fonte: Própria

Figura 38 – Potência e irradiância ao longo do dia para o *cluster* 0 do mês de maio



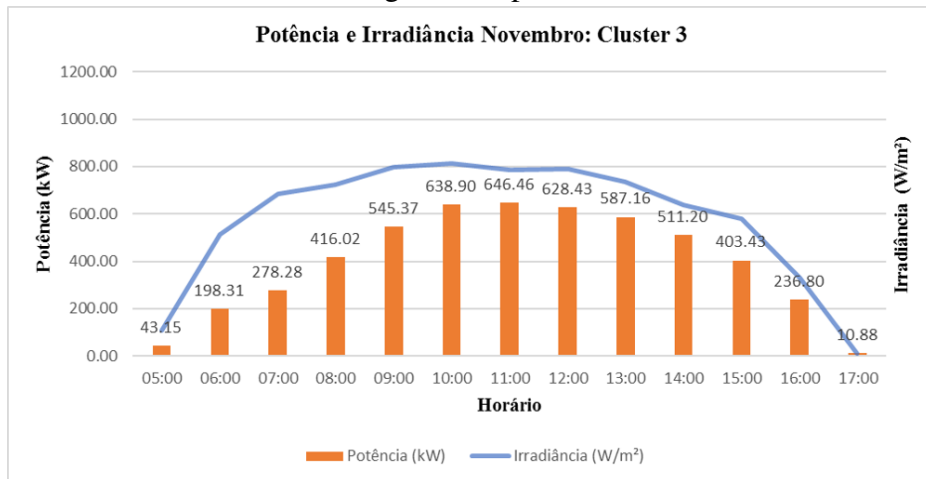
Fonte: Própria

Figura 39 – Potência e irradiância ao longo do dia para o *cluster* 0 do mês de agosto



Fonte: Própria

Figura 40 – Potência e irradiância ao longo do dia para o *cluster* 3 do mês de novembro



Fonte: Própria

Muitos fatores podem resultar na variação do padrão de comportamento dos inversores de sistemas fotovoltaicos, localizados em um mesmo complexo. Alguns desses fatores são: posicionamento das células solares, desgaste dos componentes, taxa de sujeira, falhas durante a operação, taxa pluviométrica, entre outros.

Nas Figuras 37 e 40, pode-se perceber que a potência possui um padrão de comportamento análogo à irradiância ao longo do dia. Por outro lado, na Figura 38, é possível perceber que a potência não acompanha o padrão de irradiância, destacando-se o período entre 10h e 12h, em que a irradiância é máxima e, ainda assim, a potência é, praticamente, constante.

Por este motivo, o estudo das informações do histórico de comportamento da planta fotovoltaica é de extrema importância para resultados mais precisos e exatos.

5 CONCLUSÕES E TRABALHOS FUTUROS

5.1 Conclusão

Em um sistema elétrico, é necessário elaborar estratégias que tornem a operação mais eficiente, ou seja, que maximizem o rendimento do sistema. Por esta razão, este trabalho buscou desenvolver uma metodologia que possibilitasse a realização de uma análise acerca do desempenho de usinas fotovoltaicas.

Nesse contexto, a metodologia escolhida consiste na execução do algoritmo K-means que, por meio do fornecimento de parâmetros de entrada, possibilita a formação de conjuntos, denominados *clusters*. Em posse dos dados de potência e de irradiância solar de 136 inversores, implementou-se, em linguagem Python, o algoritmo citado. Além disso, é importante salientar que o número ótimo de *clusters* foi determinado por intermédio da utilização do método Elbow. Desta forma, formaram-se 4 grupos de inversores de comportamentos diferentes entre si.

Com os resultados obtidos, foi possível identificar que, nem sempre, o aumento da magnitude de irradiância solar implica em um aumento da potência fornecida pelo inversor. Observando-se as tabelas 6 a 9, pode-se perceber que a irradiância dos *clusters* é maior para os meses de agosto e novembro, pois, no Ceará, o segundo semestre é mais ensolarado, destacando-se os meses de agosto a outubro. Isso explica a maior irradiância indicada pelos *clusters* em agosto. Em fevereiro, a irradiância é menor do que em novembro, sendo esta situação invertida com relação aos valores de potência, pois a maior taxa de pluviosidade do mês de fevereiro, em relação ao mês de novembro, possibilita uma limpeza natural dos módulos fotovoltaicos, elevando a taxa de conversão de energia. Além disso, o alto índice de ventos, no segundo semestre, favorece o transporte de particulados que se acumulam nos painéis, diminuindo, por exemplo, a taxa de conversão de novembro em relação a fevereiro. Por fim, as figuras 37 a 40 ilustram as curvas de potência e irradiância dos clusters, ratificando o comportamento esperado, de acordo com as características das células fotovoltaicas discutidas na Seção 2.2.2.

O tratamento dos dados, fornecidos pelo SCADA, representou a principal dificuldade encarada nesse trabalho, pois, com o intuito de não gerar resultados inconsistentes, deve-se realizar o polimento da base de dados, excluindo informações incoerentes, como erros de leitura. Diante disso, julga-se que esse trabalho contribui para a difusão das ideias e conhecimentos relacionados aos conceitos de clusterização, bem como da utilização desta técnica na análise de desempenho de sistemas fotovoltaicos.

5.2 Trabalhos Futuros

Com o intuito de ratificar as ideias discutidas nesse trabalho, abordando outras metodologias, e, além disso, expandir o estudo para outras aplicações, os seguintes tópicos são deixados como sugestões para trabalhos futuros:

- Utilização de outros parâmetros, como sujidade e regime de manutenção, na implementação do método K-means.
- Desenvolvimento de outras metodologias de agrupamento, como algoritmos hierárquicos, para comparação de resultados.
- Utilização do método K-means para gerar curvas centroides, visando avaliar a dispersão de potência e a dispersão entre os centroides, similar a (CHRISTOFORIDIS *et al.*, 2013).
- Análise das possíveis causas que diminuem a eficiência de módulos fotovoltaicos, a exemplo do estudo de sujidade, como descrito em (REZENDE, 2018).

REFERÊNCIAS

- ARORA, P.; DEEPALI; VARSHNEY, S. Analysis of k-means and k-medoids algorithm for big data. **Procedia Computer Science**, Nagpur, INDIA, 2016. 1st International Conference on Information Security Privacy 2015. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050916000971>>.
- BOYLESTAD, R. L.; NASHELSKY, L. **Dispositivos eletrônicos: e teoria de circuitos**. 11. ed. São Paulo: Pearson Education do Brasil, 2013.
- BRITO, M. A. G. d. **Inversores integrados monofásicos e trifásicos para aplicações fotovoltaicas: técnicas para obtenção de mppt, detecção e proteção de ilha, sincronização e paralelismo com a rede de distribuição de energia elétrica**. Tese (Doutorado) — Universidade Estadual Paulista, 2013.
- CAVALCANTE, D. B. Bacharelado em Engenharia Elétrica, **Análise de desempenho de parques eólicos por meio de clusterização de aerogeradores**. 2020.
- CHRISTOFORIDIS, G.; PAPADOPOULOS, T.; PANAPAKIDIS, I.; PAPAGIANNIS, G. Pv power clustering as a means to evaluate energy storage options. In: . [S.l.: s.n.], 2013.
- CLAUDIO, J. B. Bacharelado em Engenharia Elétrica, **Análise de padrões de consumidores de energia fotovoltaica por meio de aprendizado não supervisionado**. 2018.
- COELHO, R. F. **Estudo dos conversores Buck e Boost aplicados ao rastreamento de máxima potência de sistemas solares fotovoltaicos**. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal de Santa Catarina, Santa Catarina, 2008.
- FENSTERSEIFER, H. **Uma abordagem de analítica visual e clusterização para avaliação da qualidade da distribuição de energia elétrica**. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal do Rio Grande do Sul, 2008.
- HOLMAN, J. P. **Heat Transfer**. 10. ed. [S.l.]: McGraw-Hill Education, 2009.
- INTRODUCAO a clusterização e os diferentes métodos. 2018. Portal Data Science. Disponível em: <<https://portaldatascience.com/introducao-a-clusterizacao-e-os-diferentes-metodos/>>. Acesso em: 29 dez. 2021.
- JAIN, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, 2010. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865509002323>>.
- JONES, A. Z. **The Photoelectric Effect**. 2020. Disponível em: <<https://www.thoughtco.com/the-photoelectric-effect-2699352>>. Acesso em: 29 dez. 2021.
- LACHI, R. L.; ROCHA, H. V. d. **Aspectos básicos de clustering: conceitos e técnicas**. [S.l.], 2005.
- LIMA, A. M. C. **Balanco de energia em painel fotovoltaico: um estudo experimental**. Dissertação (Mestrado em Engenharia Mecânica) — Universidade de Brasília, Brasília - DF, 2019.
- MORAES, C. **Célula fotovoltaica: tudo que você precisa saber**. 2020. Disponível em: <<https://eletronicadepotencia.com/celula-fotovoltaica/>>. Acesso em: 29 dez. 2021.

- MOREIRA, J. R. S. **Energias renováveis, geração distribuída e eficiência energética**. 1. ed. Rio de Janeiro: LTC, 2019.
- NETO, M. R. B.; CARVALHO, P. C. M. d. **Geração de Energia Elétrica: Fundamentos**. 1. ed. São Paulo: Érica, 2012.
- OLIVEIRA, T. B. S. d. **Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, 2008.
- PARK, S.; PARK, Y. B. Photovoltaic power data analysis using hierarchical clustering. In: **2018 International Conference on Information Networking (ICOIN)**. [S.l.: s.n.], 2018.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, 2011.
- PINHO, J. T.; GALDINO, M. A. **Manual de engenharia para sistemas fotovoltaicos**. Rio de Janeiro, 2014.
- REZENDE, V. G. S. de. **Análise das perdas de produtividade em geradores fotovoltaicos por efeito de sujidade**. Dissertação (Mestrado em Ciências da Energia) — Instituto de Energia e Ambiente - USP, 2018.
- SANTANA, F. **Entenda o algoritmo k-means e saiba como aplicar essa técnica**. 2017. Disponível em: <<https://minerandodados.com.br/entenda-o-algoritmo-k-means/>>. Acesso em: 29 dez. 2021.
- SEDRA, A. S.; SMITH, K. C. **Microeletrônica**. 5. ed. São Paulo: Pearson Prentice Hall, 2007.
- SEMICONDUCTORES. 2007. EIECTRONICA PT. Disponível em: <<https://www.electronica-pt.com/semicondutores>>. Acesso em: 29 dez. 2021.
- SEMICONDUCTORES. 2018. MSPC. Disponível em: <<https://www.mspc.eng.br/dir60/semic01.php>>. Acesso em: 29 dez. 2021.
- SMITH, C. J.; FORSTER, P. M.; CROOK, R. Global analysis of photovoltaic energy output enhanced by phase change material cooling. **Applied Energy**, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261914003237>>.
- YOUNG, H. D.; FREEDMAN, R. A. **Física IV: Sears e Zemansky: ótica e física moderna**. 14. ed. São Paulo: Pearson Education do Brasil, 2016.

APÊNDICE A – CÓDIGO PYTHON K-MEANS

Código-fonte 1 – Código Python K-Means

```
1
2 #BIBLIOTECAS
3 import numpy as np
4 import pandas as pd
5 import seaborn as sb
6 from matplotlib import pyplot as plt
7 from sklearn.cluster import KMeans
8 %matplotlib inline
9
10 #BASE DE DADOS A ANALISAR
11 inversores = pd.read_csv ('/content/novembro-data.csv', sep
    = ';')
12 inversores_data = inversores.drop (['Elemento'], axis = 1)
13 inversores_data.head()
14
15 #PLOTAGEM DOS DADOS BRUTOS
16 plt.figure(figsize=(16,8))
17 plt.title ('Dados Brutos - Novembro 2020')
18 plt.scatter (inversores_data["Irradiancia"],
    inversores_data["Potencia (kW)"])
19 plt.xlabel ('Irradiancia (W/m )')
20 plt.ylabel ('Potencia (kW)')
21 plt.show()
22
23 #CLUSTERIZANDO
24 kmeans = KMeans(n_clusters = 4, max_iter=300, init = '
    random',) #Cria o modelo para K = 4
25 kmeans.fit (inversores_data) #Agrupa os elementos em 4
    clusters
```

```
26 y_kmeans = kmeans.predict (inversores_data)
27
28 #CLUSTERS
29 centers = kmeans.cluster_centers_
30 plt.figure(figsize=(16,8))
31 plt.title ('Clusters - Noviembre 2020')
32 plt.scatter (inversores_data["Irradiancia"],
               inversores_data["Potencia (kW)"], c = y_kmeans, s = 50)
33 plt.scatter (centers[:, 0], centers[:, 1], c = 'red', s =
               100)
34 plt.xlabel ('Irradiancia (W/m )')
35 plt.ylabel ('Potencia (kW)')
36 plt.show()
37
38 #DADOS DOS CLUSTERS
39 print(y_kmeans)
40 inversores ['K_Clusters'] = kmeans.labels_
41 inversores
42
43 #DADOS DOS CENTRODIES
44 print(kmeans.cluster_centers_)
```

APÊNDICE B – CÓDIGO PYTHON ELBOW

Código-fonte 2 – Código Python Elbow

```
1
2 #METODO ELBOW - Within-Cluster Sums of Squares
3 wcss = []
4 K = range (1, 13)
5 for k in K:
6 kmeansModel = KMeans(n_clusters=k, max_iter=300)
7 kmeansModel.fit(inversores_data) #Aplicando o k-means ao
   conjunto da base de dados
8 wcss.append(kmeansModel.inertia_)
9
10
11 #GRAFICO METODO ELBOW - Within-Cluster Sums of Squares
12 plt.figure (figsize=(16,8))
13 plt.plot(K, wcss, 'bx-')
14 plt.title ('Metodo Elbow - Novembro 2020')
15 plt.xlabel ('Numero de Clusters - K')
16 plt.ylabel ('Within-Cluster Sums of Squares')
17 plt.show()
```