

INSTITUTO FEDERAL DE SANTA CATARINA

LUCAS COELHO RAUPP

**Desenvolvimento de uma Metodologia  
Utilizando Aprendizado de Máquina para  
Estimativa de Perdas Energéticas em Usinas  
Fotovoltaicas por Falhas em *Trackers***

São José - SC

fevereiro/2025

# **DESENVOLVIMENTO DE UMA METODOLOGIA UTILIZANDO APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE PERDAS ENERGÉTICAS EM USINAS FOTOVOLTAICAS POR FALHAS EM *TRACKERS***

Monografia apresentada ao Curso de Engenharia de Telecomunicações do campus São José do Instituto Federal de Santa Catarina para a obtenção do diploma de Engenheiro de Telecomunicações.

Orientador: Prof. Mario de Noronha Neto,  
Dr.

Coorientador: Prof. Richard Demo Souza,  
Dr.

São José - SC

fevereiro/2025

Lucas Coelho Raupp

Desenvolvimento de uma Metodologia Utilizando Aprendizado de  
Máquina para Estimativa de Perdas Energéticas em Usinas  
Fotovoltaicas por Falhas em *Trackers*

Este trabalho foi julgado adequado para obtenção do título de Engenheiro de Telecomunicações, pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, e aprovado na sua forma final pela comissão avaliadora abaixo indicada.

São José - SC, 25 de fevereiro de 2025:

---

**Prof. Mario de Noronha Neto, Dr.**  
Orientador  
Instituto Federal de Santa Catarina

---

**Prof. Richard Demo Souza, Dr.**  
Coorientador  
Universidade Federal de Santa Catarina

---

**Prof. Arliones Stevert Hoeller Junior,  
Dr.**  
Instituto Federal de Santa Catarina

---

**Prof. Roberto Wanderley da Nóbrega,  
Dr.**  
Instituto Federal de Santa Catarina

*Este trabalho é dedicado à minha família,  
que nunca deixou de me apoiar.*

# AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão, em primeiro lugar, à minha família, que esteve presente em todos os momentos da minha jornada acadêmica. Agradeço especialmente aos meus pais e a minha namorada, que tantas vezes abriram mão de momentos de lazer para me auxiliar nos estudos.

Agradeço também aos meus colegas de curso e aos parceiros de projetos de pesquisa e desenvolvimento no Inova IFSC. Em especial, meu agradecimento vai para Andrey Adriano da Rosa e Fausto Cristiano, que estiveram ao meu lado durante todos os cinco anos dessa jornada, me apoiando, motivando e dando forças para seguir em frente, mesmo nos momentos em que parecia quase impossível.

Sou profundamente grato ao meu orientador, Mario de Noronha Neto, e ao meu coorientador, Richard Demo Souza, por acreditarem em meu potencial e me abrirem as portas para o projeto de Pesquisa e Desenvolvimento da empresa ENGIE Brasil Energia S.A., em parceria com o IFSC e a UFSC, regulado pela Agência Nacional de Energia Elétrica (ANEEL) (PD-00403-0056/2023), que resultou neste trabalho de conclusão de curso. Agradeço também pelo suporte e orientação fundamentais durante esse marco tão importante da minha trajetória acadêmica.

Por fim, expresso minha gratidão especial a todos os professores que guiaram minha trajetória até aqui. Além de transmitirem seus conhecimentos, foram fundamentais para que eu pudesse expandir minha visão de mundo e crescesse como pessoa e profissional.

*Há uma força motriz mais poderosa que o vapor,  
a eletricidade e a energia atômica: a vontade.*

*Albert Einstein*

# RESUMO

Diante da crescente demanda por fontes de energia renováveis, a energia solar tem ganhado destaque global por sua eficiência e sustentabilidade. Para maximizar a sua geração, usinas de grande porte utilizam *trackers*, dispositivos que ajustam a inclinação dos painéis solares conforme a posição do sol. No entanto, anomalias como o travamento desses dispositivos comprometem o desempenho das usinas, reduzindo a geração e demandando ações corretivas por parte dos operadores. Este trabalho propõe uma metodologia para estimar o percentual de perdas energéticas causadas por esse problema em uma usina solar de grande porte localizada no Brasil. O método foi desenvolvido em quatro etapas: tratamento e geração de dados, construção de uma tabela de perdas, criação de um modelo de *Machine Learning*, e aplicação em cenários reais. Como resultado, obteve-se uma regressão polinomial de segundo grau capaz de estimar perdas passadas e futuras de maneira simples e rápida, com precisão média de 86% quando comparada com a tabela de perdas gerada previamente, demonstrando o potencial do método tanto em aplicações acadêmicas quanto corporativas.

**Palavras-chave:** Energia solar. Usina fotovoltaica de grande porte. *Tracker*. *Machine Learning*. Eficiência energética.

# ABSTRACT

Given the growing demand for renewable energy sources, solar energy has gained global prominence for its efficiency and sustainability. To maximize generation, large-scale power plants employ *trackers*, devices that adjust the tilt of solar panels according to the sun's position. However, anomalies such as the malfunctioning of these devices compromise plant performance, reducing energy generation and requiring corrective actions by operators. This work proposes a methodology to estimate the percentage of energy losses caused by this issue in a large-scale solar power plant located in Brazil. The method was developed in four stages: data processing and generation, construction of a loss table, creation of a *Machine Learning* model, and application in real scenarios. As a result, a second-degree polynomial regression was obtained, capable of estimating past and future losses in a simple and fast manner, with an average accuracy of 86% when compared to the previously generated loss table, demonstrating the method's potential for both academic and corporate applications.

**Keywords:** Solar energy. Large-scale solar plant. Tracker. Machine Learning. Energy efficiency.

# LISTA DE ILUSTRAÇÕES

Figura 1 – Funcionamento de uma usina fotovoltaica . . . . .	17
Figura 2 – Exemplo de uma curva de funcionamento de um <i>tracker</i> . . . . .	18
Figura 3 – Gráfico MPPT . . . . .	19
Figura 4 – <i>Machine Learning Pipeline</i> . . . . .	23
Figura 5 – Diagrama da estrutura da geração e tratamento dos dados . . . . .	25
Figura 6 – Diagrama da geração dos dados . . . . .	26
Figura 7 – Diagrama do pré-processamento da GTI . . . . .	27
Figura 8 – Séries temporais dos piranômetros . . . . .	28
Figura 9 – Gráfico de Gantt dos piranômetros . . . . .	29
Figura 10 – Comparação entre as curvas GTI e GHI . . . . .	29
Figura 11 – Diagrama da identificação de <i>trackers</i> travados em dias com GHI indisponível . . . . .	30
Figura 12 – Classificação dos sensores com GHI indisponível . . . . .	30
Figura 13 – Diagrama da identificação de <i>trackers</i> travados em dias com GHI disponível . . . . .	31
Figura 14 – Classificação dos sensores com GHI disponível . . . . .	32
Figura 15 – GTI teórica . . . . .	33
Figura 16 – Comparação entre potências . . . . .	34
Figura 17 – Diagrama da criação da tabela de perdas . . . . .	35
Figura 18 – Diagrama do cálculo do CSI . . . . .	35
Figura 19 – Comparação entre a GHI medida e a teórica . . . . .	36
Figura 20 – Diagrama do cálculo da perda teórica e geração da tabela . . . . .	36
Figura 21 – Perda por ângulo para dias com CSI de 0,94 e 0,34, respectivamente . . . . .	38
Figura 22 – Perdas por intervalo de CSI . . . . .	38
Figura 23 – Perdas por quantil de CSI . . . . .	39
Figura 24 – Perdas por ângulo para CSI's específicos . . . . .	40
Figura 25 – Diagrama da estrutura da geração da regressão e aplicação do método . . . . .	41
Figura 26 – Diagrama da geração da regressão . . . . .	41
Figura 27 – Classificação de dados usando K-NN . . . . .	45
Figura 28 – Perda estimada diária por <i>trackers</i> travados . . . . .	47
Figura 29 – Blocos informativos . . . . .	47

# LISTA DE TABELAS

Tabela 1 – Tabela de perdas para um dia com CSI alto . . . . .	37
Tabela 2 – Tabela de perdas para um dia com CSI baixo . . . . .	37
Tabela 3 – Tabela comparativa entre os modelos . . . . .	46

# LISTA DE ABREVIATURAS E SIGLAS

**CA** Corrente Alternada.

**CC** Corrente Contínua.

**CSI** *Clear Sky Index*.

**DPS** Dispositivo de Proteção Contra Surtos.

**EQM** Erro Quadrático Médio.

**GHI** *Global Horizontal Irradiance*.

**GTI** *Global Tilted Irradiance*.

**IDE** Ambiente de Desenvolvimento Integrado.

**IFSC** Instituto Federal de Santa Catarina.

**IIQ** Intervalo Interquartil.

**K-NN** *K-Nearest Neighbors*.

$L_e$  Perda Estimada.

$L_i$  Limite Inferior.

$L_s$  Limite Superior.

$L_t$  Perda Teórica.

**ML** *Machine Learning*.

**MPPT** *Maximum Power Point Tracking*.

**MSE** *Mean Squared Error*.

**ONS** Operador Nacional do Sistema Elétrico.

$P_{GHI}$  Potência Teórica Calculada Considerando a GHI.

$P_{GTI}$  Potência Teórica Calculada Considerando a GTI.

$Q_1$  Primeiro Quartil.

$Q_3$  Terceiro Quartil.

**R<sup>2</sup>** Coeficiente de Determinação.

**RMSE** *Root Mean Squared Error*.

**SGD** *Stochastic Gradient Descent*.

**UFSC** Universidade Federal de Santa Catarina.

**VM** *Virtual Machine*.

# LISTA DE SÍMBOLOS

$A_a$  precisão média.

$CSI$  *Clear Sky Index*.

$GHI_m$  *Global Horizontal Irradiance* medida.

$GHI_t$  *Global Horizontal Irradiance* teórica.

$IIQ$  intervalo interquartil.

$L_e$  perda estimada.

$L_i$  limite inferior.

$L_s$  limite superior.

$L_t$  perda teórica.

$n$  número total de amostras.

$P_{GTI}$  potência teórica baseada na *Global Tilted Irradiance*.

$P_N$  potência teórica com *trackers* travados em um ângulo “N”.

$Q_1$  primeiro quartil.

$Q_3$  terceiro quartil.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>1.1</b>	<b>Objetivo geral</b>	<b>15</b>
<b>1.2</b>	<b>Objetivos específicos</b>	<b>16</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>17</b>
<b>2.1</b>	<b>Geração de energia fotovoltaica</b>	<b>17</b>
2.1.1	Componentes de uma usina	18
2.1.1.1	<i>Tracker</i>	18
2.1.1.2	<i>Stringbox</i>	18
2.1.1.3	Inversor	18
2.1.2	Cenário da geração fotovoltaica no mundo	19
<b>2.2</b>	<b>Análise e processamento dos dados em sistemas fotovoltaicos</b>	<b>19</b>
2.2.1	Séries temporais	20
2.2.2	Processamento dos sensores	20
2.2.2.1	Piranômetros e potência	20
2.2.2.2	<i>Setpoint</i>	21
2.2.2.3	Angulação dos <i>trackers</i>	21
<b>2.3</b>	<b>Simulação de potência em sistemas fotovoltaicos</b>	<b>21</b>
<b>2.4</b>	<b>AutoML</b>	<b>22</b>
<b>2.5</b>	<b>Ferramentas utilizadas</b>	<b>23</b>
2.5.1	<i>Docker</i>	23
2.5.2	<i>Jupyter Notebook</i>	24
2.5.3	TPOT	24
2.5.4	<i>Streamlit</i>	24
<b>3</b>	<b>GERAÇÃO E TRATAMENTO DOS DADOS</b>	<b>25</b>
<b>3.1</b>	<b>Geração dos dados</b>	<b>25</b>
3.1.1	Limpeza e tratamento dos dados	25
3.1.2	Pré-processamento da curva GTI	26
3.1.2.1	Separação do <i>Data Frame</i>	26
3.1.2.2	Classificação e remoção de sensores “indisponíveis”	27
3.1.2.3	Identificação de <i>trackers</i> travados	28
3.1.2.4	Cálculo da GTI de referência	33
3.1.3	Geração das potências teóricas	33
<b>3.2</b>	<b>Criação e análise da tabela de perdas</b>	<b>34</b>

3.2.1	Cálculo do CSI . . . . .	35
3.2.2	Determinação da perda teórica e geração da tabela . . . . .	36
3.2.3	Análise do comportamento das perdas . . . . .	37
<b>4</b>	<b>MODELAGEM DA REGRESSÃO E APLICAÇÃO DA METODOLOGIA . . . . .</b>	<b>41</b>
<b>4.1</b>	<b>Modelagem da regressão . . . . .</b>	<b>41</b>
4.1.1	Remoção de <i>outliers</i> . . . . .	41
4.1.2	Separação do conjunto de dados . . . . .	42
4.1.3	Geração de diferentes modelos de regressão . . . . .	42
4.1.4	Comparação entre os modelos . . . . .	45
<b>4.2</b>	<b>Aplicação da metodologia . . . . .</b>	<b>46</b>
4.2.1	Cálculo das perdas . . . . .	46
<b>5</b>	<b>CONCLUSÕES . . . . .</b>	<b>48</b>
<b>5.1</b>	<b>Trabalhos futuros . . . . .</b>	<b>49</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>50</b>
	<b>APÊNDICES . . . . .</b>	<b>54</b>
	<b>APÊNDICE A – REPOSITÓRIO DO PROJETO . . . . .</b>	<b>55</b>
	<b>APÊNDICE B – TRECHO DE GERAÇÃO DA SÉRIE TEMPORAL DE UM TRACKER . . . . .</b>	<b>56</b>

# 1 INTRODUÇÃO

A energia solar é uma fonte de energia limpa e sustentável, sendo o segmento energético que mais cresce no Brasil e no mundo (IEA, 2023b). Para usufruir desse recurso, usinas fotovoltaicas de grande porte foram implantadas, por exemplo, no estado de Minas Gerais e no Nordeste, regiões próximas da linha do Equador e com menor cobertura florestal se comparada com o restante do território brasileiro. Em outubro de 2023, a energia solar representava 5,27% da matriz energética brasileira, totalizando 10,4 GW (GOV.BR, 2023).

Comparativamente à Alemanha, um dos líderes mundiais em energia solar, as regiões menos ensolaradas do Brasil podem gerar mais energia do que as áreas mais ensolaradas do país europeu (TONOLO, 2019). No entanto, a eficiência das usinas está sujeita a múltiplas variáveis, como temperatura, sujeira nos módulos fotovoltaicos e sombreamento, representando perdas na potência total que podem chegar até 15% (TONOLO, 2019). Sendo assim, a capacidade de identificar essas anomalias e quantificar suas respectivas perdas energéticas de forma automatizada, possibilita a redução das perdas na geração, a priorização de investimentos e um planejamento mais preciso da energia gerada, melhorando a eficiência operacional e otimizando todo o processo.

Segundo Raptor Maps (2023), os principais fatores de perda em usinas solares de larga escala decorrem de quatro categorias de anomalias: *strings*, inversores, *trackers* e combinadores. Entre essas, as anomalias em *trackers* são as mais comuns, representando aproximadamente 26,86% de todas as anomalias.

Dado que as falhas em *trackers* são comuns em usinas de grande porte, quantificar o percentual de perdas energéticas decorrentes delas, torna-se essencial para o planejamento de geração e para a tomada de decisões. Conhecer quanto da geração de energia é comprometida por falhas em *trackers* permite aos gestores avaliar a viabilidade econômica de investir em novas tecnologias ou na substituição dos componentes que falham de forma recorrente.

## 1.1 Objetivo geral

Com base neste contexto, este trabalho tem como objetivo estimar o percentual de perdas energéticas em uma usina solar de grande porte devido a falhas em *trackers*.

## 1.2 Objetivos específicos

Para alcançar o objetivo geral, serão aplicados os seguintes objetivos específicos:

- Desenvolver uma metodologia para estimativa de perdas;
- Aplicar modelos de *Machine Learning* (ML) nesta metodologia;
- Implementar um *dashboard* para visualização dos resultados.

## 2 FUNDAMENTAÇÃO TEÓRICA

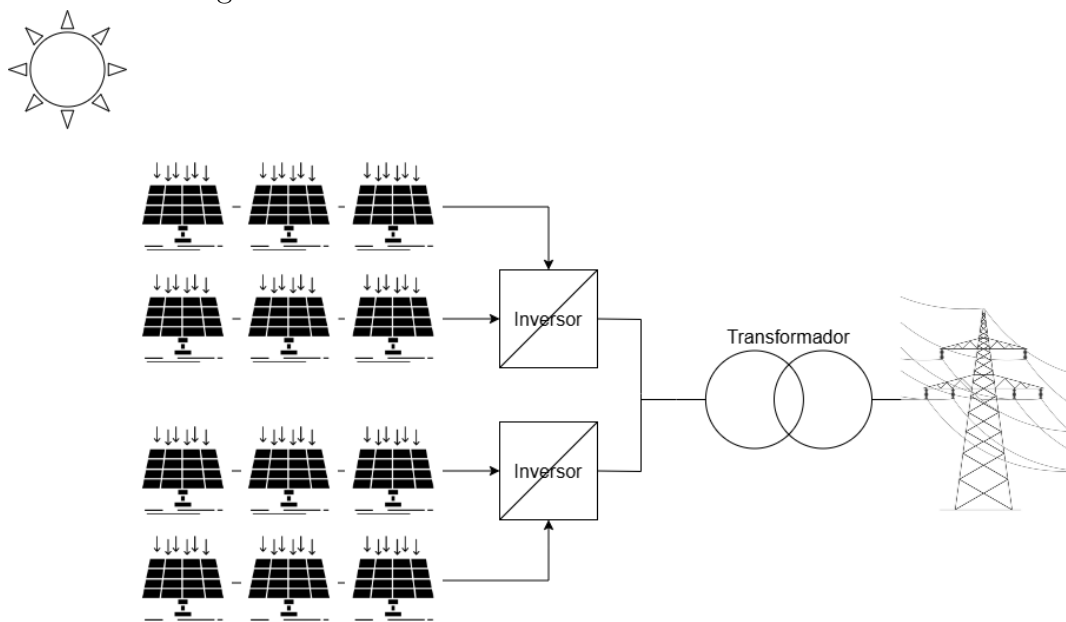
Este capítulo tem o objetivo de apresentar e fundamentar os conceitos e tecnologias que serão utilizados neste projeto, contextualizando e ilustrando seus objetivos.

### 2.1 Geração de energia fotovoltaica

Com base no princípio do efeito fotoelétrico, em que o material absorve fótons e libera elétrons, as usinas solares utilizam um dispositivo semicondutor conhecido como célula fotovoltaica para converter a energia presente nos fótons em energia elétrica. Esse processo ocorre quando os elétrons são excitados da camada de valência para a camada de condução, gerando eletricidade na forma de **Corrente Contínua (CC)** (TONOLO, 2019).

Para aumentar a potência das células, já que individualmente produzem entre 1 e 2 W, elas são conectadas em cadeias para formar unidades maiores chamadas painéis ou módulos fotovoltaicos (SETO, [s.d.]). Esses painéis, por sua vez, são conectados a inversores, que convertem a **CC** em **Corrente Alternada (CA)**. Entretanto, nesta etapa, a tensão gerada está abaixo da necessária para a rede de distribuição. Por isso, são incorporados transformadores para elevar a tensão média da **CA**, que então é distribuída por meio das linhas de transmissão, como ilustrado na **Figura 1**.

Figura 1 – Funcionamento de uma usina fotovoltaica



Fonte: Elaborada pelo autor.

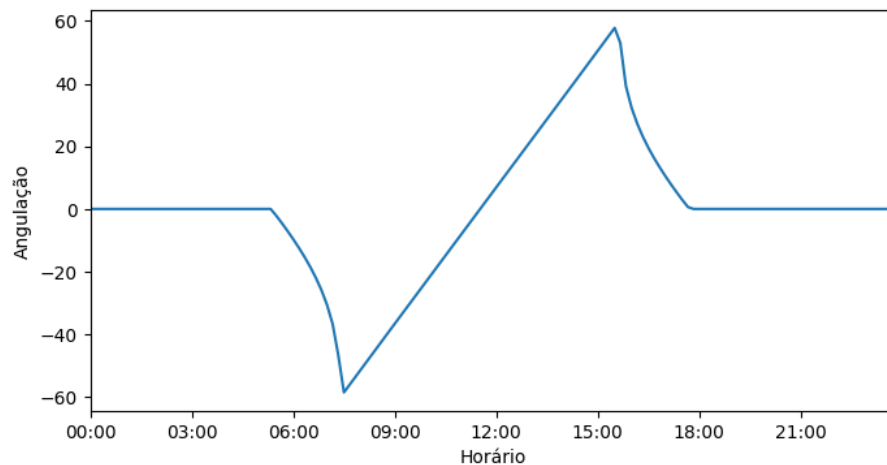
### 2.1.1 Componentes de uma usina

Conforme apontado por [Fernández et al. \(2021\)](#), uma usina solar de grande porte é constituída por centenas de módulos fotovoltaicos agrupados em série (*strings*), inversores, estruturas de sustentação (*trackers*), transformadores e sistemas de supervisão e controle.

#### 2.1.1.1 Tracker

Além de serem responsáveis pela sustentação dos módulos, os *trackers* ou rastreadores solares, possuem o papel de posicionar os mesmos, de maneira a maximizar sua exposição aos raios solares. Uma usina com *trackers* operando corretamente, ou seja, mantendo a face do módulo perpendicular aos raios solares ao longo do dia, como ilustrado na [Figura 2](#), pode aumentar seu rendimento anual em até 50% ([ALLAMEHZADEH, 2016](#)).

Figura 2 – Exemplo de uma curva de funcionamento de um *tracker*



Fonte: Elaborada pelo autor.

#### 2.1.1.2 Stringbox

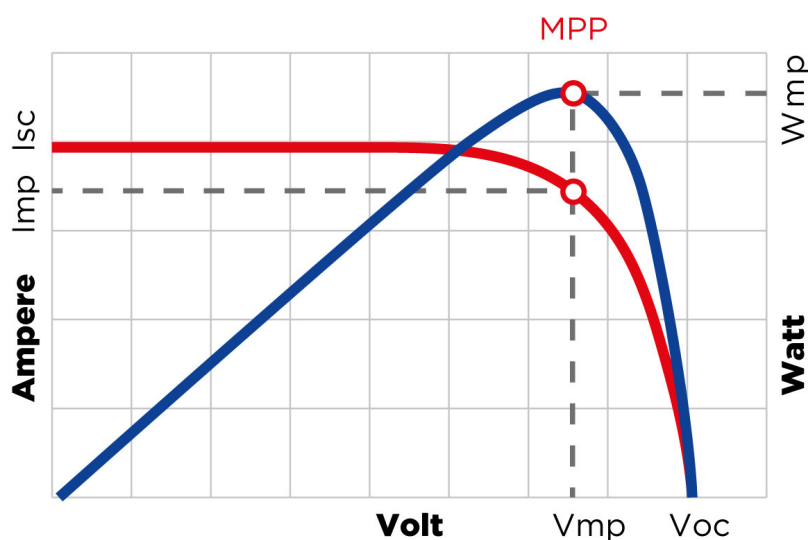
Na configuração padrão, conhecida como série-paralelo, um conjunto de *strings* é conectado em paralelo aos barramentos positivo e negativo de um componente de proteção de CC chamado *stringbox*. Este componente é composto por um quadro, chamado Dispositivo de Proteção Contra Surtos (DPS), que protege o sistema contra distúrbios causados por descargas atmosféricas, chaves seccionadoras que possibilitam o isolamento do circuito, permitindo que os operadores o manipulem com segurança, fusíveis para proteção contra sobrecorrente e curtos-circuitos, além de cabos e conectores ([FERNÁNDEZ et al., 2021](#)).

#### 2.1.1.3 Inversor

Conectado a um agrupamento de *stringboxes*, o inversor é um “(...) conversor estático de potência que converte a corrente contínua do gerador fotovoltaico em corrente

alternada apropriada para a utilização pela rede elétrica (...)” (ABNT, 2013). Dentre os modos de operação dos inversores, o mais comum é o baseado no *Maximum Power Point Tracking* (MPPT). Utilizando algoritmos que consideram variações de irradiação solar e temperatura, o dispositivo rastreia o ponto de maior tensão e corrente elétrica em cada momento (SCARABELOT, 2020). Na prática, como mostra a Figura 3, o inversor monitora a tensão e a corrente dos módulos solares, ajustando-se para maximizar a produção de energia. Em condições variáveis de luz ou temperatura, o MPPT adapta esses parâmetros, assegurando a máxima eficiência do sistema.

Figura 3 – Gráfico MPPT



Fonte: Nastec ([s.d.])

### 2.1.2 Cenário da geração fotovoltaica no mundo

Com um aumento de 26% em 2022, a produção fotovoltaica registrou o maior crescimento absoluto de geração entre todas as fontes renováveis, totalizando quase 1.300 TWh (IEA, 2023a). A liderança desse avanço é atribuída principalmente à China, com uma geração de cerca de 393 GW, seguida pelos Estados Unidos da América, com 113 GW, e o Japão, com 78,8 GW. Mesmo não estando entre os primeiros colocados, o Brasil conseguiu alcançar a oitava posição um ano após entrar no ranque dos dez maiores produtores, com 24 GW (BRASIL, 2023).

## 2.2 Análise e processamento dos dados em sistemas fotovoltaicos

Considerando que a informação presente em uma usina é representada por um conjunto de séries temporais, é necessário que haja um tratamento entre as etapas de coleta e consumo dos dados, de forma a evitar que leituras errôneas prejudiquem o restante da análise.

### 2.2.1 Séries temporais

Uma série temporal é definida como um conjunto de observações, ordenadas no tempo e registradas periodicamente, sobre uma variável. Portanto, a análise de seu gráfico é usada para identificar padrões não aleatórios, uma vez que comportamentos observados no eixo Y se relacionam com o incremento do tempo no eixo X (REIS, 2023).

De acordo com o modelo clássico, todas as séries temporais são compostas por quatro padrões (REIS, 2023). O primeiro deles, a tendência, é definido como o comportamento da série a longo prazo, podendo ser representado pelo comportamento da curva de geração de uma usina.

O segundo padrão é a variação cíclica, composta por flutuações periódicas nos valores da variável, com duração superior a um ano. Essas variações podem ser resultado de fenômenos climáticos como o El Niño. Em contrapartida, há as variações sazonais, que, ao contrário das cíclicas, possuem duração inferior a um ano e se repetem anualmente. Isso pode ser observado, por exemplo, pela dicotomia entre as estações de verão e inverno em usinas localizadas no estado de Minas Gerais, onde há um verão chuvoso e um inverno seco.

Por fim, há as variações irregulares, que são flutuações aleatórias resultantes de fatores externos, como a presença de nuvens cobrindo a usina ao longo do dia.

### 2.2.2 Processamento dos sensores

Dentre todos os sensores presentes em uma usina solar de grande porte, destacam-se os piranômetros, a potência gerada, o *setpoint* e a angulação dos *trackers*.

#### 2.2.2.1 Piranômetros e potência

Os piranômetros são sensores utilizados para medir a radiação solar incidente e, em parques solares, são usados para medir dois tipos distintos de irradiância: a *Global Horizontal Irradiance (GHI)* e a *Global Tilted Irradiance (GTI)*. A GHI é medida por piranômetros acoplados a painéis fixados no ângulo de  $0^\circ$ , permitindo verificar o comportamento do Sol ao longo do dia. Já a GTI é a medida que de fato impacta na geração da usina, uma vez que seus sensores estão presentes em painéis móveis que seguem a trajetória do Sol, mantendo-se perpendiculares a ele.

Como essas medições são suscetíveis a influências externas, gerando oscilações na série temporal, é possível realizar uma média móvel para suavizar as curvas. Além disso, é importante aplicar um filtro para eliminar possíveis erros de leitura em períodos em que não há incidência solar.

Quanto a potência, ela pode ser medida no lado CC ou CA, dependendo de ser

calculada antes ou depois dos inversores. E, assim como os piranômetros, ela também sofre influência de fatores externos, permitindo que o mesmo processo de análise anteriormente mencionado seja repetido para este sensor.

### 2.2.2.2 *Setpoint*

Para manter a estabilidade da rede elétrica e garantir que a oferta corresponda à demanda, o [Operador Nacional do Sistema Elétrico \(ONS\)](#) define um valor de referência de geração para cada usina no país. Esse limiar, denominado *setpoint*, não possui um valor fixo e está altamente relacionado ao *Constrained-OFF*.

Segundo [ANEEL \(2019\)](#), o *Constrained-OFF* é a redução da geração devido a questões operacionais impostas pelo [ONS](#). Esse fenômeno pode ocorrer por diversos fatores, como a redução da capacidade da rede ou a diminuição da demanda elétrica nacional no referido intervalo de tempo.

Uma vez que a redução do *setpoint* limita a geração da usina, dias com ocorrência de *Constrained-OFF* precisam ser removidos das análises de perda. Isso é necessário porque essa redução pode ser confundida com possíveis falhas operacionais, distorcendo os resultados das análises.

### 2.2.2.3 *Angulação dos trackers*

Como abordado na Subseção 2.1.1.1, a angulação dos *trackers* é ajustada conforme a posição do Sol ao longo do dia. No entanto, alguns *trackers* podem acabar travados em uma posição específica, provocando uma queda na geração da usina. Para ilustrar, um *tracker* travado na posição de  $-60^\circ$  tende a deixar o módulo acoplado a ele em uma posição sombreada no período da tarde, reduzindo drasticamente a captação da irradiância solar e, conseqüentemente, a eficiência da geração de energia.

## 2.3 Simulação de potência em sistemas fotovoltaicos

Como dados de geração em usinas de larga escala são informações frequentemente confidenciais, especialmente porque as empresas que operam essas usinas estão, em sua maioria, listadas na bolsa de valores, o laboratório de pesquisa e desenvolvimento da Administração Nacional de Segurança Nuclear do Departamento de Energia dos Estados Unidos, [Sandia National Laboratories](#), desenvolveu em 2013 um conjunto de ferramentas de código aberto, denominada **pvlib** ([ANDERSON et al., 2023](#)).

O **pvlib** é uma biblioteca Python que busca simular o desempenho de sistemas e componentes em usinas solares. Baseando-se em mais de 100 modelos físicos e empíricos da literatura científica, incluindo algoritmos de posicionamento solar, modelos de irradiância,

temperatura e elétricos, essa biblioteca oferece desde implementações individuais de baixo nível até fluxos de alto nível que conectam os modelos completos de sistemas fotovoltaicos. Além disso, o **pvlib** permite buscar conjuntos de dados meteorológicos para auxiliar o desenvolvedor na modelagem (ANDERSON et al., 2023).

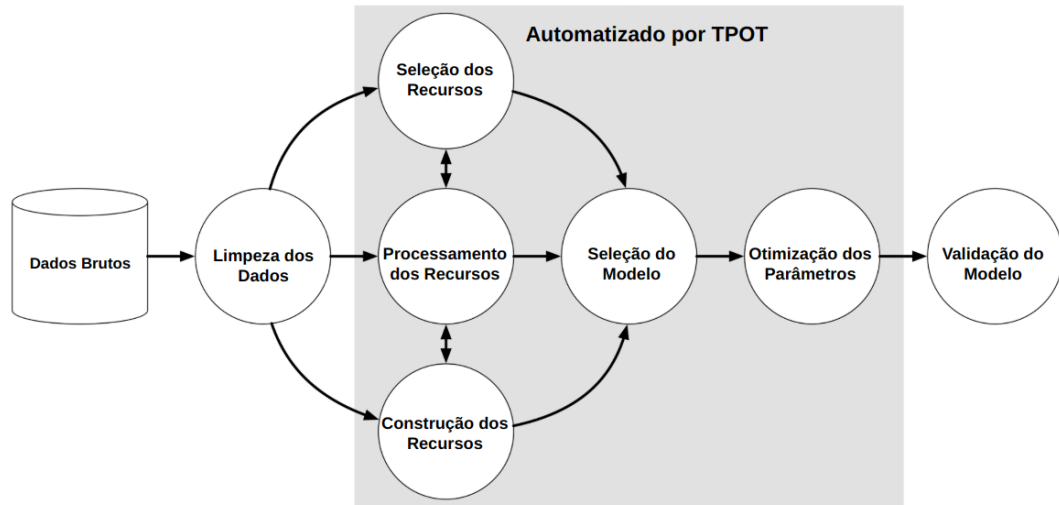
Com o objetivo de desenvolver uma biblioteca mais granular e específica, as instituições de ensino e pesquisa Instituto Federal de Santa Catarina (IFSC) e Universidade Federal de Santa Catarina (UFSC), juntamente com a empresa ENGIE, formaram uma parceria. Essa biblioteca, que futuramente será disponibilizada com uma licença BSD-3 e cujo nome ainda não está definido, visa permitir a simulação de geração de uma usina fotovoltaica, incorporando diferentes tipos de falha.

Por exemplo, com ela seria possível visualizar a geração em um cenário onde metade dos módulos solares estão acoplados a *trackers* travados na posição  $-60^\circ$  e um terço desses módulos está sujo. Simultaneamente, na parte com *trackers* funcionais, poderia haver um problema no barramento dos módulos. Todos esses são defeitos plausíveis em usinas fotovoltaicas de grande porte e a capacidade de estimar as perdas causadas por cada uma dessas falhas, ou por várias ocorrendo simultaneamente, aprimoraria a precisão das análises futuras.

## 2.4 AutoML

O *machine learning* automatizado, também conhecido como AutoML, é o processo de automatizar tarefas iterativas no desenvolvimento de modelos de ML. Com ele, processos como limpeza de dados, seleção de recursos, escolha de modelos e análise de resultados são realizados de forma automática, eliminando a necessidade de intervenção constante do desenvolvedor (IBM, [s.d.]).

Para isso, a ferramenta separa os dados de forma aleatória em dois conjuntos: treinamento e teste, onde o conjunto de treinamento possui um tamanho superior. Em seguida, ela cria múltiplos *pipelines* em paralelo, cada um com diferentes parâmetros e algoritmos, treinando-os com base no primeiro conjunto. Após a execução, esses *pipelines* recebem uma pontuação de treinamento e os que obtêm os melhores resultados são ajustados com novos parâmetros para uma nova rodada de treinamento. Esse processo é repetido até que os critérios de saída da predição estejam próximos dos valores de teste, resultando em um modelo final, como pode ser conferido na Figura 4. Outro ponto importante é que, devido à popularidade dessa técnica, diversas bibliotecas foram desenvolvidas, como TPOT, AutoKeras, H2O, entre outras.

Figura 4 – *Machine Learning Pipeline*

Fonte: Adaptado de TPOT ([s.d.])

## 2.5 Ferramentas utilizadas

Neste trabalho, serão utilizadas ferramentas para facilitar e gerenciar a implementação da proposta, cada uma delas sendo associada a uma necessidade específica da estrutura do projeto.

### 2.5.1 *Docker*

Promovendo maior flexibilidade e desempenho quase nativo em comparação com *Virtual Machines* (VMs) (COMBE; MARTIN; PIETRO, 2016), o uso de contêineres tem se tornado cada vez mais necessário no mundo do desenvolvimento, pois a manutenção e atualização do *software* se tornam muito mais simples e menos custosas.

Os contêineres, como o próprio nome sugere, são ambientes isolados utilizados para empacotar aplicações. Cada contêiner contém as ferramentas essenciais para executar a aplicação a que está associado, garantindo que ela funcione corretamente mesmo quando o código é executado em um ambiente diferente, como em uma nova máquina.

Para gerar um contêiner, é preciso ter um arquivo executável contendo todas as bibliotecas, dependências e arquivos necessários para sua execução. Esse arquivo é chamado de imagem e pode ser baixado do site [Docker Hub](#) ou criado a partir de uma imagem modificada.

Por serem ambientes isolados, é possível ter diversos contêineres em um mesmo computador, cada um utilizando versões diferentes da mesma biblioteca. Ou seja, é possível ter várias aplicações distintas, cada uma utilizando uma versão diferente da biblioteca **pandas** do Python, sem que uma interfira na outra ou mesmo que o dispositivo precise ter um interpretador da linguagem instalado.

Assim como um navio cargueiro é capaz de transportar múltiplas cargas, uma máquina pode possuir diferentes contêineres. Por conta disso, é necessário haver uma ferramenta para gerenciá-los, e assim surgiu o *Docker*. O *Docker* implementa uma camada de virtualização nos contêineres, realizando o mesmo papel de uma *VM*, porém, como só contém os dados necessários para rodar o programa, é mais leve e utiliza menos recursos. Além disso, ele gera uma interface de comunicação entre o contêiner e o sistema hospedeiro, controlando o uso de recursos como RAM, processamento e *bitrate*.

### 2.5.2 *Jupyter Notebook*

Na parte de *ML*, será adotada a linguagem de programação Python, utilizando o Ambiente de Desenvolvimento Integrado (IDE) *Jupyter Notebook*. Com ele, é possível criar documentos interativos compostos por blocos de código, textos formatados, imagens, gráficos e equações matemáticas. Além disso, cada bloco de código funciona de forma independente, permitindo que, uma vez executado, seu processo fique armazenado em memória, evitando a necessidade de reexecutá-lo.

Devido a essa funcionalidade, processos lentos e custosos de *ML* podem ser rodados apenas uma vez, permitindo que diferentes abordagens, pautadas nos resultados do treinamento, sejam realizadas de forma mais eficiente.

### 2.5.3 TPOT

Dentre as bibliotecas voltadas para o AutoML, a que será usada neste trabalho é a TPOT. Ela é uma biblioteca baseada no scikit-learn, cujo objetivo é automatizar o processo de seleção do melhor modelo de *ML*, simplificando a busca pelas características e parâmetros ideais para o conjunto de dados analisado (TPOT, [s.d.]).

Por usar algoritmos de pesquisa estocástica, o TPOT se baseia na aleatoriedade para a otimização de recursos e modelos de aprendizado de máquina. Dessa forma, sua execução pode levar de algumas horas a até mesmo dias, dependendo do tamanho e complexidade dos dados que estão sendo trabalhados.

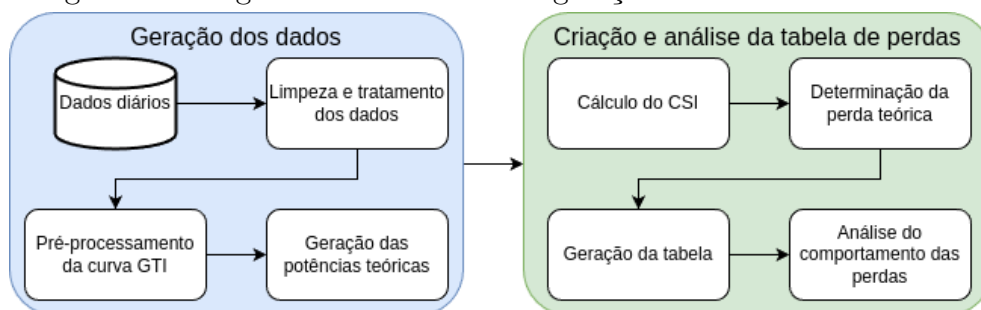
### 2.5.4 *Streamlit*

Para a visualização dos resultados, será empregada a biblioteca Python, *Streamlit*. Essa ferramenta permite que o usuário insira dados, execute modelos e avalie os resultados em uma página *web*, sem necessitar de conhecimentos prévios em programação *front-end*. Além disso, ela possui *widgets* integrados que simplificam a criação de painéis interativos e ferramentas de exploração de dados, ideais para o desenvolvimento baseado em *ML* (SARANGPURE et al., 2023).

# 3 GERAÇÃO E TRATAMENTO DOS DADOS

A metodologia proposta será dividida em quatro partes principais: a geração dos dados, a criação e análise de uma tabela de perdas, a subsequente utilização dessa tabela para modelagem da regressão e a aplicação da metodologia. As primeiras duas partes, bem como suas etapas de desenvolvimento, podem ser conferidas na [Figura 5](#).

Figura 5 – Diagrama da estrutura da geração e tratamento dos dados



Fonte: Elaborada pelo autor.

Vale ressaltar que serão consideradas três classificações de dados ao longo deste capítulo: real, teórico e estimado. Os dados reais são obtidos por meio de medições realizadas por sensores nas usinas. O dado teórico é calculado por meio de equações matemáticas, como por exemplo a potência obtida por meio da biblioteca em desenvolvimento mencionada na [Seção 2.3](#), que utiliza modelos e características de usinas para realizar os cálculos. Por fim, os dados estimados são resultados de previsões geradas por bibliotecas de AutoML.

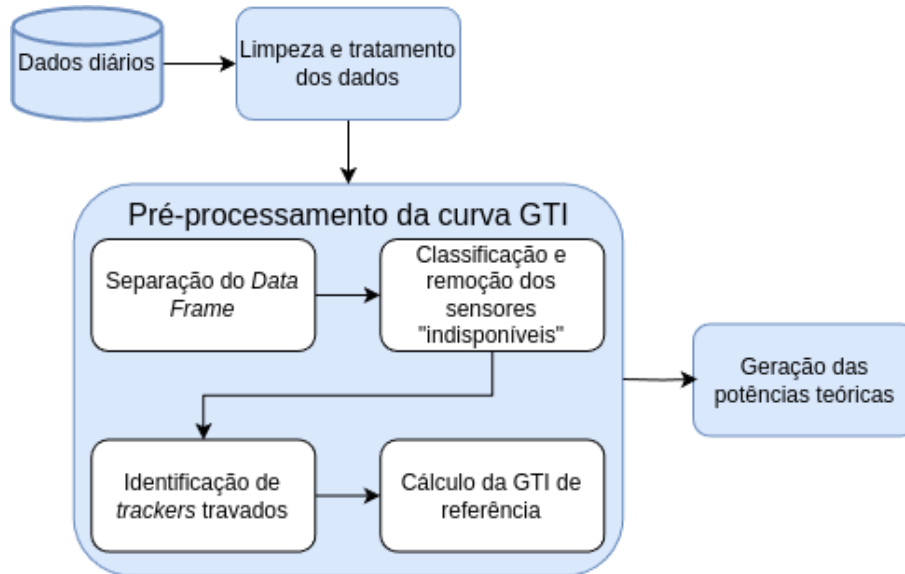
## 3.1 Geração dos dados

Como o principal objetivo deste projeto é estimar as perdas causadas por *trackers* que deixaram de seguir a posição do Sol, a primeira parte envolve a coleta e o tratamento das variáveis, seguido pelo pré-processamento da curva GTI e por fim, na geração da potência teórica para as usinas analisadas, conforme ilustrado na [Figura 6](#).

### 3.1.1 Limpeza e tratamento dos dados

Para evitar inconsistências nas análises e modelagens futuras, é necessário realizar um pré-processamento dos dados, o que inclui os passos de limpeza, filtragem e organização dos mesmos.

Figura 6 – Diagrama da geração dos dados



Fonte: Elaborada pelo autor.

Inicialmente, as variáveis GTI, GHI, velocidade do vento e temperatura ambiente dos módulos passam por um processo de limpeza para remoção de dados ausentes ou inválidos, resultando na exclusão de 22 dias ao longo do intervalo de 1 ano analisado. Adicionalmente, aplica-se uma média móvel para suavizar as curvas de irradiância, mitigando as variações causadas por ruídos atmosféricos e pela presença de nuvens.

Posteriormente, as irradiâncias são filtradas para zerar valores fora do período de incidência solar, removendo possíveis medições errôneas. Esse processo utiliza a função “get\_clearsky” da biblioteca **pvl** (JENSEN et al., 2023), que, com base em latitude, longitude, altitude e fuso horário do parque solar, calcula a curva GHI ideal, permitindo estimar o nascer e pôr do Sol. Assim, todas as amostras fora desses horários são zeradas.

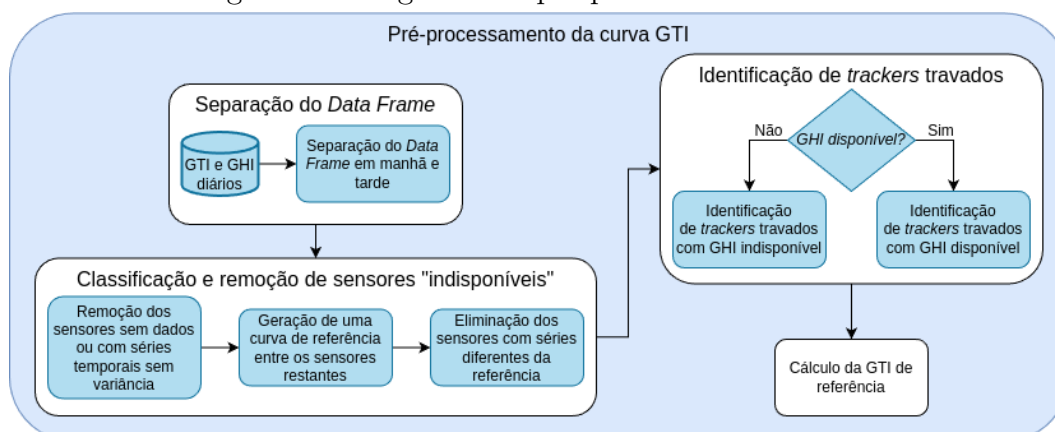
### 3.1.2 Pré-processamento da curva GTI

Em uma usina fotovoltaica de grande porte, a presença de dezenas de piranômetros espalhados por toda a extensão pode gerar discrepâncias nas medições. Esses erros podem ocorrer devido a falhas nos sensores, *trackers* travados em determinada angulação, ou diferenças nas leituras causadas por nuvens localizadas. Para evitar com que esses problemas sejam incorporados no cálculo da geração do parque, é necessário identificar e descartar equipamentos que apresentam este comportamento. O diagrama desta etapa, pode ser visualizado na Figura 7.

#### 3.1.2.1 Separação do *Data Frame*

Para aprimorar a precisão e a modularidade na classificação dos piranômetros, considerando que as variações nas suas séries temporais se acentuam pela manhã e à

Figura 7 – Diagrama do pré-processamento da GTI



Fonte: Elaborada pelo autor.

tarde, a análise foi segmentada nesses dois períodos. Com isso, sensores que apresentem problemas em apenas um dos intervalos ainda podem ser considerados válidos no outro.

Tendo em vista que o período de incidência solar varia ao longo dos dias, definir um horário fixo para seu início e fim poderia resultar em perda ou adição desnecessária de dados à análise. Para evitar isso, os horários de início da manhã e de fim da tarde são calculados ao fornecer a localização geográfica da usina para a biblioteca **pvl**. Quanto aos limites restantes, ambos são valores fixos, estabelecidos para ocorrer em torno das 12:00 horas.

### 3.1.2.2 Classificação e remoção de sensores “indisponíveis”

A remoção dos piranômetros com erro de leitura ocorre em dois passos: remoção dos sensores sem dados ou com baixa dispersão em torno do período e remoção daqueles cujo comportamento difere de um valor base. No primeiro passo, verifica-se se algum dos dispositivos analisados apresenta dados faltantes. Por exemplo, considerando o período da manhã, com início às 06:00 horas e uma taxa amostral de 10 minutos, a quantidade de leituras realizadas por cada sensor, deve ser igual a 37. Caso um piranômetro não atenda a esse critério, ele é classificado como “indisponível” e removido da análise.

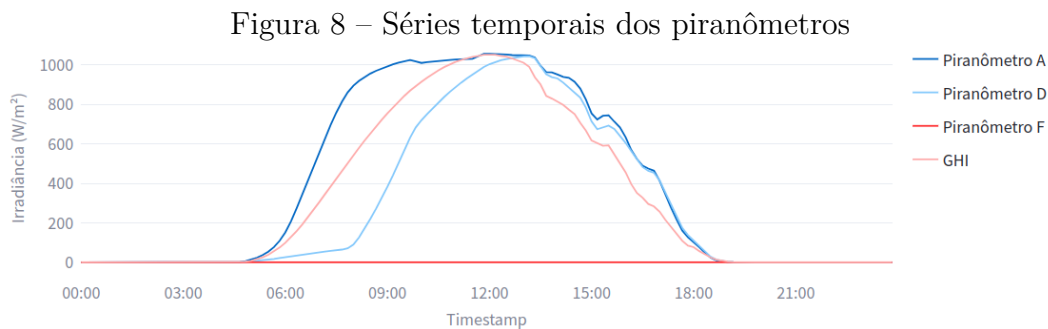
Segundo Bland e Altman (1996), um desvio padrão baixo indica que os valores de um conjunto de dados estão próximos da média, demonstrando menor dispersão e maior concentração. Desse modo, o desvio padrão ponderado pelo tempo é comparado a um valor de referência. Caso ele se mostre inferior a esse valor, há uma indicação de que o sensor apresenta um comportamento menos variável do que o esperado ou até constante, sinalizando problemas nas medições.

No próximo passo, uma curva teórica é gerada com base nos sensores não descartados para cada intervalo de tempo. Foram consideradas duas opções para estabelecer

essa série: a média e a mediana. Contudo, em dias ensolarados, as leituras associadas a sensores acoplados a *trackers* funcionais podem apresentar uma diferença significativa em relação àquelas de dispositivos travados em  $0^\circ$ . Nesse cenário, a mediana tenderia a comportamentos semelhantes aos sensores do primeiro caso, o que poderia comprometer o processo de filtragem. Por essa razão, optou-se pela média, uma vez que ela representa melhor a concentração dos dados de uma distribuição, servindo como ponto de equilíbrio das frequências (CARVALHO, 2011).

Posteriormente, o comportamento de cada piranômetro é comparado com a curva de referência gerada. Caso as séries temporais de algum sensor se afastem significativamente da média ou apresentem um comportamento que não reflita o padrão observado, o dispositivo é classificado como “indisponível” e descartado da análise.

Um exemplo da aplicação desses dois passos pode ser observado na Figura 8. Nota-se que o “Piranômetro F”, representado pela curva vermelha, manteve suas medições em  $0 \text{ W/m}^2$  durante os dois períodos analisados, indicando a ausência de leituras ao longo do dia. Por sua vez, o “Piranômetro D” (curva azul clara), embora tenha realizado medições, apresenta um comportamento distinto na parte da manhã, mantendo-se abaixo da GHI durante todo o período.



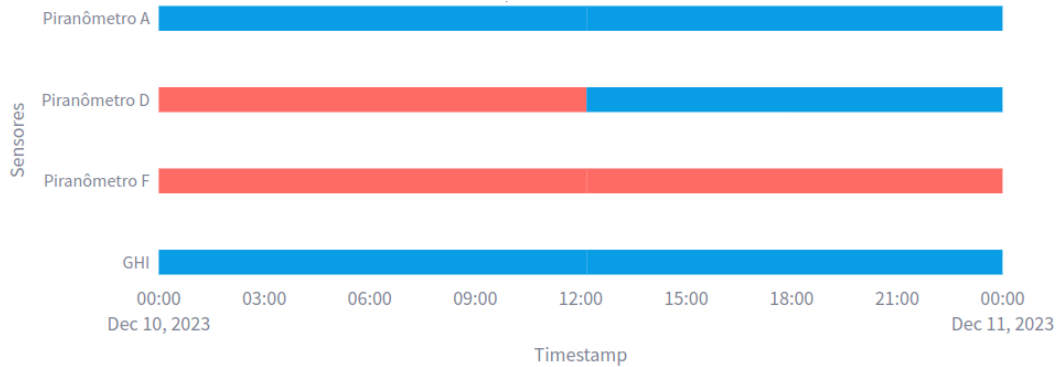
Fonte: Elaborada pelo autor.

O reflexo das observações feitas anteriormente pode ser visualizado na Figura 9. Nesse gráfico de Gantt, os piranômetros “D” e “F” possuem barras vermelhas que indicam sua indisponibilidade. O piranômetro “D” apresenta essa condição apenas no período da manhã, enquanto o “F” permanece indisponível durante o dia inteiro. Em contraste, os demais sensores estão disponíveis e são representados pelas barras azuis no gráfico.

### 3.1.2.3 Identificação de *trackers* travados

Conforme discutido na Subseção 2.1.1.1, os rastreadores solares desempenham a função de posicionar os módulos fotovoltaicos de forma a otimizar sua exposição à luz solar. Em dias sem a presença de nuvens, os piranômetros acoplados a *trackers* em funcionamento geralmente atingem seus valores máximos nos primeiros momentos da manhã e mantêm esses níveis até o final da tarde. Em contraste, sensores acoplados a *trackers*

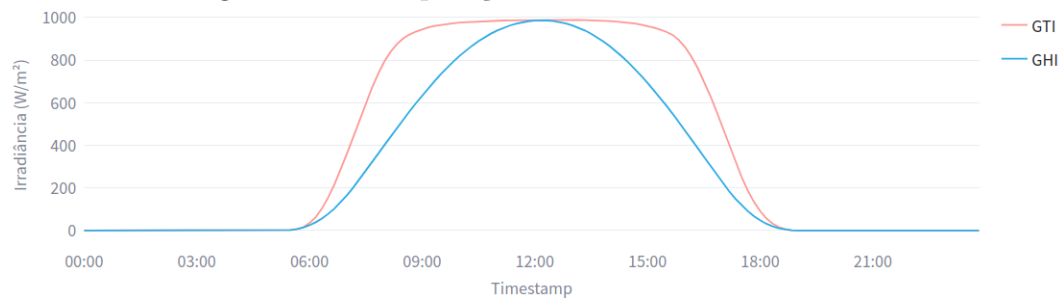
Figura 9 – Gráfico de Gantt dos piranômetros



Fonte: Elaborada pelo autor.

travados em  $0^\circ$ , como os responsáveis pela medição da GHI, apresentam um comportamento com um aumento gradual pela manhã, um pico próximo ao meio-dia, seguido de uma queda igualmente gradual no período da tarde. Essa diferença de comportamento pode ser observada na Figura 10.

Figura 10 – Comparação entre as curvas GTI e GHI



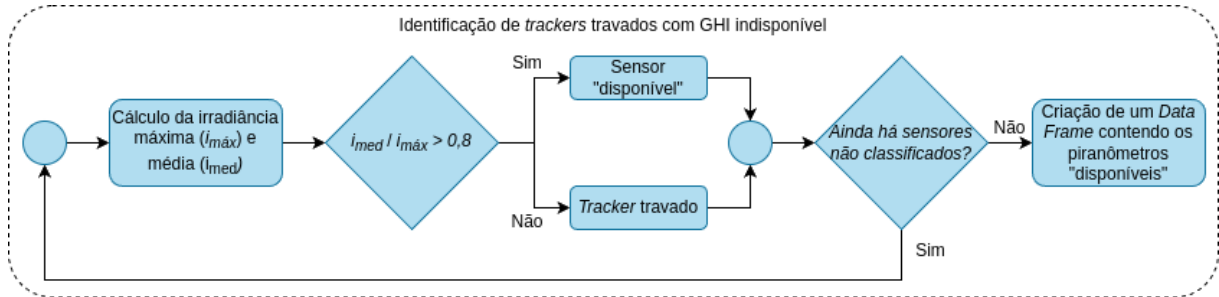
Fonte: Elaborada pelo autor.

Levando em consideração que dispositivos de medição acoplados a *trackers* travados podem introduzir erros no cálculo da geração teórica do parque solar, é necessário identificá-los e removê-los da análise. Para isso, o início desse passo envolve a avaliação da integridade da GHI, utilizando o método do desvio padrão ponderado pelo tempo, similar ao processo descrito na Subseção 3.1.2.2.

Em cenários onde a GHI não está disponível, a identificação de *trackers* travados é realizada por meio da relação entre a média e o valor máximo das leituras de cada sensor, conforme Figura 11. Quando essa relação excede o valor de 0,8, o rastreador é classificado como “disponível”. Esse limiar foi definido com base em análises heurísticas, que indicaram que, para rastreadores travados em  $0^\circ$ , essa relação tende a apresentar valores abaixo desse limite.

Um exemplo desse método pode ser visualizado na Figura 12, onde as curvas da GHI (verde claro) e do “Piranômetro I” (verde escuro) não apresentaram medições durante todo o dia. Vale destacar que a curva verde escuro não aparece, pois está sobreposta pela

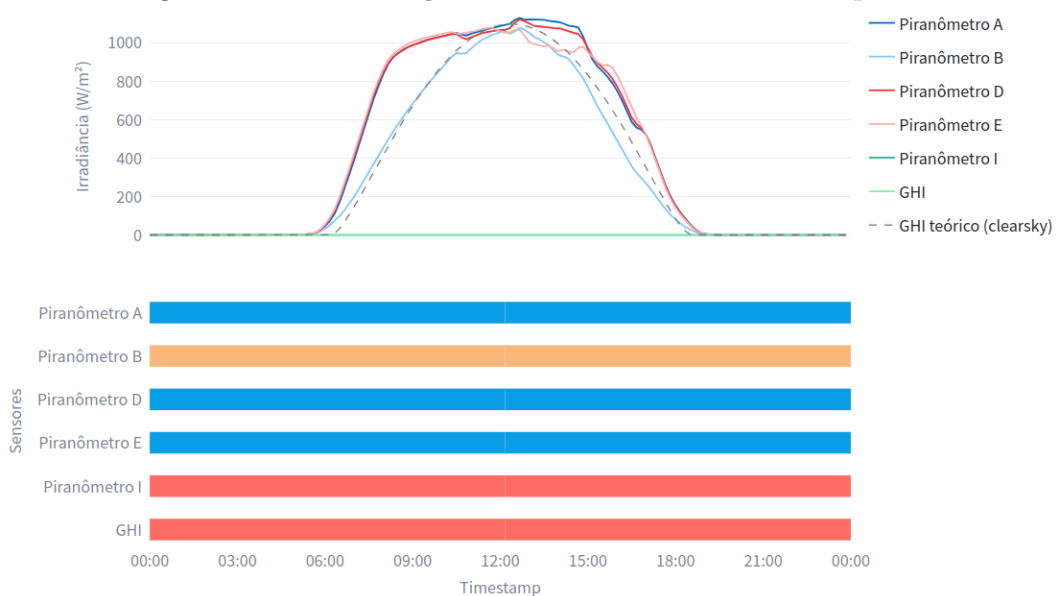
Figura 11 – Diagrama da identificação de *trackers* travados em dias com GHI indisponível



Fonte: Elaborada pelo autor.

verde claro ao longo de todo o período. Enquanto isso, os piranômetros “A”, “D” e “E”, passaram por todos os critérios de avaliação e foram classificados como “disponível”. P

Figura 12 – Classificação dos sensores com GHI indisponível



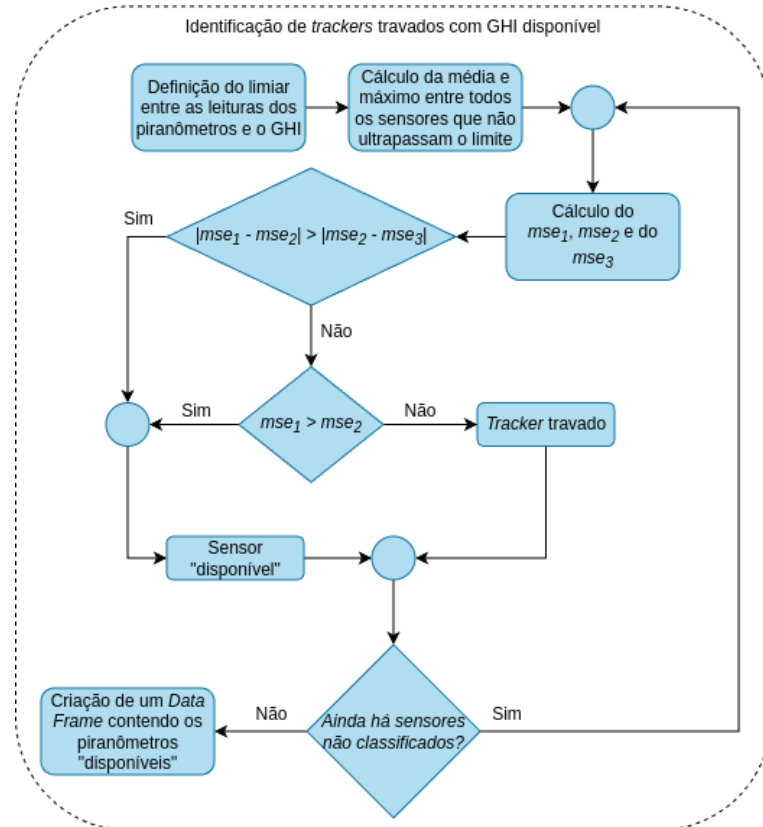
Fonte: Elaborada pelo autor.

Como pode ser observado, o “Piranômetro B” (azul claro) foi classificado como “*stow*” (acoplado a um *tracker* travado em  $0^\circ$ ), representado pela barra laranja no gráfico de disponibilidade. Diferente dos sensores “A”, “D” e “E”, que passaram por todos os critérios de avaliação, ele não apresenta picos pela manhã ou tarde, uma característica típica de dispositivos que ficaram perpendiculares à irradiância solar. Outro fator que reforça sua classificação é a semelhança com a série temporal tracejada em preto, que representa o valor de uma GHI ideal que não considera a presença de nuvens, obtida com a biblioteca **pvl**.

Nos casos em que a GHI está disponível, sua inclusão no método de classificação dos *trackers* travados tende a aumentar a precisão da identificação. Este passo é separado em dois subprocessos, conforme ilustrado pela Figura 13. No primeiro, é definido um

limiar máximo de proximidade entre a série temporal dos piranômetros e a curva GHI. Sensores que não ultrapassem esse limite são usados para calcular duas curvas teóricas: a média e o valor máximo amostra a amostra, para cada intervalo de tempo analisado.

Figura 13 – Diagrama da identificação de *trackers* travados em dias com GHI disponível



Fonte: Elaborada pelo autor.

Com as curvas definidas, o próximo subprocesso consiste em relacioná-las de modo a classificar o sensor. Para isso, foi escolhido o **Erro Quadrático Médio (EQM)**, ou *Mean Squared Error (MSE)* em inglês, uma métrica estatística que quantifica a diferença entre os valores previstos por um modelo e os valores observados (TAUSIESAKUL; ASAVAS-KULKIET, 2023). No contexto da classificação dos dispositivos, o EQM será utilizado para calcular as distâncias entre as séries temporais. Esse cálculo será aplicado a três conjuntos de dados. O primeiro, chamado de  $mse_1$ , baseia-se na relação entre o piranômetro analisado e a curva GHI, refletindo o quão próximo o dispositivo está dessa referência. Um valor baixo de  $mse_1$  indica que as medições estão alinhadas com as registradas pela GHI.

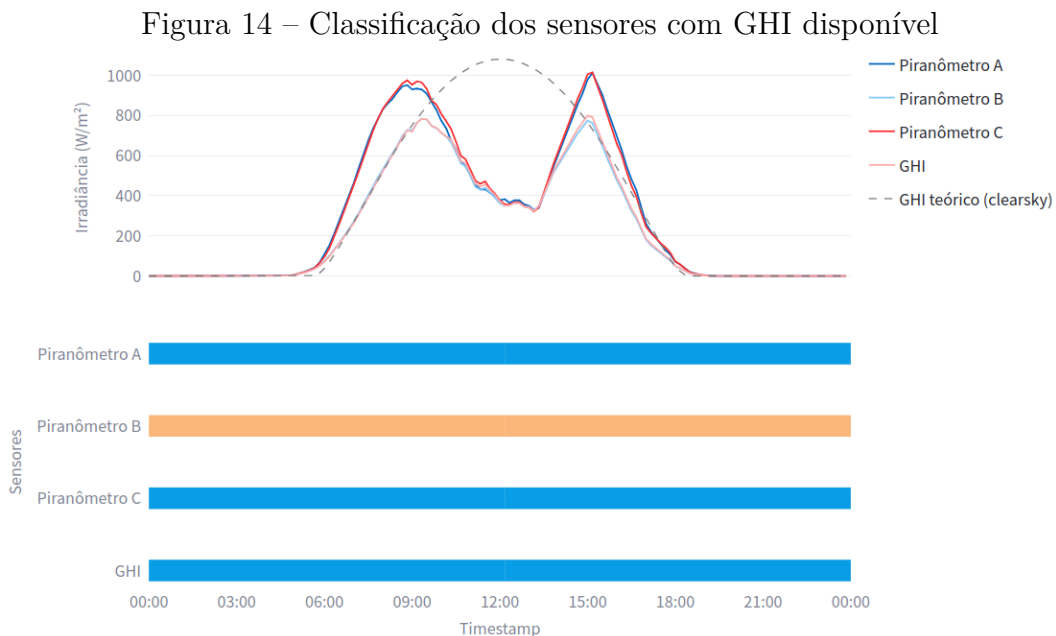
O segundo conjunto,  $mse_2$ , é o EQM calculado entre o dispositivo e o valor máximo de irradiância medido entre todos os piranômetros. Um  $mse_2$  baixo indica que o equipamento está registrando valores altos de irradiância, sugerindo que ele pode estar exposto a condições ideais. Já o terceiro conjunto,  $mse_3$ , avalia a diferença entre o piranômetro e a média das medições de irradiância de todos os dispositivos. Um  $mse_3$  baixo indica que

o equipamento está alinhado com a maioria dos outros, enquanto um  $mse_3$  alto sugere uma discrepância em relação à média.

Para que um dispositivo seja considerado “disponível”, ele precisa atender uma entre duas condições. Em uma delas, a diferença absoluta entre o  $mse_1$  e o  $mse_3$  é comparada com a diferença absoluta entre o  $mse_2$  e o  $mse_3$ . Se a primeira diferença for superior a segunda, então o equipamento está mais alinhado com a média ou o valor máximo do que com a GHI, indicando que tanto ele quanto o seu dispositivo de sustentação estão operando corretamente.

A outra opção aplica-se a cenários em que apenas um piranômetro está funcional. Nesse caso, tanto o valor máximo quanto a média seriam iguais às leituras desse dispositivo, fazendo com que a diferença entre o  $mse_2$  e o  $mse_3$  seja sempre zero, classificando-o erroneamente como “stow”. Para evitar esse erro, é verificado se o  $mse_1$  é maior que o  $mse_2$ . Como o  $mse_2$  será igual a zero nesse cenário, essa condição será verdadeira.

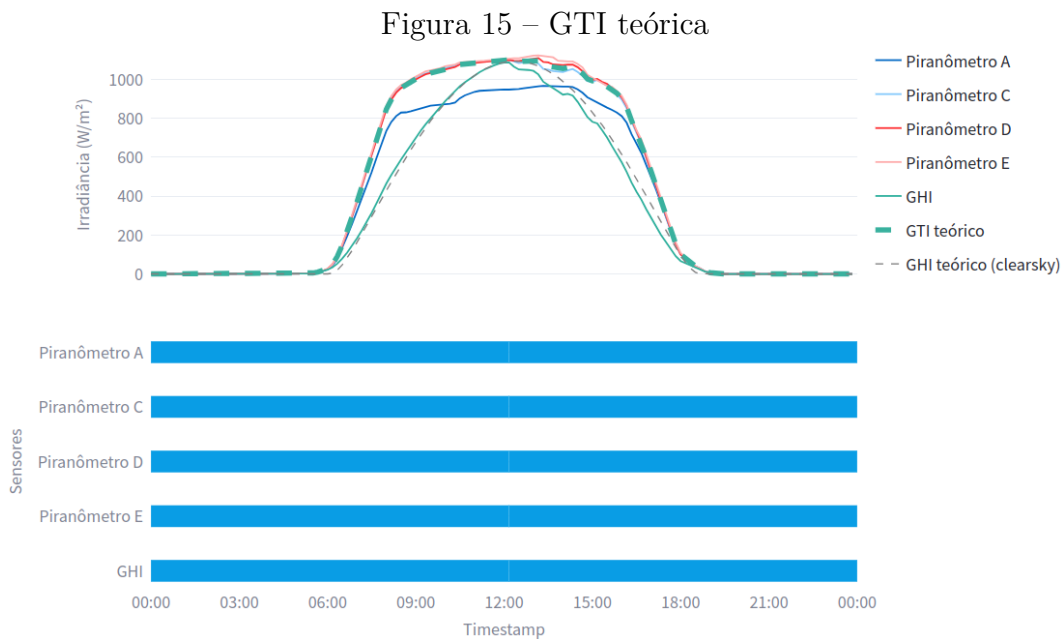
Um exemplo da classificação em dias com GHI pode ser visualizado na Figura 14. Nela, a série temporal do “Piranômetro B” (verde) é praticamente idêntica ao GHI, o que indica que, apesar de o sensor estar funcional, seu *tracker* está travado em  $0^\circ$ . Uma observação importante sobre esse dia é a queda nas medições de todos os dispositivos por volta das 12:00. Esse comportamento é típico da presença de nuvens localizadas sobre os medidores.



Fonte: Elaborada pelo autor.

### 3.1.2.4 Cálculo da GTI de referência

Após a classificação dos sensores, o próximo passo é gerar uma série temporal que sintetize as leituras dos piranômetros funcionais acoplados a dispositivos de sustentação que estão efetivamente acompanhando a posição solar. Essa série será utilizada para representar o comportamento da GTI de referência da usina. Novamente, duas curvas foram consideradas: a média e a mediana. No entanto, diferentemente da escolha feita na Subseção 3.1.2.2, a curva selecionada desta vez foi a mediana, pois ela representa o valor central do conjunto de dados, sendo mais imparcial e resistente a *outliers* (IEZZI; HAZZAN; DEGENSZAJN, 2013). Um exemplo dela pode ser visto na Figura 15, representada pelo tracejado verde.



Fonte: Elaborada pelo autor.

### 3.1.3 Geração das potências teóricas

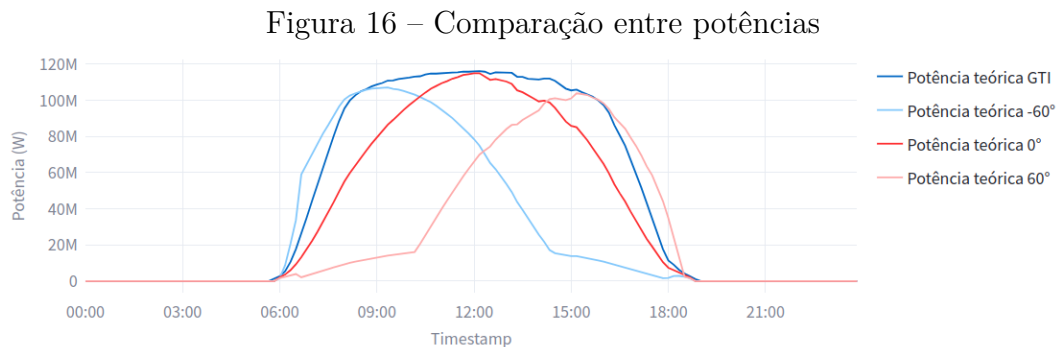
Nesta etapa, os valores de velocidade do vento, temperatura ambiente e GTI de referência, são passados como entrada para a biblioteca em desenvolvimento que, por sua vez, calcula a geração de energia da usina para as condições e período fornecidos, considerando que todos os painéis e *trackers* estão operando com as mesmas condições.

Além do cálculo da Potência Teórica Calculada Considerando a GTI ( $P_{GTI}$ ), a biblioteca mencionada na seção 2.3 permite modelar o comportamento da geração da usina em cenários onde diferentes *trackers* estão travados em posições distintas. No entanto, simular todas as possíveis combinações de travamento seria inviável. Por isso, optou-se por simplificar o modelo, assumindo que todos os *trackers* da usina estariam travados na mesma inclinação. Dessa forma, a estimativa de perda decorrente desse problema é feita

através de uma regressão considerando essa condição, e posteriormente, o valor é ajustado de acordo com a porcentagem real de dispositivos travados.

O intervalo escolhido para realizar este experimento foi de  $-60^\circ$  a  $60^\circ$ , correspondendo aos valores máximo e mínimo que os dispositivos de sustentação podem operar. No total, foram realizadas 25 simulações com *trackers* travados, com variações ocorrendo a cada  $5^\circ$ . Além delas, também foi realizada a simulação da  $P_{GTI}$  para a usina analisada.

Um exemplo dessas simulações pode ser conferido na [Figura 16](#). Nela, é possível visualizar a  $P_{GTI}$  representada pela cor azul escuro, a Potência Teórica Calculada Considerando a GHI ( $P_{GHI}$ ) em vermelho, e as potências nos limites extremos citados anteriormente. Observe que na curva azul claro (*trackers* travados em  $-60^\circ$ ), a série temporal apresenta um comportamento coerente com a curva azul escuro na parte da manhã, com um pico de geração no início deste período. No entanto, à medida que o tempo avança, devido ao dispositivo de sustentação estar travado, sua curva apresenta uma queda acentuada, atingindo o menor valor no final da tarde, quando a irradiação solar não incide mais sobre os módulos fotovoltaicos. Já a curva rosa (*trackers* travados em  $60^\circ$ ) apresenta o comportamento oposto.

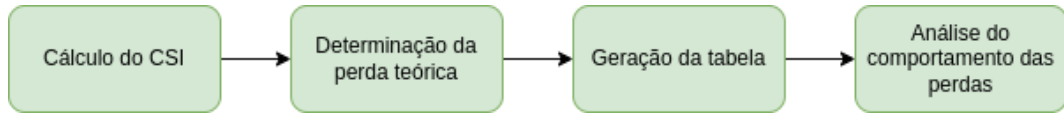


Fonte: Elaborada pelo autor.

## 3.2 Criação e análise da tabela de perdas

Após obter as potências necessárias, a próxima parte, ilustrada pela [Figura 17](#), consiste em gerar uma tabela que relacione a condição climática do dia, representada por um índice denominado *Clear Sky Index* (CSI), a angulação de travamento do *tracker*, e a Perda Teórica ( $L_t$ ) de energia. O processo será dividido em quatro partes: cálculo do CSI, determinação da  $L_t$  diária, geração da tabela e, por fim, a análise do comportamento das perdas.

Figura 17 – Diagrama da criação da tabela de perdas

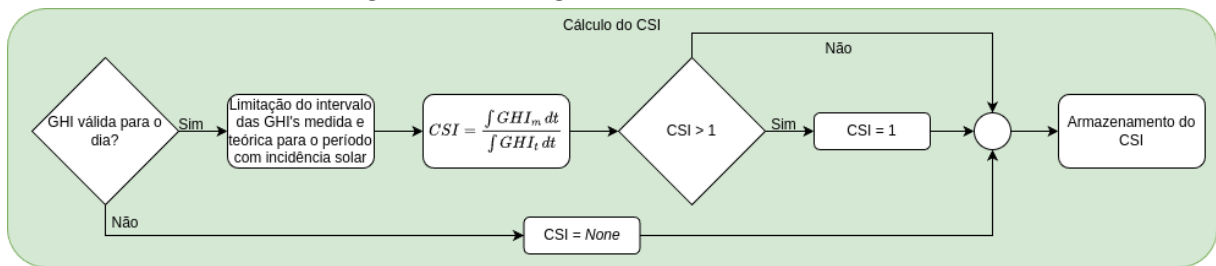


Fonte: Elaborada pelo autor.

### 3.2.1 Cálculo do CSI

Como a condição climática impacta diretamente na geração da usina, variando entre máximos em dias ensolarados e mínimos em dias extremamente nublados, essa informação não poderia ser omitida na análise de perdas. Mas, para incluí-la na tabela, é necessário mensurá-la, e para isso, foi criado o “índice de céu limpo”, ou *Clear Sky Index*, conforme Figura 18.

Figura 18 – Diagrama do cálculo do CSI



Fonte: Elaborada pelo autor.

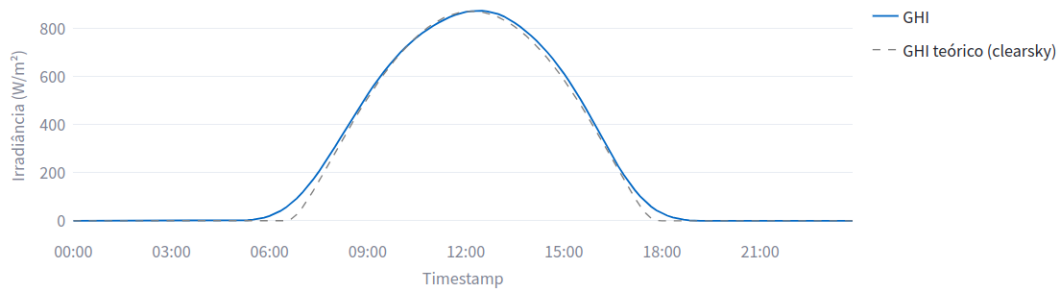
Para calcular esse índice, são necessários três elementos: a **GHI** medida na usina, sua classificação diária (“disponível” ou “indisponível”) e a **GHI** teórica gerada pela biblioteca **pvl**. A classificação é necessária porque, em dias onde algum período não apresenta valores válidos de **GHI**, o **CSI** não seria confiável, exigindo sua exclusão da análise.

Levando em conta apenas os dias com **GHI** válidos, o cálculo do **CSI** é realizado de acordo com a relação:

$$CSI = \frac{\int GHI_m dt}{\int GHI_t dt} \quad (3.1)$$

Observe que o **CSI** é um valor contínuo, a princípio, limitado entre 0 e 1. Porém, como a **GHI** teórica é uma estimativa aproximada, é possível obter valores de **CSI** superiores a 1. Um exemplo disso está na Figura 19, que apresenta um índice de 1,04. Para evitar distorções no limite superior do intervalo, esses índices são arredondados para 1.

Figura 19 – Comparação entre a GHI medida e a teórica

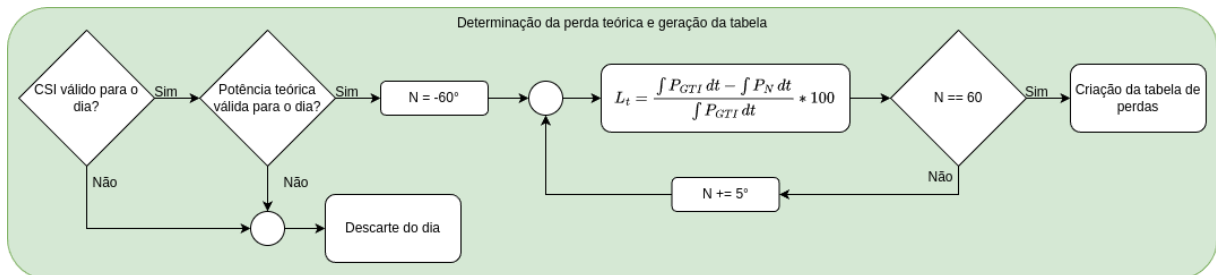


Fonte: Elaborada pelo autor.

### 3.2.2 Determinação da perda teórica e geração da tabela

Após a obtenção do CSI, a próxima etapa consiste em calcular a  $L_t$  correspondente a cada um dos 25 ângulos selecionados anteriormente, conforme Figura 20. No entanto, antes desse cálculo, é necessário verificar se a  $P_{GTI}$  é válida para o dia em questão. Isso é feito removendo dias em que, em pelo menos um dos períodos analisados, não havia nenhum piranômetro classificado como “disponível”.

Figura 20 – Diagrama do cálculo da perda teórica e geração da tabela



Fonte: Elaborada pelo autor.

Com a garantia de que a potência pode ser utilizada, a  $L_t$  é então calculada como:

$$L_t = \frac{\int P_{GTI} dt - \int P_N dt}{\int P_{GTI} dt} \cdot 100 \quad (3.2)$$

onde “N” representa o ângulo em que o *tracker* está travado.

Com base nas informações reunidas até o momento, é possível gerar uma tabela que relaciona o dia, o CSI, a angulação de travamento dos *trackers*, e a  $L_t$  associada a esse problema. A Tabela 1 apresenta um exemplo para um dia com CSI 0,9, com os ângulos variando em intervalos de 20°. Observe que, à medida que os ângulos se aproximam dos extremos (limites de operação dos *trackers*), há um aumento na  $L_t$ , uma vez que os módulos fotovoltaicos gradualmente deixam de captar uma parcela da irradiância solar.

Entretanto, ao gerar a Tabela 2, que retrata um dia com CSI baixo, o aumento percentual visto anteriormente não se mantém. Nesse cenário específico, as angulações

Data	CSI	Angulação (°)	Perda teórica (%)
03/12/2023	0,9	-60	42,84
03/12/2023	0,9	-40	34,05
03/12/2023	0,9	-20	27,36
03/12/2023	0,9	0	24,48
03/12/2023	0,9	20	28,03
03/12/2023	0,9	40	35,37
03/12/2023	0,9	60	44,92

Tabela 1 – Tabela de perdas para um dia com CSI alto

abaixo de 20° chegam a apresentar ganhos, indicados pelos valores negativos na coluna de perdas, atingindo seu pico no ângulo de -40°, com um ganho percentual de 22,04%. Isso sugere que, conforme apontado por McIntosh, Abbott e Sudbury (2022), em condições de alta concentração de nuvens, determinadas angulações podem favorecer a captação de energia.

Data	CSI	Angulação (°)	Perda teórica (%)
19/12/23	0,21	-60	-19,00
19/12/23	0,21	-40	-22,04
19/12/23	0,21	-20	-14,90
19/12/23	0,21	0	-0,72
19/12/23	0,21	20	20,42
19/12/23	0,21	40	40,04
19/12/23	0,21	60	50,22

Tabela 2 – Tabela de perdas para um dia com CSI baixo

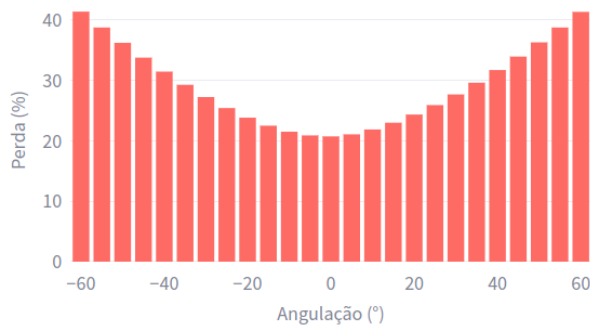
### 3.2.3 Análise do comportamento das perdas

As observações discutidas na Subseção 3.2.2 são exemplificadas na Figura 21, que apresenta dados para dois dias com CSI de 0,94 e 0,34, respectivamente. No gráfico do dia ensolarado (CSI de 0,94), as perdas percentuais variam de 20,7% com os *trackers* na posição 0° a 41,31% no ângulo de -60°. Observa-se também que as perdas aumentam à medida que os ângulos se aproximam dos polos. Já no dia com baixo CSI, a usina passa a apresentar ganhos entre os ângulos de 5° e 45°, ângulos estes que, devido ao comportamento das nuvens, proporcionaram uma melhor captação da luz solar.

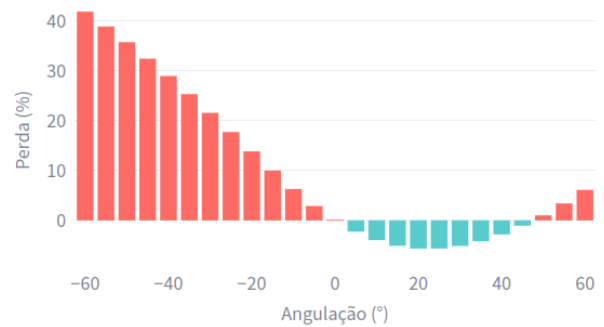
Para entender melhor o comportamento das perdas em relação ao CSI, seus valores foram organizados em intervalos constantes de CSI com variação de 0,1. A Figura 22 apresenta uma comparação da perda média para três ângulos específicos: -60°, 60° e 0°. Contudo, como cada intervalo de CSI abrange uma quantidade diferente de dias, os resultados não possuem o mesmo peso. Por exemplo, para o ângulo de -60°, não houve dias com CSI nos dois primeiros intervalos no período de um ano, enquanto os intervalos

Figura 21 – Perda por ângulo para dias com CSI de 0,94 e 0,34, respectivamente

Perda por angulação (21/04/2023)



Perda por angulação (28/11/2023)

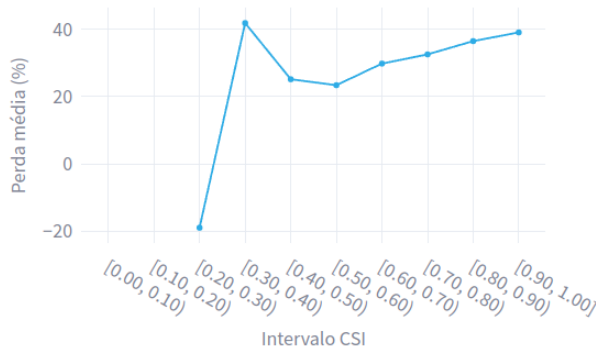


Fonte: Elaborada pelo autor.

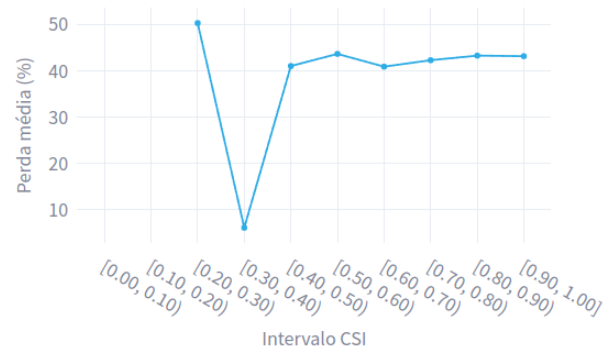
de [0,2, 0,3) e [0,3, 0,4) englobaram apenas um dia cada. Em contraste, as perdas médias entre [0,8, 0,9) e [0,9, 1,0] foram calculadas com base em 115 e 122 dias, respectivamente.

Figura 22 – Perdas por intervalo de CSI

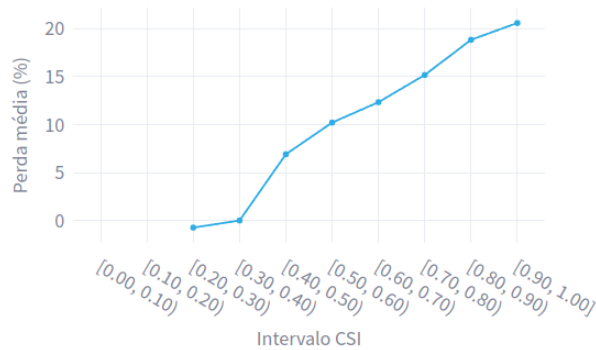
Perda por intervalo de 0.1 para um ângulo de -60°



Perda por intervalo de 0.1 para um ângulo de 60°



Perda por intervalo de 0.1 para um ângulo de 0°

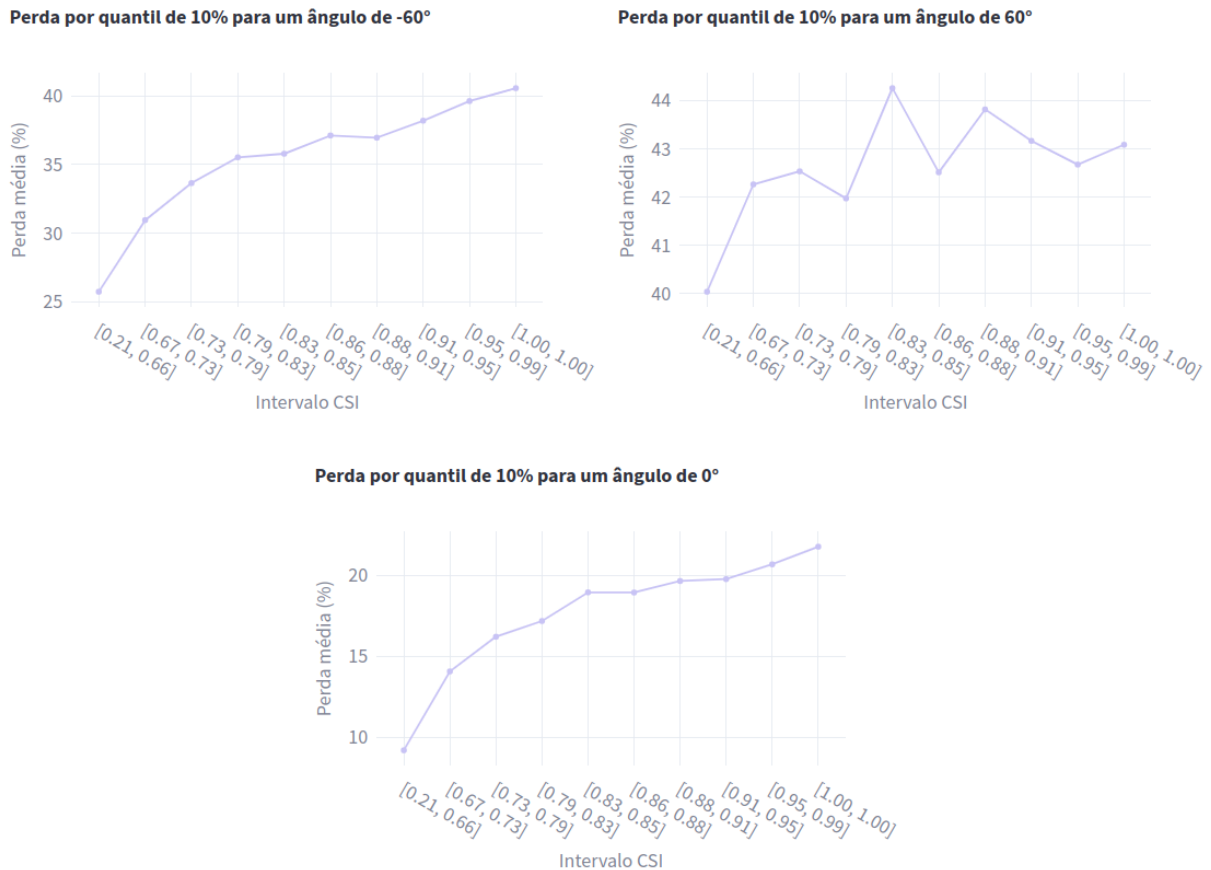


Fonte: Elaborada pelo autor.

Com o intuito de equilibrar o peso dos resultados, a tabela foi agrupada em quantis de 10%, conforme ilustrado na Figura 23. Dessa forma, embora o tamanho dos intervalos

varie, as perdas médias são calculadas com a mesma quantidade de dias, eliminando intervalos vazios e reduzindo a influência de *outliers*. Nesse contexto, observou-se um aumento nas perdas à medida que o valor de *CSI* cresce. A única exceção foi observada para o ângulo de  $60^\circ$ , onde dois picos ocorrem, um entre  $[0,83, 0,85]$  e outro entre  $[0,88, 0,91]$ . Em ambos os casos, a perda média diminui logo no intervalo seguinte.

Figura 23 – Perdas por quantil de *CSI*

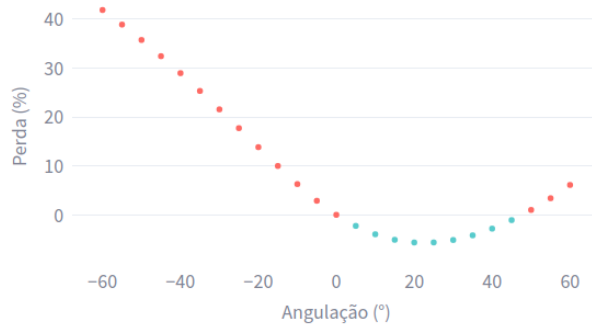


Fonte: Elaborada pelo autor.

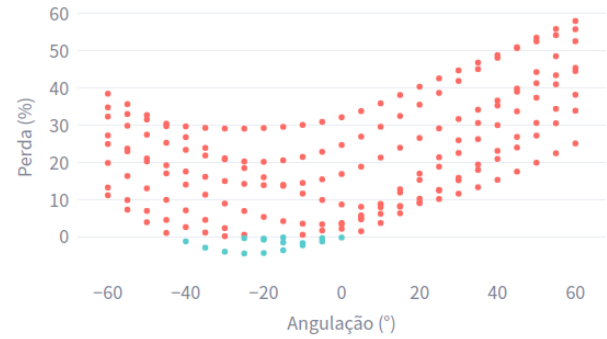
Em seguida, foi analisado o comportamento das perdas para cada um dos 25 ângulos, considerando apenas os dias em que o *CSI* assume um dos quatro valores específicos: 0,3, 0,5, 0,7 e 1. Na Figura 24, é possível observar que, conforme apontado por Aia (2004), índices baixos apresentam perdas com comportamentos mais irregulares e caóticos. À medida que o *CSI* se aproxima de 1, indicando condições de céu limpo, as perdas tendem a seguir um padrão, assumindo uma forma parabólica. Além disso, observa-se momentos de ganho de geração para todos os gráficos, exceto no caso de *CSI* igual a 1. Outro ponto relevante é que, em todo o intervalo analisado, apenas um dia correspondeu a um *CSI* igual a 0,3, por conta disso, não foi possível observar múltiplos comportamentos de perda.

Figura 24 – Perdas por ângulo para CSI's específicos

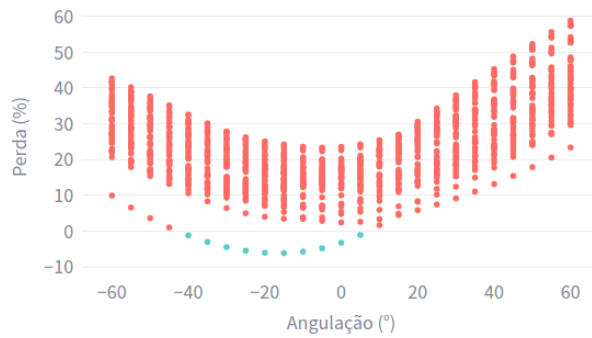
Perda por angulação para CSI 0.3



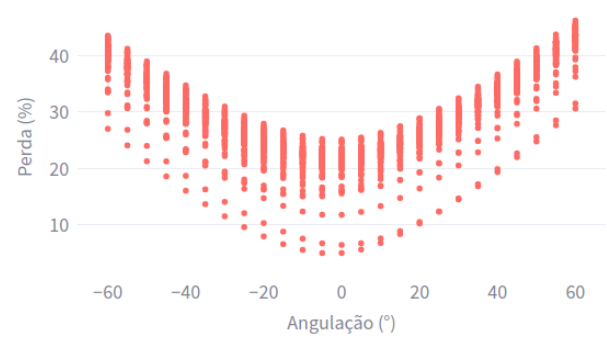
Perda por angulação para CSI 0.5



Perda por angulação para CSI 0.7



Perda por angulação para CSI 1

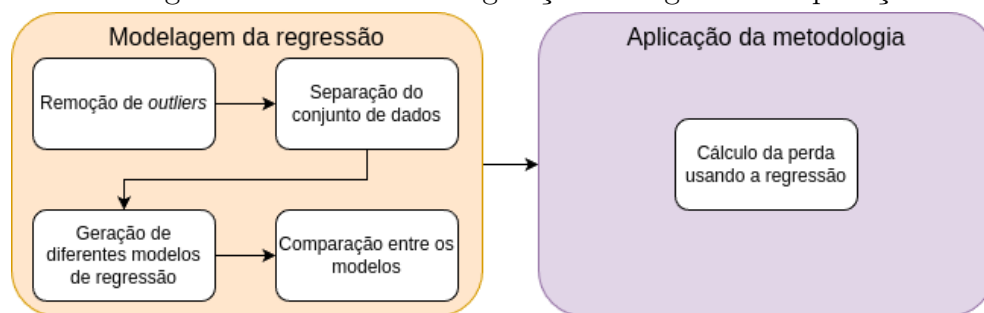


Fonte: Elaborada pelo autor.

# 4 MODELAGEM DA REGRESSÃO E APLICAÇÃO DA METODOLOGIA

Após gerar e tratar todos os dados necessários, culminando na criação da tabela de perdas, a metodologia requer ainda duas etapas para sua conclusão: a modelagem de uma regressão e sua aplicação em um cenário prático, conforme ilustrado na [Figura 25](#).

Figura 25 – Diagrama da estrutura da geração da regressão e aplicação do método

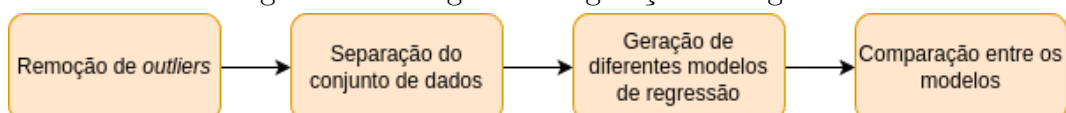


Fonte: Elaborada pelo autor.

## 4.1 Modelagem da regressão

Esta etapa consiste na identificação de um método de regressão que seja ao mesmo tempo adequado ao conjunto de dados e de fácil interpretação pelos operadores da usina, que podem não possuir conhecimentos avançados em programação. Assim, a análise será limitada a modelos e bibliotecas que gerem expressões matemáticas para as regressões, excluindo abordagens mais complexas, como redes neurais. O processo pode ser visualizado no diagrama apresentado na [Figura 26](#).

Figura 26 – Diagrama da geração da regressão



Fonte: Elaborada pelo autor.

### 4.1.1 Remoção de outliers

Em estatística, pontos fora da curva, ou *outliers*, são valores que se desviam do padrão observado em uma série de dados. Eles representam exceções à normalidade e podem distorcer as conclusões de uma análise. Um conjunto pode conter múltiplos *outliers*

(SOARES; SIQUEIRA, 1999), que podem surgir devido a fatores como erros em sensores, falta de calibração ou mudanças externas, como condições climáticas. A ausência de um tratamento adequado para esses pontos pode resultar em propagação de erros ao longo da análise, ocasionando problemas e anomalias (SOARES; SIQUEIRA, 1999).

Com o objetivo de eliminar valores atípicos, foi aplicado à tabela de perdas o método do Intervalo Interquartil (IIQ), uma medida estatística que indica a dispersão dos dados em um conjunto, sendo definido como a diferença entre o Terceiro Quartil ( $Q_3$ ) e o Primeiro Quartil ( $Q_1$ ) (PINHEIRO et al., 2013). A partir do IIQ, calcula-se o Limite Inferior ( $L_i$ ) e o Limite Superior ( $L_s$ ) utilizando as seguintes equações:

$$L_i = Q_1 - 1,5 \cdot IIQ \quad (4.1)$$

$$L_s = Q_3 + 1,5 \cdot IIQ \quad (4.2)$$

esses limites atuam como um critério de filtragem, removendo os registros associados aos conjuntos CSI, angulação e  $L_t$  que estiverem fora desse intervalo.

#### 4.1.2 Separação do conjunto de dados

Conforme indicado por Guimarães, Meireles e Almeida (2019), o treinamento de um modelo de ML requer a separação dos dados em dois conjuntos: treinamento e teste. Essa abordagem permite avaliar a capacidade dos modelos testados de se adequarem a dados desconhecidos. Além disso, o uso de um conjunto de teste possibilita a obtenção de métricas que auxiliam as bibliotecas de AutoML a selecionar o modelo mais adequado para os dados analisado.

Para garantir a representatividade sazonal no conjunto de teste, foi escolhido aleatoriamente um mês entre cada estação do ano (verão, outono, inverno e primavera), com os meses restantes sendo utilizados para o treinamento. Essa abordagem assegura que os dados de teste reflitam padrões sazonais, como a variação na incidência solar e os períodos de chuva, evitando que fiquem restritos a um único período, o que poderia mascarar o desempenho do modelo em diferentes épocas do ano.

#### 4.1.3 Geração de diferentes modelos de regressão

O primeiro modelo selecionado foi a regressão linear, um método que busca minimizar a soma dos quadrados dos resíduos entre os valores observados no conjunto de dados e aqueles previstos por uma aproximação linear (SCIKIT-LEARN, [s.d.]b). Em essência, a regressão linear fornece uma equação que estima o valor esperado de uma variável dependente, no caso, a Perda Estimada ( $L_e$ ), com base nos valores das variáveis

independentes, *CSI* e angulação. Utilizando a biblioteca **sklearn**, foram calculados os coeficientes e o intercepto da regressão, resultando na seguinte equação:

$$L_e = 19.43 \cdot CSI + 0.06 \cdot Angulação + 10.37 \quad (4.3)$$

Na sequência, aplicou-se uma regressão polinomial, técnica estatística que modela a relação entre variáveis independentes e uma variável dependente utilizando um polinômio (KUNZEL et al., [s.d.]). Para isso, utilizou-se o **TPOT**, responsável por identificar os valores que melhor se adequam ao conjunto de dados. O treinamento seguiu os seguintes parâmetros:

- 50 gerações, cada uma com uma população de tamanho 50;
- Uma semente fixa para geração de números aleatórios, garantindo reprodutibilidade do modelo;
- Validação do treinamento pelo método **K-Fold** com 5 partições;
- As regressões polinomiais poderiam assumir graus 2, 3 ou 4.

Segundo **scikit-learn** ([s.d.]a), o **K-Fold** é uma técnica de treinamento e avaliação que divide o conjunto de dados em “K” partições, ou *folds*. Em cada iteração, o modelo é treinado em  $K - 1$  *folds* e avaliado no *fold* restante. Esse processo se repete para cada partição, garantindo que todos os dados sejam usados tanto para treinamento quanto para validação. No contexto do **TPOT**, essa abordagem permite selecionar o modelo com melhor desempenho em cada iteração, servindo de base para aprimorar as gerações subsequentes.

Após a etapa de treinamento, foi encontrada a seguinte equação polinomial de grau 2:

$$\begin{aligned} L_e = & 3.74 + 16.77 \cdot CSI + 0.22 \cdot Angulação \\ & + 1.68 \cdot CSI^2 - 0.19 \cdot CSI \cdot Angulação \\ & + 0.01 \cdot Angulação^2 \end{aligned} \quad (4.4)$$

Na próxima etapa de treinamento, a biblioteca **TPOT** foi utilizada novamente, mantendo os mesmos parâmetros definidos anteriormente, mas removendo a limitação a um único modelo. Isso permitiu que a biblioteca explorasse diferentes combinações de técnicas, resultando na seleção de um modelo híbrido onde os dados passam por uma transformação linear usando o *Stochastic Gradient Descent* (SGD) e são classificados usando o *K-Nearest Neighbors* (K-NN).

O **SGD** é um método de otimização utilizado em aprendizado de máquina para ajustar os parâmetros de um modelo com o objetivo de minimizar a função de perda. Ele opera selecionando aleatoriamente uma amostra do conjunto de dados, calculando o gradiente da função de perda em relação a essa amostra e atualizando os parâmetros do modelo com base no gradiente calculado. Esse processo é repetido até alcançar a convergência. Diferentemente do *Gradient Descent* tradicional, que utiliza todo o conjunto de dados para calcular o gradiente, o **SGD** trabalha com pequenos subconjuntos de dados, reduzindo assim o custo computacional por iteração (SCIKIT-LEARN, [s.d.]d).

Para o modelo encontrado, foram definidos os seguintes parâmetros:

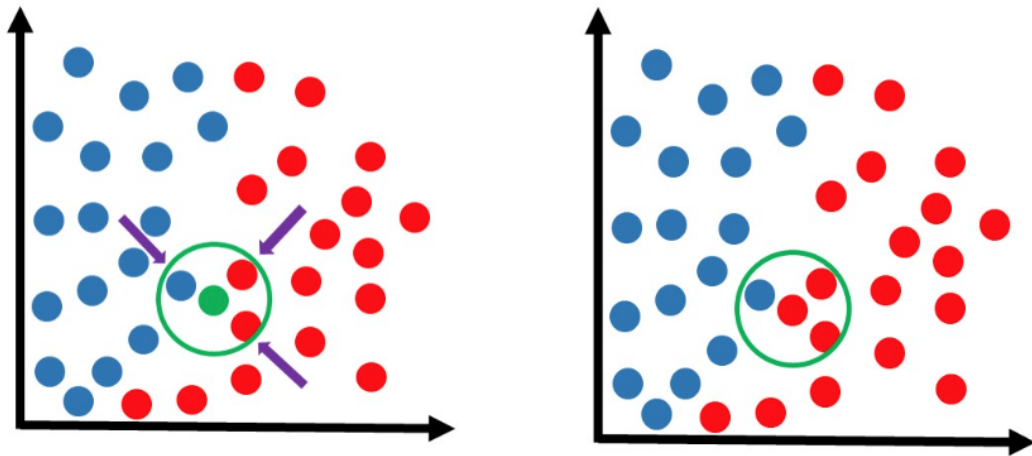
- `alpha = 0,01`: Constante multiplicativa do termo de regularização ( $\alpha$ ) igual a 0,01, que controla a penalização dos coeficientes do modelo;
- `eta0 = 1,0`: Taxa de aprendizado inicial ( $\eta_0$ ) definida como 1,0;
- `fit_intercept = False`: O modelo não adiciona um intercepto à regressão, assumindo que os dados já estão centralizados;
- `l1_ratio = 0,5`: Combinação de 50% de regularização com norma  $L_1$  (absoluta) e 50% com norma  $L_2$  (euclidiana quadrada);
- `learning_rate = invscaling`: A taxa de aprendizado diminui de forma inversamente proporcional ao número de iterações;
- `loss = epsilon_insensitive`: Utiliza uma função de perda que desconsidera erros inferiores a um valor  $\epsilon$  configurado no treinamento;
- `penalty = elasticnet`: A penalização combina normas  $L_1$  e  $L_2$ , equilibrando esparsidade e estabilidade do modelo;
- `power_t = 0,1`: Define o ritmo de decaimento da taxa de aprendizado como 0,1.

Já o **K-NN** é um método de aprendizado de máquina que classifica amostras com base na média entre seus “K” vizinhos mais próximos já classificados (SCIKIT-LEARN, [s.d.]c). Por exemplo, considere dois conjuntos de dados, representados por círculos vermelhos e azuis. Quando um novo dado ainda não classificado é inserido no algoritmo, com  $K = 3$ , o modelo calcula a média das classes dos três vizinhos mais próximos. A classificação final do novo dado é determinada pela classe predominante entre esses vizinhos, como ilustrado na Figura 27.

No modelo gerado, os seguintes parâmetros foram identificadas:

- `n_neighbors = 70`:  $K = 70$ ;

Figura 27 – Classificação de dados usando K-NN



Fonte: Adaptado de Didática Tech ([s.d.])

- $p = 2$ : A distância entre os pontos foi calculada utilizando a distância euclidiana;
- $weights = uniform$ : Todos os vizinhos tiveram o mesmo peso na classificação da amostra analisada.

#### 4.1.4 Comparação entre os modelos

Para selecionar o modelo mais adequado descrito na Subseção 4.1.3, duas métricas foram adotadas: o Coeficiente de Determinação ( $R^2$ ) e o *Root Mean Squared Error* (RMSE).

O  $R^2$  é uma medida estatística utilizada para validar a qualidade do ajuste de uma regressão, variando entre 0 e 1. Um valor 0 indica que o modelo não é capaz de explicar a variação dos dados, enquanto um valor 1 representa uma adequação perfeita ao conjunto analisado (QUININO; REIS, 2011).

Já o RMSE mede a diferença entre os valores previstos pelo modelo e os valores reais observados, refletindo o quão próximo o modelo está dos dados reais. Valores maiores de RMSE indicam previsões menos precisas, enquanto valores menores sugerem maior confiabilidade no modelo (BICKEL; DOCKSUM, 2015).

Com base nos resultados apresentados na Tabela 3, foi possível comparar os modelos avaliados. O pior desempenho, considerando ambas as métricas, foi da regressão linear, o que levou à sua eliminação imediata.

Já a regressão polinomial e o modelo híbrido que combina SGD com K-NN apresentaram valores idênticos para o  $R^2$ , o que exigiu critérios adicionais para o desempate. Nesse caso, dois fatores foram considerados: o RMSE e a complexidade do modelo.

A regressão polinomial foi escolhida por apresentar um RMSE menor, ainda que

por uma margem reduzida, e por ser mais simples, permitindo sua representação através de uma equação de fácil interpretação. Em contraste, o modelo híbrido não oferece uma forma direta de representar a regressão, tornando-o mais complexo e menos transparente.

Modelo	R <sup>2</sup>	RMSE
Regressão Linear	0,21	8,77
Regressão Polinomial	0,73	5,08
SGD + K-NN	0,73	5,14

Tabela 3 – Tabela comparativa entre os modelos

Como último teste, a equação polinomial foi aplicada a todos os pares de  $CSI$  e angulação presentes no conjunto de dados utilizado. A  $L_e$  obtida a partir dessa aplicação foi então comparada à  $L_t$ , através da seguinte equação:

$$A_a = 100 - \frac{1}{n} \cdot \sum_{i=1}^n |L_t - L_e| \quad (4.5)$$

com isso, foi possível obter uma precisão média de aproximadamente 86%.

## 4.2 Aplicação da metodologia

Para esta etapa, é necessário um novo conjunto de dados: o perfil de indisponibilidade dos *trackers* da usina. Esse relatório contém o percentual de dispositivos de sustentação que permaneceram travados diariamente, bem como seus ângulos de travamento. Esses dados são usados então, para determinar quais angulações a equação polinomial deve calcular e como a  $L_e$  deve ser ponderada para refletir o cenário real da usina.

Por exemplo, suponha um dia com  $CSI = 0,8$ , em que 10% dos *trackers* estavam travados em 0°, e 2% em 20°. Substituindo esses valores na equação, observa-se que a  $L_e$  correspondente a esses ângulos é de 18,23% e 23,59%, respectivamente. Contudo, a equação considera que todos os *trackers* estavam travados nesses ângulos. Para representar o cenário do dia, é necessário ponderar esses valores pelos percentuais reais de dispositivos travados.

Assim, as perdas ponderadas seriam 1,82% (para os *trackers* travados em 0°) e 0,47% (para os travados em 20°), totalizando uma perda de 2,29% para a usina nesse cenário hipotético.

### 4.2.1 Cálculo das perdas

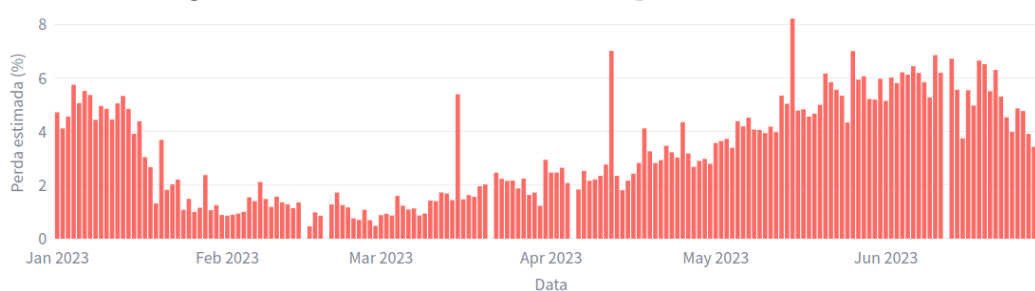
Como o relatório do perfil de indisponibilidade dos *trackers* já inclui a data, o ângulo de travamento dos dispositivos de sustentação e o percentual correspondente de dispositivos nesse estado, o cálculo da  $L_e$  para cada dia do relatório torna-se direto. Basta

utilizar o valor do **CSI**, previamente calculado na Subseção 3.2.1, substituir as variáveis independentes na equação, ponderar os resultados pelos percentuais de indisponibilidade de cada ângulo e somar os valores obtidos.

O processo descrito acima resulta na geração da **Figura 28**. Três observações devem ser destacadas sobre ela:

1. O relatório de indisponibilidade de *trackers* utilizado neste trabalho abrange um período limitado, compreendido entre janeiro e junho de 2023;
2. Dias em que não foi possível determinar o valor do **CSI** foram excluídos da análise devido à ausência de dados necessários para a equação;
3. Observa-se também que os meses com a menor perda (fevereiro, com média de 1,14%) e a maior perda (junho, com 5,37%) correspondem aos meses com os menores e maiores percentuais médios de indisponibilidade, registrando 5,44% e 26,37%, respectivamente.

Figura 28 – Perda estimada diária por *trackers* travados



Fonte: Elaborada pelo autor.

Por fim, a **Figura 29** sintetiza as informações obtidas para o período de janeiro a junho de 2023. Durante esse intervalo, a  $L_e$  foi de 3,28% e o percentual de travamento dos *trackers* ficou em 15,96%.

Figura 29 – Blocos informativos



Fonte: Elaborada pelo autor.

# 5 CONCLUSÕES

Este trabalho teve como objetivo destacar a relevância da análise da eficiência energética em usinas fotovoltaicas de grande porte, com ênfase nas perdas de geração causadas por falhas nos dispositivos de rastreamento solar (*trackers*), a anomalia mais recorrente nessas usinas. A abordagem foi realizada por meio do desenvolvimento de uma metodologia que combina a análise de dados de uma usina real localizada no Brasil com métodos de [ML](#).

A parte inicial deste projeto, voltada à análise e tratamento dos dados de irradiância, foi essencial para o desenvolvimento como um todo. Por meio dela, foi possível identificar e desconsiderar medições inconsistentes provenientes de piranômetros, causadas por falta de calibração ou fatores externos, como o travamento do *tracker* ao qual estavam acoplados. Esse processo resultou na geração de uma [GTI](#) de referência, com o objetivo de minimizar a influência de *outliers*, representando de forma mais precisa os valores de irradiância medidos pelos sensores distribuídos por toda a extensão da usina. Além disso, essa série temporal de referência, combinada com a [GHI](#), foi utilizada para calcular as potências teóricas empregadas nos modelos de [ML](#).

Na sequência, a elaboração de uma tabela que correlaciona dados de [CSI](#), angulação e perda teórica para diversos ângulos dentro dos limites operacionais de um *tracker*, com variação de  $5^\circ$  entre os pontos, foi essencial para simplificar a quantidade de simulações, reduzindo o total para apenas 25. Esse processo permitiu entender como as perdas de geração se comportam à medida que tanto o [CSI](#) quanto a angulação variam, proporcionando uma visão mais clara do impacto dessas variáveis.

Dado o foco deste trabalho na implementação da metodologia no ambiente corporativo, além de sua relevância acadêmica, foi necessário optar por um método de [ML](#) de fácil compreensão. Assim, a escolha por bibliotecas de AutoML, que utilizam modelos estatísticos facilmente representáveis e compreensíveis por profissionais sem profundo conhecimento em programação, mostrou-se altamente eficaz, apresentando uma precisão média de 86%, quando comparada com a tabela de perdas gerada previamente. Essa abordagem conciliou a simplicidade demandada pelo contexto corporativo com a precisão exigida no cenário acadêmico.

Em resumo, embora a metodologia demande mais etapas iniciais em comparação ao cálculo direto de perdas pela biblioteca, sua simplicidade e rapidez após a definição da regressão compensam essa diferença. Para estimar a perda de um dia específico, a metodologia requer apenas o [CSI](#) e o perfil de indisponibilidade dos *trackers*. Isso a torna capaz de estimar perdas para dias futuros, desde que o [CSI](#) e o perfil de indisponibilidade

sejam informados.

## 5.1 Trabalhos futuros

Considerando que a metodologia foi desenvolvida com a premissa de que todos os painéis solares da usina possuem a mesma eficiência, um trabalho futuro poderia explorar a ponderação das perdas em cenários onde o travamento de *trackers* específicos impacta de forma diferenciada a geração de energia. Essa abordagem permitiria identificar e priorizar dispositivos cuja falha tem maior influência na eficiência global da usina, proporcionando uma análise mais refinada e alinhada à realidade operacional.

Uma segunda recomendação é aplicar o método em outra usina fotovoltaica de grande porte, que possua componentes e uma estrutura distintas da usina analisada neste estudo. Isso permitirá validar a aplicabilidade e a eficácia da metodologia em diferentes cenários operacionais.

Por fim, recomenda-se explorar o uso de outra biblioteca para a geração dos dados teóricos. Embora isso possa diminuir a precisão alcançada, há potencial para tornar o método mais genérico e flexível, ampliando sua aplicabilidade em outros contextos e configurações.

# REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. *Tratamento regulatório à situação de Constrained-off de usinas eólicas*: Relatório de análise de impacto regulatório nº 001/2019-srg/aneel. 1. ed. Brasília, 2019. 10 p. (48500.006218/2017-10). Disponível em: [https://antigo.aneel.gov.br/pt/web/guest/audiencias-publicas-antigas?p\\_p\\_id=participacaopublica\\_WAR\\_participacaopublicaportlet&p\\_p\\_lifecycle=2&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_cacheability=cacheLevelPage&p\\_p\\_col\\_id=column-2&p\\_p\\_col\\_pos=1&p\\_p\\_col\\_count=2&\\_participacaopublica\\_WAR\\_participacaopublicaportlet\\_ideDocumento=32971&\\_participacaopublica\\_WAR\\_participacaopublicaportlet\\_tipoFaseReuniao=fase&\\_participacaopublica\\_WAR\\_participacaopublicaportlet\\_jspPage=%2Fhtml%2Fpp%2Fvisualizar.jsp#:~:text=O%20Constrained%2Doff%20caracteriza%2Dse,operativas%20que%20impe%C3%A7%C3%A3o](https://antigo.aneel.gov.br/pt/web/guest/audiencias-publicas-antigas?p_p_id=participacaopublica_WAR_participacaopublicaportlet&p_p_lifecycle=2&p_p_state=normal&p_p_mode=view&p_p_cacheability=cacheLevelPage&p_p_col_id=column-2&p_p_col_pos=1&p_p_col_count=2&_participacaopublica_WAR_participacaopublicaportlet_ideDocumento=32971&_participacaopublica_WAR_participacaopublicaportlet_tipoFaseReuniao=fase&_participacaopublica_WAR_participacaopublicaportlet_jspPage=%2Fhtml%2Fpp%2Fvisualizar.jsp#:~:text=O%20Constrained%2Doff%20caracteriza%2Dse,operativas%20que%20impe%C3%A7%C3%A3o). Acesso em: 11 jun. 2024. 21

AIA, A. B. Sobre nuvens e relógios: O par determinístico/aleatório. In: UNIVERSIDADE DE SÃO PAULO. 2004. Disponível em: <https://www.nilsonjosemachado.net/20040917.pdf>. Acesso em: 17 nov. 2024. 39

ALLAMEHZADEH, H. Solar energy overview and maximizing power output of a solar array using sun trackers. In: *2016 IEEE Conference on Technologies for Sustainability (SusTech)*. [s.n.], 2016. p. 14–19. Disponível em: <https://ieeexplore.ieee.org/document/7897136>. Acesso em: 08 mai. 2024. 18

ANDERSON, K. S. et al. pvlb python: 2023 project update. *Journal of Open Source Software*, The Open Journal, v. 8, n. 92, p. 5994, 2023. Disponível em: <https://doi.org/10.21105/joss.05994>. Acesso em: 20 jun. 2024. 21, 22

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 16149:2013*: Sistemas fotovoltaicos (fv) - características da interface de conexão com a rede elétrica de distribuição. [S.l.], 2013. 12 p. Disponível em: <https://www.studocu.com/pt-br/document/centro-universitario-do-sudeste-mineiro/centro-universitario-jorge-amado/nbr-16149-sistemas-fotovoltaicos-fv/77912319>. Acesso em: 08 mai. 2024. 19

BICKEL, P. J.; DOCKSUM, K. A. *Mathematical Statistics: Basic Ideas and Selected Topics*. 2. ed. [S.l.]: Chapman and Hall/CRC, 2015. v. 1. 576 p. ISBN 978-1315369266. 45

BLAND, J. M.; ALTMAN, D. G. Measurement error. *BMJ*, Londres, v. 312, 1996. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2351401/pdf/bmj00548-0038.pdf>. Acesso em: 18 set. 2024. 27

BRASIL, M. *Veja como o Brasil se tornou um dos dez maiores geradores solares do mundo*. 2023. epbr. Disponível em: <https://epbr.com.br/veja-como-o-brasil-se-tornou-um-dos-dez-maiores-geradores-solares-do-mundo>. Acesso em: 07 mai. 2024. 19

CARVALHO, J. I. F. de. *Média aritmética nos livros didáticos nos anos finais do ensino fundamental*. Dissertação (Mestrado) — Universidade Federal de Pernambuco,

- Recife, 2011. Disponível em: [https://repositorio.ufpe.br/bitstream/123456789/3728/1/arquivo2878\\_1.pdf](https://repositorio.ufpe.br/bitstream/123456789/3728/1/arquivo2878_1.pdf). Acesso em: 18 set. 2024. 28
- COMBE, T.; MARTIN, A.; PIETRO, R. D. To docker or not to docker: A security perspective. *IEEE Cloud Computing*, v. 3, n. 5, p. 54–62, 2016. Disponível em: <https://ieeexplore.ieee.org/document/7742298>. Acesso em: 08 mai. 2024. 23
- DIDÁTICA TECH. *O que é e como funciona o algoritmo KNN?* [s.d.]. Disponível em: <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-knn/>. Acesso em: 30 nov. 2024. 45
- FERNÁNDEZ, P. de A. et al. An iot open source platform for photovoltaic plants supervision. *International Journal of Electrical Power & Energy Systems*, v. 125, p. 106540, 2021. ISSN 0142-0615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0142061520310279>. Acesso em: 27 abr. 2024. 18
- GOV.BR. *Energia solar: usinas centralizadas ultrapassam os 5% na matriz elétrica brasileira, com 10,4 GW instalados*. 2023. Disponível em: <http://tiny.cc/mullxz>. Acesso em: 30 mar. 2024. 15
- GUIMARÃES, L. M. S.; MEIRELES, M. R. G.; ALMEIDA, P. E. M. de. Avaliação das etapas de pré-processamento e de treinamento em algoritmos de classificação de textos no contexto da recuperação da informação. *Perspectivas em Ciência da Informação*, Escola de Ciência da Informação da UFMG, v. 24, n. 1, p. 169–190, Jan 2019. ISSN 1413-9936. Disponível em: <https://doi.org/10.1590/1981-5344/3505>. Acesso em: 29 nov. 2024. 42
- IEZZI, G.; HAZZAN, S.; DEGENSZAJN, D. M. *Fundamentos de Matemática Elementar*. SARAIVA S.A. Livreiros Editores, 2013. v. 11. 260 p. Disponível em: <https://barbosadejesu.wordpress.com/wp-content/uploads/2021/09/fundamentos-da-matematica-elementar-11.pdf>. Acesso em: 18 set. 2024. 33
- INTERNATIONAL BUSINESS MACHINES CORPORATION. *O que é AutoML?* [s.d.]. Disponível em: <https://www.ibm.com/br-pt/topics/automl#:~:text=IBM-,O%20que%20%C3%A9%20AutoML%3F,componente%20do%20pipeline%20de%20ML>. Acesso em: 20 jun. 2024. 22
- INTERNATIONAL ENERGY AGENCY. *Solar PV*. 2023. Disponível em: <https://www.iea.org/energy-system/renewables/solar-pv>. Acesso em: 07 mai. 2024. 19
- INTERNATIONAL ENERGY AGENCY. *World Energy Outlook 2023*. 2023. Disponível em: <https://www.iea.org/reports/world-energy-outlook-2023>. Acesso em: 29 out. 2024. 15
- JENSEN, A. R. et al. pvlib iotools—open-source python functions for seamless access to solar irradiance data. *Solar Energy*, v. 266, p. 112092, 2023. ISSN 0038-092X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038092X23007260>. Acesso em: 20 jun. 2024. 26
- KUNZEL, B. et al. *REGRESSÃO POLINOMIAL*. [s.d.]. Disponível em: [https://edisciplinas.usp.br/pluginfile.php/8083873/mod\\_resource/content/1/Regressao%20Polinomial.pdf](https://edisciplinas.usp.br/pluginfile.php/8083873/mod_resource/content/1/Regressao%20Polinomial.pdf). Acesso em: 29 nov. 2024. 43

MCINTOSH, K. R.; ABBOTT, M. D.; SUDBURY, B. A. The optimal tilt angle of monofacial and bifacial modules on single-axis trackers. *IEEE Journal of Photovoltaics*, v. 12, n. 1, p. 397–405, 2022. ISSN 2156-3381. Disponível em: <https://ieeexplore-ieee-org.ez130.periodicos.capes.gov.br/document/9631233>. Acesso em: 13 nov. 2024. 37

NASTEC. [s.d.]. Disponível em: <https://nastec.eu/en/solar-energy-products/vasco-solar/mppt-always-the-maximum-power-available/>. Acesso em: 08 mai. 2024. 19

PINHEIRO, J. et al. *Probabilidade e Estatística: Quantificando a Incerteza*. [S.l.]: Elsevier Brasil, 2013. ISBN 978-8535259308. 42

QUININO, R. C.; REIS, E. A. O coeficiente de determinação  $R^2$  como instrumento didático para avaliar a utilidade de um modelo de regressão linear múltipla. p. 14, 2011. Disponível em: [https://www.est.ufmg.br/portal/wp-content/uploads/2023/01/RTP\\_01\\_2011.pdf](https://www.est.ufmg.br/portal/wp-content/uploads/2023/01/RTP_01_2011.pdf). Acesso em: 01 ago. 2024. 45

RAPTOR MAPS. Global solar report 2023 edition. 2023. Disponível em: <https://raptormaps.com/resources/2023-global-solar-report>. Acesso em: 17 jun. 2024. 15

REIS, M. M. Análise de séries temporais. Florianópolis, p. 55, 2023. Disponível em: <https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>. Acesso em: 30 mai. 2024. 20

SARANGPURE, N. et al. Automating the machine learning process using pycaret and streamlit. In: *2023 2nd International Conference for Innovation in Technology (INOCON)*. [s.n.], 2023. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/10101357>. Acesso em: 08 mai.2024. 24

SCARABELOT, L. T. *Influência do fator de dimensionamento de inversor no desempenho termoenergético de sistemas fotovoltaicos de geração distribuída*. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2020. Disponível em: <https://repositorio.ufsc.br/handle/123456789/219475?show=full>. Acesso em: 08 mai. 2024. 19

SCIKIT-LEARN. *KFold*. [s.d.]. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.KFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html). Acesso em: 29 nov. 2024. 43

SCIKIT-LEARN. *Linear Models*. [s.d.]. Disponível em: [https://scikit-learn.org/stable/modules/linear\\_model.html#linear-models](https://scikit-learn.org/stable/modules/linear_model.html#linear-models). Acesso em: 29 nov. 2024. 42

SCIKIT-LEARN. *Nearest Neighbors*. [s.d.]. Disponível em: <https://scikit-learn.org/stable/modules/neighbors.html>. Acesso em: 30 nov. 2024. 44

SCIKIT-LEARN. *Stochastic Gradient Descent*. [s.d.]. Disponível em: <https://scikit-learn.org/1.5/modules/sgd.html#stochastic-gradient-descent>. Acesso em: 30 nov. 2024. 44

SOARES, J. F.; SIQUEIRA, A. L. *Introdução A Estatística Médica*. Belo Horizonte: COOPMED, 1999. v. 1. ISBN 978-8585002558. Disponível em: [https://est.ufmg.br/~enricoc/pdf/medicina/Soares\\_siqueira.pdf](https://est.ufmg.br/~enricoc/pdf/medicina/Soares_siqueira.pdf). Acesso em: 27 nov. 2024. 42

SOLAR ENERGY TECHNOLOGIES OFFICE. *Solar Photovoltaic Technology Basics*. [s.d.]. Disponível em: <https://www.energy.gov/eere/solar/solar-photovoltaic-technology-basics>. Acesso em: 27 abr. 2024. 17

TAUSIESAKUL, B.; ASAVASKULKIET, K. An approximation of focuss mean squared error. In: *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. [s.n.], 2023. p. 231–236. Disponível em: <https://ieeexplore-ieee-org.ez130.periodicos.capes.gov.br/document/10202042>. Acesso em: 04 out. 2024. 31

TONOLO, É. A. Análise dos fatores de perdas nos sistemas fotovoltaicos da UTFPR Campus Curitiba. 2019. Disponível em: [https://repositorio.utfpr.edu.br/jspui/bitstream/1/4664/1/CT\\_PPGSE\\_M\\_Tonolo%2C%20%C3%89dwin%20Augusto\\_2019.pdf](https://repositorio.utfpr.edu.br/jspui/bitstream/1/4664/1/CT_PPGSE_M_Tonolo%2C%20%C3%89dwin%20Augusto_2019.pdf). Acesso em: 02 abr. 2024. 15, 17

TPOT. [s.d.]. Disponível em: <https://epistasislab.github.io/tpot/>. Acesso em: 20 jun. 2024. 23, 24

# Apêndices

# APÊNDICE A – REPOSITÓRIO DO PROJETO

O link abaixo fornece acesso ao repositório contendo os códigos desenvolvidos neste trabalho.

<https://github.com/lucasraupp/TCC>

# APÊNDICE B – TRECHO DE GERAÇÃO DA SÉRIE TEMPORAL DE UM *TRACKER*

Código B.1 – Geração e *plot* do comportamento de um *tracker* ao longo de um dia

```
1 import matplotlib.dates as mdates
2 import matplotlib.pyplot as plt
3 import pandas as pd
4 from pvlib import solarposition, tracking
5
6 latitude = -5.130642726593804
7 longitude = -37.047300940169734
8 tz = "Brazil/East"
9
10 begin = pd.Timestamp("2024-01-01")
11 end = pd.Timestamp(f"{begin} 23:50:00")
12
13 times = pd.date_range(
14     begin,
15     end,
16     freq="10min",
17     tz=tz,
18 )
19
20 solpos = solarposition.get_solarposition(times, latitude, longitude)
21
22 truetracking_angles = tracking.singleaxis(
23     apparent_zenith=solpos["apparent_zenith"],
24     apparent_azimuth=solpos["azimuth"],
25     axis_tilt=0,
26     axis_azimuth=180,
27     max_angle=90,
28     backtrack=True, # for true-tracking
29     gcr=0.5,
30 )
31
32 truetracking_position = truetracking_angles["tracker_theta"].fillna(0)
33 Timestamp = truetracking_position.index
34
35 df_trackers = pd.DataFrame(
36     {"Timestamp": Timestamp, "truetracking_position": truetracking_position}
37 )
38 df_trackers["Timestamp"] = df_trackers["Timestamp"].dt.tz_localize(None)
```

```
39 df_trackers = df_trackers.set_index("Timestamp")
40
41 fig = plt.figure(figsize=(8, 4))
42 ax = fig.add_subplot(111)
43 ax.plot(df_trackers.index, df_trackers["truetracking_position"])
44
45 ax.xaxis.set_major_formatter(mdates.DateFormatter("%H:%M"))
46 ax.set_xlim(begin, end)
47
48 ax.set_xlabel("Horário")
49 ax.set_ylabel("Angulação")
```