



**UNIVERSIDADE ESTADUAL DE CAMPINAS**

**FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO**

**JORGE WILLIAMS ARAGUNDI RODRIGUEZ**

**MODELO CA PARA A ALOCAÇÃO ÓTIMA DE GERAÇÃO FOTOVOLTAICA EM REDES DE DISTRIBUIÇÃO CONSIDERANDO INCERTEZAS**

**CAMPINAS**

**2022**

**JORGE WILLIAMS ARAGUNDI RODRIGUEZ**

**MODELO CA PARA A ALOCAÇÃO ÓTIMA DE GERAÇÃO FOTOVOLTAICA EM  
REDES DE DISTRIBUIÇÃO CONSIDERANDO INCERTEZAS**

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na área de Energia Elétrica.

**Orientador: Prof. Dr. Carlos Alberto de Castro Junior**

**Coorientador: Prof. Dr. Miguel Alberto Torres Rodriguez**

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO JORGE WILLIAMS ARAGUNDI RODRIGUEZ, E ORIENTADA PELO PROF. DR. CARLOS ALBERTO DE CASTRO JUNIOR

**CAMPINAS  
2022**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Elizangela Aparecida dos Santos Souza - CRB 8/8098

Ar12m Aragundi Rodriguez, Jorge Williams, 1967-  
Modelo CA para alocação ótima de geração fotovoltaica em redes de distribuição considerando incertezas. / Jorge Williams Aragundi Rodriguez. – Campinas, SP : [s.n.], 2022.

Orientador: Carlos Alberto de Castro Junior.  
Coorientador: Miguel Alberto Torres Rodriguez.  
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Sistemas de energia solar fotovoltaica. 2. Energia solar. 3. Desenvolvimento energético- Modelo matemático. 4. Energia elétrica- Sistemas. I. Castro Junior, Carlos Alberto de, 1960-. II. Torres Rodriguez, Miguel Alberto, 1987-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** AC model for optimal allocation of photovoltaic generation in distribution networks considering uncertainties

**Palavras-chave em inglês:**

Photovoltaic solar energy systems

Solar energy

Energy development- Mathematical model

Electrical energy- Systems

**Área de concentração:** Energia Elétrica

**Titulação:** Doutor em Engenharia Elétrica

**Banca examinadora:**

Carlos Alberto de Castro Junior [Orientador]

Luiz Carlos Pereira da Silva

Lia Toledo Moreira Mota

Bibiana Maitê Petry Ferraz

Carmen Karina Vaca Ruiz

**Data de defesa:** 06-05-2022

**Programa de Pós-Graduação:** Engenharia Elétrica

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-6010-4886>

- Currículo Lattes do autor: <http://lattes.cnpq.br/1261874828887206>

## COMISSÃO JULGADORA – TESE DOUTORADO

**Candidato:** Jorge Williams Aragundi Rodriguez RA: 142362

**Data da Defesa:** 6 de maio de 2022

**Título da Tese:** "Modelo CA para a alocação ótima de geração fotovoltaica em redes de distribuição considerando incertezas".

Prof. Dr. Carlos Alberto de Castro Junior

Prof. Dr. Luiz Carlos Pereira da Silva

Profa. Dra. Lia Toledo Moreira Mota

Dra. Bibiana Maitê Petry Ferraz

Dra. Carmen Karina Vaca Ruiz

A ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós- Graduação da Faculdade de Engenharia Elétrica e de Computação.

*Dedico esta tese ao meu pai Jorge (in memoriam) e a todos os membros da minha família cujo incentivo e apoio foram essenciais nesta longa jornada: a Elsita, minha esposa Cristina, minha filha Karen, meu filho Pedro, meus irmãos e irmãs, sem vocês não seria possível alcançar este objetivo.*

## AGRADECIMENTOS

A Deus em primeiro lugar, que me deu forças quando necessitei, e pela felicidade de estar aqui no fim deste trabalho.

À UNICAMP, pela oportunidade de realização desta tese.

Ao meu orientador Professor Dr. Carlos Castro, que me acolheu e acreditou neste trabalho, pela orientação, ajuda e incentivo fornecido durante o desenvolvimento desta tese.

Ao meu coorientador Professor Dr. Miguel Torres, pela ajuda, apoio e orientação, e transmissão de seus conhecimentos ao longo deste trabalho.

Ao meu pai Jorge, que foi sempre meu exemplo a seguir, infelizmente o senhor não vai estar fisicamente, mas vai estar sempre em nossos corações.

A minha família, Elsita, meus irmãos, minhas irmãs, pelo afeto, carinho e incentivo permanente.

A minha esposa Cristina, minha filha Karen, meu filho Pedro, pelo amor, carinho e apoio, fundamentais a longo desta caminhada.

A todos os professores do Programa de Pós-Graduação em Engenharia Elétrica da FEEC, pela dedicação ao curso que me foi oferecido.

A todos os funcionários da Pós-Graduação da FEEC, pelo auxílio e esclarecimento.

A todos os colegas de mestrado e doutorado, pelo convívio, companheirismo e amizade ao longo desta jornada, especialmente a meus amigos do LADIME.

À ESPOL, pela confiança, apoio e ajuda permanente para a realização deste trabalho.

À SENESCYT, pela bolsa de estudos concedida.

“A educação é o grande motor do desenvolvimento pessoal”

(Nelson Mandela)

## RESUMO

A inserção de unidades de geração distribuída, cuja fonte primária de energia vem de recursos renováveis como a eólica e a solar, vem ganhando cada vez mais relevância. Este trabalho visa minimizar o investimento na instalação de unidades de geração fotovoltaica e a redução das perdas de energia em redes de distribuição através de um modelo convexo não linear, o qual foi resolvido utilizando uma linguagem de otimização algébrica. Neste trabalho, foi considerada a variabilidade das fontes de energia solar, utilizando uma técnica de aprendizado de máquina. O algoritmo *K-means* foi utilizado para tratar os dados de entrada para o modelo, e um algoritmo foi desenvolvido para resolver o problema de otimização, sendo possível obter a localização ótima dos painéis fotovoltaicos bem como o número de unidades, buscando minimizar as perdas de potência ativa na rede de distribuição. Os resultados obtidos mostram a efetividade com que a inteligência artificial pode lidar com a variabilidade das grandezas do modelo de otimização de geração distribuída. O método proposto foi avaliado através de simulações utilizando dados de redes da literatura e redes de distribuição reais.

**Palavras-chave:** Aprendizagem de máquina. Algoritmo *K-means*. Redução de perdas. Geração distribuída. Energia solar, Geração fotovoltaica.

## ABSTRACT

*The insertion of distributed generation units, whose primary energy source comes from renewable resources such as wind and solar, has been gaining more and more relevance. This work aims to minimize the investment in photovoltaic generation and energy losses reduction in distribution networks through a non-linear convex model, which was solved by an algebraic optimization language. In this work, the solar energy source variability was considered, by incorporating the concept of machine learning. The K-means algorithm was used to process the input data for the model, and an algorithm was developed to solve the optimization problem in order to obtain the optimal location of photovoltaic panels as well as the number of units, aiming to minimize the active power losses in the distribution network. The obtained results show that artificial intelligence can deal with the variability of distributed generation optimization model. The proposed method was evaluated through simulations involving both test and real-life distribution networks.*

**Keywords:** *Machine Learning. K-mean Algorithm. Losses Reduction. Distributed Generation. Solar Energy, Photovoltaic Generation.*



## LISTA DE FIGURAS

<b>Figura 2.1</b> – Fluxograma do método proposto. Fonte: Autor.....	23
<b>Figura 2.2</b> – Grandezas dentro do sistema de distribuição radial. [42].....	25
<b>Figura 2.3</b> – Representação gráfica de um modelo com região factível não convexa do problema de minimização com duas variáveis de decisão. [42].....	31
<b>Figura 2.4</b> – Relaxação convexa com gap de relaxação zero. [42].....	32
<b>Figura 2.5</b> – Relaxação convexa com gap de relaxação não zero. [42].....	32
<b>Figura 2.6</b> – Solução do modelo (48)-(51) não relaxado. [42].....	33
<b>Figura 2.7</b> – Solução do modelo (48)-(49);(51)-(52) relaxado. [42].....	34
<b>Figura 2.8</b> – Solução do modelo (53)-(57) não relaxado. [42].....	34
<b>Figura 2.9</b> – Solução do modelo (53)-(54);(56)-(58) relaxado. [42].....	35
<b>Figura 2.10</b> – Circuito ilustrativo do SDEE. [42].....	36
<b>Figura 2.11</b> – Solução do modelo (60)-(62) não relaxado. [42].....	36
<b>Figura 2.12</b> – Solução do modelo (60)-(61); (63) relaxado. [42].....	37
<b>Figura 2.13</b> – A aprendizagem de máquina utilizada em diferentes campos. [49].....	41
<b>Figura 2.14</b> – Modelo de aprendizagem supervisionado. [49].....	42
<b>Figura 2.15</b> – Modelo de aprendizagem não supervisionado. [49].....	43
<b>Figura 2.16</b> – Modelo de aprendizagem reforçada. [49].....	43
<b>Figura 2.17</b> – Algoritmo de agrupamento <i>K-means</i> . [37].....	47
<b>Figura 2.18</b> – Dados agrupados e dados não agrupados. [36].....	48
<b>Figura 2.19</b> – Representação técnica de agrupamento hierárquico aglomerado. [36].....	49
<b>Figura 2.20</b> – Dendograma. [36].....	51
<b>Figura 2.21</b> – Gráfico do Cotovelo Fonte: Autor. ....	52
<b>Figura 2.22</b> – Fluxograma geral de modelagem dos módulos solares. Fonte: Autor. ....	54
<b>Figura 2.23</b> – Modelo do módulo solar utilizando o Simulink. Fonte: Autor. ....	54
<b>Figura 2.24</b> – Dados fornecidos pela empresa de distribuição de energia elétrica [60]. ....	58
<b>Figura 2.25</b> – Código de dados semiprocessados. [60].....	59
<b>Figura 2.26</b> – Código para <i>upload</i> de dados semiprocessados [60]. ....	59
<b>Figura 2.27</b> – Código para unificar os dados semiprocessados [60]. ....	60
<b>Figura 2.28</b> – Código para o filtrado de dados [60]. ....	60
<b>Figura 2.29</b> – Código para a criação do histograma [60]. ....	61
<b>Figura 2.30</b> – Histograma [60]. ....	61

<b>Figura 2.31</b> – Código para eliminação de dados ou leituras repetidas [60].	62
<b>Figura 2.32</b> – Código para filtragem de <i>outliers</i> [60].	62
<b>Figura 2.33</b> – Código para normalização de dados [60].	63
<b>Figura 2.34</b> – Código para usar o algoritmo <i>K-means</i> [60].	64
<b>Figura 2.35</b> – Código para criar <i>clusters</i> com $K=3$ [60].	68
<b>Figura 2.36</b> – Resumo da estratificação com $K=3$ [60].	68
<b>Figura 3.1</b> – Código para o processamento de dados da aprendizagem de máquina. Fonte: Autor.	71
<b>Figura 3.2</b> – Dados de um dia para aplicação da normalização. Fonte: Autor.	71
<b>Figura 3.3</b> – Base de dados para obter os valores de irradiação e temperatura. Fonte: Autor.	72
<b>Figura 3.4</b> – Leitura e processamento dos dados de irradiação e temperatura. Fonte: Autor.	73
<b>Figura 3.5</b> – Normalização dos dados no horário especificado. Fonte: Autor.	73
<b>Figura 3.6</b> – Código para obtenção do cotovelo. Fonte: Autor.	74
<b>Figura 3.7</b> – Algoritmo para obter dias representativos ( <i>K-means</i> ). Fonte: Autor.	74
<b>Figura 3.8</b> – Identificação das horas de cada <i>cluster</i> . Fonte: Autor.	75
<b>Figura 3.9</b> – Obtenção dos dados de irradiação de cada <i>cluster</i> . Fonte: Autor.	75
<b>Figura 3.10</b> – Técnica do cotovelo. [64]	76
<b>Figura 3.11</b> – Agrupamento de amostras aleatórias de acordo com dois <i>clusters</i> . [65]	77
<b>Figura 3.12</b> – Dados brutos fornecidos pela empresa de distribuição do setor elétrico. Fonte: Autor.	79
<b>Figura 3.13</b> – Cotovelo para todos os dados históricos. Fonte: Autor.	79
<b>Figura 3.14</b> – Diagrama de dispersão para todos os dados históricos. Fonte: Autor.	80
<b>Figura 3.15</b> – Cotovelo para dados históricos limitados das 6h às 18h. Fonte: Autor.	81
<b>Figura 3.16</b> – Diagrama de dispersão para dados históricos limitados das 6h às 18h. Fonte: Autor.	81
<b>Figura 3.17</b> – Cotovelo para dados históricos com a presença de chuva. Fonte: Autor.	82
<b>Figura 3.18</b> – Diagrama de dispersão para os meses com chuva. Fonte: Autor.	83
<b>Figura 3.19</b> – Diagrama de dispersão para os meses sem chuva. Fonte: Autor.	83
<b>Figura 3.20</b> – Modelo célula solar (irradiação solar e temperatura na entrada). Fonte: Autor.	84
<b>Figura 3.21</b> – Curva P-V do módulo solar. Fonte: Autor.	85
<b>Figura 3.22</b> – Curva I-V do módulo solar. Fonte: Autor.	85
<b>Figura 3.23</b> – Sistema de 22 barras. Fonte: Autor.	87

## LISTA DE TABELAS

<b>Tabela 2.1</b> – Amostras para o exemplo simples do funcionamento do <i>K-means</i> .....	65
<b>Tabela 2.2</b> – Centroides para o exemplo. ....	66
<b>Tabela 2.3</b> – Distâncias das amostras até o centroide.....	66
<b>Tabela 2.4</b> – Centroides recalculados. ....	66
<b>Tabela 2.5</b> – Resumo da segunda iteração.....	67
<b>Tabela 2.6</b> – Resumo da terceira iteração.....	67
<b>Tabela 3.1</b> – Exemplo do processamento de dados e sua normalização. ....	72
<b>Tabela 3.2</b> – Resultados considerando todos os dados históricos. ....	80
<b>Tabela 3.3</b> – Resultados para os dados históricos limitados às 6h-18h.....	82
<b>Tabela 3.4</b> – Resultados para dados históricos de acordo com a presença de chuva. ....	84
<b>Tabela 3.5</b> – Resultados para rede teste de 22 barras, $k=2$ , sem alocação de plantas solares [13]. .....	87
<b>Tabela 3.6</b> – Resultados para rede teste de 22 barras, $k=2$ , possibilidade de alocar 22 plantas solares [13]. ....	87
<b>Tabela 3.7</b> – Resultados para rede teste de 22 barras, $k=2$ , possibilidade de alocar plantas solares em todas as barras, exceto na 134 e na 118.....	88
<b>Tabela 3.8</b> – Resultados para rede teste de 22 barras, $k=3$ , caso base [13]. ....	88
<b>Tabela 3.9</b> – Resultados para rede teste de 22 barras, $k=2$ , caso com demanda diminuída.....	89
<b>Tabela 3.10</b> – Resultados para rede teste de 22 barras, $k=2$ , menor demanda, limite de corrente aumentado no ramal 103-107. ....	89
<b>Tabela 3.11</b> – Resultados para rede teste de 22 barras, $k=2$ , menor demanda, menor fluxo corrente ramal 103-107 e 78-90.....	90
<b>Tabela 3.12</b> – Resultados para rede teste de 22 barras, $k=2$ , menor demanda, magnitude fixa ( $0,97 \cdot V_{nom}$ ) na barra 90.....	90
<b>Tabela 3.13</b> – Resultados para rede real de 137 barras $k=2$ .....	91
<b>Tabela 3.14</b> – Resultados para rede real de 137 barras $k=3$ .....	92
<b>Tabela A.1</b> – Informações das barras para o sistema teste de 22 barras.....	104
<b>Tabela A.2</b> – Informações dos circuitos para o sistema teste de 22 barras. ....	105

## LISTA DE ABREVIATURAS E SIGLAS

### *Índices e conjuntos*

<b><math>D</math></b>	Horizonte em anos
<b><math>C</math></b>	Conjunto de circuitos
<b><math>N</math></b>	Número de barras
<b><math>T</math></b>	Conjunto de intervalos
<b><math>i, j</math></b>	Índice de barras
<b><math>ij</math></b>	Índice de circuitos
<b><math>k</math></b>	Índice de módulos fotovoltaicos
<b><math>t</math></b>	Índice de tempo
<b><math>L</math></b>	Ramos
<b><math>\Omega_b</math></b>	Conjunto de nós
<b><math>\Omega_d</math></b>	Conjunto de níveis de demanda

### *Parâmetros*

<b><math>\gamma_{t,d}^G</math></b>	Custo de energia no instante <b><math>t</math></b> no ano <b><math>d</math></b>
<b><math>\gamma^{PV}</math></b>	Custo de instalação dos módulos PV
<b><math>\Delta_d</math></b>	Duração de anos (8760h)
<b><math>\Delta_t</math></b>	Duração do intervalo de tempo <b><math>t</math></b>
<b><math>\phi_{min}, \phi_{max}</math></b>	Limites do fator de potência da subestação
<b><math>\varphi^{PV}</math></b>	Fator de potência do módulo fotovoltaico
<b><math>f_{t,d}^{Gpv}</math></b>	Fator de geração fotovoltaica no ano <b><math>d</math></b> e no instante <b><math>t</math></b>
<b><math>I_{max}</math></b>	Magnitude máxima de corrente permitida no circuito
<b><math>IC^{PV}</math></b>	Máximo investimento em módulos fotovoltaicos
<b><math>N^{PV}</math></b>	Número de módulos fotovoltaicos
<b><math>P_{i,t,d}^{Ld}</math></b>	Demanda de potência ativa demandada na barra <b><math>i</math></b> , no instante <b><math>t</math></b> e no ano <b><math>d</math></b>
<b><math>P^{-PV}</math></b>	Capacidade de potência ativa do módulo PV
<b><math>Q_{i,t,d}^{Ld}</math></b>	Demanda de potência reativa demandada na barra <b><math>i</math></b> , no instante <b><math>t</math></b> e no ano <b><math>d</math></b>
<b><math>R_{ij}, X_{ij}, Z_{ij}</math></b>	Resistência, indutância, impedância no circuito <b><math>ij</math></b>
<b><math>V_{min}, V_{max}</math></b>	Limites de tensão

$V_{nom}$	Magnitude de tensão nominal
$PV$	Número de painéis fotovoltaicos
$\theta$	Ângulo de fase nodal

*Variáveis contínuas*

$I_{ij,t,d}^{sqr}$	Quadrado da magnitude do fluxo de corrente no circuito $ij$
$P_{ij,t,d}$	Fluxo de potência ativa no circuito $ij$ no instante $t$ e no ano $d$
$P_{i,t,d}^{PV}$	Geração de energia fotovoltaica na barra $i$ , no instante $t$ e no ano $d$
$P_{t,d}^S$	Potência fornecida pela subestação no instante $t$ e no ano $d$
$Q_{ji,t,d}$	Fluxo de potência reativa no circuito $ij$ no instante $t$ e no ano $d$
$Q_{i,t,d}^{PV}$	Potência reativa fornecida pelos módulos fotovoltaicos
$Q_{t,d}^S$	Potência reativa fornecida pela subestação
$V_{i,t,d}^{sqr}$	Quadrado da magnitude da tensão na barra $i$ , no instante $t$ e no ano $d$
$\delta r_{i,j}$	Custo monetário por perda técnica em linhas

*Variáveis binárias*

$z_{i,k}$	Variável binária que define os arranjos fotovoltaicos $k$ a serem instalados na barra $i$
-----------	---

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 OBJETIVO DA PESQUISA .....	17
1.2 JUSTIFICATIVA DO PROJETO .....	17
1.3 REVISÃO BIBLIOGRÁFICA .....	18
1.4 CONTRIBUIÇÃO.....	21
1.5 ORGANIZAÇÃO.....	22
<b>2 METODOLOGIA</b> .....	<b>23</b>
2.1 MODELO CLÁSSICO DE FLUXOS EM RAMOS PARA O FLUXO DE POTÊNCIA CA PARA A ALOCAÇÃO DE FONTES FOTOVOLTAICAS EM REDES DE DISTRIBUIÇÃO .....	24
<b>2.1.1 MODELO DE FLUXO EM RAMOS PARA O FLUXO DE POTÊNCIA CA NÃO LINEAR PARA SISTEMAS DE DISTRIBUIÇÃO [42]</b> .....	<b>24</b>
<b>2.1.2 TROCA DE VARIÁVEL PARA <math>I_{ij}, d_2</math> E <math>V_i, d_2</math> DO MODELO DE FLUXO EM RAMOS [42], [45]</b> .....	<b>29</b>
<b>2.1.3 MODELO CA EQUIVALENTE (PROGRAMAÇÃO CÔNICA DE SEGUNDA ORDEM) PARA O MODELO DE FLUXO EM RAMOS PARA SDEE</b> .....	<b>31</b>
<b>2.1.4 IDENTIFICAÇÃO DO PARÂMETRO <math>f_t, d_{Gpv}</math> PERTENCENTE AO MODELO (65) – (80)</b> .....	<b>39</b>
2.2 OBTENÇÃO DE PARÂMETROS DETERMINÍSTICOS .....	40
<b>2.2.1 OBTENÇÃO DOS AGRUPAMENTOS (CLUSTERS)</b> .....	<b>44</b>
<b>2.2.2 ALGORITMO K-MEANS</b> .....	<b>45</b>
<b>2.2.3 ALGORITMO HIERÁRQUICO</b> .....	<b>47</b>
2.2.3.1 AGRUPAMENTO HIERÁRQUICO AGLOMERADO .....	48
2.2.3.2 AGRUPAMENTO HIERÁRQUICO DE DIVISÃO [36] .....	50
2.2.3.3 AGRUPAMENTO HIERÁRQUICO: RESUMO [36].....	50
<b>2.2.4 MÉTODO DO COTOVELO</b> .....	<b>52</b>
2.3 MODELAGEM DE MÓDULOS SOLARES USANDO SIMULINK [57] .....	53
2.4 APLICAÇÕES DE <i>CLUSTERING</i> EM SISTEMAS ELÉTRICOS DE POTÊNCIA .....	55
<b>2.4.1 DETERMINAÇÃO DA TARIFA PARA O SERVIÇO DE ENERGIA ELÉTRICA NAS EMPRESAS DE DISTRIBUIÇÃO</b> .....	<b>55</b>
<b>2.4.2 PROGNÓSTICO DA DEMANDA</b> .....	<b>56</b>
<b>2.4.3 RESPOSTA DA DEMANDA</b> .....	<b>56</b>

<b>2.4.4 CLASSIFICAÇÃO DE USUÁRIOS QUE NÃO TÊM MEDIDORES INTELIGENTES.....</b>	<b>56</b>
<b>2.4.5 EXEMPLO DE APLICAÇÃO DO ALGORITMO <i>K-MEANS</i> PARA A ESTRATIFICAÇÃO DE CONSUMIDORES.....</b>	<b>57</b>
2.4.5.1 PRÉ-PROCESSAMENTO DE DADOS [60] .....	57
2.4.5.2 NORMALIZAÇÃO DE DADOS [60].....	63
2.4.5.3 ALGORITMO USADO EM DETALHE [60] .....	63
2.4.5.4 ESTRATIFICAÇÃO DOS USUÁRIOS APLICADA À EMPRESA DE DISTRIBUIÇÃO [60] .....	67
<b>3 RESULTADOS.....</b>	<b>70</b>
3.1 IMPLEMENTAÇÃO DO MÉTODO <i>K-MEANS</i> .....	70
<b>3.1.1 NORMALIZAÇÃO DOS DADOS PARA O PROCESSAMENTO DOS DADOS HISTÓRICOS .....</b>	<b>70</b>
<b>3.1.2 PROGRAMAÇÃO DO ALGORITMO <i>K-MEANS</i> .....</b>	<b>72</b>
<b>3.1.3 OBTENÇÃO DE DIAS REPRESENTATIVOS .....</b>	<b>75</b>
3.2 MODELOS MATEMATICOS USADOS PARA INSERIR OS RESULTADOS DA APRENDIZAGEM DE MÁQUINA.....	78
3.3 SIMULAÇÕES E RESULTADOS.....	78
3.3.1 RESULTADOS USANDO O ALGORITMO <i>K-MEANS</i> .....	78
3.3.2 OBTENÇÃO DA GERAÇÃO DAS FONTES FOTOVOLTAICAS USANDO SIMULINK DO MATLAB.....	84
3.3.3 RESULTADOS PARA O MODELO MATEMÁTICO DE OTIMIZAÇÃO .....	86
<b>4 CONCLUSÕES E RECOMENDAÇÕES.....</b>	<b>93</b>
<b>REFERÊNCIAS .....</b>	<b>95</b>
<b>APÊNDICE A – PUBLICAÇÕES RESULTANTES DO TRABALHO DE PESQUISA.....</b>	<b>101</b>
<b>ANEXOS.....</b>	<b>102</b>

## 1 INTRODUÇÃO

Atualmente, as fontes renováveis de energia desempenham um papel muito importante na redução das emissões de carbono para mitigar os problemas climáticos. Na área de sistemas elétricos de potência, são considerados diferentes modelos matemáticos que buscam otimizar as condições de operação da rede por meio da redução das perdas técnicas e pela introdução dos diferentes tipos de geração de energias renováveis [1]. Esses modelos matemáticos são geralmente não lineares, e a utilização de solucionadores comerciais que obtêm a solução ótima global de um modelo de otimização matemática exige que sejam utilizadas técnicas de linearização [2].

A inclusão do aprendizado de máquina para a modelagem de sistemas elétricos visa desenvolver técnicas e processos que permitam a um computador ou algoritmo computacional generalizar uma ideia a partir dos exemplos que são fornecidos [3]. Em [4] definiu-se o aprendizado de máquina como "um campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados".

Há vários algoritmos de aprendizado de máquina propostos na literatura. Neste trabalho de pesquisa são utilizados algoritmos de agrupamento, focando principalmente no algoritmo hierárquico que particiona o conjunto de dados por níveis, de forma que dois grupos sejam unidos ou divididos em cada nível de grupos do nível anterior [5]. Assim, é feito uso do algoritmo *K-means* como detalhado em [6].

Neste trabalho são determinados os dias representativos primordiais nos modelos de longo prazo para o sistema elétrico, para os quais se utilizam aplicações de aprendizado de máquina através do algoritmo *K-means* [6], que permitem introduzir variáveis determinísticas em um modelo de linearização matemática, partindo de variáveis estocásticas como a temperatura e irradiação solar para uma fonte de energia renovável.

Este trabalho visa minimizar o investimento na instalação de unidades de geração fotovoltaica e a redução de perdas de potência em uma rede de distribuição. Para isso, foi desenvolvida a formulação de um modelo convexo não linear, o qual foi resolvido utilizando a linguagem de otimização AMPL [7].

Foi desenvolvido um algoritmo para resolver o problema de otimização e obter a localização ótima dos painéis fotovoltaicos, bem como o número de unidades necessárias.

O método proposto foi avaliado por meio de simulações utilizando redes teste e redes reais, cujos dados foram fornecidos pela empresa do setor elétrico.



## 1.1 OBJETIVO DA PESQUISA

O objetivo desta pesquisa é propor um modelo CA convexo cônico de segunda ordem inteiro misto para resolver o problema de localização ótima de painéis fotovoltaicos, considerando a aleatoriedade e variabilidade da fonte solar, visando minimizar o investimento na instalação de unidades de geração fotovoltaica e a redução de perdas de potência ativa nas redes de distribuição. Após formular um modelo CA não linear para obter o ponto de operação de redes de distribuição radiais, são aplicadas técnicas de linearização e convexidade ao modelo proposto, com vistas à obtenção de uma solução ótima global para o problema ou equivalente. Neste problema, mediante o uso de aprendizagem de máquina obtêm-se parâmetros determinísticos, a partir de valores estocásticos, relacionados aos painéis solares a serem inseridos no modelo matemático, e assim determinar o impacto das fontes renováveis nas perdas técnicas na rede de distribuição.

## 1.2 JUSTIFICATIVA DO PROJETO

O aspecto econômico-financeiro da maioria das distribuidoras de energia elétrica é um fator determinante na tomada de decisões de curto, médio e longos prazos. Por outro lado, as perdas de energia foram e continuarão a ser métricas utilizadas para obter melhoria no desempenho das redes de distribuição. A redução de perdas pode ser obtida, por exemplo por meio de reconfiguração da rede [8],[9] ou do chaveamento de bancos de capacitores [10], [11].

A localização ideal e operação ótima da geração distribuída para minimizar as perdas têm despertado o interesse de pesquisadores nos últimos anos. Os estudos encontrados na literatura abordam o problema sob os pontos de vista da minimização das: perdas de energia [12], do custo de energia fornecida pela subestação em paralelo com a alocação dos bancos de capacitores fixos ou desconectáveis considerando custos de investimento em painéis solares [13], além de outras propostas levantadas em [14].

A minimização de perdas de energia é amplamente utilizada quando se considera redes passivas (sem geração distribuída). Esta abordagem considera um único nível de carga tornando impossível determinar o impacto atual das formas variáveis de geração distribuída (eólica e fotovoltaica). A própria variabilidade das cargas implica redução das perdas devido ao tamanho e localização ótimos de uma unidade firme (gerador distribuído a gás) durante a demanda

máxima. Esta abordagem foi levantada por meio do uso de índices de impacto [15], [16], e metaheurísticas [17], [18].

No início do atual milênio foram apresentadas abordagens relacionadas aos impactos dos geradores distribuídos nas redes de distribuição com o objetivo de minimizar perdas técnicas. Neste sentido, em [19] é apresentada uma técnica baseada em “*Tabu Search*” para resolver o modelo de otimização para minimização das perdas técnicas mediante a alocação e dimensionamento dos geradores distribuídos. Outra abordagem similar foi apresentada em [20] com uma técnica baseada no enfoque multiobjetivo para permitir ao planejador do sistema decidir o melhor compromisso entre os custos de melhorar a rede, perdas de potência, energia não fornecida e o custo da energia requerida pelos consumidores servidos [20].

Portanto, a consideração da geração distribuída e a minimização de perdas são imperativas na modelagem de sistemas de distribuição de energia elétrica usando as equações matemáticas que as regem. Ressalta-se que essas equações possuem uma característica não linear. No entanto, existem certas abordagens que permitem obter uma excelente aproximação linear dos pontos de operação do sistema modelado e em análise.

A aplicação de técnicas que permitem abordar a questão estocástica dos parâmetros de entrada para fontes renováveis de geração, no caso a irradiação solar e a temperatura, pode melhorar os resultados obtidos nos modelos clássicos de localização de painéis solares. Nesse sentido, a aplicação de *clusters* [21] permite tratar as grandezas estocásticas fora do modelo e, após obter os representantes de cada parâmetro em cada época do ano, estas são incluídas no modelo matemático como parâmetros determinísticos.

É de extrema importância integrar as práticas das empresas do setor elétrico e os avanços acadêmicos e científicos em relação a esta linha de pesquisa. Em particular, empresas de distribuição do Equador têm interesse em aplicar a metodologia proposta em suas redes.

### 1.3 REVISÃO BIBLIOGRÁFICA

A geração distribuída consiste na produção de energia elétrica ao nível da distribuição de forma descentralizada. As fontes utilizadas não são necessariamente renováveis, no entanto, existem estudos voltados para a inserção de energias renováveis por meio da geração distribuída [22], [23], [24]. Uma das características desse tipo de sistema é que a capacidade de geração nos pontos onde as fontes estão localizadas não é alta, mas oferece a vantagem de reduzir as perdas na distribuição [12], [25].

A redução das perdas de potência é um aspecto de constante preocupação por parte das empresas do setor. Existem duas abordagens para isso, respectivamente a redução das perdas de energia e das perdas de potência [12]. Os diferentes métodos que buscam esses dois objetivos podem ser divididos em métodos de otimização tradicionais [23], [24], [26], métodos iterativos [22], [25], [26] e métodos de otimização modernos (que utilizam aprendizagem de máquinas) [12], [26], [21].

As abordagens baseadas em modelos tradicionais de otimização são descritas em [26], onde é proposto um modelo não linear que é resolvido por um fluxo de potência ótimo CA, e busca determinar a capacidade máxima de geração distribuída que pode ser conectada à rede mediante o uso da administração ativa de rede (ANM pelas suas siglas no inglês *Active Network Management*).

Em 2015, foi abordada a inserção dos geradores distribuídos para a minimização do perfil da tensão no alimentador de distribuição mediante a alocação e dimensionamento dos geradores distribuídos [24].

Para melhorar a alocação de geradores distribuídos visando minimizar as perdas técnicas, em [23] é apresentada uma modelagem híbrida para reduzir as perdas dos sistemas de distribuição, obtendo o tamanho e a localidade ótima dos geradores, resolvida pelo método de pontos interiores.

Em [27] é determinada a localização ótima de um gerador distribuído cuja capacidade é de 2,3 MW, porém, sem considerar o custo de operação. Para isso, foi utilizado o pacote computacional NEPLAN [28] que utiliza vários métodos de otimização. Deve-se notar que há poucas referências a este tipo de estudo na literatura.

Enquanto isso, técnicas de otimização modernas podem ser encontradas em [22], [25], [29] onde são apresentados algoritmos genéticos para a solução de modelos buscando reduzir as perdas de potência e energia. Além do algoritmo genético também são propostos em [30] algoritmos como de “*simulated annealing*”, em [31] é apresentada a técnica de busca de entorno variável descendente (BEVD), e o algoritmo genético híbrido como “*Cuckoo Search*” [32] e “*hybrid particle swarm optimization-gravitational search algorithm*” [33]. Estes algoritmos foram implementados e usados em [34] para a alocação e o dimensionamento ótimo dos geradores distribuídos.

Uma proposta não muito explorada na literatura é a inserção de aprendizagem não supervisionada como parte fundamental para trabalhar com as naturezas estocásticas dos parâmetros dos geradores distribuídos renováveis.

As fontes renováveis de geração distribuída dependem de parâmetros estocásticos como: velocidade de vento, temperatura do ar, irradiação solar, variabilidade da carga. Estas informações requeridas pelos geradores distribuídos foram abordadas com simulações de Monte Carlo [35].

Em estudos como [12], [26], a consideração dos perfis de vento e demanda é feita pela utilização de técnicas metaheurísticas. Estas técnicas são utilizadas separadamente, buscando-se modelar a variabilidade que existe na inserção de fontes renováveis de energia.

Uma ideia interessante consiste no desenvolvimento de técnicas que permitam transformar a natureza estocástica destes parâmetros relacionados às fontes renováveis distribuídas em uma natureza determinística. Como é exposto em [3] e [36], é possível agrupar dados históricos em *clusters* que serão representantes dos dados históricos do banco de dados. Com este antecedente é possível obter um representante de cada *cluster*, sendo possível obter tantos representantes quanto *clusters* após as técnicas de agrupamento mediante técnicas de aprendizagem de máquina terem sido empregadas.

Esta ideia de agrupamento foi trabalhada pelos autores em [1], onde foi avaliado o banco de dados históricos das demandas de potência colocadas em séries temporais. Considerando as curvas de demanda dos dias úteis e os finais de semana, os autores obtiveram um representante “dia representativo” que sintetiza as informações de todas as cargas dos dias úteis e também o “dia representativo” que sintetiza as informações de todas as demandas dos finais de semana.

Seguindo esta técnica, em [21] foi considerada uma abordagem para tratar a intermitência dos geradores distribuídos renováveis ao trabalhar com uma seleção de “dias representativos” mediante um agrupamento em *cluster* usando algoritmos com metodologia hierárquica aplicada ao planejamento da expansão dos sistemas de geração. Uma desvantagem em [21] é o aumento do tempo e esforço computacional além do fato de que algoritmos hierárquicos possuem alto nível de complexidade.

Para superar a desvantagem das técnicas de agrupamento aplicadas em [21], pode ser usada outra técnica de agrupamento mais simples, do ponto de vista computacional, como a *K-means* [37] [38]. Esta técnica possui como ideia principal a definição dos centroides  $K$  (que corresponde ao número de agrupamentos) e, em seguida, tomar cada ponto da base de dados e situá-lo na classe do centroide mais próximo. No passo seguinte, o centroide de cada grupo é recalculado para distribuir todos os objetos de acordo com o centroide mais próximo. O processo é repetido até que não haja mudança nos grupos de uma etapa para a próxima.

É possível adaptar a técnica de agrupamento mediante *K-means* [37] [38] aos dados históricos que contêm os parâmetros requeridos pelas fontes de geração distribuída renovável,

e assim que estiveram determinados os  $K$  grupos, continuar com a obtenção do representante de cada grupo. Estes representantes serão chamados de “dias representativos”.

Em [12] conclui-se que, quando se tem um único objetivo de minimizar as perdas de energia, há a tendência de se comprometer a capacidade potencial de geração renovável que poderia ser conectada às redes de distribuição. Ao invés disso, a redução deve ser encontrada mediante uma estrutura regulatória em que perdas excessivas são penalizadas e o benefício líquido entre a energia de baixo carbono e as perdas é maximizado.

Analisando a informação de [1] em modelos de longo prazo para sistemas de energia, é essencial selecionar corretamente os dias representativos, a fim de reduzir a resolução temporal e manter os requisitos computacionais do modelo ao mínimo.

Em [21] critérios e métricas são propostos para avaliar a representatividade, bem como é apresentada uma nova abordagem baseada em otimização para selecionar um conjunto representativo de dias, avaliando e comparando com várias abordagens.

Um modelo de programação não linear inteiro misto (PNLIM) é proposto em [13], que é reformulado para um problema de programação linear inteira mista (PLIM) usando técnicas de linearização. Este modelo PLIM aproximado é implementado na linguagem matemática AMPL [7], enquanto o solucionador comercial CPLEX [39] é usado para obter soluções ótimas globais.

## 1.4 CONTRIBUIÇÃO

Neste trabalho de pesquisa, é proposto um modelo matemático CA para a obtenção da localização ótima de unidades de geração fotovoltaica em redes de distribuição elétrica, visando minimizar as perdas de potência na rede de distribuição. As principais contribuições deste trabalho são:

- a) Apresentação de um novo modelo convexo cônico de segunda ordem integrado que permite alocar unidades de geração fotovoltaica, considerando a característica estocástica nos parâmetros de entrada para fontes fotovoltaicas usando aprendizagem de máquinas.
- b) A inclusão do aprendizado de máquina para a modelagem de um sistema de energia elétrica, que utiliza algoritmos de agrupamento, com foco principalmente no reconhecimento de padrões nos níveis de irradiação solar e temperatura ambiente em diferentes épocas do ano.

- c) Implementação do algoritmo *K-means* para determinar os dias representativos.
- d) Uso das informações obtidas do algoritmo *K-means* como dados de entrada para o modelo de otimização proposto.
- e) Inserção no modelo da aleatoriedade e variabilidade das fontes fotovoltaicas, permitindo a melhoria na precisão da resposta obtida.

## 1.5 ORGANIZAÇÃO

Este trabalho é estruturado da maneira descrita a seguir. No Capítulo 2 a metodologia aplicada nesta pesquisa é apresentada, sendo dividida em duas partes. A primeira está relacionada à descrição do modelo matemático usado para a alocação ótima e dimensionamento das fontes renováveis solares no sistema de distribuição considerando aos parâmetros relacionados à produção dos geradores fotovoltaicos determinísticos. A segunda parte consiste em aplicar a técnica de *clustering*, especificamente o algoritmo de *K-means*, aos dados históricos dos parâmetros relacionados (como a irradiação solar e a temperatura ambiente) à produção de geradores fotovoltaicos para obter um melhor valor determinístico a partir do comportamento estocástico evidenciado na data histórica. Além disso, o modelo de otimização toma como parâmetro de entrada a resposta do algoritmo *K-means* para encontrar uma solução para a localização de painéis solares.

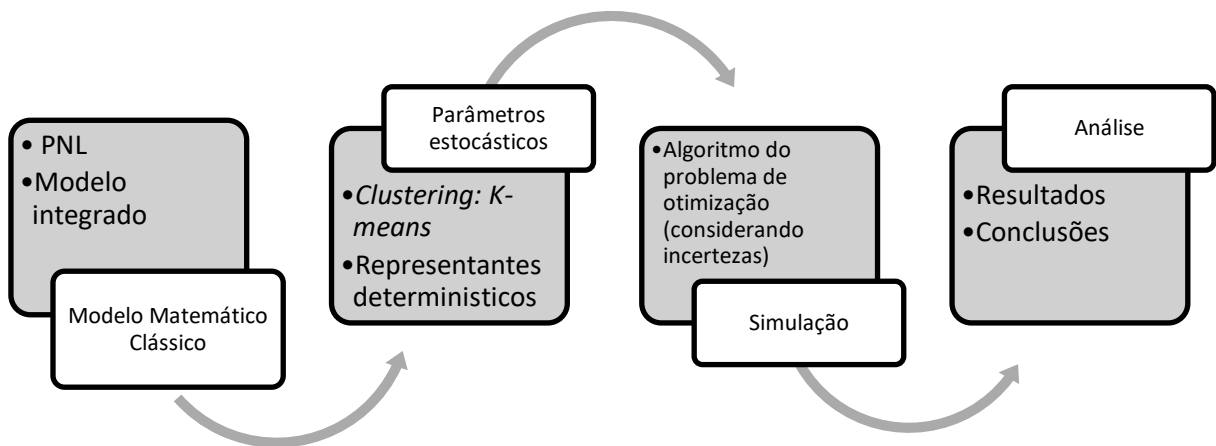
No Capítulo 3 são apresentadas as simulações e resultados em que as informações dos dados obtidos do algoritmo *K-means* foram utilizadas para inserir o modelo convexo não linear levantado, que é resolvido utilizando a linguagem de otimização AMPL [7], com a qual o problema de otimização proposto e as localizações ideais dos painéis fotovoltaicos são obtidos, bem como o número de unidades necessárias.

Finalmente, o Capítulo 4 contém a conclusão do presente trabalho de pesquisa e possíveis trabalhos futuros.

## 2 METODOLOGIA

A proposta deste trabalho, em termos gerais, está dividida em duas partes. A primeira está relacionada à descrição do modelo matemático usado para a alocação ótima e dimensionamento das fontes renováveis solares [13], [40] no sistema de distribuição considerando parâmetros determinísticos relacionados à produção de energia fotovoltaica. A segunda parte consiste em aplicar a técnica de *clustering*, especificamente o algoritmo de *K-means*, aos dados históricos dos parâmetros relacionados à produção de geradores fotovoltaicos (como a irradiação solar [41] e a temperatura ambiente) para obter valores determinísticos a partir de seus comportamentos estocásticos evidenciado na data histórica. Na Figura 2.1 é apresentado o fluxograma da metodologia proposta.

**Figura 2.1**



**Figura 2.1** – Fluxograma do método proposto. Fonte: Autor.

Da metodologia apresentada na Figura 2.1 são apresentados e usados os modelos matemáticos clássicos para a alocação e dimensionamento dos geradores fotovoltaicos partindo de um problema de Programação Não Linear Inteiro Misto (PNLIM), considerando um modelo equivalente cônico de segunda ordem neste ponto, cuja motivação e estruturação serão apresentadas em detalhe adiante. Foram realizadas simulações para a solução do problema de otimização obtendo a localização ótima dos painéis fotovoltaicos, bem como o número de unidades necessárias atendendo os parâmetros relacionados à produção das fontes renováveis de forma determinística. No segundo bloco obtém-se as condições dos parâmetros estocásticos mediante a aplicação das técnicas de aprendizagem de máquina. O foco deste método consiste, em primeiro lugar, na determinação dos parâmetros estocásticos necessários para geração

fotovoltaica, e, na sequência, a metodologia proposta é aplicada onde os representantes são obtidos, ou seja, os valores determinísticos dos parâmetros de geração fotovoltaica. O algoritmo de *K-means* é usado para obter um representante determinístico a partir do comportamento estocástico dos parâmetros envolvidos na análise. Em seguida, são feitas simulações usando o modelo matemático com valores determinísticos provenientes do algoritmo *K-means*. Finalmente, tem-se a etapa de análise dos resultados das simulações para assim obter as conclusões.

Neste capítulo, a ideia é construir gradativamente o método proposto, por meio da descrição detalhada de cada etapa que constituirá o modelo integrado final com representantes determinísticos mais robustos para obter a geração fotovoltaica a ser envolvida no modelo de otimização.

## 2.1 MODELO CLÁSSICO DE FLUXOS EM RAMOS PARA O FLUXO DE POTÊNCIA CA PARA A ALOCAÇÃO DE FONTES FOTOVOLTAICAS EM REDES DE DISTRIBUIÇÃO

Uma descrição detalhada, do modelo matemático baseado em fluxo em ramos aplicado ao sistema de distribuição para o fluxo de potência CA e também do modelo equivalente CA cônico de segunda ordem (que também é chamado de equivalente [42]), é apresentada em [13]. As seções a seguir contêm uma breve descrição do processo através do qual o modelo clássico não linear vai se transformar no modelo equivalente (sendo este último o modelo a ser usado para resolver o problema proposto neste trabalho considerando as técnicas e metodologias de *clustering* e o algoritmo de *K-means*).

### 2.1.1 MODELO DE FLUXO EM RAMOS PARA O FLUXO DE POTÊNCIA CA NÃO LINEAR PARA SISTEMAS DE DISTRIBUIÇÃO [42]

De acordo com [42], para Sistemas de Distribuição de Energia Elétrica (SDEE) radiais operando em regime permanente, normalmente se considera as seguintes hipóteses:

- As demandas são representadas como potências ativa e reativa constantes;
- O SDEE é balanceado e representado por um equivalente monofásico.

A Figura 2.2 apresenta todas as grandezas típicas do sistema de distribuição radial operando em regime permanente. São representadas três barras conectadas por duas linhas de distribuição.



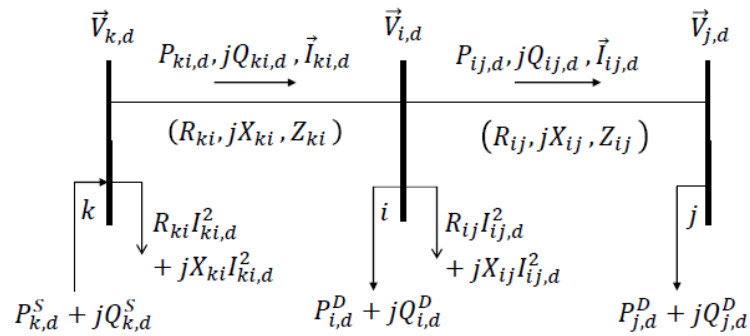


Figura 2.2 – Grandezas dentro do sistema de distribuição radial. [42]

Da Figura 2.2, na qual é apresentado um SDEE radial, pode se obter as equações de balanço de potência ativa e reativa como segue:

$$\sum_{ji \in L} P_{ji,d} - \sum_{ij \in L} (P_{ij,d} + R_{ij} I_{ij,d}^2) + P_{i,d}^S = P_{i,d}^D \quad (1)$$

$$\sum_{ji \in L} Q_{ji,d} - \sum_{ij \in L} (Q_{ij,d} + X_{ij} I_{ij,d}^2) + Q_{i,d}^S = Q_{i,d}^D \quad (2)$$

$$\forall i \in \Omega_b, \forall d \in \Omega_d$$

Onde  $\Omega_b$  é o conjunto de nós,  $\Omega_d$  é o conjunto de níveis de demanda.  $R_{ij} I_{ij,d}^2$  e  $X_{ij} I_{ij,d}^2$  são as perdas de potências ativa e reativa do circuito  $ij$ , respectivamente, no nível de demanda  $d$ .

A queda de tensão no ramo  $ij$  é definida por:

$$\vec{V}_{i,d} - \vec{V}_{j,d} = \vec{I}_{ij,d} (R_{ij} + jX_{ij}) \quad (3)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

em que  $\Omega_L$  é o conjunto de ramos, e  $\vec{I}_{ij,d}$  pode ser calculada no nó  $j$  usando a equação (4).

$$\vec{I}_{ij,d} = \left( \frac{P_{ij,d} + jQ_{ij,d}}{\vec{V}_{j,d}} \right)^* \quad (4)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

É possível substituir (4) em (3), obtendo-se:

$$(\vec{V}_{i,d} - \vec{V}_{j,d}) \vec{V}_{j,d}^* = (P_{ij,d} - jQ_{ij,d})(R_{ij} + jX_{ij}) \quad (5)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

Considerando que  $\vec{V}_{i,d} = V_{i,d}(\cos \theta_{i,d} + j \sin \theta_{i,d})$ ,  $\vec{V}_{j,d} = V_{j,d}(\cos \theta_{j,d} + j \sin \theta_{j,d})$  e  $\theta_{ij,d} = \theta_{i,d} - \theta_{j,d}$ , então (5) pode ser escrita como:

$$V_{i,d}V_{j,d}[\cos \theta_{ij,d} + j \sin \theta_{ij,d}] - V_{j,d}^2 = (P_{ij,d} - jQ_{ij,d})(R_{ij} + jX_{ij}) \quad (6)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

em que  $V_{i,d}$  e  $\theta_{i,d}$  representam, a magnitude e o ângulo de fase da tensão na barra  $i$  no nível de demanda  $d$  respectivamente. É possível identificar as partes real e imaginária da equação (6) e se tem:

$$V_{i,d}V_{j,d} \cos \theta_{ij,d} = V_{j,d}^2 + (R_{ij}P_{ij,d} + X_{ij}Q_{ij,d}) \quad (7)$$

$$V_{i,d}V_{j,d} \sin \theta_{ij,d} = (X_{ij}P_{ij,d} - R_{ij}Q_{ij,d}) \quad (8)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

Somando os quadrados de (7) e (8) e utilizando a relação trigonométrica  $\sin^2(\theta_{ij,d}) + \cos^2(\theta_{ij,d}) = 1$ , obtém-se:

$$[V_{i,d}^2 - 2(R_{ij}P_{ij,d} + X_{ij}Q_{ij,d})]V_{j,d}^2 - Z_{ij}^2(P_{ij,d}^2 + Q_{ij,d}^2) - V_{j,d}^4 = 0 \quad (9)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

É destacável notar nesta modelagem [43] que:

- A diferença angular entre as tensões  $\theta_{ij,d}$  é eliminada;
- A equação (9) consiste em um polinômio de quarto grau que calcula a queda da magnitude de tensão no ramo  $ij$  no nível de demanda  $d$ ;
- Pode-se obter a magnitude da tensão do último nó  $V_{j,d}$  em termos da magnitude do nó inicial  $V_{i,d}$ , dos fluxos de potência ativa  $P_{ij,d}$  e reativa  $Q_{ij,d}$ , e dos parâmetros elétricos do ramo  $ij$ .

É possível representar (9) usando duas equações de segundo grau. De (4), o quadrado da magnitude de corrente  $I_{ij,d}$  é mostrada em (10) como segue:

$$I_{ij,d}^2 = \frac{P_{ij,d}^2 + Q_{ij,d}^2}{V_{j,d}^2} \quad (10)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

Substituindo (10) em (9), é possível escrever:

$$[V_{i,d}^2 - 2(R_{ij}P_{ij,d} + X_{ij}Q_{ij,d})] - Z_{ij}^2 I_{ij,d}^2 - V_{j,d}^2 = 0 \quad (11)$$

$$\forall ij \in \Omega_L, \forall d \in \Omega_d$$

Note que:

- Substituindo (10) em (11) obtém-se (9).

- Assim como em (9), a diferença angular entre as tensões  $\theta_{ij,d}$  é eliminada;
- A equação (11) é um polinômio de segundo grau que calcula a queda da magnitude de tensão no circuito  $ij$  no nível de demanda  $d$ ;
- É possível obter a magnitude da tensão do nó final  $V_{j,d}$  em termos da magnitude do nó inicial  $V_{i,d}$ , dos fluxos de potência ativa  $P_{ij,d}$  e reativa  $Q_{ij,d}$ , da magnitude de corrente  $I_{ij,d}$ , e dos parâmetros elétricos do ramo  $ij$ .

Finalmente, o sistema de equações algébricas não lineares que representa a operação em regime permanente de um sistema de distribuição radial, e que são frequentemente utilizados no método de varredura [44], é:

$$\sum_{ji \in L} P_{ji,d} - \sum_{ij \in L} (P_{ij,d} + R_{ij} I_{ij,d}^2) + P_{i,d}^S = P_{i,d}^D \quad (12)$$

$$\sum_{ji \in L} Q_{ji,d} - \sum_{ij \in L} (Q_{ij,d} + X_{ij} I_{ij,d}^2) + Q_{i,d}^S = Q_{i,d}^D \quad (13)$$

$$\forall_i \in \Omega_b, \forall_d \in \Omega_d$$

$$I_{ij,d}^2 = \frac{P_{ij,d}^2 + Q_{ij,d}^2}{V_{j,d}^2} \quad (14)$$

$$[V_{i,d}^2 - 2(R_{ij} P_{ij,d} + X_{ij} Q_{ij,d})] - Z_{ij}^2 I_{ij,d}^2 - V_{j,d}^2 = 0 \quad (15)$$

$$\forall_{ij} \in \Omega_L, \forall_d \in \Omega_d$$

Se a magnitude da tensão na subestação for conhecida (como normalmente é assumido no problema de fluxo de carga) o número de equações é igual ao número de incógnitas, e este sistema possui solução única.

Em [13], há uma representação matemática exata para o fluxo de carga CA baseada em ramos com finalidade de minimizar as perdas técnicas de sistemas de distribuição, determinando o número e o tamanho ótimo das unidades fotovoltaicas a serem inseridas no sistema teste sob análise. Adicionalmente, os autores analisam o efeito de inserir equipamentos de armazenamento de energia elétrica como baterias além de fontes de potência reativa fixa (bancos de capacitores).

A presente proposta de pesquisa focará na inserção ótima das fontes fotovoltaicas em sistemas de distribuição sem considerar possibilidade de armazenamento de energia, e somente focando o impacto das próprias fontes renováveis na operação do sistema de distribuição. Essas equações matemáticas são uma representação das equações do fluxo de carga não linear (16) – (29) que poderiam ser utilizadas na formulação, por exemplo, a fim de minimizar as perdas de

potência da rede, adicionando à função objetivo (16) o custo de energia no início do alimentador que sai da subestação de distribuição e o custo de investimento e instalação das fontes fotovoltaicas nas barras do sistema de distribuição, descrito pelas seguintes equações:

$$\min \sum_{d \in D} \sum_{t \in T} \Delta_d \gamma_{t,d}^G P_{t,d}^S + \sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} + \sum_{d \in D} \sum_{ij \in L} \sum_{t \in T} \delta r_{i,j} I_{ij,t,d}^2 \quad (16)$$

Sujeito a:

$$\sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} \leq IC^{PV} \quad (17)$$

$$\sum_{ji \in L} P_{ji,t,d} - \sum_{ji \in L} (P_{ij,t,d} + R_{ij} I_{ij,t,d}^2) + P_{t,d}^S + P_{i,t,d}^{PV} = P_{i,t,d}^{Ld} \quad (18)$$

$$\sum_{ji \in L} Q_{ji,t,d} - \sum_{ji \in L} (Q_{ij,t,d} + X_{ij} I_{ij,t,d}^2) + Q_{t,d}^S + Q_{i,t,d}^{PV} = Q_{i,t,d}^{Ld} \quad (19)$$

$$V_{i,t,d}^2 - 2[(R_{ij} P_{ij,t,d} + X_{ij} Q_{ij,t,d}) - Z_{ij}^2 I_{ij,t,d}^2] - V_{j,t,d}^2 = 0 \quad (20)$$

$$V_{j,t,d}^2 I_{ij,t,d}^2 = P_{ij,t,d}^2 + Q_{ij,t,d}^2 \quad (21)$$

$$V_{min} \leq V_{i,t,d} \leq V_{max} \quad (22)$$

$$0 \leq I_{ij,t,d} \leq I_{max} \quad (23)$$

$$-P_{t,d}^S \tan(\cos^{-1} \phi_{min}) \leq Q_{t,d}^S \leq P_{t,d}^S \tan(\cos^{-1} \phi_{max}) \quad (24)$$

$$\sum_{t \in T} e^p \Delta_d P_{t,d}^S \leq PE \quad (25)$$

$$P_{i,t,d}^{PV} = \sum_{k=1}^{PV} k z_{i,k} P^{-PV} f_{t,d}^{Gpv} \quad (26)$$

$$|Q_{i,t,d}^{PV}| = P_{i,t}^{PV} \tan(\cos^{-1} \phi^{PV}) \quad (27)$$

$$\sum_{k=1}^{PV} z_{i,k} \leq 1 \quad (28)$$

$$\sum_{i \in N} \sum_{k=1}^{PV} z_{i,k} \leq N^{PV} \quad (29)$$

Segundo [13], a queda de tensão é estabelecida apropriadamente no modelo (16)-(29).

A função objetivo está relacionada com três aspectos: o custo pela energia fornecida pela

subestação de distribuição, o custo de instalação das fontes renováveis e o componente das perdas técnicas do sistema de distribuição. A equação (17) representa o comportamento do investimento máximo em fontes renováveis. As restrições (18) e (19) representam os balanços de potência ativa e reativa respectivamente. Na equação (20) é possível determinar a queda de tensão em cada ramo do sistema de distribuição. A restrição (21) apresenta a magnitude do fluxo de corrente no ramo pertencente ao sistema de distribuição sob análise. As equações (22), (23), (24) e (25) representam os limites da magnitude de tensão, fluxo de corrente pelos ramos, injeção de potência reativa da subestação e magnitude máxima da potência ativa injetada pela subestação. Nas restrições (26) e (27) é possível determinar as potências ativas e reativas injetadas pelas fontes fotovoltaicas, respectivamente. Finalmente em (28) e (29) é apresentado o comportamento da variável binária de decisão associada à inserção de fonte fotovoltaica na barra e os valores máximos de módulos por cada fonte renovável.

É importante notar que, no modelo de fluxo de ramos, as magnitudes das correntes nos ramos ( $I_{ij,d}$ ) e das tensões nos nós ( $V_{i,d}$ ) aparecem apenas nas formas  $I_{ij,d}^2$  e  $V_{i,d}^2$  respectivamente. Na seguinte secção será proposta uma troca de variável com o objetivo de linearizar a função objetivo (16) e as restrições (18) - (21), e assim melhorar o tempo e esforço computacional.

### 2.1.2 TROCA DE VARIÁVEL PARA $I_{ij,d}^2$ E $V_{i,d}^2$ DO MODELO DE FLUXO EM RAMOS [42], [45]

De acordo com [42], as equações (1)-(2) e (10)-(11) representam a operação em regime permanente de um SDEE radial, e as equações [45], (16)-(29) que representam a alocação e dimensionamento de fontes renováveis solares, são não lineares. Portanto, pode-se aplicar uma técnica de troca de variáveis para linearizar a função objetivo (16) e as restrições (1)-(2), (10)-(11), (18)-(21). Com isso, melhora-se o tempo e esforço computacional.

O modelo (16)-(29) é não linear, pois apresenta variáveis elevadas ao quadrado nas restrições (18)-(20) e produtos de variáveis (21). Entretanto, é possível obter outro modelo com uma única restrição não linear (21), com poucas aproximações, a partir deste modelo.

Como o modelo apresenta correntes e tensões, ao quadrado, é feita uma troca de variáveis, como se mostra em (30) e (31) para as equações (16), (18)-(20) e (22)-(23):

$$V_{i,t,d}^{sqr} = V_{i,t,d}^2, \quad V_{i,t,d}^{sqr} \geq 0 \quad (30)$$

$$I_{ij,t,d}^{sqr} = I_{ij,t,d}^2, \quad I_{ij,t,d}^{sqr} \geq 0 \quad (31)$$

O modelo atualizado, com as trocas de variáveis nas equações respectivas, é apresentado a seguir somente as equações que mudaram, ou seja, as equações (17) e (24)-(29) ficam iguais.

O novo modelo é expresso por:

$$\min \sum_{d \in D} \sum_{t \in T} \Delta_d \gamma_{t,d}^G P_{t,d}^S + \sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} + \sum_{d \in D} \sum_{ij \in L} \sum_{t \in T} \delta r_{i,j} I_{ij,t,d}^{sqr} \quad (32)$$

Sujeito a:

$$\sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} \leq IC^{PV} \quad (2) \quad (33)$$

$$\sum_{ji \in L} P_{ji,t,d} - \sum_{ji \in L} (P_{ij,t,d} + R_{ij} I_{ij,t,d}^{sqr}) + P_{t,d}^S + P_{i,t,d}^{PV} = P_{i,t,d}^{Ld} \quad (34)$$

$$\sum_{ji \in L} Q_{ji,t,d} - \sum_{ji \in L} (Q_{ij,t,d} + X_{ij} I_{ij,t,d}^{sqr}) + Q_{t,d}^S + Q_{i,t,d}^{PV} = Q_{i,t,d}^{Ld} \quad (35)$$

$$V_{i,t,d}^{sqr} - 2[(R_{ij} P_{ij,t,d} + X_{ij} Q_{ij,t,d}) - Z_{ij}^2 I_{ij,t,d}^{sqr}] - V_{j,t,d}^{sqr} = 0 \quad (36)$$

$$V_{j,t,d}^{sqr} I_{ij,t,d}^{sqr} = P_{ij,t,d}^2 + Q_{ij,t,d}^2 \quad (37)$$

$$V_{min}^2 \leq V_{i,t,d}^{sqr} \leq V_{max}^2 \quad (38)$$

$$0 \leq I_{ij,t,d}^{sqr} \leq I_{max}^2 \quad (39)$$

$$-P_{t,d}^S \tan(\cos^{-1} \phi_{min}) \leq Q_{t,d}^S \leq P_{t,d}^S \tan(\cos^{-1} \phi_{max}) \quad (40)$$

$$\sum_{t \in T} e^p \Delta_d P_{t,d}^S \leq PE \quad (41)$$

$$P_{i,t,d}^{PV} = \sum_{k=1}^{PV} k z_{i,k} P^{-PV} f_{t,d}^{Gpv} \quad (42)$$

$$|Q_{i,t,d}^{PV}| = P_{i,t}^{PV} \tan(\cos^{-1} \phi^{PV}) \quad (43)$$

$$\sum_{k=1}^{PV} z_{i,k} \leq 1 \quad (44)$$

$$\sum_{i \in N} \sum_{k=1}^{PV} z_{i,k} \leq N^{PV} \quad (45)$$

$$V_{i,t,d}^{sqr} \geq 0 \quad (46)$$

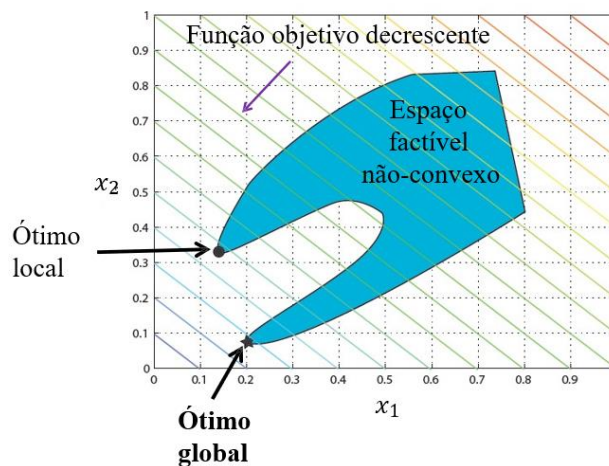
$$I_{ij,t,d}^{sqr} \geq 0 \quad (47)$$

Neste modelo a equação (37) ainda é não linear, pois contém termos quadráticos e o produto de duas variáveis de decisão. Depois de resolver o problema de programação não linear, é possível obter as magnitudes das correntes nos ramos  $I_{ij,d}$  e das tensões nos nós  $V_{i,d}$  por meio do valor positivo da raiz quadrada dos valores das variáveis lineais  $I_{ij,t,d}^{sqr}$  e  $V_{i,t,d}^{sqr}$  respectivamente.

Na seção seguinte é apresentada a abordagem de tratamento da não linearidade da equação (37) mediante a técnica de programação cônica de segunda ordem para o modelo de fluxo em ramo dos SDEE.

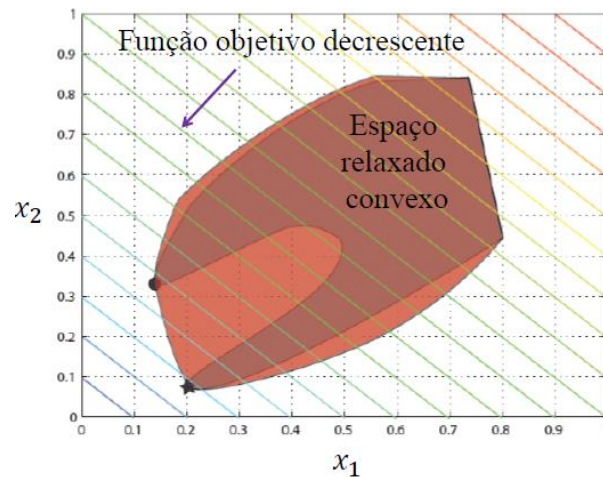
### 2.1.3 MODELO CA EQUIVALENTE (PROGRAMAÇÃO CÔNICA DE SEGUNDA ORDEM) PARA O MODELO DE FLUXO EM RAMOS PARA SDEE.

Ao trabalhar com modelos matemáticos é muito importante identificar a região factível do problema assim como o comportamento das variáveis de decisão independentes que afetam a função objetivo. A Figura 2.3 apresenta graficamente um problema de minimização com duas variáveis independentes.

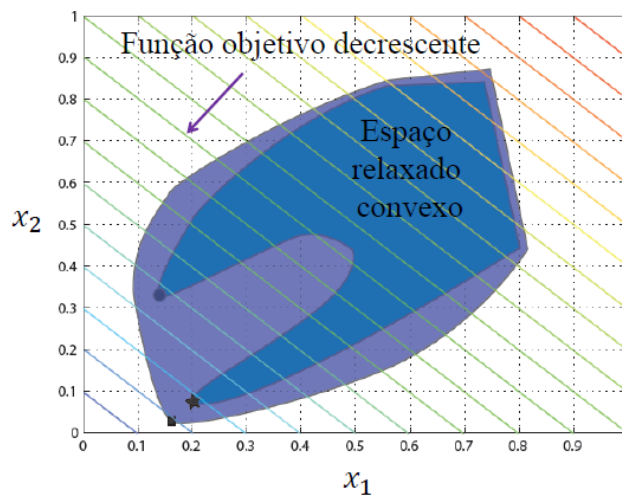


**Figura 2.3** – Representação gráfica de um modelo com região factível não convexa do problema de minimização com duas variáveis de decisão. [42]

Da Figura 2.3 pode-se obter a solução do problema de otimização, o qual pode levar a uma solução ótima local ou a um ótimo global. É possível, graficamente fazer uma relaxação convexa com gap de relaxação zero e outra como gap de relaxação não zero, como é apresentado nas Figura 2.4 e Figura 2.5, respectivamente.



**Figura 2.4** – Relaxação convexa com gap de relaxação zero. [42]



**Figura 2.5** – Relaxação convexa com gap de relaxação não zero. [42]

A Figura 2.4 mostra que a solução ótima global do problema relaxado com gap igual a zero corresponde exatamente à solução ótima do problema original. Por outro lado, a solução ótima do problema relaxado com gap não nulo apresenta uma solução ótima diferente do ótimo global do problema original. No caso, a solução ótima do problema relaxado apresenta um valor de função objetivo menor que o valor da função objetivo para a solução ótima do problema original.



Para compreender melhor a técnica da programação cônica de segunda ordem em [42] são apresentados a seguir dois exemplos.

**Exemplo 1:**

$$\min x_2 \quad (48)$$

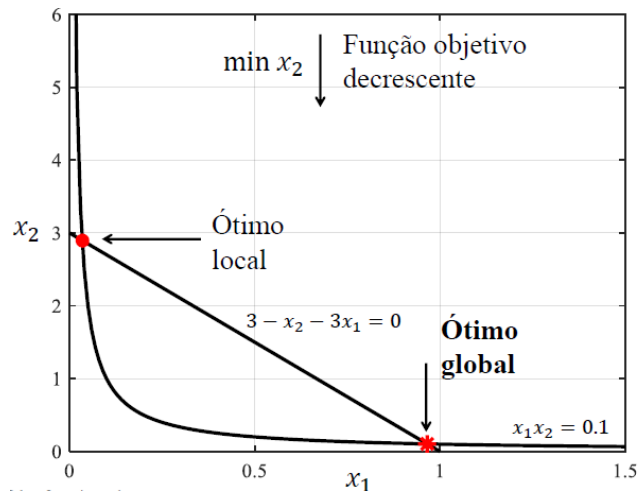
Sujeito a:

$$3 - x_2 - 3x_1 = 0 \quad (49)$$

$$x_2 x_1 = 0,1 \quad (50)$$

$$x_2, x_1 \geq 0 \quad (51)$$

Graficamente é possível obter a solução do problema (48)-(51) apresentado na Figura 2.6, onde a solução ótima local e global é encontrada. Note que os pontos ótimos local e global constituem o espaço factível para o problema, que é claramente desconexo.



**Figura 2.6** – Solução do modelo (48)-(51) não relaxado. [42]

Propõe-se relaxar a restrição (50), transformando-a na equação obtendo-se um problema de programação não linear convexo.

$$\min x_2 \quad (48)$$

Sujeito a:

$$3 - x_2 - 3x_1 = 0 \quad (49)$$

$$x_2, x_1 \geq 0 \quad (51)$$

$$x_2 x_1 \geq 0,1 \quad (52)$$

Graficamente é possível obter a solução do problema (48)-(49); (51)-(52) apresentado na Figura 2.7, onde a solução ótima global é encontrada. Nesta nova situação, os pontos ótimos local e global pertencem a um espaço factível conexo (em vermelho).

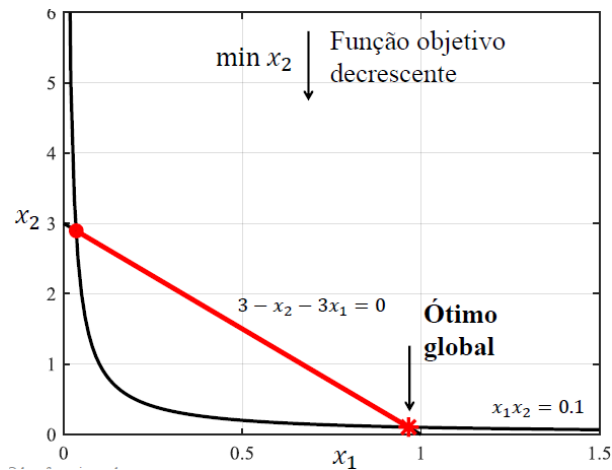


Figura 2.7 – Solução do modelo (48)-(49);(51)-(52) relaxado. [42]

### Exemplo 2:

$$\min x_2 \quad (53)$$

Sujeito a:

$$3 - x_2 - 3x_1 = 0 \quad (54)$$

$$x_2 x_1 = x_3^2 \quad (55)$$

$$-0,4 \leq x_3 \leq 0,4 \quad (56)$$

$$x_2, x_1 \geq 0 \quad (57)$$

Graficamente é possível obter a solução do problema (53)-(57) apresentado na Figura 2.8, onde as soluções ótimas locais e a solução ótima global são encontradas. A Figura 2.8 mostra que as regiões factíveis são desconexas (em vermelho).

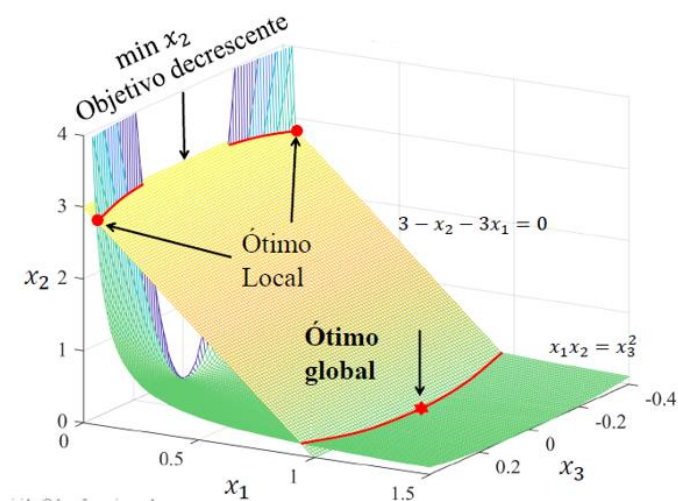


Figura 2.8 – Solução do modelo (53)-(57) não relaxado. [42]

Propõe-se relaxar a restrição (55), transformando-a na equação (58), obtendo-se assim um problema de programação não linear convexo.

$$\min x_2 \quad (53)$$

Sujeito a:

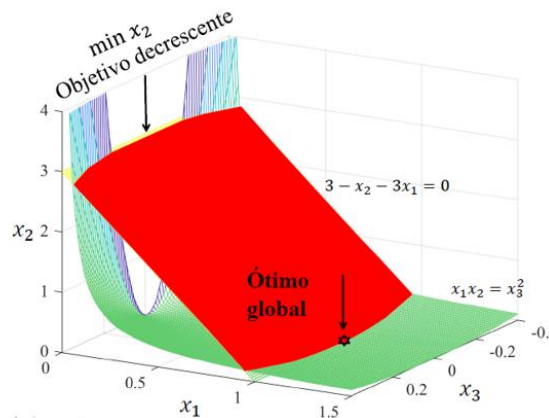
$$3 - x_2 - 3x_1 = 0 \quad (54)$$

$$-0,4 \leq x_3 \leq 0,4 \quad (56)$$

$$x_2, x_1 \geq 0 \quad (57)$$

$$x_2 x_1 \geq x_3^2 \quad (58)$$

Graficamente é possível obter a solução do problema (53)-(54);(56)-(58) apresentado na Figura 2.9, onde a solução ótima global é encontrada. Nesta figura, o espaço de soluções é conexo.

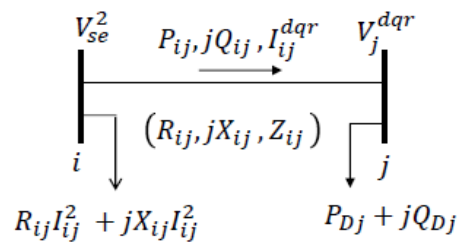


**Figura 2.9** – Solução do modelo (53)-(54);(56)-(58) relaxado. [42]

De forma geral a restrição (59) é chamada de restrição de cônica de segunda ordem rotacionada:

$$2x_1x_2 \geq \sum_{j=3}^n x_j^2; \quad x_1, x_2 \geq 0 \quad (59)$$

Para obter uma relação para os SDEE, é usado a seguir o circuito com dois nós e uma linha da Figura 2.10.



**Figura 2.10** – Circuito ilustrativo do SDEE. [42]

Da Figura 2.10, o problema de programação não linear é dado por:

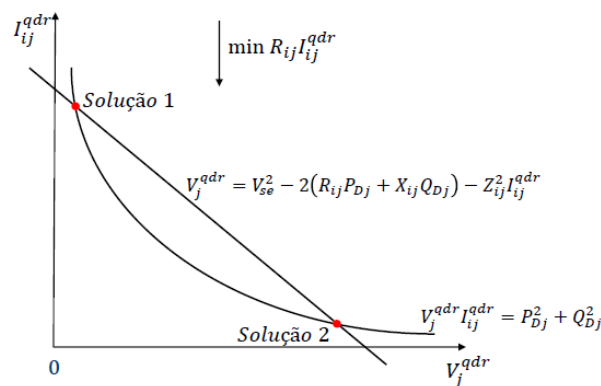
$$\min v = R_{ij}I_{ij}^{dqr} \quad (60)$$

Sujeito a:

$$V_{se}^2 - 2(R_{ij}P_{Dj} + X_{ij}Q_{Dj}) - Z_{ij}^2 I_{ij}^{dqr} - V_j^{dqr} = 0 \quad (61)$$

$$V_j^{dqr} I_{ij}^{dqr} = P_{Dj}^2 + Q_{Dj}^2 \quad (62)$$

Graficamente é possível obter a solução do problema (60)-(62) apresentado na Figura 2.11, onde a solução ótima local e global é encontrada.



**Figura 2.11** – Solução do modelo (60)-(62) não relaxado. [42]

Pode-se relaxar a restrição (62), transformando-a na equação (63), obtendo-se assim um problema de programação não linear convexo.

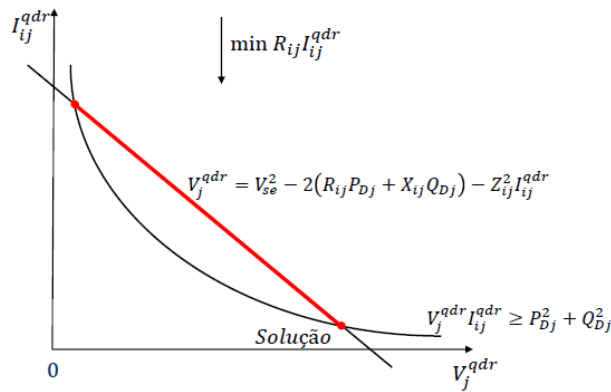
$$\min v = R_{ij}I_{ij}^{dqr} \quad (60)$$

Sujeito a:

$$V_{se}^2 - 2(R_{ij}P_{Dj} + X_{ij}Q_{Dj}) - Z_{ij}^2 I_{ij}^{dqr} - V_j^{dqr} = 0 \quad (61)$$

$$V_j^{dqr} I_{ij}^{dqr} \geq P_{Dj}^2 + Q_{Dj}^2 \quad (63)$$

Graficamente é possível obter a solução do problema (60)-(61); (63) apresentado na Figura 2.12, onde a solução ótima global é encontrada.



**Figura 2.12** – Solução do modelo (60)-(61); (63) relaxado. [42]

De acordo com [46], [47], [48], podem ser consideradas as seguintes quatro características:

- Minimização das perdas de potência ativas no sistema;
- As resistências nos circuitos são diferentes de zero;
- A operação radial dos sistemas de distribuição;
- A não negatividade das variáveis  $V_{i,t,d}^{sqr}$  e  $I_{ij,t,d}^{sqr}$ .

Com isso é possível relaxar a restrição não linear:

$$V_{j,t,d}^{sqr} I_{ij,t,d}^{sqr} = P_{ij,t,d}^2 + Q_{ij,t,d}^2 \quad (37)$$

pela restrição cônica de segunda ordem rotacionada:

$$V_{j,t,d}^{sqr} I_{ij,t,d}^{sqr} \geq P_{ij,t,d}^2 + Q_{ij,t,d}^2 \quad (64)$$

O modelo atualizado, com as trocas de variáveis nas equações respectivas e a programação cônica de segunda ordem, é apresentado a seguir:

$$\min \sum_{d \in D} \sum_{t \in T} \Delta_d \gamma_{t,d}^G P_{t,d}^S + \sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} + \sum_{d \in D} \sum_{ij \in L} \sum_{t \in T} \delta r_{i,j} I_{ij,t,d}^{sqr} \quad (65)$$

Sujeito a:

$$\sum_{k=1}^{PV} \sum_{i \in N} \gamma^{PV} k z_{i,k} \leq IC^{PV} \quad (2) \quad (66)$$

$$-P_{t,d}^S \tan(\cos^{-1} \varphi_{min}) \leq Q_{t,d}^S \leq P_{t,d}^S \tan(\cos^{-1} \varphi_{max}) \quad (67)$$

$$\sum_{t \in T} e^p \Delta_d P_{t,d}^S \leq PE \quad (68)$$

$$P_{i,t,d}^{PV} = \sum_{k=1}^{PV} k z_{i,k} P^{-PV} f_{t,d}^{Gpv} \quad (69)$$

$$|Q_{i,t,d}^{PV}| = P_{i,t}^{PV} \tan(\cos^{-1} \varphi^{PV}) \quad (70)$$

$$\sum_{k=1}^{PV} z_{i,k} \leq 1 \quad (71)$$

$$\sum_{i \in N} \sum_{k=1}^{PV} z_{i,k} \leq N^{PV} \quad (72)$$

$$\sum_{ji \in L} P_{ji,t,d} - \sum_{ji \in L} (P_{ij,t,d} + R_{ij} I_{ij,t,d}^{sqr}) + P_{t,d}^S + P_{i,t,d}^{PV} = P_{i,t,d}^{Ld} \quad (73)$$

$$\sum_{ji \in L} Q_{ji,t,d} - \sum_{ji \in L} (Q_{ij,t,d} + X_{ij} I_{ij,t,d}^{sqr}) + Q_{t,d}^S + Q_{i,t,d}^{PV} = Q_{i,t,d}^{Ld} \quad (74)$$

$$V_{i,t,d}^{sqr} - 2[(R_{ij} P_{ij,t,d} + X_{ij} Q_{ij,t,d}) - Z_{ij}^2 I_{ij,t,d}^{sqr}] - V_{j,t,d}^{sqr} = 0 \quad (75)$$

$$V_{min}^2 \leq V_{i,t,d}^{sqr} \leq V_{max}^2 \quad (76)$$

$$0 \leq I_{ij,t,d}^{sqr} \leq I_{max}^2 \quad (77)$$

$$V_{i,t,d}^{sqr} \geq 0 \quad (78)$$

$$I_{ij,t,d}^{sqr} \geq 0 \quad (79)$$

$$V_{j,t,d}^{sqr} I_{ij,t,d}^{sqr} \geq P_{ij,t,d}^2 + Q_{ij,t,d}^2 \quad (80)$$

Assim, o modelo de fluxo em ramos pode ser calculado usando um problema de programação cônica de segunda ordem (PCSO) e resolvido usando *solvers* comerciais, como o CPLEX [39].

Em [42], afirma-se que é possível mostrar que na solução ótima do problema de PCSO, a restrição cônica de segunda ordem rotacionada está ativa e é igual à restrição não linear.

Em outras palavras, o modelo cônico de segunda ordem não é um modelo aproximado do modelo não linear, é um modelo equivalente. Ambos os problemas calculam o mesmo ponto de operação em regime permanente de um sistema de distribuição radial.

O modelo resultante é um problema de programação cônica de segunda ordem (PCSO) para o modelo de fluxo em ramos e é equivalente do problema de PNL original. Não houve um aumento no número de restrições, nem do número das variáveis contínuas.

Pode haver problemas de imprecisão no cálculo de  $I_{ij,t,d}^{sqr}$  quando a resistência  $R_{ij}$  apresenta um valor muito pequeno. Por isso é sempre importante verificar o cálculo da corrente usando a expressão não linear original. Com as hipóteses apresentadas no início da seção, conclui-se que:

- O modelo de PCSO pode ser utilizado para modelar outros problemas de otimização; e
- É possível resolvê-lo utilizando técnicas de otimização clássica existentes.

Aproveitando estas funcionalidades e/ou vantagens, na presente pesquisa será aplicada a modelagem de PCSO para alocar e dimensionar fontes renováveis fotovoltaicas para minimizar as perdas técnicas de SDEE.

#### 2.1.4 IDENTIFICAÇÃO DO PARÂMETRO $f_{t,d}^{Gpv}$ PERTENCENTE AO MODELO (65) – (80)

Na equação (69) o produto  $P^{-PV} f_{t,d}^{Gpv}$  representa a produção de potência ativa que pode ser fornecida pela fonte distribuída renovável solar à rede do SDEE. Essa multiplicação está relacionada ao valor máximo de produção da fonte renovável  $P^{-PV}$  e à eficiência do painel solar  $f_{t,d}^{Gpv}$ .

No modelo de otimização de PCSO esses dois fatores são inseridos no modelo como parâmetros determinísticos, e são conhecidos a priori mediante estudos feitos por especialistas em geração solar distribuída.

Seguindo a metodologia proposta na Figura 2.1, na próxima seção será apresentada a metodologia para considerar, mediante as técnicas de aprendizagem de máquina, especificamente com o algoritmo de *K-means*, as incertezas dos parâmetros ambientais que afetam diretamente a produção da geração distribuída solar.

Um dos objetivos do presente trabalho é considerar estas incertezas, processando-as e inserindo-as na análise de alocação e dimensionamento de fontes solares renováveis distribuídas, simultaneamente minimizando as perdas técnicas na rede do SDEE.

Como foi descrito na seção 1.3, existem alguns modelos estocásticos que permitem considerar as incertezas dentro dos modelos matemáticos de otimização aplicados aos sistemas elétricos de potência. A ideia principal deste trabalho de pesquisa é incorporar estas incertezas sem resultar em um modelo pesado do ponto de vista computacional. Assim, os valores relacionados à produção das fontes distribuídas renováveis são representados por grandezas determinísticas, levando à necessidade de um estágio prévio de gerenciamento de dados históricos a fim de considerar suas incertezas.

## 2.2 OBTENÇÃO DE PARÂMETROS DETERMINÍSTICOS

Em 1959, o cientista da IBM Arthur Samuel escreveu um programa para jogar damas. Para melhorá-lo, fez o programa jogar consigo mesmo, milhares de vezes. O programa conseguiu melhorar seu desempenho por meio da experiência, portanto, o programa aprendeu e assim nasceu a aprendizagem de máquina [4].

A aprendizagem de máquina ou *Machine Learning* (ML) estuda o aprendizado de máquina a partir de dados (*data-driven*) para ser capaz de fazer previsões precisas a partir de observações com dados anteriores [4].

A aprendizagem de máquina é um método científico que permite o uso de computadores e outros dispositivos com capacidade computacional para que eles aprendam a extrair os padrões e relacionamentos que existem em conjuntos de dados por conta própria. Esses padrões podem então ser usados para prever comportamentos e na tomada de decisões [49]. Ainda em [49] é definida a aprendizagem de máquina como um campo em inteligência artificial, onde as máquinas podem "aprender" com elas próprias, sem serem explicitamente programadas por seres humanos. Ao analisar dados anteriores, chamados "dados de treinamento", o modelo de aprendizado de máquina forma padrões e usa esses padrões para aprender e fazer previsões futuras.

A Figura 2.13 ilustra como o aprendizado de máquina pode ser usado de várias maneiras relacionadas em vários campos [41], e essas aplicações continuarão a aumentar à medida que a importância do uso de dados seja compreendida em todos os setores. Portanto, a forma como



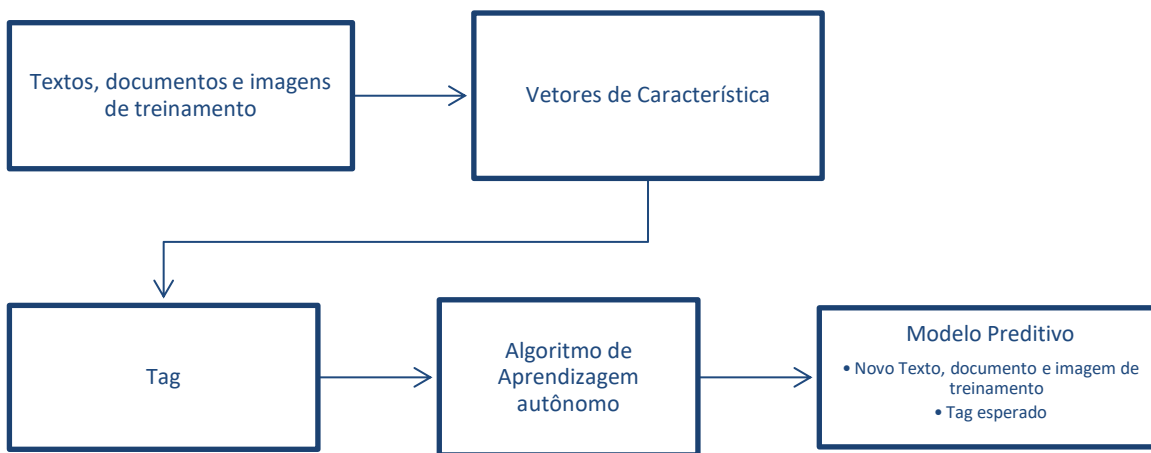
as informações são interpretadas é capaz de auxiliar na tomada de decisões e resulta em uma solução mais eficiente para os problemas [50].



**Figura 2.13** – A aprendizagem de máquina utilizada em diferentes campos. [49]

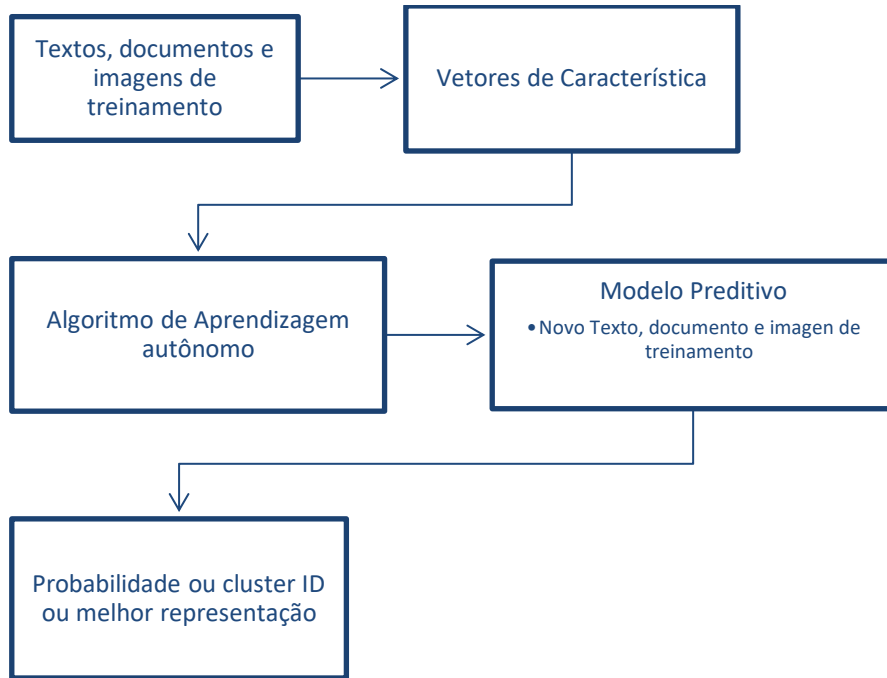
A classificação automática de objetos ou dados é um dos objetivos do aprendizado de máquina. De acordo com [49] é possível considerar três tipos de algoritmos:

- **Aprendizagem supervisionada:** Neste tipo de aprendizagem, a máquina aprende a partir de um conjunto de casos previamente rotulados por um especialista ou de forma semiautomática com base nos dados. Ou seja, para um conjunto de dados de entrada sabe-se antecipadamente os dados corretos de saída. Na aprendizagem supervisionada, o objetivo é aprender uma maneira de mapear as entradas para as saídas, cujos valores corretos são fornecidos por um supervisor. Na Figura 2.14, tem-se o modelo de aprendizado supervisionado. A máquina aprende a partir dos dados rotulados, ou seja, os dados para os quais a variável de resultado é conhecida, e faz previsões para esse resultado em novos conjuntos de dados.



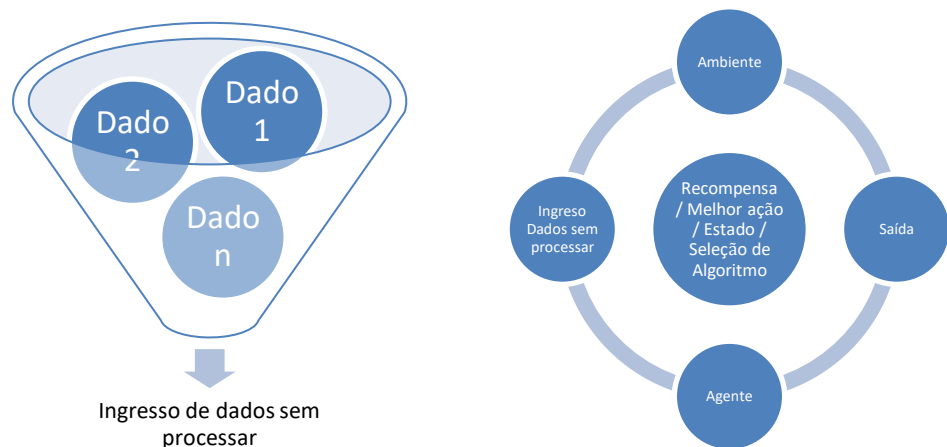
**Figura 2.14** – Modelo de aprendizagem supervisionado. [49]

- **Aprendizagem não supervisionada:** Aprendizagem não supervisionada é aquela que não requer qualquer rotulação prévia dos casos. É baseada nos dados à medida que os recebe e tem por objetivo determinar relações de semelhança, diferença ou associação. Na aprendizagem não supervisionada para um conjunto de dados de entrada, não sabemos os dados de saída com antecedência. A Figura 2.15 apresenta de forma esquemática os passos seguidos para representar o modelo de aprendizagem não supervisionado. O objetivo da aprendizagem não supervisionada é encontrar as regularidades ou padrões na entrada.



**Figura 2.15** – Modelo de aprendizagem não supervisionado. [49]

- **Aprendizagem reforçada:** Em alguns aplicativos, a saída do sistema é uma sequência de ações. Nesse caso, uma única ação não é importante; o importante é a política, que é a sequência de ações corretas para atingir a meta, ou seja, uma ação é boa se faz parte de uma boa política. Nesse caso, o programa de aprendizado de máquina deve ser capaz de avaliar a validade das políticas e aprender com as sequências de boas ações anteriores para gerar uma política. A Figura 2.16 ilustra o processo de aprendizagem reforçada.



**Figura 2.16** – Modelo de aprendizagem reforçada. [49]

### 2.2.1 OBTENÇÃO DOS AGRUPAMENTOS (*CLUSTERS*)

*Clustering* é um processo importante na aprendizagem de máquina. Este processo realiza uma ação fundamental que permite aos algoritmos de aprendizagem automatizada serem treinados e compreenderem adequadamente os dados com os quais realizam as suas atividades [41]. Este processo ajuda as máquinas a construir recursos de análise rapidamente, em grandes volumes e com o mínimo de erros possível. A seguir, esse importante processo de estudar e agrupar dados para aproveitá-los ao máximo nos projetos será apresentado com maior detalhe.

O objetivo principal do *Clustering* é agrupar conjuntos de objetos não rotulados, a fim de construir subconjuntos de dados conhecidos como *Clusters*. Cada *cluster* é formado por uma coleção de objetos ou dados que são semelhantes entre si em termos gerais, mas que possuem elementos diferenciais em relação a outros objetos pertencentes ao conjunto de dados, e que podem formar um *cluster* independente [41].

Este processo é classificado como um processo de aprendizado de máquina não supervisionado. Graças à sua implementação, um determinado sistema pode analisar os dados, realizar a tarefa e encontrar possíveis erros no seu funcionamento. O *clustering*, neste caso, serve para segmentar os dados em grupos de dimensões semelhantes com base nas características para facilitar esse processo [41].

Os processos de *clustering* em ML apresentam um nível de dificuldade significativo, pois, dependendo dos critérios e regras definidas para gerar o cluster, ele será eficiente ou não para o objetivo que se pretende alcançar. Em primeiro lugar, para realizar um processo de *clustering* deve-se definir o número de clusters que se deve ter no conjunto de dados.

Ao definir uma métrica ou um nível de erro dentro do modelo, pode-se definir níveis aceitáveis de falha para posteriormente aplicar as fórmulas dos algoritmos correspondentes com os dados que o modelo possui. Ao determinar o erro geral do modelo, ele deve ser incorporado ao algoritmo de treinamento de ML. Então, executa-se um processo iterativo para encontrar todas as combinações de erros que possam existir no modelo [41]. Este processo deve ser repetido continuamente até que o algoritmo possa entender completamente os possíveis erros que são gerados nos conjuntos de dados.

Além disso, existem alguns outros métodos de agrupamento de dados que podem ser funcionais para este tipo de projeto, dentre eles se tem o algoritmo *K-means*, que é um algoritmo hierárquico.

### 2.2.2 ALGORITMO *K-MEANS*

O algoritmo de agrupamento *K-means*, desenvolvido em [51], é um dos mais simples e populares algoritmos de aprendizado de máquina não supervisionado. O termo não supervisionado se refere à inferência de conjuntos de dados usando apenas vetores de entrada sem referir-se a resultados conhecidos. O objetivo do referido algoritmo é agrupar dados semelhantes e descobrir padrões subjacentes, procurando assim um número fixo de *clusters* em um conjunto de dados [37].

Segundo [52] “o objetivo do *K-means* é simples: agrupar pontos de dados semelhantes e descobrir padrões subjacentes. Para atingir este objetivo, *K-means* procura um número fixo ( $K$ ) de *clusters* em um conjunto de dados”.

Em [53] as incertezas inerentes à presença de recursos energéticos distribuídos e à demanda de energia elétrica são representadas por cenários de operação, selecionados com o algoritmo *K-means (clustering)*, visando preservar a cronologia dos dados históricos, bem como considerar a simultaneidade entre diferentes cargas e unidades de geração, com um número reduzido de avaliações necessárias para simular a operação da rede. Ainda em [53], afirma-se que o algoritmo *K-means* tem a capacidade de processar grandes conjuntos de dados com implementação simples e com rápida convergência

Seu processo é baseado na identificação de  $K$  centroides, e na atribuição de cada ponto do conjunto de dados ao *cluster* mais próximo, além de manter os centroides tão pequenos quanto possível. Este algoritmo é iterado até que os centroides se estabilizem (seus valores não mudam) ou após um número definido de iterações [37].

Este tipo de algoritmo é fácil de programar, é útil quando se deseja segmentar grandes volumes de dados, é computacionalmente mais barato que o algoritmo hierárquico e tem sido usado em bancos de dados reais com bons resultados. Para o algoritmo *K-means*, o número de *clusters* a serem formados deve ser conhecido com antecedência. Além disso, os *clusters* dependem em grande medida dos valores com os quais o algoritmo é inicializado, e um resultado que é ideal localmente, mas não globalmente pode ser obtido [37].

O agrupamento é feito minimizando a soma das distâncias entre cada objeto e o centroide de seu grupo ou *cluster*. A distância quadrática é frequentemente usada. O algoritmo é dividido em quatro estágios [37],[54]:

- **Estágio 1:** Arbitre o número de *clusters*  $K$ ;

- **Estágio 2:** Inicializar os centros dos *clusters* (centroides) de maneira aleatória. Escolha  $K$  objetos aleatórios  $\{x_i, i \in [1, K]\}$  do conjunto de dados para formar  $K$  *clusters* iniciais. Cada *cluster*  $K_i$  contém apenas o objeto  $x_i$ . O objeto  $x_i$  é definido como o centroide do *cluster*  $K_i$ ;
- **Estágio 3:** Aloque todos os demais dados/objetos ao *cluster* mais próximo, utilizando a Distância Euclidiana de todos os dados para cada ponto do centroide;
- **Estágio 4:** Recalcule cada centroide do *cluster* com os membros atuais do *cluster*. O centroide é a média de todos os dados/objetos dentro de um determinado *cluster*;
- **Estágio 5:** Reatribua cada objeto usando o novo centroide do *cluster*;
- **Estágio 6:** Se algum objeto foi realocado no Estágio 5, volte para o Estágio 3. Caso contrário, pare.

Embora o algoritmo sempre termine, a solução ótima não é garantida. Na verdade, o algoritmo é muito sensível à escolha aleatória dos  $K$  centroides iniciais. Esta é a razão pela qual o algoritmo *K-means* é usado inúmeras vezes no mesmo conjunto de dados para tentar minimizar este efeito, sabendo que centros iniciais mais espaçados levam a melhores resultados.

Fundamentalmente, o algoritmo *K-means* resolve um problema de otimização, minimizando uma função objetivo baseada na soma das distâncias quadráticas de cada objeto ao centroide de seu *cluster*. A escolha dos  $K$  objetos, seus centroides e a determinação de  $K$  depende dos seguintes fatores [37], [54]:

- Se  $K$  é muito pequeno, poderá haver a necessidade de fazer o agrupamento dos *clusters*;
- Se  $K$  é muito grande, alguns centros poderão ficar órfãos, ou sem agrupamento;
- O valor de  $K$  pode ser determinado por alguma heurística.

Portanto, para atingir um  $K$  ótimo ou uma aproximação conclusiva, deve-se realizar vários testes com os dados, a fim de analisar os resultados e fazer uma melhor estimativa da variável  $K$ .

Na Figura 2.17 apresenta-se um diagrama com os estágios a seguir no algoritmo *K-means*.

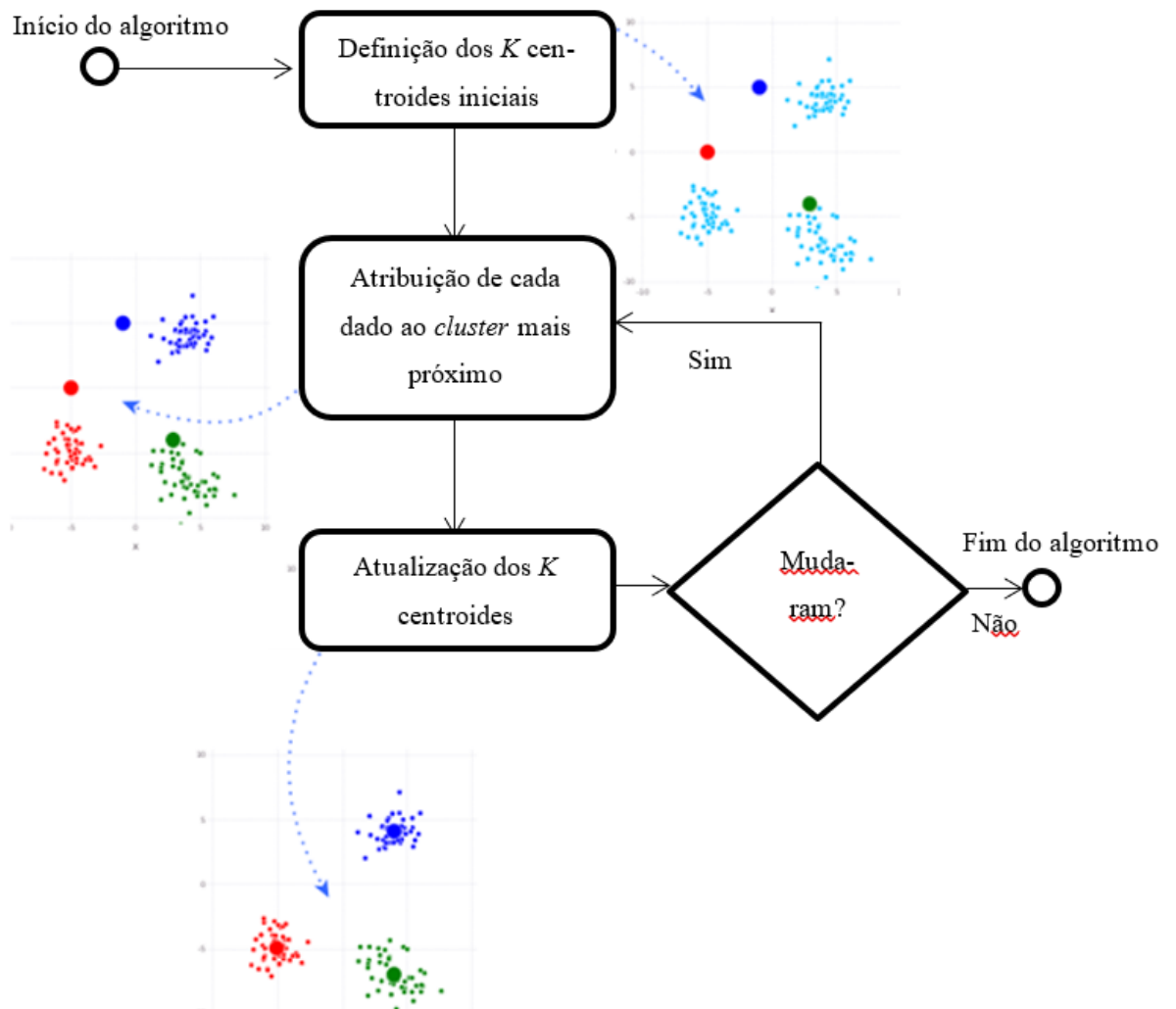
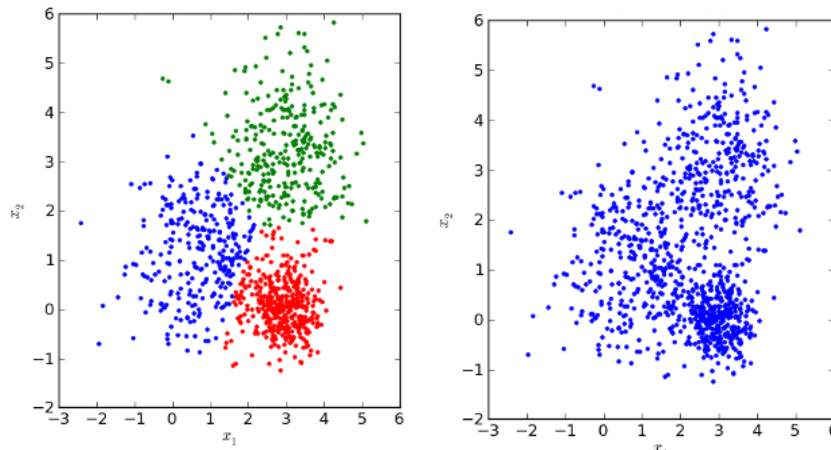


Figura 2.17 – Algoritmo de agrupamento *K-means*. [37]

### 2.2.3 ALGORITMO HIERÁRQUICO

A técnica de agrupamento hierárquico [36] é uma das técnicas de agrupamento mais populares em aprendizagem de máquina. O agrupamento é baseado na extração de agrupamentos naturais de objetos de dados semelhantes. A Figura 2.18 apresenta os gráficos dos dados quando estes estão agrupados (esquerda) e não agrupados (direita).



**Figura 2.18** – Dados agrupados e dados não agrupados. [36]

Existem algumas ideias gerais que ocorrem com bastante frequência em relação ao agrupamento:

- Os *clusters* devem estar naturalmente presentes nos dados;
- *Clustering* ou agrupamento deve descobrir padrões ocultos nos dados;
- Os pontos de dados dentro do *cluster* devem ser semelhantes;
- Os pontos de dados em dois grupos diferentes não devem ser semelhantes.

O algoritmo hierárquico é baseado no uso das técnicas de *clustering* para encontrar uma hierarquia de *clusters*, onde essa hierarquia se assemelha a uma estrutura de árvore, chamada de dendrograma. De um modo geral, existem duas maneiras de agrupar pontos de dados com base na estrutura e no desempenho algorítmico: o aglomerado e o de divisão.

### 2.2.3.1 AGRUPAMENTO HIERÁRQUICO AGLOMERADO

Em uma abordagem de *clustering bottom-up*, cada amostra é tratada como um único *cluster* e, em seguida, pares de *clusters* são sucessivamente mesclados, ou agrupados, até que todos os *clusters* tenham sido mesclados em um só.

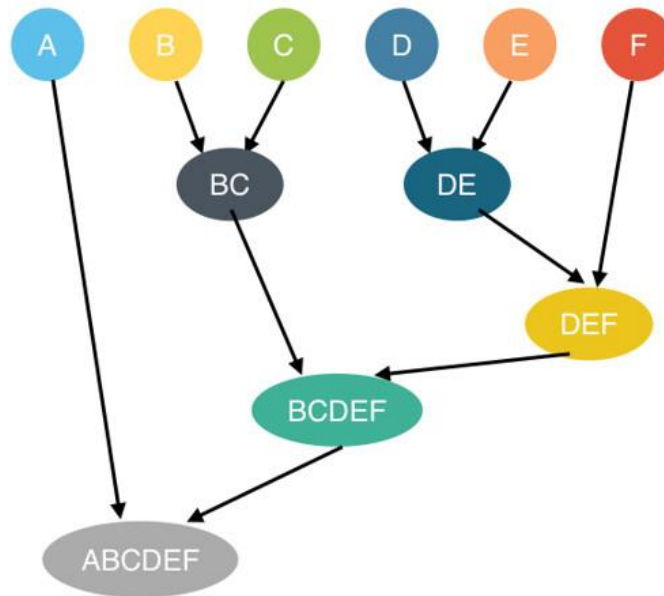
Nesta técnica, cada ponto de dados é inicialmente considerado como um *cluster* individual. Em cada iteração, grupos semelhantes são mesclados com outros grupos até que um ou mais grupos  $K$  sejam formados. O algoritmo básico é simples:

- Calcule a matriz de proximidade;
- Deixe cada ponto de dado ser um *cluster*;



- Repetir: mesclar os dois *clusters* mais próximos e atualizar a matriz de proximidade;
- Repetir os passos acima até que apenas um único *cluster* permaneça.

A operação chave é o cálculo da proximidade de dois *clusters*. Para ilustrar o processo, a Figura 2.19 mostra uma representação, ou dendrograma da técnica de agrupamento hierárquico aglomerado. Um dendrograma é um diagrama em forma de árvore que registra as sequências de fusões. Suponha que se tem seis pontos de dados: A, B, C, D, E, F.



**Figura 2.19** – Representação técnica de agrupamento hierárquico aglomerado. [36]

**Estágio 1:** Na etapa inicial, são calculadas as proximidades entre os indivíduos e são considerados os seis pontos de dados como *clusters* individuais;

**Estágio 2:** Grupos semelhantes se fundem e formam um novo grupo. Considera-se que os *clusters* (BC) e (DE) são semelhantes e que se fundem na etapa 2. Agora permanecem quatro grupos que são: A, (BC), (DE), F;

**Estágio 3:** É recalculada a proximidade entre os novos *clusters* e se mescla os *clusters* semelhantes para formar novos *clusters*: A, (BC), (DEF);

**Estágio 4:** Calcular a proximidade entre os novos *clusters*. Os *clusters* (DEF) e (BC) são semelhantes e foram mesclados para formar um novo *cluster*. Agora se tem 2 *clusters* restantes: A, (BCDEF);

**Estágio 5:** Finalmente todos os *clusters* são mesclados e formam um único *cluster*.

### 2.2.3.2 AGRUPAMENTO HIERÁRQUICO DE DIVISÃO [36]

Em um agrupamento *top-down* ou de divisão, um único agrupamento de todas as amostras é dividido recursivamente em dois agrupamentos cujos elementos são menos semelhantes até que haja um agrupamento para cada observação. Em palavras simples, pode-se dizer que o agrupamento hierárquico de divisão é exatamente o oposto do agrupamento hierárquico aglomerado. Esta técnica não é tão usada quanto a anterior, então, será apresentado apenas um resumo de seu funcionamento.

No agrupamento hierárquico de divisão, considera-se todos os pontos de dados como um único *cluster* e, a cada iteração, os pontos de dados do *cluster* que não são semelhantes são separados. Cada ponto de dados desanexado é considerado um *cluster* individual. No final, sobram  $n$  grupos. Como divide-se *clusters* individuais em  $n$  *clusters*, isso é chamado de *clustering* hierárquico de divisão.

### 2.2.3.3 AGRUPAMENTO HIERÁRQUICO: RESUMO [36]

A principal operação no agrupamento hierárquico é combinar repetidamente os dois *clusters* mais próximos em um *cluster* maior.

#### **Medidas de relacionamento**

Antes de qualquer agrupamento é necessário determinar a matriz de proximidade entre pontos usando uma função de distância. A matriz é então atualizada para mostrar a distância entre cada *cluster*. Existem várias maneiras de medir a ligação entre os *clusters* para decidir as regras de agrupamento, que muitas vezes são chamadas de métricas de ligação (do inglês *linkage metrics*), como por exemplo [55]:

- *Single/Min*: Mede a similaridade de dois *clusters* como igual ao mínimo da similaridade entre os pontos mais próximos um do outro. O benefício dessa métrica é que ela pode lidar com formas não padronizadas de *clusters*, mas não pode separar os *clusters* adequadamente se houver ruído;
- *Complete/maximum*: é o oposto da abordagem mínima, pois a similaridade dos dois *clusters* é igual ao máximo da semelhança entre os pontos dentro de cada um dos *clusters*. Isso funciona bem em separar *clusters* se houver ruído entre

eles, mas é tendencioso para *clusters* circulares e tende a quebrar grandes *clusters*;

- *Average*: Esta métrica de ligação usa a média das distâncias de todas as observações dos dois conjuntos. Isso funciona bem na separação de *clusters* se houver ruído entre os clusters;
- *Wards*: Essa abordagem de cálculo de similaridade entre os dois *clusters* é a mesma que a média do grupo, exceto que calcula a soma dos quadrados das distâncias. Essencialmente, isso minimiza a variação dos *clusters* que estão sendo mesclados. Os benefícios desta métrica é que ela se sai bem na separação de *clusters* se houver ruído, mas é tendenciosa para grupos circulares de *clusters*.

### Dendrograma

Os resultados do agrupamento hierárquico podem ser exibidos usando um dendrograma. O dendrograma pode ser interpretado com o auxílio da Figura 2.20.

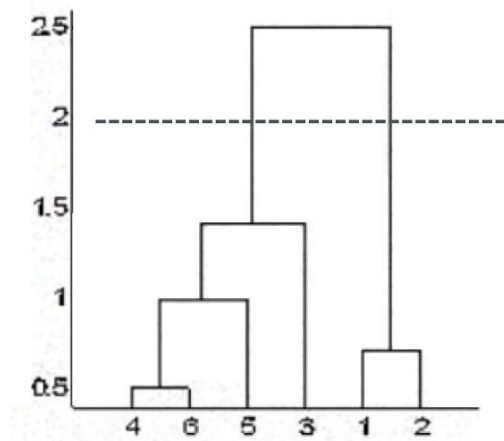


Figura 2.20 – Dendrograma. [36]

Da Figura 2.20, no início 4 e 6 são combinados em um grupo, por exemplo, grupo 1, pois eram os mais próximos em distância, seguido pelos pontos 1 e 2, por exemplo, grupo 2. Depois disso, 5 se fundiu no mesmo grupo 1 seguido por 3, resultando em dois grupos. Por fim, os dois *clusters* são mesclados em um único *cluster* e o processo de clusterização termina.

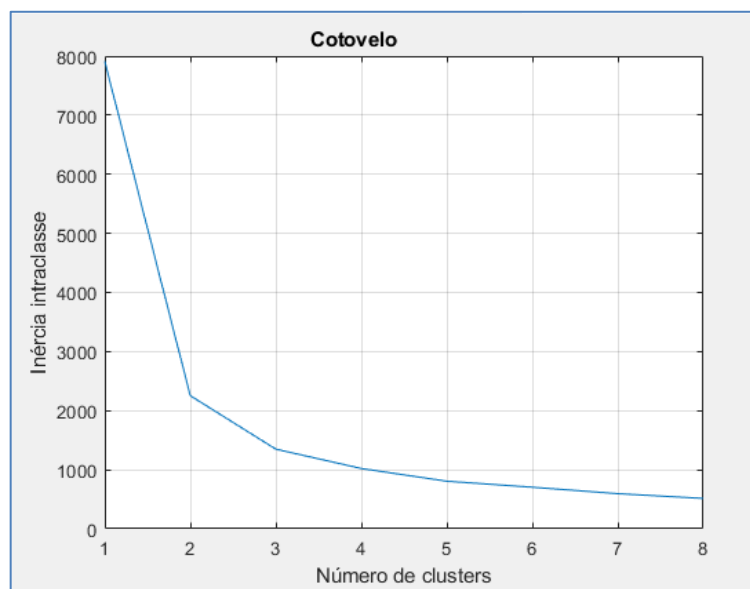
A decisão sobre o número de *clusters* que podem representar melhor os diferentes grupos podem ser escolhidos observando o dendrograma. O menor número de *clusters* é o número de linhas verticais no dendrograma cortado por uma linha horizontal que pode cruzar a distância máxima verticalmente sem interceptar um *cluster*. No caso anterior, estará entre as alturas 1,5

e 2,5. Se o corte for feito conforme mostrado, vai se acabar com apenas dois *clusters*, 1-2 e 3-4-5-6.

Não é necessário fazer um corte apenas nesses locais, podendo-se escolher qualquer ponto como ponto de corte dependendo de quantos *clusters* são desejados. Por exemplo, cortar abaixo de 1,5 e acima de 1 resultará em três *clusters*, correspondendo aos objetos 4-5-6, 3 e 1-2. Esta não é uma regra exata e fácil para decidir o número de *clusters*. Também pode-se considerar representar graficamente o método da silhueta, ou o método do cotovelo ou qualquer outro que mostre a variação do erro com o número de *clusters* e vai se decidir pelo método no qual o erro é menor.

## 2.2.4 MÉTODO DO COTOVELO

O método do cotovelo [56] é provavelmente o método mais conhecido, em que a soma total dos quadrados das distâncias entre cada dado ao centroide dentro do *cluster* é calculada e plotada. Na sequência, procura-se por uma mudança na inclinação de íngreme para rasa, um cotovelo, para determinar o número ideal de *clusters*, conforme ilustra a Figura 2.21.



**Figura 2.21** – Gráfico do Cotovelo Fonte: Autor.

O método da curva de cotovelo é útil porque mostra como o aumento do número de *clusters* ajuda a separar os *clusters* de uma maneira significativa. O método do cotovelo é bastante direto, mas uma solução ingênua baseada na variação *intracluster*. A estatística de lacuna é um método mais sofisticado de lidar com dados que têm uma distribuição sem agrupamento

óbvio [56]. O método do cotovelo é usado neste trabalho porque a característica do problema de agrupamento abordado é óbvia, baseado na observação do rendimento da solução proposta para as duas estações do ano (com chuva e sem chuva).

O método do cotovelo funciona como descrito a seguir. Escolhe-se um  $K$  tal que a variação de inércia para  $K$  e  $(K+1)$  esteja abaixo de certo limiar. Isso é visível como um cotovelo como se explica a seguir [56]:

- O erro quadrático para cada ponto é o quadrado da distância do ponto de sua representação, ou seja, seu centro previsto do *cluster*;
- A soma da pontuação dos erros quadrados é a soma desses erros quadráticos para todos os pontos;
- Qualquer métrica de distância pode ser usada como, por exemplo, distância euclidiana.

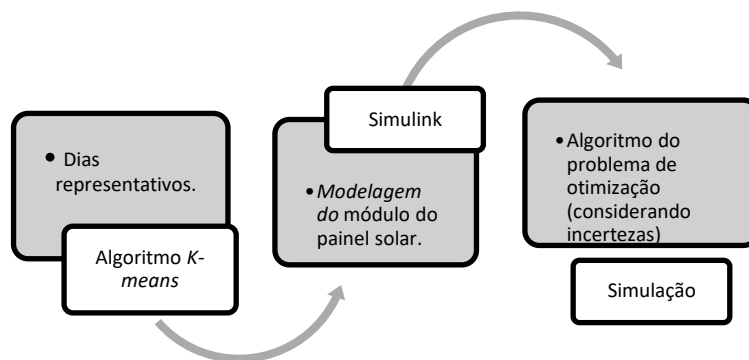
A aplicação do método do cotovelo para a obtenção do número ideal de *clusters* pode ser feita da seguinte maneira:

- I. Calcule o algoritmo de agrupamento para diferentes valores de  $K$ . Por exemplo, variando  $K$  de 1 a 10 grupos;
- II. Para cada  $K$ , calcule a soma total do quadrado das distâncias entre cada dado ao centroide dentro do *cluster*;
- III. Desenhe a curva da soma dos erros quadráticos de acordo com o número de grupos  $K$ .

A localização de uma curva, cotovelo, no gráfico é geralmente considerada um indicador do número apropriado de grupos. O nome cotovelo deve-se à forma do gráfico que é gerado e, o ponto onde começa a sua estabilização ou diminui a variação, indica onde se encontra o  $K$  ideal, o que faz com que a inércia (soma das distâncias quadradas de cada objeto no *cluster* a seu centroide que em inglês corresponde ao WCSS – *within-clusters sum-of-squares*) se estabilize o que implica que não vai ser mais necessário tentar melhorar porque a inércia intraclasse se estabilizou.

### 2.3 MODELAGEM DE MÓDULOS SOLARES USANDO SIMULINK [57]

A Figura 2.22 apresenta o fluxograma geral para a modelagem de módulos solares.

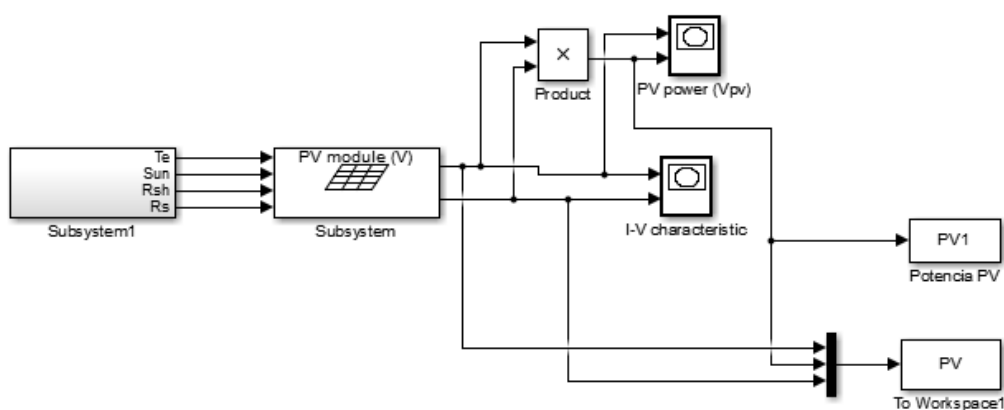


**Figura 2.22** – Fluxograma geral de modelagem dos módulos solares. Fonte: Autor.

Segundo a Figura 2.22, uma vez determinados os dias representativos (representantes de cada grupo) necessários para o modelo do sistema elétrico, os quais foram obtidos por meio de aplicações de aprendizado de máquina por meio do algoritmo *K-means* [6], que nos permitiu introduzir variáveis determinísticas em um modelo de linearização matemática, partindo de variáveis estocásticas como temperatura e irradiação solar.

Pode-se ir para o próximo passo e utilizar o programa Simulink do MATLAB [57] para modelar o módulo do painel solar a ser utilizado, a partir do qual serão obtidos os valores da potência do módulo solar. Com os resultados obtidos com a aplicação do algoritmo *K-means* e do modelo Simulink, pode-se finalmente modelar e resolver o problema de otimização utilizando o AMPL [7].

A Figura 2.23 mostra o modelo feito em Simulink para a obtenção da potência máxima fornecida por uma célula solar.



**Figura 2.23** – Modelo do módulo solar utilizando o Simulink. Fonte: Autor.

Para o modelo apresentado na Figura 2.23, é necessário ter os parâmetros de entrada que são obtidos aplicando o algoritmo *K-means*.

O bloco (*Subsystem*), chamado de “PV module (V)”, representa uma única célula solar. Assim, a célula é o elemento mais simples, e a conexão de algumas delas em arranjos série/paralelo representam as fontes solares alocadas pela distribuidora. A seguinte denominação será utilizada nos próximos capítulos: a unidade mais simples é a célula, um conjunto de células forma um módulo, um conjunto de módulos forma um painel, um conjunto de painéis forma um arranjo solar, e, finalmente, um conjunto de arranjos forma uma planta fotovoltaica.

## 2.4 APLICAÇÕES DE *CLUSTERING* EM SISTEMAS ELÉTRICOS DE POTÊNCIA

Com a maior disponibilidade de informações obtidas por medições, os estudos de carga no nível residencial e comercial são utilizados para melhorar os perfis de carga [58] e também para melhorar a caracterização do usuário que demanda a carga.

A característica das cargas residencial e comercial é representada em termos de número de usuários (ou contas) como aquela que têm o maior número de consumidores finais totais. Além disso, a carga residencial difere das cargas comercial e industrial porque apresenta maior heterogeneidade entre os usuários, não podendo ser agrupada por semelhanças socioeconômicas ou localização geográfica, uma vez que os padrões de consumo mudam de uma residência para outra.

Esta variabilidade de padrões do consumo torna muito importante a aplicação de técnicas de agrupamento ou *clustering*, que pode ser utilizada em diferentes aplicações dentro do setor elétrico.

### 2.4.1 DETERMINAÇÃO DA TARIFA PARA O SERVIÇO DE ENERGIA ELÉTRICA NAS EMPRESAS DE DISTRIBUIÇÃO

A determinação das tarifas para o serviço de energia elétrica nas empresas de distribuição é normalmente realizada com base em considerações econômicas e, se for necessário, com a aplicação de subsídios a determinados grupos de consumidores considerando características sociodemográficas [58].

Nesse sentido, técnicas de agrupamento podem ser utilizadas para maximizar os benefícios anuais a partir do agrupamento de consumidores, propondo novas alternativas como a criação de taxas múltiplas para os grupos formados, ou seja, busca atribuir a taxa mais adequada para cada grupo de clientes com características semelhantes.

É muito importante destacar que o número de grupos ou *clusters* pode ser estabelecido pela distribuidora ou comercializadora de energia, mas também pode ser estabelecido mediante a utilização de algoritmos de *clustering*.

#### **2.4.2 PROGNÓSTICO DA DEMANDA**

As aplicações de *clustering* na previsão de carga são baseadas em uma técnica chamada *Cluster-based Aggregate Forecasting (CBAF)*, que consiste em fazer grupos ou *clusters* de usuários residenciais, prevendo a carga para cada *cluster* em intervalos de 1 a 24 horas e, em seguida, calculando a carga total prevista [59]. Esta metodologia depende do número de *clusters* usados e do número total de usuários residenciais.

#### **2.4.3 RESPOSTA DA DEMANDA**

Assim como no caso da determinação de tarifas, os métodos de agrupamento podem ajudar a identificar, com base nos padrões de consumo, aqueles grupos para os quais os incentivos de redução do consumo podem ser direcionados para reduzir a demanda nos horários de pico e melhorar a estabilidade do sistema. Para esta abordagem, algumas metodologias ou técnicas de mineração de dados contam com curvas de carga, enquanto outros estudos a segmentação demográfica foi realizada [58].

#### **2.4.4 CLASSIFICAÇÃO DE USUÁRIOS QUE NÃO TÊM MEDIDORES INTELIGENTES**

Uma das aplicações mais recentes nos SDEE é a aplicação de *clustering* com base em dados obtidos de medidores inteligentes para poder agrupar aqueles usuários para os quais ainda não existem informações de medidores inteligentes. A partir dos *clusters* criados, a metodologia



permite que seja determinado um perfil de carga característico de cada *cluster*, que posteriormente é comparado com o perfil de carga dos usuários não agrupados ou daqueles que possuem pouca informação devido às limitações de medição [58].

#### **2.4.5 EXEMPLO DE APLICAÇÃO DO ALGORITMO *K-MEANS* PARA A ESTRATIFICAÇÃO DE CONSUMIDORES**

Na seção 2.2 foi apresentada a teoria para obter valores determinísticos ou valores representativos de grupos de dados históricos divididos em diferentes grupos de acordo com seus padrões característicos. A seguir é apresentado o procedimento para aplicar o algoritmo de *K-means* para encontrar padrões de consumo energético e, assim, poder estratificar consumidores residenciais da empresa de distribuição [60].

##### **2.4.5.1 PRÉ-PROCESSAMENTO DE DADOS [60]**

Para iniciar o trabalho de *clustering*, é importante receber os dados de trabalho e adaptá-los às necessidades de estratificação dos dados, ou seja, obter as informações realmente importantes que permitirão categorizar os elementos do *cluster*.

O presente exemplo tem como objetivo estratificar os usuários da empresa de distribuição de energia elétrica com base no seu consumo histórico de energia, por isso é importante obter uma diferenciação de cada usuário e as informações sobre o consumo de energia em kWh para cada um dos meses dentro da análise.

##### **Dados obtidos da empresa de distribuição de energia elétrica**

A empresa de distribuição de energia elétrica disponibilizou as informações de seus usuários durante 18 meses, correspondendo a todo o ano de 2019 e primeiro semestre de 2020, em arquivos do tipo apresentado na Figura 2.24.

A1	GYE,"200020924433","BTCSR0010","MARIA ANGELA FARIÑA O QUINDE","ETAPAS\MZ#785\SOL#05 Y COOP.SAN FRANCISCO PB																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	GYE,"200020924433","BTCSR0010","MARIA ANGELA FARIÑA O QUINDE","ETAPAS\MZ#785\SOL#05 Y COOP.SAN FRANCISCO PB PASCUALES","0915737332","0997190854","616163.6828","9773502.567","0409M035","GUAYAQUIL","PASCUALES",																										
2	GYE,"200023643121","BTCSR0010","FRANKLIN JUBER MOSQUERA CHEME","MZ# 2516, SOL# 18 Y . PB MUCHO LOTE# 4","0800307175","0911105968","619962.4533","9768636.947","0416M017","GUAYAQUIL","PASCUALES",																										
3	GYE,"200016953842","BTCSR0010","JULIA MARIA GOMEZ PEÑA AFIEL","BLQ#1\MZ#666\SOL#24 Y . PB BASTION POPULAR","0909962078","0994484191","618901.1214","9768161.427","0409M054","GUAYAQUIL","PASCUALES",																										
4	GYE,"200017875440","BTCSR0010","CLARA IVONNE CASSANELLO GARCES","MZ# 5224, SOL# 21 Y URB.MI LOTE PB VILLA BONITA","0930614417","0979857048","613841.4155","9771994.034","0409M028","GUAYAQUIL","PASCUALES",																										
5	GYE,"200020285041","BTCSR0010","ANA MARITZA SANCHEZ ANTEPARA","6TA.CALLE 0218 Y 2DA.-4TA.AVENIDAS PB PASCUALES","0915629687","0980063991","619002.6318","9771217.335","0412M062","GUAYAQUIL","PASCUALES",																										
6	GYE,"200015996362","BTCSR0010","MARTHA JULIA IZQUIERDO CEDILLO","MZ# 184, SOL# 02 Y COOP.LOS VERGELES PB VERGELES","0923241954","621504.781","9768549.833","0412M052","GUAYAQUIL","PASCUALES",																										
7	GYE,"200020537961","BTCSR0010","ROMELIA GUALE MENOSCAL","BLQ#4\MZ#700\SOL#17 Y . PB BASTION POPULAR","0910712777","619744.7063","9767857.663","0410M058","GUAYAQUIL","PASCUALES",																										
8	GYE,"200017580370","BTCSR0010","SANDY MARIANA GONZALEZ SANCHEZ","MZ# 2313, SOL# 11 Y COOP.TIWINZA PB","0930063177","0990064379","614327.2048","9768233.776","0413M095","GUAYAQUIL","PASCUALES",																										
9	GYE,"200015915651","BTCSR0010","MARTHA LILIANA CALLE LOJA","MZ# 960, SOL# 21 Y CIUDAD DEL RIO PB CDLA.MAGISTERIO","0909049488","0994432834","621315.4353","9771346.453","0409M002","GUAYAQUIL","PASCUALES",																										
10	GYE,"200022974816","BTCSR0010","WELLINGTON ALBERTO BRIONES LEON","MZ# 2516, SOL# 19 Y . PB MUCHO LOTE# 4","0905388525","0989297165","619958.3693","9768652.593","0416M017","GUAYAQUIL","PASCUALES",																										
11	GYE,"200021734773","BTCSR0010","FELIX SEGUNDO ALAVA ANCHUNDIA","BLQ#3\MZ#730\SOL#03 Y . PB BASTION POPULAR","0908253669","0994444647","619629.5193","9768273.32","0410M055","GUAYAQUIL","PASCUALES",																										
12	GYE,"200020828253","BTCSR0010","SERGIO MACARIO FLORES MENA","MZ# 58, SOL# 34 Y . PB CDLA.ORQUIDEAS","0901343897","620730.2494","9768986.612","0412M003","GUAYAQUIL","PASCUALES",																										
13	GYE,"200022970376","BTCSR0010","EDISON OMAR ARELLANO CARBO","MZ# 2189, SOL# 17 Y ESPAÑA-A-MUCHO LOTE PB CDLA.MALLORCA","0916417868","620296.1879","9770057.996","0416M004","GUAYAQUIL","PASCUALES",																										
14	GYE,"200023600816","BTCSR0010","FERMINA J SANTANA CORONEL","MZ# 2223, SOL# 20 Y ESPAÑA-A-MUCHO LOTE PB CDLA.MADRID","1203917289","0994423163","620318.4023","9769900.484","0416M002","GUAYAQUIL","PASCUALES",																										
15	GYE,"200019834247","BTCSR0010","DIANA LEON NADER","MZ# 1032, SOL# 11 Y . PB CDLA.ORQUIDEAS","1201201637","620574.2931","9769354.705","0412M008","GUAYAQUIL","PASCUALES",																										
16	GYE,"200018725008","BTCSR0010","LORENA LIZ AMAIQUEMA ONOFRE","MZ# 97, SOL# 3 Y COOP.LOS VERGELES PB VERGELES","0914429014","0997602836","621980.4745","9769233.758","0412M043","GUAYAQUIL","PASCUALES",																										
17	GYE,"200016216679","BTCSR0010","CECILIA C AGUIRRE TROYA","MZ# 1374, SOL# 10 Y COOP.LOS VERGELES PB VERGELES","1202703565","621995.2531","9769357.544","0412M045","GUAYAQUIL","PASCUALES",																										
18	GYE,"200020599664","BTCSR0010","MARTHA EUFEMIA VALLEJO PAEZ","MZ# 54, SOL# 13 Y . PB CDLA.ORQUIDEAS","0902073592","0995341951","620672.7514","9768970.378","0412M002","GUAYAQUIL","PASCUALES",																										
19	GYE,"200023768837","BTCSR0010","NELSON P ALVARADO FLORES","MZ# 2281, SOL# 18 Y ESPANA-MUCHO LOTE PB CDLA VALENCIA","0911645638","0994853296","619961.6043","9769622.537","0416M009","GUAYAQUIL","PASCUALES",																										
20	GYE,"200020948879","BTCSR0010","ROSA MARISOL GONZALEZ VERA","BLQ#7A\MZ#1017,SL#15 Y . PB BASTION POPULAR","0919934943","0993660978","618661.0366","9769270.549","0412M019","GUAYAQUIL","PASCUALES",																										
21	GYE,"200019360599","BTCSR0010","NARCISA DE JESUS MONTOYA ZAMBRANO","BLQ#22\MZ#1442,SL#14 Y . PB FLOR DE BASTION","0912884210","614638.5773","9769867.322","0411M045","GUAYAQUIL","PASCUALES",																										

Figura 2.24 – Dados fornecidos pela empresa de distribuição de energia elétrica [60].

O arquivo mostrado na Figura 2.24 tem extensão “.csv”, disponibilizado pela empresa de distribuição de energia elétrica para o primeiro mês de 2019. Este arquivo contém uma grande quantidade de informações desnecessárias para cada usuário, do ponto de vista da estratificação como: endereço do usuário, número de identificação, telefone, entre outros. As informações devem ser refinadas para que possam ser processadas, evitando arquivos demasiadamente grandes, com informações irrelevantes.

### **Crerios de processamentos de dados e metodologia aplicada**

Desses arquivos é extraída a informação única de cada usuário, que neste caso será o código de conta do contrato fornecido pela empresa para cada usuário. A informação do consumo de energia de cada mês registrado por leituras devidamente efetuadas pelos colaboradores da concessionária é a informação relevante para o agrupamento, razão pela qual será a informação que deve ser extraída de cada um dos arquivos disponibilizados. Para extrair essas informações, um código em Matlab [57] é apresentado na Figura 2.25.

```

%% Close the text file.
fclose(fileID);
frmt0='%*s %[^\\n\\r]';
consumo=[];
codigo=[];
for i= 1:length(dataArray{1,2})
    linea=dataArray{1,2}{i};
    if length(strfind(linea,'@'))>0
        linea=string(textscan(linea,frmt0,'Delimiter','@'));
    end
    filtr1=textscan(linea,frmt0,'Delimiter','->');
    filtr2=textscan(string(filtr1),frmt0,'Delimiter','0');
    filtr3=textscan(string(filtr2),frmt0,'Delimiter','1');
    filtr4=textscan(string(filtr3),'%*s %q %*[^\\n\\r]','Delimiter','');
    cons=textscan(string(filtr4),'%f %*[^\\n\\r]','Delimiter','');
    cod=textscan(string(dataArray{1,1}{i}),'%f %*[^\\n\\r]','Delimiter','');

    codigo(i,1)=cell2mat(cod);
    consumo(i,1)=cell2mat(cons);
end
result=[codigo consumo];

clearvars filename delimiter formatSpec fileID linea filtr1 filtr2 filtr3 filtr4 cons ans frmt0 cod i;
dlmwrite('C:\Users\kevin\Downloads\DATOS_CNEI\DatosSemiproces\pre202006.csv',result,'precision','%2f')

```

**Figura 2.25** – Código de dados semiprocessados. [60]

A Figura 2.25 apresenta o código [60] em Matlab [57] usado para extrair o código e consumo de cada usuário. A ideia é criar um arquivo com duas colunas: a primeira tem o código e a segunda tem o consumo mensal. Desta forma, de cada arquivo de cada mês é criado um novo arquivo o qual apresenta apenas estas duas informações, fornecendo dezoito novos arquivos que para os fins deste exemplo serão denominados "Dados Semiprocessados".

Uma vez realizado o pré-processamento, é necessário juntar esses dezoito arquivos em um único arquivo, para o qual é utilizado um novo código no Matlab [57].

As Figura 2.26 e 2.27 mostram o código desenvolvido. Seus resultados são apresentados na Figura 2.28.

```

1 - clc
2 - clear all
3
4 - ruta= 'C:\Users\alber\Downloads\DatosSemiproces\DatosSemiproces\pre';
5 - an19=(201901:01:201912);
6 - an20=(202001:01:202006);
7 - meses=[an19 an20];

```

**Figura 2.26** – Código para *upload* de dados semiprocessados [60].

A Figura 2.26 apresenta o código para fazer *upload* os arquivos que têm os dados semiprocessados criados anteriormente com o objetivo de unificá-los.

```

9 - fileID = fopen(strcat(ruta,num2str(meses(1)),'.csv'),'r');
10 - datacell = textscan(fileID, '%f%f%[\n\r]', 'Delimiter', ',');|
11 - fclose(fileID);
12 - A=cell2mat(datacell);
13 - Enel9=table(A(:,1),A(:,2),'VariableNames',{'Codigo' 'Cons_201901'});
14 - clearvars fileID datacell A
15 - GranM=Enel9;
16 - columnas={'Cons_201901'};
17 - for i =2:length(meses)
18 -     fileID = fopen(strcat(ruta,num2str(meses(i)),'.csv'),'r');
19 -     datacell = textscan(fileID, '%f%f%[\n\r]', 'Delimiter', ',');
20 -     fclose(fileID);
21 -     col_name=strcat('Cons_',num2str(meses(i)));
22 -     columnas(i)=col_name;
23 -     B=cell2mat(datacell);
24 -     mes=table(B(:,1),B(:,2),'VariableNames',{'Codigo' col_name});
25 -     clearvars fileID datacell B
26 -
27 -     GranM = outerjoin(GranM,mes,'MergeKeys',true);
28 - end
29 - clearvars i Enel9 an19 an20 colname
30 - idx_miss= ismissing(GranM(:,columnas));
31 - GranM(:,columnas)(idx_miss) = 0;

```

**Figura 2.27** – Código para unificar os dados semiprocessados [60].

A Figura 2.27 apresenta o código que faz a leitura de cada um dos 18 arquivos e retorna uma matriz completa com 19 colunas, a primeira contendo os códigos de cada usuário já existente desde janeiro de 2019 e os novos usuários adicionados dos outros meses, colunas 2 até 19 têm o consumo de cada usuário para cada um dos 18 meses.

É natural que existam usuários com valores não tomados em todos os meses, pois são contas que foram adicionadas após a data de início da análise. Uma vez que esses valores estranhos são contraproducentes para o trabalho de agrupamento, é necessário aplicar um filtro às informações conforme mostrado na Figura 2.28.

```

32 %% Filtrado de Datos
33 - idx_inc_rows=sum(idx_miss,2)>0;% Indices de filas (usuarios) que le faltan datos (NaN)
34 - Users_dat_incomp=sum(idx_inc_rows);% Numero de usuarios con datos incompletos(NaN)
35 - Users_dat_compl=height(GranM)-Users_dat_incomp;% Numero de usuarios con datos completos, pero con 0
36 - M_no_miss=GranM;
37 - M_no_miss(idx_inc_rows,:) = []; %Tabla sin los usuarios que tienen NaN
38 - idx_nonzero=table2array(GranM(:,columnas))>0;
39 - num_med=sum(idx_nonzero,2);% Numero de mediciones por usuario
40 - idx_compl_rows=num_med<18;% Indices de filas(usuarios) que tienen 18 mediciones diferente de 0
41 - M_no_zero=GranM;
42 - M_no_zero(idx_compl_rows,:)=[];%Tabla sin los usuarios que tienen algún 0 o NaN

```

**Figura 2.28** – Código para o filtrado de dados [60].

O código mostrado na Figura 2.28 usa a matriz obtida anteriormente e elimina usuários com dados perdidos e medições nulas, deixando apenas aqueles que têm suas 18 medições completas e com os valores de medições reais.

Esses usuários são os que serão utilizados para o agrupamento, sendo um total de 498.991, para mostrar o número de usuários com medições completas, foi elaborado um código que é apresentado na Figura 2.29.

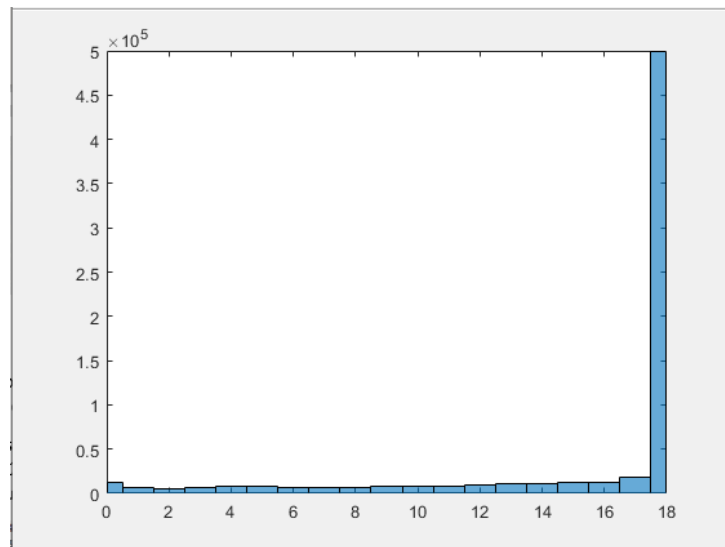
```

44 %% Resumen de mediciones
45 - Med=histogram(num_med,'BinLimits',[0,max(num_med)])
46 - Dist_med=histcounts(num_med,'BinLimits',[0,max(num_med)]);
47 - T_med=table((0:1:max(num_med))',Dist_med,'VariableNames',{'Total Mediciones' 'Total Usuarios'});

```

**Figura 2.29** – Código para a criação do histograma [60].

A Figura 2.29 mostra o código que conta o número de usuários que têm medições completas, que estão perdendo uma medição, e que estão perdendo duas medições. A resposta deste processo é apresentada no histograma da Figura 2.30.



**Figura 2.30** – Histograma [60].

No histograma apresentado na Figura 2.30 é possível observar a comparação entre os usuários com medidas faltantes, onde vemos que a maioria possui as 18 medidas completas, de forma que os usuários eliminados no agrupamento representam um número reduzido em relação aos dados iniciais globais. Isso confirma que os usuários presentes na matriz resultante dada pela filtragem dos dados são uma amostra significativa (amostra representa 75% dos usuários finais) para a estratificação dos usuários.

A seguir, deve-se eliminar os usuários que têm todas as suas medições repetidas, e o código mostrado na Figura 2.31 é usado para este fim.

```

48     %% Eliminacion de datos repetidos
49 -   varianzas=var(table2array(M_no_zero(:,columnas)),0,2);
50 -   idx_var_rp=varianzas<=0.1;
51 -   M_no_repeat=M_no_zero;
52 -   M_no_repeat(idx_var_rp,:)=[];
53 -   M_rep=M_no_zero{idx_var_rp,:};

```

**Figura 2.31** – Código para eliminação de dados ou leituras repetidas [60].

Para eliminar esses valores é utilizado o critério de variância, conforme mostrado no código da Figura 2.31. A variância é calculada para as linhas da matriz, ou seja, para cada um dos usuários, então com este valor para cada usuário, aqueles que têm uma variação menor ou igual a 0,1 são considerados.

A matriz chamada “M\_no\_repeat” usa a matriz sem dados igual a zero e compara esta matriz com a variância dos usuários e elimina aqueles que estão fora do parâmetro inserido, esta variância poderia ser reajustada para aceitar valores maiores ou menores.

A importância de eliminar esses usuários com dados repetidos reside no fato de que, embora possa parecer normal, à primeira vista, que um usuário consuma a mesma coisa todos os meses, isso não corresponde à realidade, pois ignora que existe um tempo de maior calor na cidade concessionada pela empresa distribuidora, onde a experiência indica que o consumo de energia cresce devido ao uso de ventiladores e condicionadores de ar. Tampouco faz sentido que em 18 meses um usuário consuma o mesmo valor de energia, já que os usuários tendem a adquirir novos equipamentos como eletrodomésticos, televisores, consoles etc.

A Figura 2.32 mostra o código usado para eliminar os *outliers*, ou seja, os valores atípicos no banco de dados históricos.

```

56     %% Filtrado
57 -   Info=table2array(M_no_repeat(:,2:19));
58 -   codigo=table2array(M_no_repeat(:,1));
59 -   [idx_outlier,lower,upper,center]=isoutlier(Info,'median',2,'ThresholdFactor',5);
60 -   num_haveoutl=sum(idx_outlier,2);
61 -   Depurado=Info;
62 -   Info=[Info,num_haveoutl];
63 -   Depurado(idx_outlier)=NaN;
64 -   F=fillmissing(Depurado,'movmean',7,2);
65 -   idx_Oneoutl=num_haveoutl<=1;
66 -   M_final=array2table([codigo(idx_Oneoutl),F(idx_Oneoutl,:)],'VariableNames',{'Codigo' columnas});
67 -   Data_Out=[codigo(~idx_Oneoutl),Info(~idx_Oneoutl,:)];

```

**Figura 2.32** – Código para filtragem de *outliers* [60].

O outlier, também conhecido como valor aberrante ou valor atípico, corresponde a uma observação inconsistente, ou que apresenta um grande afastamento das demais da série. Por exemplo, informações referentes a um mês que não fazem sentido em relação àquelas de meses

anteriores e posteriores. A função *isoutlier* do Matlab é utilizada para identificá-lo e retirá-lo da matriz, para se ter um melhor desempenho na análise dos resultados.

#### 2.4.5.2 NORMALIZAÇÃO DE DADOS [60]

Após eliminar os valores nulos, repetições e *outliers*, pode-se redimensionar os dados. Isso é chamado de normalização e o código correspondente é apresentado na Figura 2.33. Neste exemplo, o agrupamento foi realizado para os dados normalizados e reais.

```

55     %% Reescalamiento
56 -   M_rescale=M_no_repeat;
57 -   for k = 1:length(columns)
58 -       col_M=table2array(M_rescale(:,k+1));|
59 -       max_value=max(col_M);
60 -       min_value=min(col_M);
61 -       col_resc=(col_M- min_value)/ (max_value - min_value);
62 -       M_rescale.(k+1)=col_resc;
63 -   end
64 -   clearvars k col_M col_resc min_value max_value

```

**Figura 2.33** – Código para normalização de dados [60].

Para isso, foi definida uma matriz “M\_rescale”, que é baseada na matriz sem outliers de dados obtidos anteriormente com o código da Figura 2.32. Na Figura 2.33 é mostrado que para cada coluna, ou seja, para cada mês, são tomados os valores máximo e mínimo de consumo de eletricidade, então cada valor é tomado e normalizado subtraindo o valor mínimo e dividindo esta diferença pela diferença entre os valores de consumo máximo e mínimo do mês. Isto é iterado até que uma matriz de dados normalizados seja obtida. Uma vez realizado este processo, se procede com o algoritmo *K-means*. A equação para normalizar é:

$$Data_{Resc} = \frac{Data_{Real} - Min}{Max - Min} \quad (81)$$

#### 2.4.5.3 ALGORITMO USADO EM DETALHE [60]

Para usar o algoritmo, primeiro define-se os dados que serão usados, e essas linhas de código são mostradas na Figura 2.34.

```

79 %% Dados para Kmeans
80 - Data=table2array(M_final(:,2:19));
81 - Data_norm=table2array(M_rescale(:,2:19));
82 - max_data=max(Data,[],1);
83 - min_data=min(Data,[],1);
84 - std_data=std(Data_norm,0,2);
85 - cons_data=mean(Data,2);

```

**Figura 2.34** – Código para usar o algoritmo *K-means* [60].

Conforme mostrado na Figura 2.34, “Data” e “Data\_norm” são definidos, matrizes que não contêm o número de identificação dos usuários, mas apenas seus consumos durante os 18 meses. Adicionalmente os dados mínimo e máximo são obtidos e definidos para trabalhar no algoritmo com a média.

### **Algoritmo no MATLAB [57]**

A seguir, será mostrado um exemplo de como funciona o algoritmo utilizado, para que o leitor conheça de forma prática o funcionamento do *K-means*, que posteriormente será aplicado diretamente como uma função Matlab [57] para a estratificação de usuários da empresa de distribuição de energia elétrica. A seguir serão descritos os passos do algoritmo de *K-means*.

#### Inicialização

Neste primeiro estágio, os  $k$  centroides iniciais para os *clusters* são selecionados aleatoriamente. No algoritmo tradicional, os  $k$  centroides são escolhidos aleatoriamente ao mesmo tempo a partir das  $k$  observações presentes no conjunto de observações  $X$ .

Matlab [57] faz uso da inicialização *K-means*, quando o algoritmo é repetido várias vezes para encontrar a solução ótima. Esta variante de inicialização consiste no seguinte:

- I. Selecione o primeiro centroide  $C_1$  aleatoriamente do conjunto de observações  $X$ , assumindo que todas as observações têm a mesma probabilidade de serem escolhidas.
- II. Calcule as distâncias  $d(x, C_1)$  de cada uma das observações  $x \in X$  ao centroide  $C_1$ .
- III. Selecione o próximo centroide  $C_2$  aleatoriamente do conjunto de observações  $X$ , assumindo agora que a probabilidade de uma observação  $x \in X$  ser escolhida é dada por:

$$P(x_m) = \frac{d^2(x_m, C_1)}{\sum_{i=1}^n d^2(x_i, C_1)} \quad (82)$$



- IV. Calcule as distâncias de cada uma das observações  $x \in X$  até o centroide recém-selecionado.
- V. Selecione o seguinte centroide  $C_j$  aleatoriamente do conjunto de observações  $X$ , assumindo agora que a probabilidade de que uma observação  $x$  seja escolhida é proporcional a:

$$P(x_m) = \frac{d^2(x_m, C_p)}{\sum_{i=1}^n d^2(x_i, C_p)} \quad (83)$$

onde o centroide  $C_p$  usado para cada  $x_m$  ou  $x_i$  será aquele que produz a menor distância entre os centroides já calculados.

- VI. Itera-se nos estágios IV e V até encontrar os  $K$  centroides.

#### Classificação

Para cada observação  $x \in X$ , calcula-se a distância aos centroides e atribui-se a observação ao centroide com a menor distância, atribuindo-a, portanto, ao aglomerado que o centroide representa.

#### Recálculo dos centroides

Os  $k$  novos centroides são recalculados tomando a média das observações de cada um dos  $k$  grupos.

#### Iteração ou convergência

É repetida a classificação e o recálculo dos centroides até a atribuição de observações a seus *clusters* não muda.

Um exemplo simples da aplicação do *cluster* é o seguinte:

“Dado o seguinte conjunto de observações de partículas aleatórias em um espaço de  $\mathbb{R}^3$ , deseja-se catalogá-las em dois *clusters*.”

**Tabela 2.1** – Amostras para o exemplo simples do funcionamento do *K-means*.

Observação	X	Y	Z
A	5,2	3,4	1,4
B	4,8	3,1	1,6
C	7,1	3,3	4,9
D	6,4	3,2	4,6
E	6,4	3,3	5,7
F	5,8	3,7	5,2

- a) Dois centroides são escolhidos aleatoriamente para os *clusters* das observações, neste caso, as observações 1 e 2 serão escolhidas como centroides (Tabela 2.2).

**Tabela 2.2** – Centroides para o exemplo.

	<b>X</b>	<b>Y</b>	<b>Z</b>
Centroide 1	5,2	3,4	1,4
Centroide 2	4,8	3,1	1,6

- b) São calculadas as distâncias das observações a cada um dos centroides, para o caso presente será utilizada a Distância Euclidiana Quadrática, que é definida como:

$$d(P_1, P_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \quad (84)$$

Calculando as distâncias com a expressão anterior e atribuindo as observações de acordo com a menor distância calculada, obtém-se as informações mostradas na Tabela 2.3.

**Tabela 2.3** – Distâncias das amostras até o centroide.

<b>observação</b>	<b>Distância até C1</b>	<b>Distância até C2</b>	<b>Cluster</b>
A	0	0,29	1
B	0,29	0	2
C	15,87	16,22	1
D	11,72	11,57	2
E	19,94	19,41	2
F	14,89	14,32	2

- c) Os centroides de cada *cluster* são recalculados, obtendo-se a média das observações que foram classificadas dentro de cada grupo, sendo os novos centroides mostrados na Tabela 2.4.

**Tabela 2.4** – Centroides recalculados.

	<b>X</b>	<b>Y</b>	<b>Z</b>
<b>Centroide 1</b>	6,15	3,35	3,15
<b>Centroide 2</b>	5,85	3,325	4,275

- d) Os passos a) e b) são repetidos, no que corresponderia à segunda iteração e observa-se se há alguma mudança na classificação das observações em cada *cluster*.

Tabela 2.5 – Resumo da segunda iteração.

Observação	X	Y	Z	Distância até C1	Distância a C2	Cluster
A	5,2	3,4	1,4	3,9675	8,69375	1
B	4,8	3,1	1,6	4,2875	8,30875	1
C	7,1	3,3	4,9	3,9675	1,95375	2
D	6,4	3,2	4,6	2,1875	0,42375	2
E	6,4	3,3	5,7	6,5675	2,33375	2
F	5,8	3,7	5,2	4,4475	0,99875	2
<b>Centroide 1</b>	6,15	3,35	3,15			
<b>Centroide 2</b>	5,85	3,36	4,28			

- e) Como se observa que há uma mudança de grupo para a observação B e C, novos centroides são calculados e uma terceira iteração é realizada.

Tabela 2.6 – Resumo da terceira iteração.

Observação	X	Y	Z	Distancia a C1	Distancia a C2	Cluster
A	5,2	3,4	1,4	0,0725	15,19125	1
B	4,8	3,1	1,6	0,0725	14,96625	1
C	7,1	3,3	4,9	15,9725	0,50125	2
D	6,4	3,2	4,6	11,5725	0,28125	2
E	6,4	3,3	5,7	19,6025	0,36625	2
F	5,8	3,7	5,2	14,5325	0,50625	2
<b>Centroide 1</b>	5	3,25	1,5			
<b>Centroide 2</b>	6,43	3,38	5,1			

Como não há alteração na classificação das observações, o algoritmo é encerrado.

#### 2.4.5.4 ESTRATIFICAÇÃO DOS USUÁRIOS APLICADA À EMPRESA DE DISTRIBUIÇÃO [60]

Com os dados obtidos, aplica-se o algoritmo de *k-means*, definindo-se primeiro *K* igual a 3, ou seja, para que sejam obtidos 3 *clusters* a depender do que for solicitado pela empresa de distribuição. A Figura 2.35 mostra o código usado para criar os *clusters*.

```

73     %% Gráfica K=3
74     [idx_clus3,Cntroid3]=kmeans(Data_norm,3);
75     Cent_real3=[];
76     for n=1:length(Cntroid3)
77         Cent_real3(:,n)=Cntroid3(:,n)*(max_data(n)-min_data(n))+min_data(n);
78     end
79     Cent_avg3=mean(Cent_real3,2);
80     Cent_std3=std(Cntroid3,0,2);
81     figure()
82     plot(cons_data(idx_clus3==1),std_data(idx_clus3==1),'r.','MarkerSize',12)
83     hold on
84     plot(cons_data(idx_clus3==2),std_data(idx_clus3==2),'b.','MarkerSize',12)
85     plot(cons_data(idx_clus3==3),std_data(idx_clus3==3),'g.','MarkerSize',12)
86     plot(Cent_avg3,Cent_std3,'kx','MarkerSize',15,'LineWidth',3)
87     legend('Cluster 1','Cluster 2','Cluster 3','Centroids')
88     title 'Cluster Assignments and Centroids'
89     hold off
90     clearvars n

```

Figura 2.35 – Código para criar *clusters* com  $K=3$  [60].

### Estratificação de usuários com dados normalizados

Na Figura 2.35 fica evidenciado o uso da função *K-means* no Matlab [57], que retorna como resultado um vetor com o número de usuários e índices, (1, 2 e 3), de acordo com o *cluster* ao qual usuário está alocado, e para cada *cluster* tem-se também seu respectivo centroide. Em seguida, identificam-se os centroides, que foram calculados como dados normalizados, retornando-os aos seus respectivos valores reais, de acordo com (54). Traça-se então gráfico que mostra os 3 *clusters* e seus respectivos centroides.

A Figura 2.36 mostra o código para o resumo do processo de estratificação para  $k = 3$ .

```

91     %% Resumen k=3
92     array_res=[];
93     for i=1:3
94         tot=length(cons_data(idx_clus3==i));
95         max_rang=round(max(cons_data(idx_clus3==i)),2);
96         min_rang=round(min(cons_data(idx_clus3==i)),2);
97         line=[i min_rang max_rang tot];
98         array_res=[array_res;line];
99     end
100     T_res3=array2table(array_res,'VariableNames',{'Cluster' 'Cons_min' 'Cons_max' 'Total_usuarios'});
101     clearvars i tot max_rang min_rang line array_res

```

Figura 2.36 – Resumo da estratificação com  $K=3$  [60].

Este resumo processa os resultados do algoritmo *K-means* para obter o número total de usuários em cada *cluster*, bem como seus valores máximos e mínimos, e armazena essas informações em uma matriz de resultados.

Como descrito no final da seção 2.1.3, é possível inserir no algoritmo de *K-means* os dados históricos, por exemplo, de velocidade de vento, temperatura e irradiação solar horizontal

numa zona de interesse para observar os padrões de comportamento destes parâmetros ao longo do ano nos últimos cinco ou dez anos para assim obter representantes (centroides) determinísticos a partir do comportamento estocástico histórico para os inserir no modelo de otimização para alocação e dimensionamento de fontes renováveis distribuídas.

No capítulo 3 é apresentada a proposta do trabalho simulada em Simulink do MATLAB [57] e no AMPL [7].

### 3 RESULTADOS

A metodologia proposta neste trabalho foi implementada utilizando o pacote computacional MATLAB [57] e o software AMPL [7]. O primeiro foi utilizado na implementação do modelo *K-means* do aprendizado de máquina, enquanto o segundo foi utilizado para a implementação do modelo de otimização. A máquina utilizada nas simulações foi um laptop core i5-4210U CPU@1.70GHz com memória RAM de 4GB.

Neste capítulo serão apresentados os diferentes resultados das metodologias considerando a variabilidade dos parâmetros das fontes fotovoltaicas como temperatura e irradiação solar. Após a obtenção e a inserção destes parâmetros (em primeira instância eles são estocásticos) na forma determinística, é feita a avaliação de desempenho do problema de otimização proposto. Dados históricos da cidade de Guayaquil, Equador, foram utilizados para aplicar as técnicas de aprendizagem de máquina e a obtenção dos dados representativos das épocas do ano.

Para a execução do projeto, os dados históricos da cidade de Guayaquil de dezembro de 2009 a julho de 2020 obtidos a partir do serviço *Solcast API Toolkit* [61], com a taxa de amostragem de 60 minutos, fornecidos por uma empresa de distribuição do setor elétrico.

#### 3.1 IMPLEMENTAÇÃO DO MÉTODO *K-MEANS*

O objetivo desta seção é determinar os dias representativos essenciais em modelos de longo prazo para sistemas de energia utilizando técnicas de aprendizado de máquina, em particular o algoritmo *K-means*, que permite a introdução de variáveis determinísticas em um modelo matemático equivalente não linear partindo de variáveis estocásticas como temperatura e irradiação solar para uma fonte de energia renovável.

##### 3.1.1 NORMALIZAÇÃO DOS DADOS PARA O PROCESSAMENTO DOS DADOS HISTÓRICOS

Foi utilizado o software MATLAB [57], e os dados foram importados de um arquivo Excel para o MATLAB [57], convertendo-os em uma matriz em que as colunas são cada uma das variáveis a serem analisadas.

### Irradiação horizontal global e temperatura do ar

É muito importante fazer um processamento de dados para após aplicar o processo de normalização de vetores. Na Figura 3.1 encontra-se o código que permite a leitura dos dados e a normalização deles.

```

1 - colmin=min(SolcastGYE2); %Se obtiene el valor mínimo por cada columna
2 - colmax=max(SolcastGYE2); %Se obtiene el valor máximo por cada columna
3 - Bcoll = rescale(SolcastGYE2, 'InputMin', colmin, 'InputMax', colmax); %Escalamiento de los valores
4

```

**Figura 3.1** – Código para o processamento de dados da aprendizagem de máquina. Fonte: Autor.

Primeiro, obtêm-se os valores mínimo e máximo de cada uma das colunas da matriz. A função "rescale" é usada para dimensionar as entradas de uma matriz dentro do intervalo [0,1] ao longo da dimensão da matriz de entrada que corresponde à forma dos valores dos parâmetros 'InputMin' e 'InputMax'.

A seguir é feito o procedimento para a obtenção desta normalização num grupo pequeno de dados:

- Na Figura 3.2 são apresentados os valores de um dia aleatório dos valores de temperatura do ar e da irradiação horizontal global.

A	B	C
HORA	AirTemp	Ghi
2	23,70	0,00
3	22,90	0,00
4	22,30	0,00
5	22,00	0,00
6	21,70	0,00
7	21,40	0,00
8	20,90	0,00
9	20,50	0,00
10	20,60	0,00
11	21,30	0,00
12	22,00	16,00
13	22,80	102,00
14	23,70	237,00
15	24,60	532,00
16	25,40	587,00
17	26,20	478,00
18	26,90	325,00
19	27,00	261,00
20	26,50	188,00
21	26,00	249,00
22	25,50	185,00
23	25,00	81,00

**Figura 3.2** – Dados de um dia para aplicação da normalização. Fonte: Autor.

- Aplicando o código apresentado na Figura 3.1 aos dados apresentados na Figura 3.2, na faixa horária das 6hs até as 18hs, são obtidos os resultados mostrados na Tabela 3.1. Considerando a equação (81).

**Tabela 3.1** – Exemplo do processamento de dados e sua normalização.

Dado Real	Dado Normalizado
21,70	0,19
21,40	0,14
20,90	0,06
20,50	0,00
20,60	0,02
21,30	0,13
22,00	0,23
22,80	0,36
23,70	0,50
24,60	0,64
25,40	0,77
26,20	0,89
26,90	1,00

### 3.1.2 PROGRAMAÇÃO DO ALGORITMO K-MEANS

Foi implementado um código para aplicar o algoritmo *K-means* apenas aos dados obtidos das 6h às 18h, horários em que há maior irradiação solar. A Figura 3.3 apresenta a base de dados que foi usada e que tem 92614 dados.

PeriodEnd	PeriodStart	Period	AirTemp	AlbedoDai	Azimuth	CloudOpaci	DewpointT	Dhi	Dni	Ebh	Ghi	GtFixedTilt	GtTracking	PrecipitableWate	RelativeHumidit	SnowDepth	Surface
2009-12-31T02:00:00Z	2009-12-31T03:00:00Z	PT60M	23.7	0.11	118.0	20.1	0	0	0	0	0	0	0	0.48.0	80.1	0.0	973.3
2009-12-31T03:00:00Z	2009-12-31T04:00:00Z	PT60M	22.9	0.11	124.0	20.1	0	0	0	0	0	0	0	0.48.3	84.2	0.0	974.3
2009-12-31T04:00:00Z	2009-12-31T05:00:00Z	PT60M	22.3	0.11	135.0	20.1	0	0	0	0	0	0	0	0.48.7	87.2	0.0	974.7
2009-12-31T05:00:00Z	2009-12-31T06:00:00Z	PT60M	22.0	0.11	155.0	20.1	0	0	0	0	0	0	0	0.49.3	89.0	0.0	974.6
2009-12-31T06:00:00Z	2009-12-31T07:00:00Z	PT60M	21.7	0.11	-176.34.2	20.2	0	0	0	0	0	0	0	0.49.9	90.8	0.0	974.5
2009-12-31T07:00:00Z	2009-12-31T08:00:00Z	PT60M	21.4	0.11	-148.21.4	20.1	0	0	0	0	0	0	0	0.50.4	92.3	0.0	974.1
2009-12-31T08:00:00Z	2009-12-31T09:00:00Z	PT60M	20.9	0.11	-131.8.5	19.8	0	0	0	0	0	0	0	0.50.7	93.5	0.0	973.6
2009-12-31T09:00:00Z	2009-12-31T10:00:00Z	PT60M	20.5	0.11	-122.0.0	19.6	0	0	0	0	0	0	0	0.51.1	94.6	0.0	973.1
2009-12-31T10:00:00Z	2009-12-31T11:00:00Z	PT60M	20.6	0.11	-117.1.0	19.7	0	0	0	0	0	0	0	0.50.5	94.6	0.0	973.2
2009-12-31T11:00:00Z	2009-12-31T12:00:00Z	PT60M	21.3	0.11	-114.1.8	20.2	0	0	0	0	0	0	0	0.49.2	93.5	0.0	973.9
2009-12-31T12:00:00Z	2009-12-31T13:00:00Z	PT60M	22.0	0.11	-113.16.9	20.7	16	0	0	16	16	16	16	47.9	92.3	0.0	974.5
2009-12-31T13:00:00Z	2009-12-31T14:00:00Z	PT60M	22.8	0.11	-114.37.1	20.8	102	0	0	102	102	102	102	47.1	88.8	0.0	974.9
2009-12-31T14:00:00Z	2009-12-31T15:00:00Z	PT60M	23.7	0.11	-116.35.5	20.7	236	2	1	237	237	237	238	46.8	83.3	0.0	975.1
2009-12-31T15:00:00Z	2009-12-31T16:00:00Z	PT60M	24.6	0.11	-121.9.0	20.6	273	364	259	532	504	604	46.6	78.1	0.0	975.4	
2009-12-31T16:00:00Z	2009-12-31T17:00:00Z	PT60M	25.4	0.11	-131.23.5	20.6	339	299	248	587	560	599	46.3	74.5	0.0	975.1	
2009-12-31T17:00:00Z	2009-12-31T18:00:00Z	PT60M	26.2	0.11	-151.44.3	20.7	379	114	99	478	468	478	46.2	72.0	0.0	974.2	
2009-12-31T18:00:00Z	2009-12-31T19:00:00Z	PT60M	26.9	0.11	175.63.6	20.9	325	0	0	325	325	325	46.0	69.7	0.0	973.4	
2009-12-31T19:00:00Z	2009-12-31T20:00:00Z	PT60M	27.0	0.11	144.69.8	21.1	261	0	0	261	261	261	46.1	70.1	0.0	972.7	
2009-12-31T20:00:00Z	2009-12-31T21:00:00Z	PT60M	26.5	0.11	127.74.6	21.3	188	0	0	188	188	188	46.5	73.0	0.0	971.9	
2009-12-31T21:00:00Z	2009-12-31T22:00:00Z	PT60M	26.0	0.11	119.55.5	21.5	249	0	0	249	249	249	46.9	76.2	0.0	971.3	
2009-12-31T22:00:00Z	2009-12-31T23:00:00Z	PT60M	25.5	0.11	115.48.6	21.6	185	0	0	185	185	185	47.0	79.2	0.0	971.2	
2010-01-01T00:00:00Z	2009-12-31T23:00:00Z	PT60M	25.0	0.11	113.45.1	21.7	81	0	0	81	81	81	47.0	82.1	0.0	971.9	
2010-01-01T00:00:00Z	2009-12-31T23:00:00Z	PT60M	24.4	0.11	113.9.5	21.8	0	0	0	6	6	6	47.0	85.1	0.0	972.5	
2010-01-01T01:00:00Z	2010-01-01T02:00:00Z	PT60M	23.7	0.11	115.35.0	21.5	0	0	0	0	0	0	47.0	87.2	0.0	973.2	
2010-01-01T02:00:00Z	2010-01-01T03:00:00Z	PT60M	22.8	0.11	119.10.1	20.8	0	0	0	0	0	0	47.0	88.6	0.0	974.0	
2010-01-01T03:00:00Z	2010-01-01T04:00:00Z	PT60M	21.9	0.11	124.0.0	20.2	0	0	0	0	0	0	47.1	90.0	0.0	974.8	
2010-01-01T04:00:00Z	2010-01-01T05:00:00Z	PT60M	21.3	0.11	135.0.0	19.8	0	0	0	0	0	0	47.2	91.3	0.0	975.0	

**Figura 3.3** – Base de dados para obter os valores de irradiação e temperatura. Fonte: Autor.



Na coluna quatro e na coluna doze da base de dados apresentada na Figura 3.3, encontram-se os valores de temperatura do ar ( $^{\circ}\text{C}$ ) e da irradiação horizontal global ( $\text{W}/\text{m}^2$ ), respectivamente.

A Figura 3.4 mostra o código, onde é lido o arquivo Excel com os respectivos dados de horas, irradiação solar e temperatura do ar.

```
filename = 'SolcastGYE2.xlsx';
Datos = xlsread(filename);

[row,col] = find(Datos(:,1)>=6 & Datos(:,1)<=18); %Registra los índices del lugar donde se cumple la condición
COLVER = Datos(row,1); %Devuelve los valores de los índices que cumplan la condición
COLVER2 = Datos(row,2);
COLVER3 = Datos(row,3);
SolcastNew=[COLVER2 COLVER3]; %Irradiación y temperatura desde las 6am hasta 18pm
```

**Figura 3.4** – Leitura e processamento dos dados de irradiação e temperatura. Fonte: Autor.

Em seguida, a partir da linha três do código apresentado na Figura 3.4, os dados fora do horário da 6h da manhã às 18h foram descartados por apresentarem menor radiação solar, obtendo-se assim a matriz SolcastNew que contém os dados de irradiação solar e temperatura do ar nos respectivos horários.

Reforça-se que os dados de entrada da planilha “SolcastGYE2.xlsx”, com a variante de obter somente os dados das 6hs até às 18hs ainda não foram normalizados, e necessário fazer um pré-processamento conforme descrito na Seção 3.1.1. A Figura 3.5 mostra o código para a normalização dos dados da matriz SolcastNew, por meio da função “rescale” que requer os valores mínimo e máximo de cada coluna.

```
colmin=min(SolcastNew); %Se obtiene el valor mínimo por cada columna
colmax=max(SolcastNew); %Se obtiene el valor máximo por cada columna
Bcoll = rescale(SolcastNew,'InputMin',colmin,'InputMax',colmax); %Escalamiento de los valores
```

**Figura 3.5** – Normalização dos dados no horário especificado. Fonte: Autor.

Conforme apresentado na Seção **¡Error! No se encuentra el origen de la referencia.**, é necessário obter o gráfico do cotovelo. A Figura 3.6 mostra o código para construção do gráfico do cotovelo que permite determinar o número de *clusters* a serem utilizados no algoritmo *K-means*.

```

%Determina la cantidad de clústeres
sa=[];
K=[];
for k=1:8
    [idx,c,sumd]=kmeans(Bcoll,k);
    sa=[sa sum(sumd)];
    K=[K k];
end
plot(K,sa) %Codo de Jambú
title('Codo de Jambú')
xlabel('Número de Clústeres')
ylabel('Inercia Intra-clases WCSS')
grid on

```

**Figura 3.6** – Código para obtenção do cotovelo. Fonte: Autor.

No estágio seguinte do processo, para obter os dias representativos, foi desenvolvido um código em MATLAB [57] relacionado com algoritmo de *K-means*. A Figura 3.7 apresenta o código para o algoritmo *K-means* e, assim, obter dias representativos dos dados, identificando seus respectivos centroides.

```

%Algoritmo de Kmeans
[idx,C]=kmeans(Bcoll,2);
figure
gscatter(Bcoll(:,1),Bcoll(:,2),idx,'bgm')
hold on
plot(C(:,1),C(:,2),'kx') %Grafica el centroide
legend('Cluster 1','Cluster 2','Cluster 3','Cluster Centroid')
title('Aplicando Kmeans')
xlabel('Temperatura del Aire')
ylabel('Irradiación Horizontal Global')

Asg=[idx COLVER SolcastNew];

```

**Figura 3.7** – Algoritmo para obter dias representativos (*K-means*). Fonte: Autor.

Continuando com o processo, foi necessário obter duas matrizes com os parâmetros mais significativos na produção da geração fotovoltaica. A Figura 3.8 mostra o código feito para obter essas duas matrizes, cada uma contendo a hora, sua irradiação solar e a temperatura do ar.

```

%IDENTIFICAR HORAS CLUSTER 2
[row2,col2] = find(Asg(:,1)==2); %Registra los índices del lugar donde se cumple la condición
HorasCluster2 = Asg(row2,1); %Devuelve los valores de los índices que cumplan la condición
HorasCluster2_1 = Asg(row2,2);
HorasCluster2_2 = Asg(row2,3);
SolcastC2=[HorasCluster2 HorasCluster2_1 HorasCluster2_2]; %Horas del cluster 2
%IDENTIFICAR HORAS CLUSTER 1
[row1,col1] = find(Asg(:,1)==1); %Registra los índices del lugar donde se cumple la condición
HorasCluster1 = Asg(row1,1); %Devuelve los valores de los índices que cumplan la condición
HorasCluster1_1 = Asg(row1,2);
HorasCluster1_2 = Asg(row1,3);
SolcastC1=[HorasCluster1 HorasCluster1_1 HorasCluster1_2]; %Horas del cluster 1

```

**Figura 3.8** – Identificação das horas de cada *cluster*. Fonte: Autor.

Finalmente na Figura 3.9 é apresentado o código para obter os dados de irradiação de cada *cluster*.

```

%C1
c_NewT1=C(1,1)*(colmax(:,1)-colmin(:,1))+colmin(:,1);
c_NewI1=C(1,2)*(colmax(:,2)-colmin(:,2))+colmin(:,2);
%C2
c_NewT2=C(2,1)*(colmax(:,1)-colmin(:,1))+colmin(:,1);
c_NewI2=C(2,2)*(colmax(:,2)-colmin(:,2))+colmin(:,2);
Centroides=[c_NewT1,c_NewI1;c_NewT2,c_NewI2]

```

**Figura 3.9** – Obtenção dos dados de irradiação de cada *cluster*. Fonte: Autor.

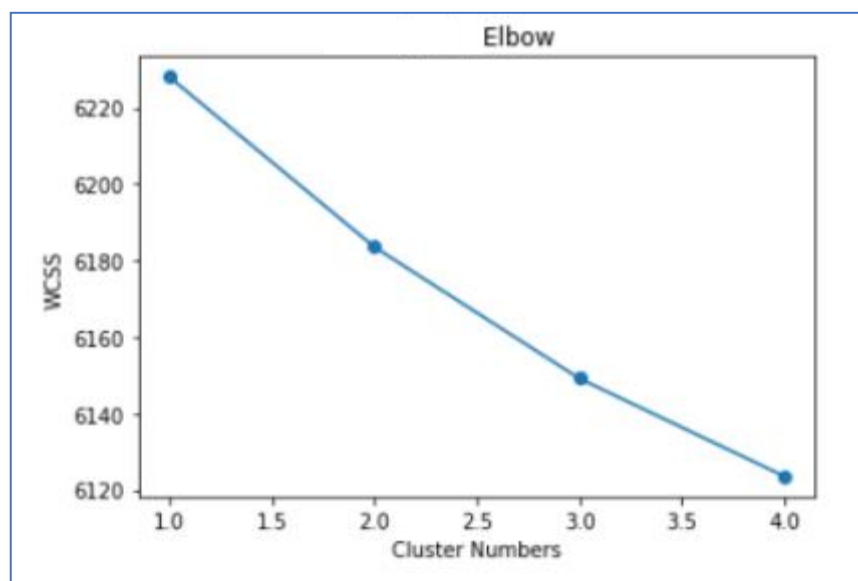
Assim, podem ser facilmente identificados as horas e os valores que compõem cada *Cluster*. Mais adiante se verificará que para o caso de estudo o *Cluster 1* segmenta as horas das 6h às 13h e o *Cluster 2* segmenta as horas das 14h às 18h.

### 3.1.3 OBTENÇÃO DE DIAS REPRESENTATIVOS

Nesta seção, será mostrado o procedimento adotado para obter os dias representativos usando o algoritmo *K-means*. O algoritmo *K-means* é um processo de aprendizagem não supervisionada usada quando se tem registros das variáveis, mas não se tem um conhecimento completo do comportamento ou da natureza delas. Este método não supervisionado realiza a segmentação dessas variáveis com base na semelhança de suas características [29].

Conforme mencionado anteriormente, antes de aplicar o algoritmo *K-means*, dados de entrada devem ser pré-processados de maneira que as variáveis sejam normalizadas, proporcionando melhores condições para a obtenção de resultados consistentes [62].

O algoritmo *K-means* parte de uma seleção inicial de centroides  $K$ , que podem ser obtidos aplicando técnicas de validação de *cluster* que nos permitem encontrar o número ótimo de partições a formar [63]. Neste trabalho foi utilizada a técnica de cotovelo que na Figura 3.10 tomada de [64] apresenta o comportamento de mudar o número de *clusters* desde 1 até 4, obtendo para cada *cluster* a soma dos quadrados dentro do *Cluster* (WCSS pelas siglas no inglês “*Within-Cluster Sum of Square*”).



**Figura 3.10** – Técnica do cotovelo. [64]

Em outras palavras WCSS é a soma da distância quadrada entre cada ponto e o centroide em um *cluster*. Quando é representado graficamente o WCSS frente ao valor  $K$ , o gráfico se parece com um cotovelo. Conforme o número de *clusters* aumenta, o valor WCSS começará a diminuir. O valor de WCSS é maior quando menor for  $K$ . Quando o gráfico é analisado, verifica-se que a curva mudará rapidamente seu comportamento em um ponto, criando assim uma forma de cotovelo. A partir deste ponto, o gráfico começa a se mover quase paralelo ao eixo X. O valor  $K$  correspondente a este ponto é o valor  $K$  ideal ou um número ótimo de *clusters*. O método do cotovelo usa os valores de inércia quando usa o algoritmo *K-means*, de modo tal que no gráfico mostrado este ponto ótimo é encontrado quando o valor de inércia intraclasse (WCSS) deixa de diminuir consideravelmente e o número de *clusters* começa a diminuir.

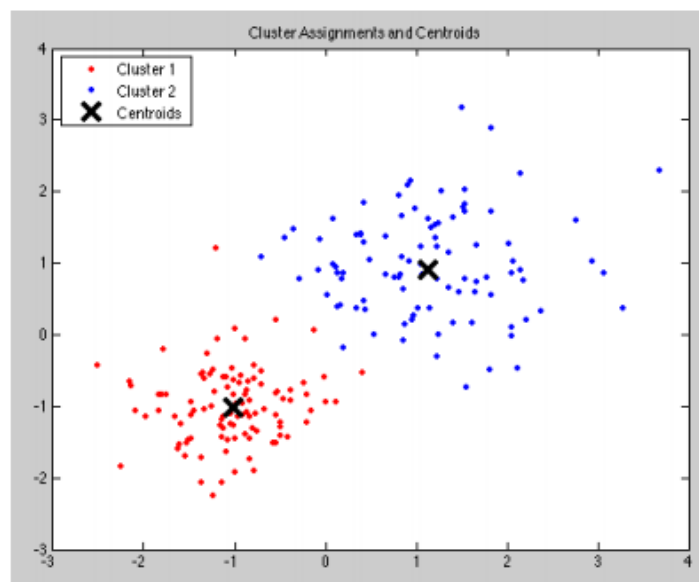
Posteriormente, para cada objeto no banco de dados, sua distância a cada centroide é calculada, o centroide mais próximo é determinado e o objeto é incorporado ao grupo relacionado àquele centroide. Após a obtenção dos respectivos grupos gerados na etapa anterior, seu centroide é recalculado. Por fim, as duas últimas etapas mencionadas são repetidas até que o

critério de convergência seja atendido, que se refere ao fato de o algoritmo convergir quando os indivíduos não estão mais sendo realocados para outro *cluster*. O processo converge quando o valor de WCSS é menor que uma tolerância que é atribuída ou que certo número de iterações é realizado [65].

Os centroides obtidos a partir do algoritmo *K-means* são aplicados aos sistemas de energia elétrica no nível de distribuição especificamente para alocar os geradores distribuídos fotovoltaicos. Neste caso, na implementação destas fontes solares, tanto os valores da irradiação global horizontal quanto da temperatura do ar são levados em consideração, visto que uma das variáveis presentes na análise é o nível de nebulosidade. O nível de nebulosidade tem um efeito significativo na previsão da temperatura porque as nuvens atuam como um regulador da quantidade de radiação solar que atinge a superfície [40].

No céu claro, a intensidade da radiação solar depende principalmente da radiação solar direta, podendo-se aproveitar ao máximo a produção de energia no painel solar, enquanto que quando o céu está nublado a produção de energia da fonte solar é menor, devido ao espalhamento da radiação [40].

Na Figura 3.11, é mostrado um agrupamento de amostras aleatórias usando dois *clusters*, que corresponde a um exemplo simples do processo de agrupamento usando o algoritmo *K-means* [66].



**Figura 3.11** – Agrupamento de amostras aleatórias de acordo com dois *clusters*. [65]

O algoritmo *K-means* encontra uma categorização que representa um ótimo local do critério escolhido, de forma que no gráfico mostrado o resultado é de dois *clusters* com seus respectivos centroides.

### 3.2 MODELOS MATEMATICOS USADOS PARA INSERIR OS RESULTADOS DA APRENDIZAGEM DE MÁQUINA

No capítulo anterior, os modelos não linear e linear foram apresentados de forma detalhada. Nesta seção, o objetivo é avaliar as técnicas da aprendizagem de máquina em termos dos desempenhos dos resultados considerando a mudança, devido ao algoritmo de *K-means*, dos parâmetros estocásticos em parâmetros determinísticos.

Os modelos usados para resolver o problema de alocação de fontes renováveis fotovoltaicas e mitigação de perdas não técnicas foram implementados na linguagem de modelagem matemática (AMPL) [7] e cada modelo foi resolvido pelo solver comercial de otimização CPLEX [39].

### 3.3 SIMULAÇÕES E RESULTADOS

Na primeira parte desta seção é apresentada a programação usando o algoritmo *K-means* para determinar os dias representativos e, na segunda parte os resultados das simulações obtidas de cada modelo matemático usando as redes testes.

#### 3.3.1 RESULTADOS USANDO O ALGORITMO *K-MEANS*

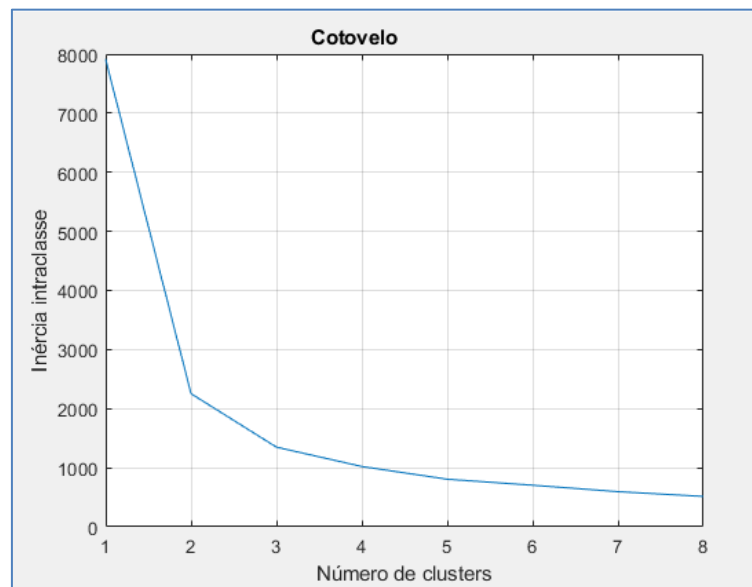
As informações fornecidas por uma empresa de distribuição do setor elétrico formam uma base de dados de irradiação e temperatura do ar, contando com um total de 92614 dados para cada parâmetro. Utilizou-se a técnica do cotovelo para determinar o valor *K* que representa a quantidade de grupos (ou *clusters*) a utilizar.

Da Figura 3.3 é possível somente obter das colunas quatro e doze os 92614 dados que apresentam os dados dos valores de temperatura do ar e da irradiação horizontal global, respectivamente, sendo estes mostrados na Figura 3.12.

PeriodEnd	PeriodStart	Period	AirTemp	Ghi
2009-12-31T02:00:00Z	2009-12-31T01:00:00Z	PT60M	23.7	0
2009-12-31T03:00:00Z	2009-12-31T02:00:00Z	PT60M	22.9	0
2009-12-31T04:00:00Z	2009-12-31T03:00:00Z	PT60M	22.3	0
2009-12-31T05:00:00Z	2009-12-31T04:00:00Z	PT60M	22.0	0
2009-12-31T06:00:00Z	2009-12-31T05:00:00Z	PT60M	21.7	0
2009-12-31T07:00:00Z	2009-12-31T06:00:00Z	PT60M	21.4	0
2009-12-31T08:00:00Z	2009-12-31T07:00:00Z	PT60M	20.9	0
2009-12-31T09:00:00Z	2009-12-31T08:00:00Z	PT60M	20.5	0
2009-12-31T10:00:00Z	2009-12-31T09:00:00Z	PT60M	20.6	0
2009-12-31T11:00:00Z	2009-12-31T10:00:00Z	PT60M	21.3	0
2009-12-31T12:00:00Z	2009-12-31T11:00:00Z	PT60M	22.0	16
2009-12-31T13:00:00Z	2009-12-31T12:00:00Z	PT60M	22.8	102
2009-12-31T14:00:00Z	2009-12-31T13:00:00Z	PT60M	23.7	237
2009-12-31T15:00:00Z	2009-12-31T14:00:00Z	PT60M	24.6	532
2009-12-31T16:00:00Z	2009-12-31T15:00:00Z	PT60M	25.4	587
2009-12-31T17:00:00Z	2009-12-31T16:00:00Z	PT60M	26.2	478
2009-12-31T18:00:00Z	2009-12-31T17:00:00Z	PT60M	26.9	325
2009-12-31T19:00:00Z	2009-12-31T18:00:00Z	PT60M	27.0	261
2009-12-31T20:00:00Z	2009-12-31T19:00:00Z	PT60M	26.5	188
2009-12-31T21:00:00Z	2009-12-31T20:00:00Z	PT60M	26.0	249
2009-12-31T22:00:00Z	2009-12-31T21:00:00Z	PT60M	25.5	185
2009-12-31T23:00:00Z	2009-12-31T22:00:00Z	PT60M	25.0	81
2010-01-01T00:00:00Z	2009-12-31T23:00:00Z	PT60M	24.4	6
2010-01-01T01:00:00Z	2010-01-01T00:00:00Z	PT60M	23.7	0
2010-01-01T02:00:00Z	2010-01-01T01:00:00Z	PT60M	22.8	0
2010-01-01T03:00:00Z	2010-01-01T02:00:00Z	PT60M	21.9	0
2010-01-01T04:00:00Z	2010-01-01T03:00:00Z	PT60M	21.3	0

**Figura 3.12** – Dados brutos fornecidos pela empresa de distribuição do setor elétrico. Fonte: Autor.

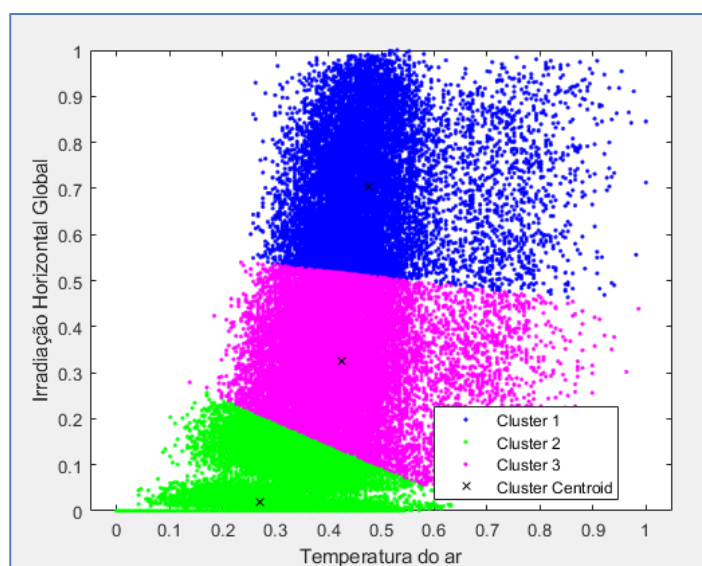
Como pode ser visto na Figura 3.13, com essa técnica avalia-se quão semelhantes os indivíduos são entre eles, e gradativamente capta-se essa informação na forma de um gráfico. Observa-se que o eixo y é a inércia intraclasse (WCSS) e o eixo x é o número de *clusters*. À medida que a inércia intraclasse diminui, o número de *clusters* aumenta. Ressalta-se que a inércia intraclasse corresponde à soma dos quadrados das distâncias entre os indivíduos que compõem cada *cluster*.



**Figura 3.13** – Cotovelo para todos os dados históricos. Fonte: Autor.

Da Figura 3.13 é possível identificar o ponto em que a inércia intraclasse (WCSS) para de diminuir drasticamente, sendo este valor  $K = 3$ , portanto, três *clusters* são obtidos.

A seguir, na Figura 3.14 se apresentam os três grupos de dados agrupados em que cada um possui seu próprio centroide que indica a localização real ou imaginária que representa o centro do grupo, portanto, são os dados representativos de cada *cluster*.



**Figura 3.14** – Diagrama de dispersão para todos os dados históricos. Fonte: Autor.

Na Tabela 3.2 observa-se que, a partir dos dados históricos completos (92614 dados), o *cluster 1* agrupa os dados entre os horários das 15h às 19h, correspondendo a 62,80% dos dados totais (equivalente ao 58162 dados). O *cluster 2* agrupa os dados da 1h às 13h, com 21,6% dos dados totais (equivalente ao 19995 dados), e o *cluster 3* com dados 13h-15h e 20h-23h, correspondendo a 15,6% dos dados totais (equivalente ao 14457 dados).

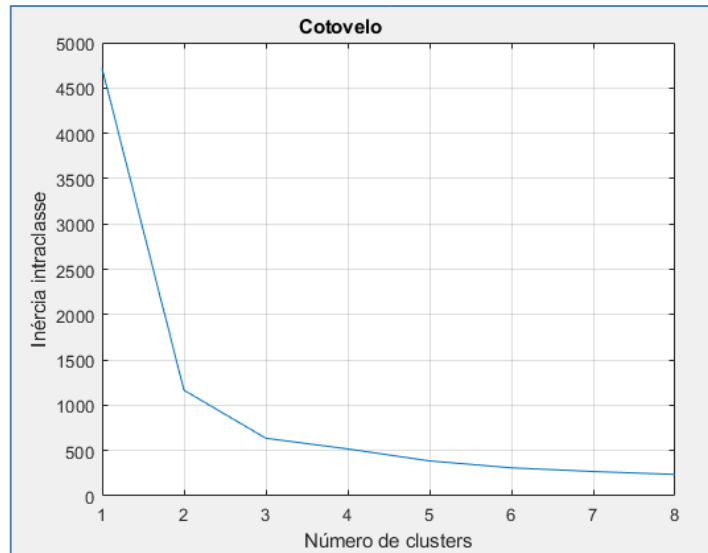
**Tabela 3.2** – Resultados considerando todos os dados históricos.

	HORÁRIO	TEMPERATURA DO AR (°C)	IRRADIAÇÃO (W/m <sup>2</sup> )
<i>Cluster 1</i>	15h-19h	21,8998	17,0319
<i>Cluster 2</i>	1h-13h	25,2195	312,0935
<i>Cluster 3</i>	13h-15h e 20h -23h	26,3191	678,5597

Analisando os resultados obtidos anteriormente, verificou-se que os horários, onde há produção solar e horários em que não há produção solar, se misturam, portanto, a próxima etapa consiste em filtrar os dados históricos, eliminando os horários em que não há luz solar, limitando os dados das 6h às 18h em que há maior irradiação solar. Essa filtragem resulta em uma quantidade de 50167 dados.

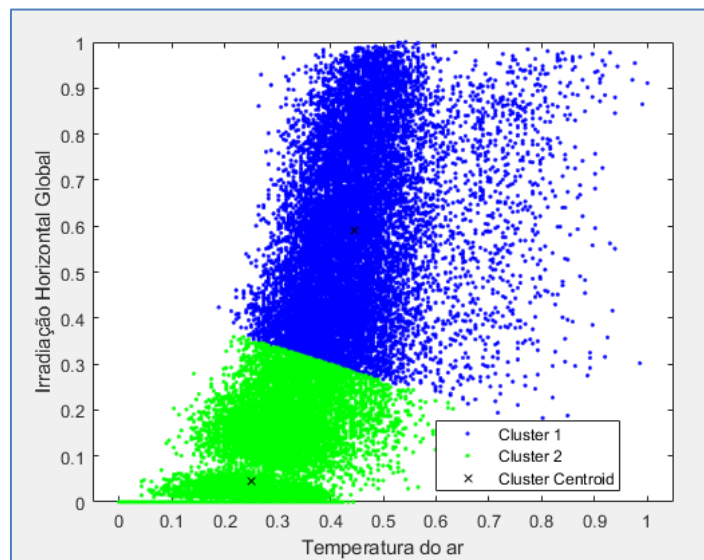


Ao realizar a depuração adequada da base de dados histórica tomando apenas as horas produtivas do painel, obteve-se a Figura 3.15, onde se observa o cotovelo, da qual se conclui que é razoável tomar dois *clusters*.



**Figura 3.15** – Cotovelo para dados históricos limitados das 6h às 18h. Fonte: Autor.

A informação obtida com o auxílio da Figura 3.15 é introduzida no algoritmo *K-means*, que resulta no diagrama de dispersão da Figura 3.16.



**Figura 3.16** – Diagrama de dispersão para dados históricos limitados das 6h às 18h. Fonte: Autor.

A Tabela 3.3 apresenta os dados representativos dessas horas que servem como parâmetro de entrada para o modelo matemático. É destacável que nesta simulação foram considerados somente os horários nos quais o sol tem participação ativa na geração fotovoltaica.

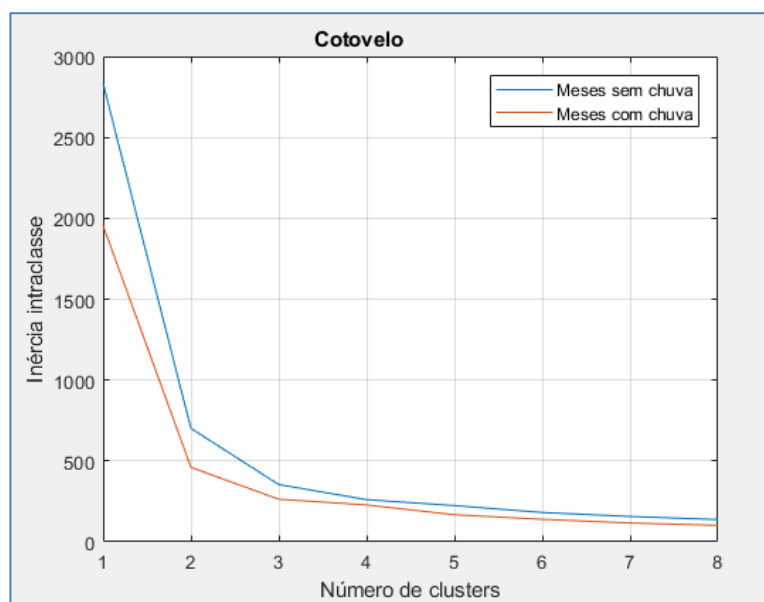
**Tabela 3.3** – Resultados para os dados históricos limitados às 6h-18h.

	HORÁRIO	TEMPERATURA DO AR (°C)	IRRADIAÇÃO (W/m <sup>2</sup> )
<i>Cluster 1</i>	6h-14h	21,3280	43,7350
<i>Cluster 2</i>	14h-18h	25,4684	570,4184

Analisando os resultados da Tabela 3.3, observa-se que, a partir de 100% dos dados históricos, obteve-se que no *cluster 1* os dados se agruparam entre os horários de 6h-14h (30,1% dos dados), e no *cluster 2* com dados entre 14h-18h (69,9% dos dados).

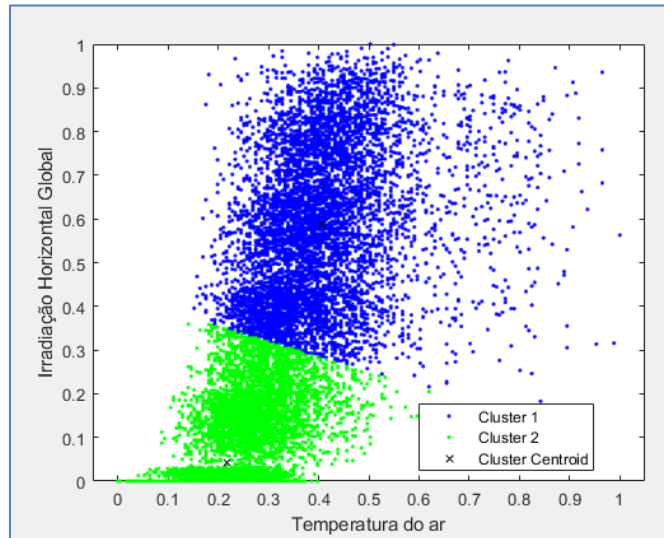
Posteriormente, considerou-se também classificar os dados do banco de dados levando em consideração os meses sem chuva e com chuva, que influenciam tanto a irradiação solar quanto a temperatura do ar. Definiu-se que os meses de janeiro a maio são meses com chuva e os meses de junho a dezembro são sem chuva. Entende-se que haverá maior utilização de irradiação solar nos painéis nos meses em que não chover.

Observa-se na Figura 3.17 que, ao se aplicar a técnica do cotovelo, tanto para a presença quanto para a ausência de chuva, existe um valor  $K$  igual a dois, sendo este um valor considerável para a aplicação do algoritmo *K-means*.

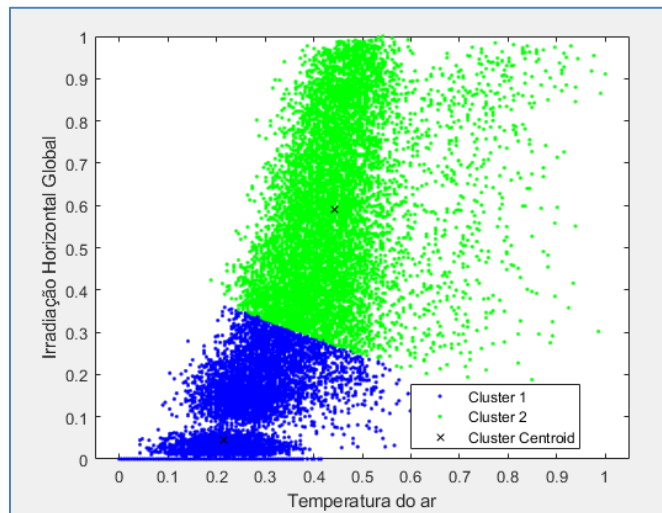
**Figura 3.17** – Cotovelo para dados históricos com a presença de chuva. Fonte: Autor.

Na Figura 3.18 e na Figura 3.19 são mostrados os respectivos diagramas de dispersão, indicando que nos meses com chuva há menos dados do que nos meses em que há presença de chuva. Portanto, nos meses sem chuva houve uma alta correlação positiva entre os parâmetros

de irradiação e temperatura do ar, enquanto nos meses com chuva há baixa correlação positiva entre os mesmos parâmetros analisados.



**Figura 3.18** – Diagrama de dispersão para os meses com chuva. Fonte: Autor.



**Figura 3.19** – Diagrama de dispersão para os meses sem chuva. Fonte: Autor.

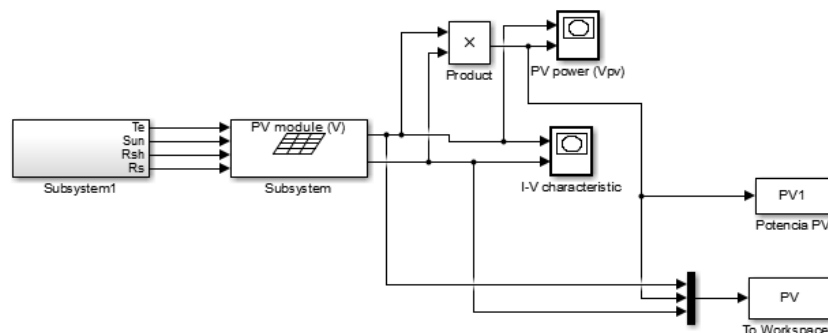
A Tabela 3.4 mostra os dois respectivos *clusters* para a segmentação dos meses em que os horários de cada *cluster* são semelhantes, mas possuem valores de irradiação diferentes para cada um deles. É destacável que estes parâmetros que consideram as variações estocásticas dos dados históricos, ao passarem pelo algoritmo de aprendizagem de máquinas, mudaram para serem parâmetros determinísticos que vão ingressar no modelo de otimização.

**Tabela 3.4** – Resultados para dados históricos de acordo com a presença de chuva.

	HORÁRIO	IRRADIAÇÃO (W/m <sup>2</sup> )	TEMPERATURA DO AR (°C)
Meses com chuva (janeiro-maio)			
<i>Cluster 1</i>	6h-13h	41,051	22,3213
<i>Cluster 2</i>	13h-18h	565,75	25,5465
Meses sem chuva (junho-dezembro)			
<i>Cluster 1</i>	6h-15h	44,2602	20,5303
<i>Cluster 2</i>	15h-18h	569,023	25,4089

### 3.3.2 OBTENÇÃO DA GERAÇÃO DAS FONTES FOTOVOLTAICAS USANDO SIMULINK DO MATLAB

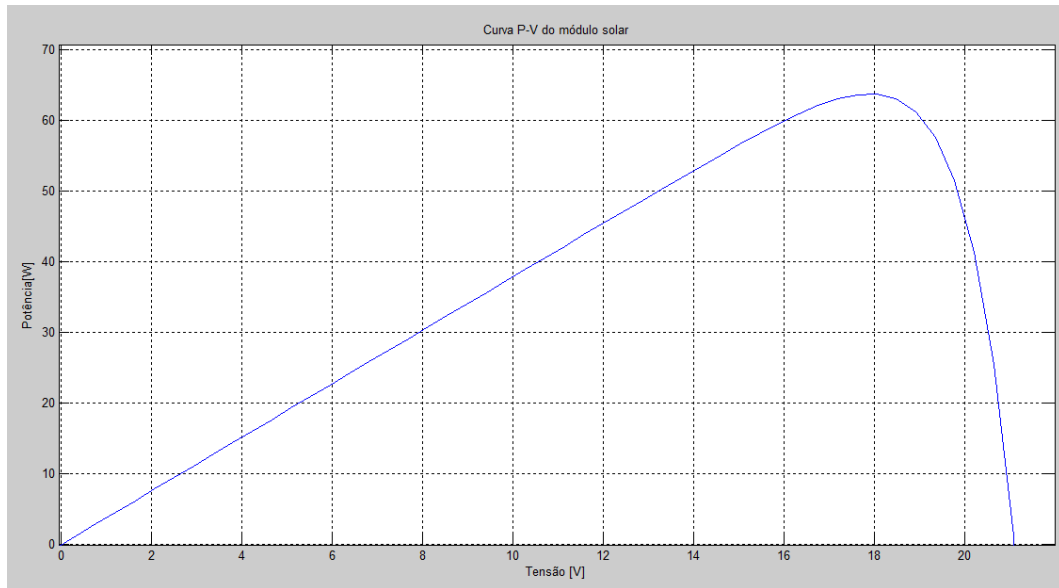
Após a obtenção dos resultados da análise utilizando o aprendizado de máquina, a célula solar a ser utilizada é modelada utilizando o programa *Simulink* do *Matlab* [57], como mostrado na Figura 3.20.



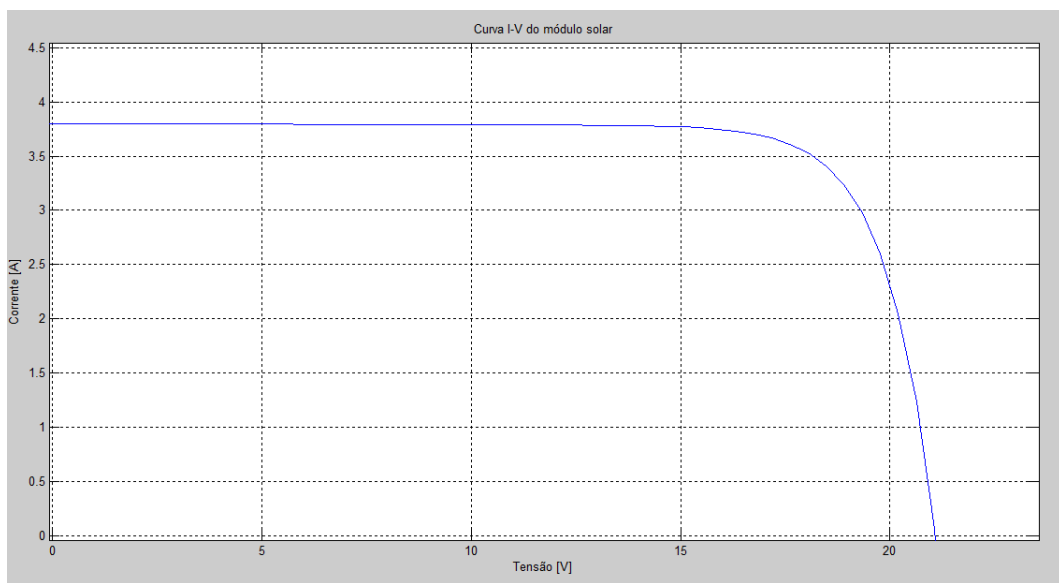
**Figura 3.20** – Modelo célula solar (irradiação solar e temperatura na entrada). Fonte: Autor.

A partir do modelo da Figura 3.20, foram obtidas as curvas características P-V e I-V mostradas na Figura 3.21 e na Figura 3.22, respectivamente.

No modelo de Simulink são inseridos no bloco chamado de *Subsystem 1* os parâmetros obtidos da aprendizagem de máquinas nos dados  $Te$ , que representa a temperatura representante obtida nos *clusters*, e de forma similar o valor de irradiação solar através do parâmetro  $Sun$ . Do modelo típico da produção de potência é obtido o valor de  $f_{t,d}^{Gpv}$  do modelo de otimização.



**Figura 3.21** – Curva P-V do módulo solar. Fonte: Autor.



**Figura 3.22** – Curva I-V do módulo solar. Fonte: Autor.

Da Figura 3.21 e da Figura 3.22 obtém-se que a potência máxima da célula solar é de 65 W, obtida em condições ideais e para este exemplo específico.

### 3.3.3 RESULTADOS PARA O MODELO MATEMÁTICO DE OTIMIZAÇÃO

Para a avaliação da metodologia foram usadas duas redes: uma rede teste de 22 barras e uma rede real do Equador de 137 barras, cujos dados foram fornecidos pela empresa distribuidora do setor elétrico.

O modelo matemático, equações (65)-(80), apresentado no capítulo anterior foi implementado no pacote computacional AMPL.

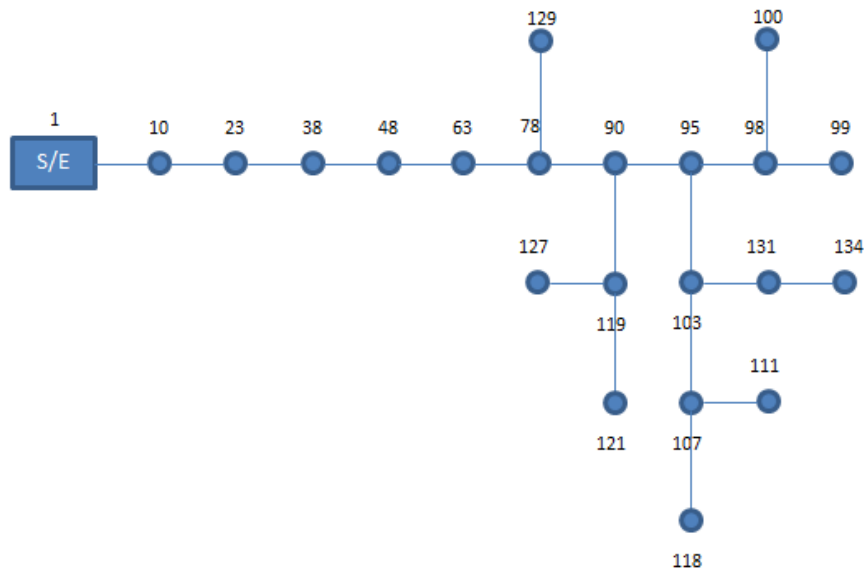
Neste trabalho, são analisados apenas dois termos da função objetivo (65) relacionados ao investimento em novas fontes renováveis e perdas técnicas, focando no impacto das próprias fontes renováveis na operação do sistema de distribuição, desprezando o custo pela entrega de energia no início do alimentador, uma vez que esta proposta busca minimizar o investimento em novos ativos (painéis solares).

Estes novos ativos consideram que o valor de potência máxima da célula solar, em condições ideais, é de 65 W. Foi selecionada uma célula solar existente no mercado com 21,5% de eficiência, e o conjunto de 18 células série forma um módulo de potência nominal de 250W. Portanto, um painel contendo dois módulos terá uma potência ideal de 500W. Todas as simulações consideraram que é possível instalar (considerando o espaço físico) plantas solares em cada barra da rede indicada pelo modelo matemático. Cada planta solar pode ter de um a cinco arranjos. Cada arranjo solar é composto por 40 painéis e o valor máximo ideal é de 20kW. Portanto, cada planta pode ter uma potência nominal máxima de saída de 100kW.

A seguir apresentam-se os resultados para a rede de pequeno porte, cujo diagrama unifilar é apresentado na Figura 3.23, para a qual foram feitas as seguintes simulações para avaliar o seu desempenho:

A informação de barras e ramos para o sistema de 22 barras é apresentada no ANEXO A na Tabela A.1 e na Tabela A.2, respectivamente. A carga total desta rede é de 6,5 MW e 2,7 MVar. A ideia neste sistema pequeno é minimizar as perdas técnicas e, simultaneamente, diminuir o custo de investimento nas fontes renováveis. O custo pela compra do painel é \$80, o investimento máximo em fontes renováveis é \$1000, o fator de potência associado ao painel solar é 0,97 e a tensão nominal é 13,8 kV, permitindo um valor máximo ou mínimo de tensão em cada barra de 14,49 e 12,42 kV respectivamente.

- Dados originais, conforme apresentados em [13].
- Com possibilidade de inserir fontes renováveis.



**Figura 3.23** – Sistema de 22 barras. Fonte: Autor.

Os valores para o parâmetro determinístico  $f_{t,d}^{Gpv}$  nos meses com chuva e sem chuva são 0,65 e 0,80 respectivamente. Estes valores foram determinados considerando a metodologia *K-means* e seus resultados incorporados ao Simulink conforme descrito na Seção 2.3.

Na Tabela 3.6 é apresentado o primeiro teste, para o qual não se permite a alocação de painéis solares, permitindo a obtenção de um ponto de operação de referência (caso base).

**Tabela 3.5** – Resultados para rede teste de 22 barras,  $k=2$ , sem alocação de plantas solares [13].

Barra	Número de arranjos	Perdas [kW]	Custo [\$]
N/A	0	364,79	21887

A seguir, acrescenta-se a possibilidade de incorporar 22 plantas solares (pode ser colocada uma em cada barra). Lembrando que cada planta solar pode ser composta por até 5 arranjos. Na Tabela 3.6 é possível observar o resultado desta simulação.

**Tabela 3.6** – Resultados para rede teste de 22 barras,  $k=2$ , possibilidade de alocar 22 plantas solares [13].

Barra	Número de arranjos	Perdas [kW]	Custo [\$]
107	2	314,09	19805
111	5		
118	3		
134	2		

Da Tabela 3.6, apesar da possibilidade de colocar 22 plantas solares, somente foram requeridas 4 plantas, com um total de 12 arranjos a serem instalados o que equivale 240kW, obtendo um menor valor de perdas técnicas no sistema, especificamente 50 kW de redução, com uma função objetivo menor ao comparar com os resultados da Tabela 3.5.

Se, por condições de disponibilidade, não fosse possível instalar arranjos nas barras 134 e 118, o modelo leva à obtenção de um novo ponto de operação ótimo que minimize as perdas técnicas e o investimento nas fontes renováveis. Na Tabela 3.7 é possível observar o resultado desta simulação.

**Tabela 3.7** – Resultados para rede teste de 22 barras,  $k=2$ , possibilidade de alocar plantas solares em todas as barras, exceto na 134 e na 118.

Barra	Número de arranjos	Perdas [kW]	Custo [\$]
107	5	314,48	19828
111	5		
131	2		

Da Tabela 3.7 o modelo encontra uma nova decisão com 3 plantas solares e em total 12 arranjos, mas as perdas técnicas aumentaram, assim como a função objetivo comparando com os resultados da Tabela 3.6.

Uma modificação no algoritmo *K-means* foi feita considerando o número de *clusters*,  $K=3$ . Os resultados com a mesma rede são apresentados na Tabela 3.8.

**Tabela 3.8** – Resultados para rede teste de 22 barras,  $k=3$ , caso base [13].

Barra	Número de arranjos	Perdas kW	Custo [\$]
107	2	316,73	19963
111	5		
118	3		
134	2		

Da Tabela 3.8 é possível obter o número recomendado para minimizar as perdas e o custo de investimento nas fontes renováveis. Neste caso foram utilizados 3 *clusters*, obtendo uma decisão de colocar os mesmos arranjos, mas os valores da função objetivo foram maiores comparados com a Tabela 3.6 ( $k=2$ ).



Com a ideia de verificar a operação lógica do modelo proposto (*k-means* e otimização), a seguir serão apresentados cenários de operação que permitirão avaliar o comportamento do modelo utilizando a rede teste de 22 barras. Em um primeiro momento, a demanda da Tabela A.1 é diminuída, multiplicando-a por 0,4. Na Tabela 3.9 são apresentados os resultados da simulação.

**Tabela 3.9** – Resultados para rede teste de 22 barras,  $k=2$ , caso com demanda diminuída.

Barra	Número de arranjos	Perdas kW	Custo [\\$]
111	3	45,78	3386
118	1		
127	2		
134	2		

Da Tabela 3.9 verifica-se que o modelo encontra uma nova decisão com 4 plantas solares e um total de 8 arranjos. Como a demanda foi diminuída, a função objetivo e as perdas são menores que aquelas mostradas na Tabela 3.6.

A seguir, o limite máximo de corrente no ramal (103,107) foi diminuído para 2 amperes. Neste sentido é esperada a inserção de planta em uma nova barra ou arranjos nas barras da Tabela 3.9, para atender à demanda nas barras 107, 111 e 118. Os resultados desta simulação são apresentados na Tabela 3.10.

**Tabela 3.10** – Resultados para rede teste de 22 barras,  $k=2$ , menor demanda, limite de corrente aumentado no ramal 103-107.

Barra	Número de arranjos	Perdas kW	Custo [\\$]
107	1	45,72	3383
111	3		
118	1		
127	2		
134	1		

Da Tabela 3.10 verifica-se que o modelo encontra uma nova decisão com 5 plantas solares e um total de 8 arranjos. A função objetivo e as perdas são um pouco menores que aquelas da Tabela 3.9.

A seguir, é mantido o limite máximo de corrente no ramal (103,107) definido na simulação anterior, e o limite máximo de corrente no ramal (78,90) foi diminuído para 35 Ampères. Neste sentido é esperada a inserção de uma nova planta porque a subestação não terá condições de atender às demandas das barras na vizinhança da barra 95 e 98. Os resultados desta simulação são apresentados na Tabela 3.11.

**Tabela 3.11** – Resultados para rede teste de 22 barras,  $k=2$ , menor demanda, menor fluxo corrente ramal 103-107 e 78-90.

Barra	Número de arranjos	Perdas kW	Custo [\\$]
100	2	40,57	3394
107	1		
111	3		
118	1		
127	2		
131	1		
134	2		

Agora o modelo encontra uma nova decisão com 7 plantas solares e um total de 12 arranjos. A função objetivo aumentou e as perdas são menores que as da Tabela 3.10. O aumento da função objetivo deve-se à inserção de mais plantas e arranjos solares, no entanto, as perdas diminuíram porque a potência demandada pelas barras na vizinhança das barras 95 e 98 já não são supridas pela subestação, e sim pela planta solar alocada na barra 100.

Finalmente, apresenta-se um último teste envolvendo a rede de 22 barras. Assumindo que, por questões de regulação, requer-se que a magnitude da tensão em uma barra do alimentador seja mantida fixa em  $0,97 \cdot V_{nom}$ . A barra eleita foi a 90, e os resultados da simulação são mostrados na Tabela 3.12.

**Tabela 3.12** – Resultados para rede teste de 22 barras,  $k=2$ , menor demanda, magnitude fixa ( $0,97 \cdot V_{nom}$ ) na barra 90.

Barra	Número de arranjos	Perdas kW	Custo [\\$]
111	3	43,81	3428
118	1		
127	1		
129	4		
134	1		

Da Tabela 3.12, nota-se que o modelo encontra uma nova solução com 5 plantas solares e um total de 10 arranjos. A função objetivo aumentou e as perdas são menores comparando com os resultados da Tabela 3.11. O aumento da função objetivo deve-se à inserção de mais plantas e arranjos solares, no entanto, as perdas diminuíram porque a potência demandada pelas barras a jusante da barra 90 já não são supridas na totalidade pela subestação, sendo parte do suprimento proveniente da planta solar com 4 arranjos alocados na barra 129.

Para a rede elétrica real de uma empresa de distribuição do Equador, foram solicitados parâmetros de funcionamento e requerimentos à empresa distribuidora de energia. A carga total desta rede é de 46,8 MW e 22,6 MVar, e as perdas de potência ativa para esta rede são de 350 kW. Na Tabela 3.13 é possível analisar o comportamento do modelo considerando as técnicas de aprendizagem de máquina a serem inseridas no modelo matemático de otimização para assim obter a localização ótima das fontes de geração solar na rede primária de distribuição. Assim mesmo, é possível implementar 5 arranjos por cada planta solar; e somente podem ser inseridas no caso base até 10 plantas solares.

**Tabela 3.13** – Resultados para rede real de 137 barras  $k=2$ .

Barra	Número de arranjos	Perdas [kW]	Custo [\$]
13	5	215,81	30811
32	4		
51	1		
53	4		
85	3		
94	3		
103	1		
105	5		
117	2		
133	4		

Da Tabela 3.13 é importante observar o desempenho da metodologia proposta neste trabalho de pesquisa considerando dois agrupamentos de acordo com o estudo de *K-means* com os horários em que se sabe que vai existir irradiação solar. De acordo com a Tabela 3.13, foram inseridos 640 kW de geração fotovoltaica, implicando em uma redução das perdas de potência de 350 kW para 215,81 kW, ou seja, uma redução de 38,34%.

Outro caso de estudo foi implementado na rede real, considerando os resultados nos agrupamentos do cotovelo  $k=3$ . A Tabela 3.14 apresenta os resultados para esta situação.

**Tabela 3.14** – Resultados para rede real de 137 barras  $k=3$ .

Barra	Número de arranjos	Perdas [kW]	Custo [\$]
13	5	236,73	32756
34	2		
53	5		
85	3		
94	3		
103	1		
105	5		
117	2		
123	2		
133	4		

Da Tabela 3.14 é importante observar o desempenho da metodologia proposta neste trabalho de pesquisa considerando dois agrupamentos de acordo com o estudo de *k-means* com os horários totais do dia. Como é considerada a alocação durante a baixa produção de irradiação solar, os parâmetros para minimizar a função objetivo são alterados, portanto, a solução obtida é pior que aquela da simulação anterior para  $k=2$ .

## 4 CONCLUSÕES E RECOMENDAÇÕES

A metodologia para a conversão dos parâmetros estocásticos em determinísticos mediante técnicas de agrupamento, mediante (*K-means*), foi integrada ao modelo matemático de otimização resultando na metodologia proposta nesta dissertação. Foi requerida a validação dos resultados dos agrupamentos analisando as respostas obtidas nas simulações do modelo de programação matemática cônico de segunda ordem.

A simplicidade de trabalhar com parâmetros determinísticos foi aproveitada e analisada na presente pesquisa, já que trabalhar com parâmetros estocásticos leva o modelo a ficar muito pesado computacionalmente.

Optou-se por uma metodologia conservadora, porque a priori se tinha conhecimento do comportamento da climatologia ao longo do ano já que o setor escolhido tem as duas épocas do ano bem definidas, baseada na identificação dos parâmetros em função das épocas do ano típico (por exemplo, com chuva e sem chuva) para avaliar o comportamento dos parâmetros e seu respectivo impacto na produção dos painéis fotovoltaicos. Para robustecer a proposta, poder-se-ia inserir a análise de agrupamento mediante modelos mistos gaussianos (GMM - Gaussian Mixture Modelling), metodologia que analisa a porcentagem ou probabilidade de ocorrência dos centroides dos agrupamentos.

É importante observar o desempenho da metodologia proposta neste trabalho de pesquisa considerando dois agrupamentos ( $k=2$ ) de acordo com o estudo de *K-means*, tem bons resultados, já que são horários em que se sabe que vai existir irradiação solar. Para o estudo considerando a alocação durante a baixa produção de irradiação solar ( $k=3$ ), os parâmetros para minimizar a função objetivo são alterados, portanto, a solução obtida é pior que aquela da simulação obtida para  $k=2$ .

A contribuição deste trabalho está relacionada com a inserção de técnicas de aprendizagem de máquinas, considerando a administração de grandes blocos de dados históricos, que permitam melhorar o desempenho dos modelos matemáticos de otimização clássicos conhecidos na literatura.

Com base nas simulações feitas, verificou-se que, com a utilização de uma técnica heurística (separando somente os meses com chuva e sem), a metodologia apresentou bons resultados em termos de obter soluções que tenham excelentes pontos de operação.

Dependendo da localização da rede da empresa de distribuição de energia elétrica, é possível expandir a análise incorporando outros fatores ambientais que podem ser atrativos no planejamento da operação, como a velocidade do vento. Por exemplo, no Equador se tem cidades que ficam na altitude acima dos 3000 metros em relação ao nível do mar, parecendo atrativo a incorporação destas tecnologias para mitigar a contaminação ambiental.

Uma melhoria rápida a ser inserida no presente trabalho seria considerar o comportamento do painel solar ao longo do tempo, considerando custos de operação e manutenção para obter uma ótima operação do sistema integral e também expandir a ideia para o planejamento da rede sob análise.

Seguindo o comportamento real das redes de distribuição, para trabalhos futuros poderia ser muito importante avaliar o comportamento da metodologia mediante (*K-means*) se a rede já contém originalmente fontes renováveis, mas deseja-se incrementar outras fontes para cobrir a demanda futura. A proposta de tornar a análise mais robusta utilizando o GMM é muito atrativa devido à resposta em termos de probabilidades de ocorrência, e deixar estas fontes existentes como parâmetros conhecidos do modelo, deixando somente uma nova análise para as possíveis novas fontes renováveis.

## REFERÊNCIAS

- [1] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf, «Carpe diem: A novel approach to select representative days for long-term power system modeling», *Energy*, vol. 112, pp. 430-442, 2016, doi: <https://doi.org/10.1016/j.energy.2016.06.081>.
- [2] M.-H. Lin, J. G. Carlsson, D. Ge, J. Shi, and J.-F. Tsai, «A Review of Piecewise Linearization Methods», *Math. Probl. Eng.*, pp. 1-8, 2013, doi: 10.1155/2013/101376.
- [3] D. Pascual, F. Pla, y S. Sánchez, «Algoritmos de agrupamiento», *Método Informático Av.*, 2007.
- [4] A. Samuel, «Some studies in machine learning using the game of checkers», *IBM J. Res. Dev.*, vol. 3, n.o 3, pp. 210-229, jul. 1959, doi: 10.1147/rd.33.0210.
- [5] R. Duda, P. Hart, and D. Stork, *Pattern Classification, Second*. John Wiley & Sons, 2000.
- [6] J. Jara estupiñán, D. Giral, y F. Martínez Santa, «Implementación de algoritmos basados en máquinas de soporte vectorial (SVM) para sistemas eléctricos: revisión de tema», *Rev. Tecnura*, vol. 20, n.o 48, pp. 149-170, 2016, doi: 10.14483/udistrital.jour.tecnura.2016.2.a11.
- [7] R. Fourer, D. Gay, and B. Kernighan, «AMPL: a modeling language for mathematical programming», *Pac. Grove ThomsonBrooksCole*, 2003.
- [8] S. Civanlar, «Distribution feeder reconfiguration for loss reduction», *IEEE Trans. Power Deliv.*, vol. 3, n.o 3, pp. 1217-1223, 1988.
- [9] D. S. Hong, «Reconfiguration of electric distribution networks for resistive line losses reduction», *IEEE Trans. Power Deliv.*, vol. 4, n.o 2, pp. 1492-1498, 1989.
- [10] M. E. Wu, «Optimal sizing of capacitors placed on aradial distribution system», *IEEE Trans. Power Deliv.*, vol. 4, n.o 1, pp. 725-734, ene. 1989.
- [11] M. E. Wu, «Optimal capacitor placement on radial dis-tribution systems», *IEEE Trans. Power Deliv.*, vol. 4, n.o 1, pp. 725-734, 1989.
- [12] L. F. Ochoa, «Minimizing Energy Losses: Optimal Accommodation and Smart Operation of Renewable Distributed Generation», *IEEE Trans. Power Syst.*, vol. 26, n.o 1, pp. 198-205, feb. 2011.
- [13] O. Melgar, M. Pourakbari, M. Lavorato, and J. Sanches, «Optimal siting and sizing of renewable energy sources, storage devices, and reactive support devices to obtain a sustainable electrical distribution systems», *Springer*, 2017.

- [14] E. El-Saadany and A. Al.Badi, «Distributed Generation: Benefits and Challenges», Int. Conf. Commun. Comput. Power, pp. 115-119, feb. 2007.
- [15] L. F. Ochoa, «Evaluating distributed generation impacts with a multi objective index», IEEE Trans. Power Deliv., vol. 21, n.o 3, pp. 1452-1458, jul. 2006.
- [16] P. C. Ramakumar, «An approach to quantify the technical benefits of distributed generation», IEEE Trans Energy Convers, vol. 19, n.o 4, pp. 764-773, dic. 2004.
- [17] C. L. T. Borges and D. M. Falcao, «Impact of distributed generation allocation and sizing on reliability, losses and voltage profile», IEEE Power Tech, vol. 2, n.o 5, 2003.
- [18] C. L. T. Falcao, «Optimal distributed generation allocation for reliability, losses, and voltage improvement», Int J ElectPower Energy Syst, vol. 28, n.o 6, pp. 413-420, 2006.
- [19] K. Nara, «Optimal allocation of dispersed generators for loss minimization», Elect EngJpn, vol. 136, n.o 2, pp. 1-8, 2001.
- [20] G. Celli, «A multiobjective evolutionary algorithm for the sizing and siting of distributed generation», IEEE Trans Power Syst, vol. 20, n.o 2, pp. 750-757, may 2005.
- [21] K. Poncelet, H. Höschle, E. Delarue, A. Virag, y W. D'haeseleer, «Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning», IEEE Trans. Power Syst., vol. 32, n.o 3, pp. 1936-1948, may 2017.
- [22] K. Praveena and A. Venkateshwar, «Power loss minimization incorporating distributed generator in distribution system using Supervised Big Bang Big Crunch Method», Int. J. Eng. Res. Gen. Sci., vol. 3, n.o 3, pp. 1077-1086, 2015.
- [23] B. Raj and N. Karthikeyan, «Loss optimization in distribution networks with distributed generation», Determ. Autom. Demands Improv. Control. Obs. Distrib. Netw., 2017.
- [24] M. Kashem, A. Negnevitsky, and G. Ledwich, «Distributed Generation for Minimization of power losses in distribution system», Aust. Res., pp. 1-8, 2006.
- [25] S. Kumar, R. Madurai, A. Singh, S. Vardhan, R. Saket, y N. Das, «Reliability Assessment of Wind-Solar PV Integrated Distribution System Using Electrical Loss Minimization Technique», energies, vol. 13, n.o 5631, pp. 1-30, 2020.
- [26] L. F. Ochoa and G. P. Harrison, «Distribution Network Capacity Assessment: Variable DG and Active Networks», IEEE Trans. Power Syst., vol. 25, n.o 1, pp. 87-95, 2010.
- [27] L. Ioan, M. Abrudeanb, and D. Bica, «Optimal Location of a Distributed Generator for Power Losses Improvement», 9th Int. Conf. Interdiscip. Eng., 2015.
- [28] A. Neplan, NEPLAN. Zurich: Neplan AG. [On line]. Available on: <https://www.neplan.ch/escompany/?lang=es>



- [29] U. Sultana, A. Khairuddin, A. Mokhtar, S. Hussain Qazi, and B. Sultana, «An optimization approach for minimizing energy losses of distribution systems based on distributed generation placement», *J. Teknol.*, vol. 79, n.o 4, pp. 87-96, 2017.
- [30] S. Kirkpatrick, C. D. Gellat, and M. P. Vecchi, «Optimization by simulated annealing», *J. Stat. Phys.*, vol. 34, n.o 5, pp. 975-986, 1983.
- [31] G. A. Glover J., «Handbook of Metaheuristics». Kluwer Academic Publishers, 2003.
- [32] W. S. Tan, M. Y. Hassan, M. S. Majid, and H. A. Rahman, «Allocation and sizing of DG using Cuckoo search Algorithm», *IEEE Int. Conf. Power Energy*, 2012.
- [33] W. S. Tan, M. Y. Hassan, and M. P. Abdullah, «Multi-distributed generation planning using hybrid particle swarm optimization-gravitational search algorithm», *IET Gener. Transm. Distrib.*, vol. 7, n.o 9, pp. 929-942, 2013.
- [34] J. M. López-Lezama, F. Villada, and N. Muñoz, «Metaheuristics Applied to the Optimal Location and Sizing of Distributed Generation in Power Distribution Systems», *Inf Tecnol*, vol. 26, n.o 4, 2005, doi: 10.4067/S0718-07642015000400018.
- [35] M. Zangiabadi, R. Feuillet, and H. Lesani, «An approach to Deterministic and Stochastic Evaluation of the Uncertainties in Distributed Generation Systems», presentado en 20th International Conference on Electricity Distribution, Prague, Czech Republic, jun. 2009. doi: 10.1049/cp.2009.1098.
- [36] AprendeIA, «Algoritmo Agrupamiento Jerárquico – Teoría». Aprende IA. Accedido: 26 de junio de 2021. [On line]. Available on: <https://aprendeia.com/algoritmo-agrupamiento-jerarquico-teoria/>
- [37] C. García Carbonero and I. Gómez Moreno, «Algoritmos de aprendizaje: KNN & K-means», Intel. En Redes Telecomunicación, 2006.
- [38] G. A. Wilkin and X. Huang, «K-Means Clustering Algorithms: Implementation and Comparison», *Second Int. Multisymposium Comput. Comput. Sci.*, Iowa City 2017.
- [39] IBM, IBM ILOG CPLEX. [En línea]. Disponible en: <https://www.ibm.com/es-es/products/ilog-cplex-optimization-studio>
- [40] D. Matuszko, «Influence of the extent and genera of cloud cover on solar radiation intensity», *Int. J. Climatol.*, pp. 2403-2414, 2012.
- [41] Graph Everywhere, «¿Qué es el Clustering?» Accessed: 26 june 2021. [On line]. Disponible em: [www.grapheverywhere.com/que-es-el-clustering](http://www.grapheverywhere.com/que-es-el-clustering)
- [42] M. J. Rider, «Estado de Operação em Regime Permanente dos Sistemas de Distribuição de Energia Elétrica», Campinas - Brasil, Maio de 2020. [On line]. Disponible em: Pagina: [www.dsee.fee.unicamp.br/~mjrider](http://www.dsee.fee.unicamp.br/~mjrider)

- [43] R. Cespedes, «New method for the analysis of distribution networks», IEEE Trans. Power Deliv., vol. 5, n.o 1, pp. 391-396, 1990.
- [44] D. Shirmohammadi, H. W. Hong, A. Semlyen, and G. X. Luo, «A compensation-based power flow method for weakly meshed distribution and transmission networks», IEEE Trans Power Syst, vol. 3, n.o 2, pp. 753-762, may 1988.
- [45] C. Villajuan, «Modelos de programação linear mista para resolver problemas de planejamento da expansão e da operação de sistemas de transmissão de energia elétrica», Dissertação de Mestrado, Faculdade de Engenharia do Câmpus de Ilha Solteira – UNESP, Ilha Solteira, 2016.
- [46] M. Farivar and H. Low, «Branch flow model Relaxations and convexification (Part I, II)», IEEE Trans Power Syst, vol. 28, n.o 3, pp. 2554-2572, 2013.
- [47] J. F. Franco, M. J. Rider, y R. Romero, «A Mixed Integer Quadratically Constrained Programming Model for the Distribution System Expansion Planning», Int J Elect Power Energy Syst, vol. 62, pp. 265-272, nov. 2014.
- [48] L. H. Macedo, J. F. Franco, M. J. Rider, and R. Romero, «Optimal Operation of Distribution Networks Considering Energy Storage Devices», IEEE Trans Smart Grid, vol. 6, n.o 6, pp. 2825-2836, nov. 2015.
- [49] A. Valdez Alvarado, «Machine Learning para todos», presentado en IV CONCITEK, Perú, 2019. doi: 10.13140/RG.2.2.13786.70086.
- [50] J. Ortiz, «Aprende Machine Learning y sus aplicaciones. 20 ejemplos». [On line]. Disponível em: <http://www.joanybelortiz.com/aplicaciones-machine-learning-ejemplos/>
- [51] J. B. MacQueen, L. M. Le Cam, and J. Neyman, «Some methods for classification and analysis of multivariate observations», Proc. Fifth Berkeley Symp. Math. Stat. Probab., vol. 1, pp. 281-297.
- [52] M. J. Garbade, «Understanding K-means Clustering in Machine Learning». junio de 2021. [En línea]. Disponível em: <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
- [53] B. P. Ferraz, «Modelo linearizado para problemas de planejamento da operação de sistemas de distribuição considerando incertezas», Doutorado em Engenharia Elétrica, -Universidade Federal do Rio Grande do Sul, Brasil, 2020.
- [54] J. Kopepasah, «Aprende Machine Learning - K-means en python paso a paso». Accessed: 26 june 2021. [On line]. Disponível em: <https://www.aprendemachinelarning.com/k-means-en-python-paso-a-paso>
- [55] P. Wilkinson, «Introduction to Hierarchical clustering (part 1 — theory, linkage and affinity)». 5 de marzo de 2021. Accedido: 17 de enero de 2022. [On line]. Disponível em:

- <https://towardsdatascience.com/introduction-to-hierarchical-clustering-part-1-theory-linkage-and-affinity-e3b6a4817702>
- [56] AprendeIA, «Seleccionar el número adecuado de clústeres». Aprende IA. Accedido: 26 de junio de 2021. [On line]. Disponible em: <https://aprendeia.com/seleccionar-el-numero-adecuado-de-clusteres/>
- [57] MATLAB and Statistics Toolbox Release. Natick, Massachusetts, United States: The MathWorks, Inc., 2020.
- [58] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, y S. P. Zhang, «A comparative study of clustering techniques for electrical load pattern segmentation», *Renew. Sustain. Energy Rev.*, vol. 120, n.o 109628, 2020.
- [59] T. K. Wijaya, M. Vasirani, and A. K. Humeau, «Cluster-based aggregate forecasting for residential electricity demand using smart meter data», *IEEE Int. Conf. Big Data Big Data*, Santa Clara 2015.
- [60] K. Mora, A. Almeida, y M. Torres, «Evaluación y propuesta de metodología aplicada a la estratificación de clientes pertenecientes al área de concesión de una Unidad de Negocio de la Corporación Nacional de Electricidad mediante la implementación de un proceso de zonificación», Proyecto Integrador previo obtención del título de Ingeniero en Electricidad, ESPOL, Guayaquil, 2021.
- [61] SOLCAST, «API Toolkit». 25 de junio de 2020. [On line]. Disponible em: <https://toolkit.solcast.com.au/live-forecast>
- [62] S. Ranka, K. A., and V. S., «An efficient k-means clustering algorithm», *Electr. Eng. Comput. Sci.*, n.o 43, 1997.
- [63] Santana y O. Fernández, «El análisis de cluster: aplicación, interpretación y validación», *Pap. Sociol.*, pp. 65-76, 1991.
- [64] L. F. Pinto y M. Sánchez, «Análisis de la aplicación de algoritmos de K-means y continuos Max Flow a la segmentación de imágenes de color», *Deposito Investig. Univ. Sevilla*, 2015.
- [65] J. B. Lara, «A Homophily-Based Rumor Propagation Model in Social Networks», *Universidad Politécnica Madr.*, 2020.
- [66] J. Pérez et al., «Mejora al algoritmo de agrupamiento K-means mediante un nuevo criterio de convergencia y su aplicación a bases de datos poblacionales de cáncer», *Liver-2do Taller Lat. Iberoam. Investig. Oper. IO Apl. Soluc. Probl. Reg.*, pp. 1-7, 2007.

## **APÊNDICES**

## APÊNDICE A – Publicações resultantes do Trabalho de Pesquisa

Os artigos científicos a seguir são resultantes deste trabalho de pesquisa.

J. ARAGUNDI, J. PEQUENA, E. RUPPERT, “Uma contribuição ao estudo do impacto de geração eólica no comportamento dos sistemas de energia elétricos,” Publicado nos anais do *XXI Congresso Brasileiro de Automática – CBA2016*, Outubro 2016.

J. PEQUENA, L. PEREIRA, J. ARAGUNDI, E. RUPPERT, “Estudios Eléctricos requeridos para una interconexión segura entre sistema eléctrico y sistema de generación eólica,” Publicado nos anais do *XXI Congresso Brasileiro de Automática – CBA2016*, Outubro 2016.

J. ARAGUNDI, J. CORDOVA, M. TORRES, C. CASTRO, “Optimal Placement and Sizing of Solar Generation Units in Distribution Systems Including an Intermittency Clustering Approach”, submetido ao *IEEE Transactions on Sustainable Energy*.

**ANEXOS**

**ANEXO A – DADOS DO SISTEMA DE 22 BARRAS**

**Tabela A.1** – Informações das barras para o sistema teste de 22 barras.

Barra	Tipo de Barra: 1(SE) / 0(Carga)	Demanda	
		Ativa (kW)	Reativa (KVAr)
1	1	178,41	76,07
10	0	973,06	414,65
23	0	794,77	338,73
38	0	483,50	206,04
48	0	712,13	303,38
63	0	729,33	310,71
78	0	425,31	167,24
90	0	88,05	38,57
95	0	237,52	101,35
98	0	3,14	1,40
99	0	21,60	9,30
100	0	172,50	73,50
103	0	167,36	71,32
107	0	281,12	117,28
111	0	222,60	96,47
118	0	73,39	31,97
119	0	66,69	28,73
121	0	27,60	11,70
127	0	157,71	66,67
129	0	345,05	159,98
131	0	173,65	73,99
134	0	165,60	70,56



**Tabela A.2** – Informações dos circuitos para o sistema teste de 22 barras.

Barra i	Barra j	Parâmetros		
		R (Ohm)	X (Ohm)	I <sub>max</sub> (A)
1	10	0,8393	0,9254	600
10	23	0,2158	0,2380	600
23	38	0,2158	0,2380	600
38	48	0,2398	0,2644	600
48	63	0,2998	0,3305	600
63	78	0,2998	0,3305	600
78	129	0,9107	0,9348	600
78	90	0,3151	0,3305	600
90	95	0,2278	0,2512	600
95	98	0,4284	0,3067	600
98	99	0,1258	0,0964	600
98	100	0,1993	0,1427	600
95	103	0,2278	0,2512	600
103	131	0,2742	0,2548	600
131	134	0,3809	0,3538	600
103	107	0,3886	0,3609	600
107	111	0,4766	0,3774	600
107	118	2,3869	1,8629	600
90	119	0,2092	0,1498	600
119	121	0,4184	0,2996	600
119	127	1,3787	1,0425	600