

Universidade Federal de Santa Catarina
Centro de Florianópolis
Departamento de Engenharia
Elétrica e Eletrônica



Bruno Castro Valle

Detecção de estado de falha em usinas fotovoltaicas utilizando
aprendizado de máquina

Florianópolis

2022

Bruno Castro Valle

**Detecção de estado de falha em usinas fotovoltaicas
utilizando aprendizado de máquina**

Trabalho de Conclusão de Curso submetido à Universidade Federal de Santa Catarina, como requisito necessário para obtenção do grau de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Richard Demo Souza

Universidade Federal de Santa Catarina
Centro de Florianópolis
Departamento de Engenharia
Elétrica e Eletrônica

Florianópolis
2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Valle, Bruno Castro

Detecção de estado de falha em usinas fotovoltaicas
utilizando aprendizado de máquina / Bruno Castro Valle ;
orientador, Richard Demo Souza, 2022.

50 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Elétrica, Florianópolis, 2022.

Inclui referências.

1. Engenharia Elétrica. 2. Energia Solar Fotovoltaica.
3. Detecção de Estado de Falha. 4. Aprendizado de Máquina.
I. Souza, Richard Demo. II. Universidade Federal de Santa
Catarina. Graduação em Engenharia Elétrica. III. Título.

Bruno Castro Valle

Detecção de estado de falha em usinas fotovoltaicas utilizando aprendizado de máquina

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Elétrica” e aceito, em sua forma final, pelo Curso de Graduação em Engenharia Elétrica.

Florianópolis, 16 de dezembro de 2022.



Documento assinado digitalmente
Miguel Moreto
Data: 18/12/2022 10:52:32-0300
CPF: ***.850.100-**
Verifique as assinaturas em <https://v.ufsc.br>

Coordenação do Curso

Banca Examinadora



Documento assinado digitalmente
Richard Demo Souza
Data: 18/12/2022 07:57:35-0300
CPF: ***.267.379-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dr. Richard Demo Souza
Orientador



Documento assinado digitalmente
Telles Brunelli Lazzarin
Data: 19/12/2022 06:50:54-0300
CPF: ***.400.719-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dr. Telles Brunelli Lazzarin
Universidade Federal de Santa Catarina



Documento assinado digitalmente
MARIO DE NORONHA NETO
Data: 19/12/2022 08:58:24-0300
CPF: 003.859.519-22
Verifique as assinaturas em <https://v.ifsc.edu.br>

Prof. Dr. Mario de Noronha Neto
Instituto Federal de Santa Catarina

Florianópolis, 2022

Dedico este trabalho a todos aqueles que, de alguma forma,
auxiliaram para a concretização desta etapa.

Agradecimentos

Gostaria de agradecer primeiramente aos meus pais por todos os ensinamentos, apoio, incentivos e principalmente pelo amor que sempre me deram. Sou e serei eternamente grato.

A minha irmã pelo apoio e amor durante toda a vida e apesar das discussões e brigas sempre que necessitei estava lá pra mim.

A minha namorada Camila por ser a melhor companheira que eu poderia ter, por ter me suportado nesse período de estresse e sempre me incentivar a crescer e melhorar, tanto como pessoa, quanto profissionalmente.

A minha família, que mesmo dos mais diversos sempre estiveram lá para me apoiar e fornecer energias positivas para o meu sucesso e futuro.

Ao professor Mário de Noronha Neto que apesar de ser de outra instituição me auxiliou, apoiou e me fez ir além para conseguir completar essa etapa de TCC da melhor maneira possível. Sou eternamente grato.

Ao professor orientador Richard Demo, que apesar das condições adversas me acolheu e me ajudou a desenvolver e escrever um tema de TCC em um período de tempo apertado.

Aos amigos de longa data que vão continuar no meu coração para o resto da vida: Rafael, Bruno, João E., João H., João P., André, Leonardo, Bernardo, Augusto, Kaique, Lobo, Tales, Pedro.

Aos meus colegas e amigos de faculdade que me apoiaram nessa caminhada: Arthur, Vitor, Ninja, Marcelo, Guilherme, Douglas, Yago, Jonas.

Aos amigos e colegas de trabalhos da Smetro Tecnologia por terem me acompanhado e ensinado diversas coisas que vou levar para a vida.

Ao Prof.Dr Adroaldo e o pessoal do laboratório MAGLAB que desenvolveram um papel importantíssimo no desenvolvimento da minha carreira e também do meu caráter.

Ao professor coordenador de TCC Pacheco que me auxiliou e tirou diversas dúvidas neste período.

Ao IFSC que me ensinou diversas coisas durante meu ensino médio profissionalizante, as quais me auxiliaram em toda minha vida até então.

E finalmente a UFSC por ter me fornecido oportunidades de aprendizado e crescimento incríveis.

"Train yourself to let go of everything you fear to lose."
(Yoda)

Resumo

Dados reportados pela Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR) demonstram um crescimento exponencial na produção de energia solar, sendo 2022 considerado o melhor ano da energia solar no Brasil. Porém, o número de sistemas fotovoltaicos monitorados não seguiu a mesma tendência de crescimento, sendo normalmente operados sem um sistema de supervisão adequado, por isso a necessidade de implementação de um sistema de detecção de estado de falha *online*.

Este trabalho de conclusão de curso propõe um estudo sobre o método de detecção de estado de falha que utiliza modelos de predição instantânea de potência CA gerada pelo sistema fotovoltaico. A pesquisa foi executada na base de dados SOLETE, disponibilizada por Pombo et al. A base contém medidas de um sistema fotovoltaico e dados meteorológicos ao longo de um ano e cinco meses. Foram utilizados três modelos de predição, denominados modelo físico, *one equation* e *random forest*. Para cada modelo foram elaboradas duas predições, uma na faixa global de irradiância e outra em faixas separadas, com o propósito de verificar se esta separação de faixas influenciará nos resultados finais. Em termos de performance o trabalho atingiu resultados semelhantes à literatura, onde os três melhores modelos atingiram uma precisão de mais de 99% na detecção de estado de falha. Os resultados demonstram que os melhores modelos foram o modelo físico e o *one equation* nas faixas separadas e o modelo *random forest* na faixa global de irradiância.

Palavras-Chave: 1.Geração. 2.Energia. 3.Fotovoltaica. 4.Falhas. 5.Detecção.

Abstract

Data reported by the Brazilian Association of Photovoltaic Solar Energy (ABSOLAR) show exponential growth in solar energy production, with 2022 being considered the best year for solar energy in Brazil. However, the number of monitored photovoltaic systems did not follow the same growth trend, being normally operated without an adequate supervisory system, hence the need to implement an online failure status detection system.

This work proposes a study on the fault state detection method that uses instantaneous prediction models of AC power generated by the photovoltaic system. The research was carried out in the SOLETE database, made available by Pombo et al. The base contains measurements of a photovoltaic system and meteorological data over a year and five months. Three prediction models were used, called physical model, one equation and random forest. For each model, two predictions were made, one in the global irradiance range and the other in separate irradiation ranges, with the purpose of verifying whether this separation of ranges will influence the final results. In terms of performance, the work reached similar results to the literature, where the three best models reached an accuracy of more than 99% in the failure state detection. The results demonstrate that the best models were the physical model and the one equation in the separate ranges and the random forest model in the global irradiance range.

Keywords: 1.Power Plant. 2.Energy. 3.Photovoltaic. 4.Fault. 5.Detection.capa

Lista de figuras

Figura 1 – Potência instalada de energia solar no Brasil entre 2017 e 2022 [1] . . .	14
Figura 2 – Diagrama esquemático de um sistema fotovoltaico.	18
Figura 3 – Possíveis gráficos dentro da ferramenta <i>matplotlib</i>	21
Figura 4 – Exemplo do algoritmo de <i>Random Forest</i>	24
Figura 5 – Organização dos dados da base de dados SOLETE.	28
Figura 6 – Informações sobre a base de dados SOLETE.	29
Figura 7 – Gráfico estilo Boxplot da potência AC e da irradiação no plano do painel fotovoltaico.	30
Figura 8 – Diagrama esquemático geral dos modelos utilizados	31
Figura 9 – Características escolhidas para treinamento e teste do modelo RF . . .	32
Figura 10 – Função que retorna a expressão de Pac para o uso da função <i>curve_fit</i>	33
Figura 11 – Detecção de falhas dos pontos no modelo MF na faixa global	36
Figura 12 – Falhas inseridas e visualizadas nos limites do modelo físico na faixa global	36
Figura 13 – Pontos normais nos limites do modelo físico na faixa global	37
Figura 14 – Detecção pela média e desvio padrão de todos os pontos no modelo físico nas faixas separadas	37
Figura 15 – Falhas inseridas e visualizadas nos limites do modelo físico nas faixas separadas	38
Figura 16 – Valores normais nos limites do modelo físico nas faixas separadas . . .	38
Figura 17 – Detecção de falhas dos pontos no modelo OE na faixa global	39
Figura 18 – Falhas inseridas e visualizadas nos limites do modelo OE na faixa global	39
Figura 19 – Valores normais nos limites do modelo físico na faixa global	40
Figura 20 – Detecção pela média e desvio padrão de todos os pontos no modelo OE nas faixas separadas	40
Figura 21 – Falhas inseridas e visualizadas nos limites do modelo OE nas faixas separadas	41
Figura 22 – Valores normais nos limites do modelo físico nas faixas separadas . . .	41
Figura 23 – Detecção de falhas dos pontos no modelo RF na faixa global	42
Figura 24 – Falhas inseridas e visualizadas nos limites do modelo RF na faixa global	42
Figura 25 – Valores normais nos limites do modelo físico nas faixas separadas . . .	43
Figura 26 – Detecção pela média e desvio padrão de todos os pontos no modelo OE nas faixas separadas	43
Figura 27 – Falhas inseridas e visualizadas nos limites do modelo RF nas faixas separadas	44
Figura 28 – Valores normais nos limites do modelo RF nas faixas separadas	44

Lista de tabelas

Tabela 1 – Exemplo de Matriz de Confusão.	25
Tabela 2 – Informações contidas no <i>datasheet</i> do inversor.	29
Tabela 3 – Informações sobre as faixas de rendimento do sistema fotovoltaico.	29
Tabela 4 – Valores de RMSE dos modelos de predição.	34
Tabela 5 – Valores dos indicadores de classificação dos três modelos na faixa global de irradiação solar.	45
Tabela 6 – Valores dos indicadores de classificação dos três modelos nas faixas separadas de irradiação solar.	46

Lista de Siglas e Abreviaturas

UFSC	<i>Universidade Federal de Santa Catarina</i>
UFMS	<i>Universidade Federal de Mato Grosso do Sul</i>
PV	<i>Fotovoltaico</i>
IA	<i>Inteligência artificial</i>
STC	<i>Condições de teste padrão</i>
RF	<i>Random Forest</i>
ML	<i>Machine Learning</i>
MAE	<i>Erro Médio Absoluto</i>
RSME	<i>Erro quadrático médio</i>
ANEEL	<i>Agência nacional de energia elétrica</i>
PR	<i>Performance Ratio</i>
OE	<i>One-Equation</i>
MF	<i>Modelo Físico</i>

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos Gerais	16
1.2	Objetivos Específicos	16
2	REVISÃO DA LITERATURA	17
2.1	Sistema Fotovoltaico	17
2.1.1	Painel Solar Fotovoltaico	17
2.1.2	Células Fotovoltaicas	18
2.1.3	Inversores Fotovoltaico	18
2.1.4	Falhas em Sistemas Fotovoltaicos	19
2.2	Análise de Dados em Aprendizado de Máquina	19
2.2.1	Tratamento de Dados	20
2.2.2	Visualização de Dados	20
2.3	Modelos de Geração	21
2.4	Inteligência Artificial	22
2.4.1	Python	23
2.4.2	Aprendizado de Máquina	23
2.4.2.1	<i>Random Forest</i>	23
2.5	Detecção de Falhas	24
2.6	Indicadores de Classificação	25
3	METODOLOGIA	27
3.1	Bases de Dados	27
3.1.1	Base de Dados da SOLETE	27
3.2	Pré-processamento da Base de Dados	28
3.3	Modelos de Detecção de Falhas	30
3.3.1	Modelo Físico	30
3.3.2	Modelo RF	31
3.3.3	Modelo OE	32
4	RESULTADOS	34
4.1	Comparando os Valores de RMSE	34
4.2	Detecção de Falha	35
4.2.1	Modelo Físico	35
4.2.1.1	Faixa Global	35
4.2.1.2	Faixas Separadas	37

4.2.2	Modelo <i>One-equation</i>	38
4.2.2.1	Faixa Global	39
4.2.2.2	Faixas Separadas	40
4.2.3	Modelo <i>Random Forest</i>	41
4.2.3.1	Faixa Global	42
4.2.3.2	Faixas Separadas	43
4.2.4	Comparando Resultados	44
5	CONCLUSÕES	47
	REFERÊNCIAS BIBLIOGRÁFICAS	48

1 Introdução

Devido à preocupações econômicas e principalmente ambientais, os sistemas de produção de energia estão passando por uma transformação que visa a migração do uso de combustíveis fósseis para o uso de tecnologias limpas. Esse é um dos tópicos bastante abordados e comentados ultimamente no setor de energia elétrica no mundo [2]. Dentro do mercado de energias renováveis, os sistemas fotovoltaicos mostraram um grande crescimento nos últimos anos, representando hoje uma tecnologia madura para produção de energia a partir de fontes renováveis e uma estratégia comum de geração de eletricidade [3].

O ano de 2022 foi considerado o melhor ano de energia solar no Brasil segundo o site O Setor Elétrico [4]. O país saltou de 7,9 GW, ao final de 2020, para 13 GW ao final de 2021, crescimento de 65%, mesmo em meio a um ano desafiador de pandemia global. Já para o ano de 2022 a mesma fonte afirma que haverá um crescimento de 105% na geração própria e 67.8% nas grandes usinas solares. Podemos verificar esse crescimento na Figura 1, reafirmando assim a grande importância dessa energia para o futuro do Brasil.

Figura 1 – Potência instalada de energia solar no Brasil entre 2017 e 2022 [1]



Fonte: Nexojornal (2022)

No entanto, o número de sistemas fotovoltaicos monitorados não seguiu a mesma tendência de crescimento, pois muitas usinas fotovoltaicas, especialmente as menores, são operadas sem um sistema de supervisão adequado [3].

Um algoritmo de detecção de falhas para sistemas fotovoltaicos pode fornecer uma estimativa precisa da produção de eletricidade sob condições normais de operação e detectar falhas no sistema ou períodos de produção de energia anormalmente baixa, quando

o sistema produz significativamente menos energia do que deveria nas condições de operação dadas [3]. Isso permite que o operador do sistema tome ações corretivas, a fim de evitar que o sistema fotovoltaico opere sob más condições por um tempo prolongado, minimizando a perda de energia gerada por essa falhas e melhorando o desempenho do sistema como um todo.

Alguns dos algoritmos de detecção de falhas relatados na literatura são baseados em simulações de circuitos elétricos de um painel fotovoltaico [5], e alguns usam análises estatísticas de diferentes medições do sistema ou valores de eficiência [6]. Métodos de análise de sinais elétricos, como a reflectometria no domínio do tempo, também foram usados para detectar sistemas fotovoltaicos defeituosos [7]. Vários algoritmos de detecção de falhas são baseados na comparação entre os valores medidos na saída do sistema e valores preditos [2].

Diferentes métodos são usados para desenvolver modelos preditivos para a produção de energia do sistema fotovoltaico. Em algumas abordagens a produção de eletricidade foi prevista com modelos paramétricos que utilizam dados do sistema e variáveis meteorológicas juntamente com parâmetros ajustáveis [8]; outras abordagens usam técnicas de inteligência artificial, como redes neurais, lógica fuzzy e sistemas especialistas [9].

Neste trabalho será reproduzido o método de detecção de falhas proposto por Platon et al.[3], que visa a implementação online, com algumas adições inspiradas no trabalho recente de Pombo et al.[2]. Como se trata de uma implementação online, o objetivo principal foi o desenvolvimento de uma abordagem prática e bastante simples (baixa complexidade) que tenha uma alta taxa de detecção de falhas e seja robusta o suficiente para lidar com sucesso com valores anormais inerentemente presentes em dados da vida real, como medições errôneas e atraso entre medições com diferentes taxas de amostragem [3]. Essa detecção se dá pela comparação entre o valor gerado e o valor predito instantâneo.

O autor do artigo [3], utilizou uma equação para prever a energia gerada, tendo como entrada os dados meteorológicos, alguns dados disponíveis no *datasheet* do inversor e também a aplicação de um algoritmo de regressão para o *fitting* de algumas constantes. Já a equação utilizada por Pombo et al.[2] utiliza apenas os dados meteorológicos e os dados disponíveis no *datasheet* do inversor, sem a necessidade de uma regressão para fazer o *fitting* de algumas constantes [8].

É importante mencionar que os trabalhos [2, 3] tem como semelhança o fato de usarem equações relativamente simples para a predição da energia gerada, mas apresentam abordagens em geral bastante diferentes. Enquanto em [3] a equação entrega o resultado final da predição, em [2] a saída da equação utilizada é apenas mais um parâmetro de entrada de um algoritmo mais complexo (o qual combina também dados meteorológicos e valores passados de energia gerada). Dentre os algoritmos mais complexos testados em [2], baseados em aprendizado de máquina, o que teve o melhor desempenho em termos de predição foi o algoritmo *random forest*.

1.1 Objetivos Gerais

O objetivo deste projeto é reproduzir, validar e complementar os diferentes métodos utilizados pelos artigos [3] e [2] de predição de potência gerada por um sistema PV e utilizar essa predição para detectar que o sistema apresenta alguma falha, porém sem classificar o tipo de falha ou identificar a sua causa.

1.2 Objetivos Específicos

- Analisar, tratar e selecionar os dados disponíveis na base de dados [10] que é utilizada pelo artigo [2];
- Verificar os resultados obtidos pelo artigo [3] com o auxílio da base de dados e de equações provindas do artigo [2];
- Utilizar, treinar e testar modelos para detectar falhas no sistema de geração fotovoltaico;

2 Revisão da Literatura

Este capítulo discute alguns dos principais conceitos tratados neste trabalho. Serão abordados de forma sucinta os sistemas fotovoltaicos, serão apresentadas as principais bibliotecas utilizadas nos resultados numéricos, bem como os conceitos de aprendizado de máquina e de análise de dados.

2.1 Sistema Fotovoltaico

Esta seção tem como intuito apresentar alguns conceitos básicos sobre um sistema fotovoltaico, o qual é composto de um ou mais painéis solares combinados com um inversor, entre outros componentes elétricos e mecânicos. A sua finalidade é gerar energia elétrica a partir da energia solar fotovoltaica, essa energia gerada é uma ótima opção para o benefício do meio ambiente, pois é considerada uma energia renovável e limpa.

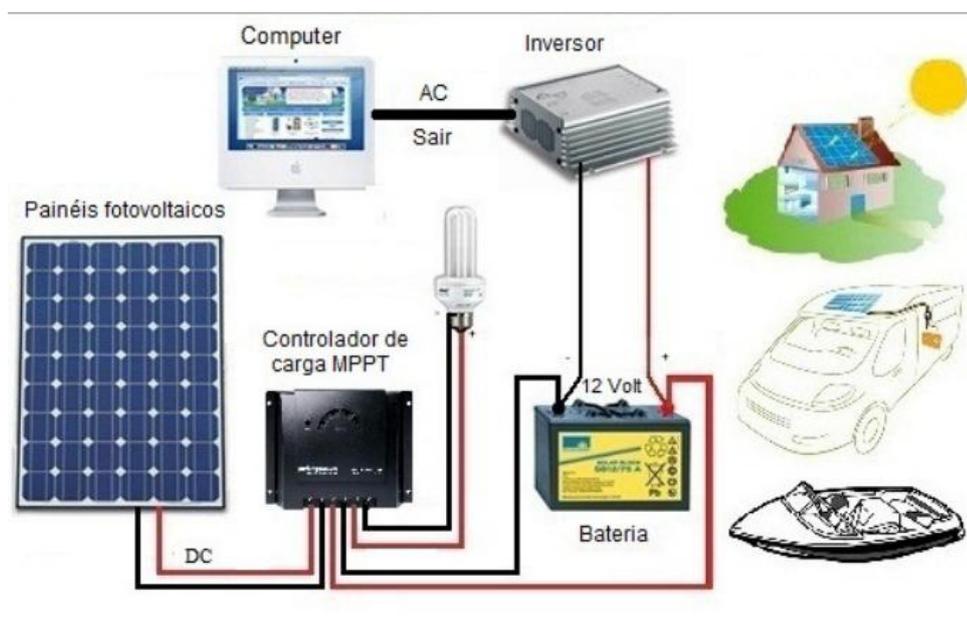
Sistemas fotovoltaicos já não são mais novidade hoje em dia. Pode até ser dito que produzir eletricidade através da energia solar foi descoberto em 1839 por Edmond Becquerel, um físico francês [11], que durante um experimento, observou que duas placas de latão imersas em um eletrólito líquido produziam eletricidade quando expostas à luz solar.

Os sistemas elétricos podem ser separados em duas classificações, sistemas isolados (*OFF GRID*, do inglês) que não dependem da rede de distribuição elétrica e os sistemas ligados à rede (*ON GRID*, do inglês) os quais dependem de um inversor de frequência para, de forma sucinta, converter a corrente contínua produzida, para corrente alternada, a qual pode ser injetada na rede elétrica convencional.

2.1.1 Painel Solar Fotovoltaico

Os painéis solares, ou módulos, utilizados nos sistemas fotovoltaicos são compostos por um conjunto de células associadas, eletricamente, em série ou paralelo, dependendo das tensões e das correntes determinadas em projeto. O conjunto desses painéis é chamado de gerador fotovoltaico, e que consiste da primeira parte do sistema fotovoltaico, ou seja, é o conjunto responsável pela captação da irradiação solar e a sua conversão em energia elétrica [12]. Pode-se ver na Figura 2 um diagrama esquemático simplificado de um sistema fotovoltaico off-grid, contendo a geração dos painéis fotovoltaicos, o controlador de carga (MPPT), uma bateria para armazenar essa carga em corrente contínua (CC) e um inversor que converte essa corrente para corrente alternada (CA) e podendo assim ser injetada no sistema elétrico e disponibilizada para o usuário.

Figura 2 – Diagrama esquemático de um sistema fotovoltaico.



Fonte: Mppt Solar (2016)

2.1.2 Células Fotovoltaicas

Os painéis fotovoltaicos são compostos de células fotovoltaicas. Uma célula fotovoltaica é onde se faz a conversão de energia solar em elétrica utilizando o efeito fotoelétrico ou fotovoltaico. Esse efeito é gerado através da absorção da luz solar, ou irradiância solar, que ocasiona uma diferença de potencial na estrutura do material semicondutor presente na célula [12]. É correto dizer que uma célula fotovoltaica não tem a capacidade de armazenar energia elétrica, ela apenas mantém um fluxo de elétrons num circuito elétrico enquanto existir a incidência de luz sobre ela [13]. Este fenômeno é chamado de “Efeito Fotovoltaico”.

Existem diversos materiais utilizados nos sistemas fotovoltaicos. O principal é o silício (Si), pois além de ser um bom material para se realizar um semicondutor, é também o segundo elemento químico mais abundante na terra [12]. Com a utilização do silício, existem algumas tecnologias aplicadas para a produção de células fotovoltaicas, algumas sendo mais eficientes e baratas do que outras, porém a explicação dessas diferenças e do funcionamento de cada tecnologia não está no escopo deste trabalho.

2.1.3 Inversores Fotovoltaico

Os inversores são dispositivos eletrônicos com a finalidade de fornecer energia elétrica em CA a partir de uma corrente elétrica em CC, devido ao fato dos painéis fotovoltaicos terem uma saída em CC e a rede elétrica utilizar CA. Além de realizar essa conversão, esse equipamento é capaz de ajustar a frequência e o nível de tensão gerada, para tornar possível a conexão do mesmo com a rede elétrica, de acordo com as normas estabelecidas

pela agência nacional de energia elétrica (ANEEL) [12].

2.1.4 Falhas em Sistemas Fotovoltaicos

As falhas mais comuns e recorrentes em sistemas fotovoltaicos podem ser divididas em duas grandes categorias: falhas presentes no lado CC ou no lado CA [14]. No estudo feito por Madeti e Singh[14] é apresentada uma discussão detalhada sobre essas falhas.

Falhas do lado CC são mais numerosas, justamente por estarem do lado que contém mais componentes e podem incluir: problemas com o algoritmo buscador de máxima potência, ou MPPT; falhas no diodo de *bypass*; falha de aterramento; falhas de arco; descasamento de células ou módulos (podendo este ser temporário ou permanente); circuito aberto; curto-circuito.

Já as falhas do lado CA são principalmente geradas por problema no(s) inversor(es) do sistema ou até na própria rede elétrica em si. Como o foco deste trabalho é a detecção de falha e não a identificação da mesma, a explicação exata sobre cada tipo de falha não se encontra no escopo, porém pode-se ter uma ideia sobre um sistema de identificação de falha em [15].

Além das falhas citadas acima, existem as falhas que podem gerar uma diminuição no rendimento do sistema por conta de fatores externos e do sensoriamento falho, os mais habituais sendo: efeitos de sombreamento no sistema; acúmulo de poeira ou neve na superfície dos painéis; e erros nas medições, podendo ser por conta de diferenças de tempo (*lagging*), erros de comunicação com os sensores e até mesmo falhas do próprio sensor [3].

2.2 Análise de Dados em Aprendizado de Máquina

É conhecido da literatura que o desempenho de um sistema de aprendizado de máquina é altamente limitado pela qualidade e quantidade de dados [16]. Embora vários pesquisadores foquem em melhorar a qualidade dos modelos de aprendizado de máquina, existe uma limitação do quanto isso impacta no sistema, sem considerar a qualidade dos dados. Um dos principais requisitos antes de consumir um conjunto de dados é estudar e entender o que são considerados dados “falhos” e remover os mesmos do conjunto. Se isso não for feito, pode resultar em análises imprecisas e decisões não confiáveis.

A importância de um conjunto de dados livre de erros também implica no entendimento do modelo, pois as tentativas de depurar e entender a saída do modelo devem ser fundamentadas na suposição que os dados estão adequadamente limpos. Para isso necessita-se realizar um tratamento de dados e uma visualização otimizada dos mesmos.

2.2.1 Tratamento de Dados

O tratamento de dados consiste de diversas técnicas que visam a remoção dos dados espúrios e a transformação destes para um melhor funcionamento no modelo de aprendizado de máquina, os tratamentos sendo:

- (i) A utilização de filtros digitais para remoção de ruídos ou suavização de curvas, no caso de dados contínuos;
- (ii) Implementações de limitações físicas, como por exemplo o sistema fotovoltaico estar produzindo uma potência qualquer mesmo não recebendo incidência solar ou até mesmo chegar a produzir uma potência maior do que a potência máxima na saída do inversor;
- (iii) A remoção de *outliers*, dados que divergem em muito do padrão de dados obtidos.

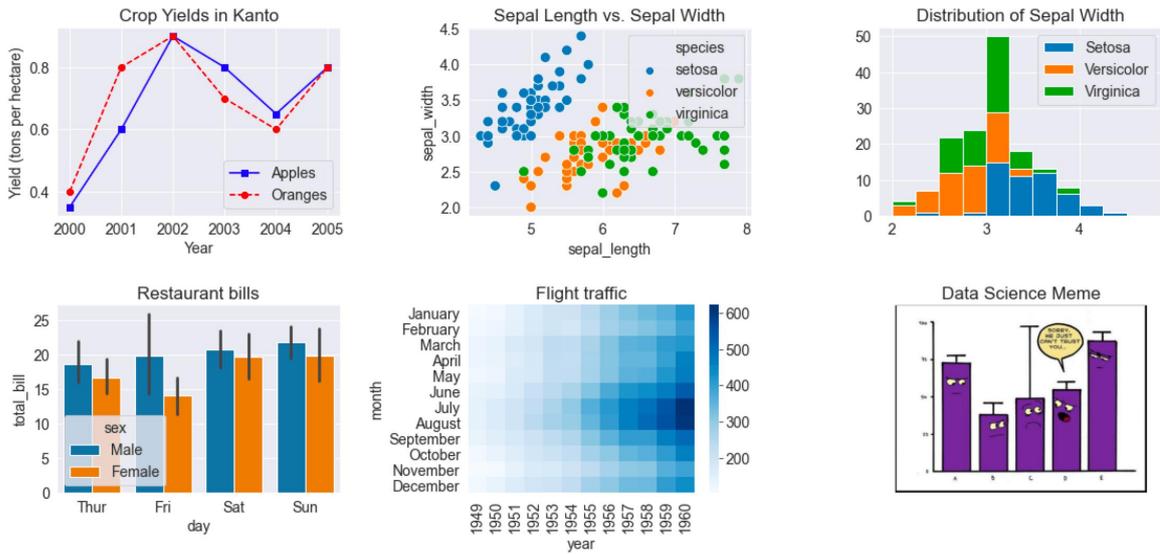
Também é possível remover ou incrementar conjuntos de dados ou características, para uma melhoria no resultado do modelo visado.

2.2.2 Visualização de Dados

Visualização de dados é uma parte extremamente importante na validação de dados e também na interpretação de resultados. Com uma boa visualização se pode verificar a existência de falhas ou inconsistências tanto nos dados iniciais, quanto nos resultados obtidos. Tendo uma visão geral de como os dados devem se comportar em períodos de funcionamento normais, é possível criar métricas e padrões para o tratamento de dados, visando melhorar a qualidade dos dados a serem utilizados nos métodos de detecção de falhas [17].

Visualizações de dados efetivas permitem ao pesquisador entender os dados trabalhados e comunicar as suas percepções e resultados para outros. Esse objetivo pode ser promovido por diversas ferramentas para especificar, por exemplo, um gráfico que forneça um bom equilíbrio entre eficiência e flexibilidade. Uma das ferramentas mais comum é a ferramenta *matplotlib*, Figura 3, uma ferramenta que está dentro do ecossistema científico da linguagem de programação Python e que será utilizada no decorrer deste projeto.

No decorrer do trabalho foram utilizadas duas importantes representações de dados, uma sendo a *fitline* e a outra sendo o *R-squared*. A *fitline* é uma equação de reta criada a partir dos dados e geralmente não passa por todos os pontos dos dados, esta representa a tendência dos dados sendo a melhor aproximação do conjunto de dados fornecido. A reta que foi utilizada no trabalho foi a reta baseada em uma regressão linear e foi desenvolvida utilizando a função *polyfit* da biblioteca *Numpy*. Já o *R-squared* é uma medida estatística em um modelo de regressão linear que determina a correlação de uma variável dependente que pode ser explicada por uma ou mais variáveis independentes, em outras palavras serve

Figura 3 – Possíveis gráficos dentro da ferramenta *matplotlib*.

Fonte: Jovian (2022)

para medir a força da relação entre seu modelo linear e as variáveis dependentes em uma escala de 0 a 100%.

2.3 Modelos de Geração

Existem vários modelos de geração de energia fotovoltaica disponíveis na literatura científica para estimar o estado operacional, variando em complexidade e precisão [2]. Estes têm o intuito de prever a produção de energia em CA do sistema fotovoltaico, usando principalmente a temperatura do painel fotovoltaico e a irradiação solar total incidente no mesmo [3]. Neste trabalho serão discutidos dois modelos genéricos, utilizados em [2] e [3].

O primeiro modelo é referente ao artigo de Platon et al.[3], um modelo mais antigo que é baseado em uma abordagem paramétrica para modelar o sistema fotovoltaico. Uma variação deste modelo foi utilizado em [3] onde a potência gerada ¹ é predita como

$$P_{ac} = G(a_1 + a_2 \times G + a_3 \times \log(G)) \times (1 + a_4(T_m - 25)), \quad (2.1)$$

onde P_{ac} é a potência CA produzida (W), G é a irradiação solar (W/m^2), T_m é a temperatura do módulo ($^{\circ}C$) e a_1, a_2, a_3 e a_4 são os coeficientes obtidos pela regressão. Esse modelo será denominado como modelo **one-equation** (OE) para a finalidade desse trabalho.

O segundo modelo tem como origem o artigo de Pombo et al.[2] e, segundo o mesmo, as principais vantagens desse método bem estabelecido são sua boa precisão, simplicidade e a

¹ Os modelos são de predição de potência, mas neste artigo é usado bastante o termo de predição de energia, uma vez que a energia e potência estão relacionadas linearmente pelo tempo

possibilidade de construí-lo usando parâmetros básicos do *datasheet* e medições meteorológicas, tornando-o adequado para ser usado rapidamente em qualquer usina fotovoltaica, sem a necessidade de uma etapa de regressão. As únicas medições de campo necessárias são irradiância, velocidade do vento e temperatura do ar. O restante das informações é coletado dos *datasheet* do sistema fotovoltaico e do inversor.

No método, usado em [2], primeiro é preciso calcular a temperatura do módulo (T_m),

$$T_m = T_a + E_{POA} \times \exp(a + b \times W_s), \quad (2.2)$$

onde T_a , W_s , a e b representam a temperatura do ar, velocidade do vento e dois coeficientes que dependem dos materiais, respectivamente. Esses dois coeficientes podem ser extraídos de [18]. E_{POA} corresponde à irradiância solar total incidente no painel fotovoltaico; como o sensor de irradiação está posicionado na mesma orientação do painel, pode-se considerar E_{POA} como a irradiação global. Em seguida é calculada a temperatura de cada célula fotovoltaica (T_c) como

$$T_c = T_m + \frac{E_{POA}}{E_{STC}} \times \Delta T, \quad (2.3)$$

onde E_{STC} representa a irradiância de referência (1000 W/m^2) e ΔT representa a diferença entre a temperatura do módulo e da célula, sendo os dois valores obtidos no *datasheet*. Após o cálculo de T_c , calcula-se a potência de saída DC como

$$P_{DC} = N_s \times N_p \times \frac{E_{POA}}{E_{STC}} \times P_{mp,STC} \times [1 + \gamma_{mp} \times (T_c - T_{STC})], \quad (2.4)$$

onde $P_{mp,STC}$, T_{STC} , γ_{mp} , N_s e N_p representam a potência de pico e temperatura do painel medidos sob condições padrão, o coeficiente de temperatura normalizado de potência de pico, o número de painéis conectados em série e paralelo, respectivamente. Todos essas variáveis estão presentes no *datasheet*. Finalmente, a eficiência do inversor é obtida a partir de uma tabela contendo 12 valores de eficiência para 12 faixas de geração de potência de saída de potência do sistema fotovoltaico.

$$P_{AC} = \begin{cases} \eta_{max} \times P_{DC}, & \text{if } \eta_{max} \times P_{DC} \leq P_{ACmax} \\ P_{ACmax}, & \text{if } \eta_{max} \times P_{DC} > P_{ACmax} \end{cases} \quad (2.5)$$

Este modelo foi denominado como **modelo físico** (MF) em [2] e também neste trabalho. O MF captura a relação entre as medições meteorológicas e a produção fotovoltaica, introduzindo efetivamente a dependência física e não-linear entre os recursos. Os autores de [2] utilizam a Equação (2.5) como entrada de um algoritmo de aprendizado de máquina, procurando aumentar ainda mais a precisão da predição da potência gerada.

2.4 Inteligência Artificial

A Inteligência Artificial é um ramo de pesquisa da ciência da computação que se propõe a estudar e criar mecanismos em sistemas computacionais que simulem a capacidade do ser

humano de pensar, demonstrando assim alguma forma de inteligência. Esses mecanismos contém sistemas que aprendem conceitos e tarefas e podem tirar conclusões úteis.

Nos últimos anos, dois ramos da IA ganharam mais popularidade, sendo eles o aprendizado de máquina (*machine learning*, do inglês) e o aprendizado profundo (*deep learning*, do inglês). O aprendizado de máquina será utilizado neste projeto, todavia, faz-se necessário o uso de uma linguagem de programação para traduzir modelos computacionais para a linguagem de máquina.

2.4.1 Python

Python é uma linguagem de programação de alto nível e de uso geral. Foi publicada em 1991 por Guido van Rossum pela primeira vez [19]. Essa linguagem tem uma ampla variedade de aplicações, podendo ser utilizada na *web*, em jogos, entre outros. Recentemente se tornou a principal ferramenta de trabalho de programação para *machine learning* e *deep learning*, justamente o motivo que levou este trabalho a utilizar Python como a linguagem de programação.

2.4.2 Aprendizado de Máquina

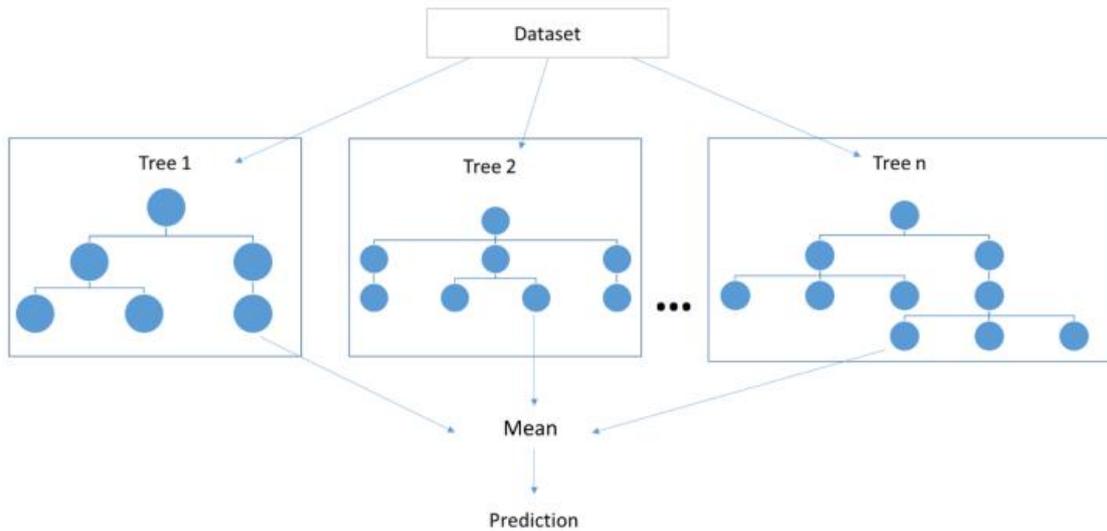
Enquanto a IA pode ser definida, amplamente, como a mimetização de habilidades humanas, o aprendizado de máquina é uma parte da IA que se especializa em treinar máquinas para aprender com dados. Com as atuais tecnologias de conectividade, como a Internet das coisas, pode-se dizer que não há recurso em maior abundância do que dados. Segundo Raschka e Mirjalili[20], na segunda metade do século 20 o aprendizado de máquina evoluiu como uma sub-área da IA com algoritmos de auto aprendizagem que derivam conhecimento a partir de dados para fazer previsões futuras diversas.

2.4.2.1 *Random Forest*

O algoritmo de *Random Forest* (RF) é uma técnica de aprendizado de máquina supervisionada [21]. Este algoritmo pode ser empregado tanto em problemas de classificação onde a saída alvo é um rótulo ou para problemas de regressão onde a saída alvo é um número.

O treinamento do algoritmo consiste na construção de uma árvore binária, conceito vindo de estruturas de dados, onde cada nó interno, fora da folha, é um teste de um atributo, cada ramificação desse nó é o resultado do teste e cada nó folha contém um rótulo da classe [21]. Um exemplo sucinto da construção de uma *Random Forest* pode ser visualizado na Figura 4.

O objetivo não é criar um único modelo de aprendizado e sim múltiplos modelos, onde a saída final será uma função de todas as saídas intermediárias de cada árvore [21].

Figura 4 – Exemplo do algoritmo de *Random Forest*

Fonte: Deep AI (2020)

É possível, também, definir o número de árvores utilizado pelo algoritmo, tipicamente entre 100 e 1000 árvores, enquanto a separação dos atributos usados em cada árvore é randômica.

Os autores de [2] realizam uma comparação de diversos algoritmos de *Machine Learning* que buscam prever a produção de energia AC do sistema fotovoltaico em questão. Eles concluíram que o *random forest* é um dos melhores algoritmos que pode ser utilizado na base de dados explorada em [2] e também utilizada neste projeto.

2.5 Detecção de Falhas

O algoritmo de detecção de falhas utilizado neste projeto foi o algoritmo de [3], o qual explora um valor de potência gerada e uma predição instantânea da potência. Platon et al.[3] fez uma divisão da potência gerada pela potência predita obtendo assim uma razão para cada medida. Com os dados utilizados no treinamento dos modelos, que não contém falhas consideráveis, eles obtém limites que supostamente delimitam dados normais de dados falhos. Esses limites são calculados utilizando as seguintes equações:

$$\text{Limite Inferior} = \mu - 3 \times \sigma \quad (2.6)$$

$$\text{Limite Superior} = \mu + 3 \times \sigma \quad (2.7)$$

onde μ e σ correspondem à média e ao desvio padrão, respectivamente, dos valores da razão entre a potência gerada e a potência predita. O intervalo de mais ou menos três desvios padrões foi escolhido pois, se os dados forem normalmente distribuídos, cerca de 99,7% dos pontos estarão dentro dos limites, sendo assim os valores encontrados fora dessa

faixa podem ser considerados valores que não seguem a distribuição estatística da maior parte dos dados e conseqüentemente seriam uma falha.

O valor da raiz quadrada do erro médio (RMSE) foi utilizado em primeira instância para avaliar a precisão dos métodos de predição de geração. Sendo este calculado como raiz quadrada da média dos quadrados do erro para cada observação de validação, normalizado para a média dos valores de potência AC medidos do conjunto de dados de teste ou validação [3].

2.6 Indicadores de Classificação

A utilização de métodos de classificação para sumarizar a performance de um modelo é possível através de indicadores que refletem diferentes aspectos sobre os resultados deste modelo. Porém para o cálculo destes aspectos, primeiro é necessário entender o conceito de “Matriz de Confusão”, ilustrada na Tabela 1 .

Tabela 1 – Exemplo de Matriz de Confusão.

		Valor Predito	
		Sim	Não
Valor Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: Autor

Tem-se as seguintes definições:

- **Verdadeiro Positivo (TP)** são os casos em que o modelo previu corretamente como classe positiva.
- **Falso Positivo (FP)** são os casos em que o modelo previu incorretamente como classe positiva.
- **Verdadeiro Negativo (TN)** são os casos em que o modelo previu corretamente como classe negativa.
- **Falso Negativo (FN)** são os casos em que o modelo previu incorretamente como classe negativa.

Com os conceitos acima, é possível classificar a performance de um modelo de acordo com os diferentes indicadores [22]:

- **Precisão.** Entre todas as classificações classe Positivo que o modelo fez, quantas estão corretas:

$$Prec = \frac{TP}{TP + FP} \quad (2.8)$$

- **Acurácia.** Entre todas as classificações tanto classe Positivo quanto classe Negativo, quantas o modelo classificou corretamente:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

- **Especificidade.** Entre todas as classificações de classe Positivo como valor esperado, quantas estão corretas:

$$Especificidade = \frac{TN}{TN + FP} \quad (2.10)$$

- **Recall.** Entre todas as classificações de classe Negativo como valor esperado, quantas estão corretas:

$$Recall = \frac{TP}{TP + FN} \quad (2.11)$$

3 Metodologia

O presente capítulo abordará a metodologia utilizada na realização deste trabalho, seguindo a ordem de elucidar as diferentes opções de bases de dados e a escolha de uma única a ser utilizada, o pré-processamento da base, análise dos dados obtidos na base dados e por fim a implementação dos diferentes modelos de detecção de falha no sistema fotovoltaico e as suas características.

3.1 Bases de Dados

No decorrer deste trabalho foram testadas algumas bases de dados porém utilizada apenas a base mais completa e detalhada, facilitando assim a aplicação e visualização do trabalho de forma geral. A base é de livre acesso (pública) até o presente momento, e consta com os seguintes dados essenciais: a produção de potência CA de um sistema fotovoltaico, informações físicas do inversor, do painel e das células, e também informações meteorológicas (temperatura ambiente e velocidade do vento). Uma reamostragem ou *resample* do período de medidas pode e foi implementada para ambas bases.

3.1.1 Base de Dados da SOLETE

A base de dados SOLETE foi a base selecionada para o desenvolvimento desse projeto, ela foi disponibilizada pelo pesquisador Pombo[10] do laboratório SYSLAB que é uma parte da empresa DTU Elektro e foi criada com o intuito de aumentar a transparência e replicabilidade [2]. O conjunto de dados é composto por um total de três arquivos, dois com uma taxa de amostragem de 5 minutos, sendo um deles o completo e outro apenas um pequeno trecho e o último um arquivo com uma taxa de amostragem de uma hora, o qual também é do período completo.

SOLETE inclui 15 meses de medidas, iniciando no dia 01/06/2018 e terminando no dia 01/09/2019. Esta base dispõe das informações de: *timestamp*, temperatura do ar, umidade relativa, pressão atmosférica, velocidade do vento, direção do vento, irradiância horizontal global, irradiância no plano do painel, e a potência ativa registrada de uma turbina eólica Gaia de 11 kW e um inversor fotovoltaico de 10 kW. Nem todos os gráficos contidos na base serão utilizados, principalmente os dados referentes a turbina eólica. Pode ser visualizada a organização do *dataset* na Figura 5.

A base de dados já fornece um conjunto de códigos em Python com diversas funções e ferramentas para auxiliar o uso da mesma, contendo funções para visualização, tratamento e expansão dos dados e até mesmo funções para treinamento de diversos algoritmos de aprendizado de máquina utilizados em [2]. Porém, o próprio autor cita que uniu o conjunto

Figura 5 – Organização dos dados da base de dados SOLETE.

	TEMPERATURE[degC]	HUMIDITY[%]	WIND_SPEED[m/s]	\
2018-06-01 04:25:00	14.700000	0.7	1.062667	
2018-06-01 04:30:00	14.753000	0.7	0.930333	
2018-06-01 04:35:00	14.892333	0.7	1.249333	
2018-06-01 04:40:00	14.803000	0.7	1.438667	
2018-06-01 04:45:00	14.946333	0.7	1.421333	

	WIND_DIR[deg]	GHI[kWh/m2]	POA Irr[kWh/m2]	P_Gaia[kW]	\
2018-06-01 04:25:00	75.906667	0.114327	0.050053	0.0	
2018-06-01 04:30:00	87.523333	0.122823	0.053703	0.0	
2018-06-01 04:35:00	101.496667	0.130153	0.055893	0.0	
2018-06-01 04:40:00	91.920000	0.136300	0.057490	0.0	
2018-06-01 04:45:00	98.243333	0.144747	0.058943	0.0	

	P_Solar[kW]	Pressure[mbar]
2018-06-01 04:25:00	0.271940	1016.659998
2018-06-01 04:30:00	0.289920	1016.746334
2018-06-01 04:35:00	0.304993	1016.786328
2018-06-01 04:40:00	0.316803	1016.860010
2018-06-01 04:45:00	0.330323	1016.906356

Fonte: Autor

de dados e os *scripts* em apenas um dia, e que por esse fato eles não são 100% livres de erros ou otimizados. Os códigos fornecidos foram utilizados como base ou parâmetro para o desenvolvimento deste trabalho.

3.2 Pré-processamento da Base de Dados

Esta seção serve para esclarecer os processos feitos inicialmente na base de dados SOLETE. Estes processos não são feitos para modificar a natureza dos dados da base, apenas limpá-la e organizá-la da melhor maneira possível, removendo dados espúrios e irregulares.

O pré processamento dos dados foi realizado sobre o arquivo que contém os dados completos com o tempo de leitura de 5 minutos. Esses dados foram transformados em um *dataframe* utilizando uma função da biblioteca Pandas, tornando possível a visualização e manipulação dos mesmos. Agora é possível visualizar quais são as colunas contidas no *dataframe* e a quantidade total de medidas do mesmo, como podemos ver na Figura 6. Também é possível reparar que o índice desta base já se encontra com o *timestamp* e no modelo padrão utilizado na biblioteca Pandas, facilitando a etapa de pré processamento.

Com o auxílio do *datasheet* do inversor foram obtidos e salvos as variáveis físicas do mesmo, podemos visualizar os dados na Tabela 2. Outro dado salvo foi a eficiência em (%) do sistema em fotovoltaico para certas faixas de produção de potência CA, como

Figura 6 – Informações sobre a base de dados SOLETE.

```
Quantidade de medidas 131617
Index(['TEMPERATURE[degC]', 'HUMIDITY[%]', 'WIND_SPEED[m/s]', 'WIND_DIR[deg]',
      'GHI[kW/m2]', 'POA_Irr[kW/m2]', 'P_Gaia[kW]', 'P_Solar[kW]',
      'Pressure[mbar]'],
      dtype='object')
DatetimeIndex(['2018-06-01 00:00:00', '2018-06-01 00:05:00',
              '2018-06-01 00:10:00', '2018-06-01 00:15:00',
              '2018-06-01 00:20:00', '2018-06-01 00:25:00',
              '2018-06-01 00:30:00', '2018-06-01 00:35:00',
              '2018-06-01 00:40:00', '2018-06-01 00:45:00',
              ...,
              '2019-08-31 23:15:00', '2019-08-31 23:20:00',
              '2019-08-31 23:25:00', '2019-08-31 23:30:00',
              '2019-08-31 23:35:00', '2019-08-31 23:40:00',
              '2019-08-31 23:45:00', '2019-08-31 23:50:00',
              '2019-08-31 23:55:00', '2019-09-01 00:00:00'],
             dtype='datetime64[ns]', length=131617, freq=None)
```

Fonte: Autor

Tabela 2 – Informações contidas no *datasheet* do inversor.

Variável	Valor
Tipo	Poly-cristaline
Estc	1000 W/m ²
Tstc	25°C
Pmp_stc	200W
Ganma_mp	-4,78E-03
Ns	18
Np	2
a	-3,56
b	-0,075
Delta_T	3

Fonte: Autor

Tabela 3 – Informações sobre as faixas de rendimento do sistema fotovoltaico.

eff_P	0	250	400	450	500	600	650	750	825	1000	1200	1600	2000	3000	4000	6000	8000	10000
eff_%	0,0	85,5	90,2	90,9	91,8	92	92,3	94	94,4	94,8	95,6	96,0	97,3	97,7	98,0	98,1	98,05	98,0

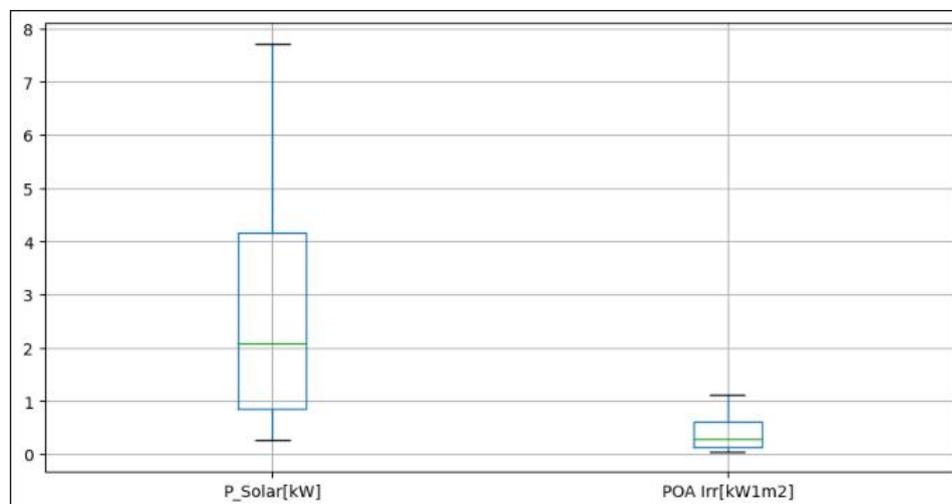
Fonte: Autor

podemos ver na Tabela 3, os quais foram retirados dos códigos disponíveis de [2].

Outra etapa foi a eliminação de observações correspondentes a níveis muito baixos de luz solar, menor de 50 W/m², conseqüentemente removendo todos os valores nulos de irradiação. Esta faixa necessita ser removida pois, segundo Platon et al.[3], a precisão da medida é significativamente reduzida dentro dela, gerando assim muitas instabilidades para os modelos. Após a eliminação, foram traçadas as medidas de potência e irradiação em um gráfico do tipo *boxplot* como mostrado na Figura 7, onde podemos perceber que

não existe nenhum *outlier*, dada a ausência de dados fora dos limites superior e inferior.

Figura 7 – Gráfico estilo Boxplot da potência AC e da irradiação no plano do painel fotovoltaico.



Fonte: Autor

3.3 Modelos de Detecção de Falhas

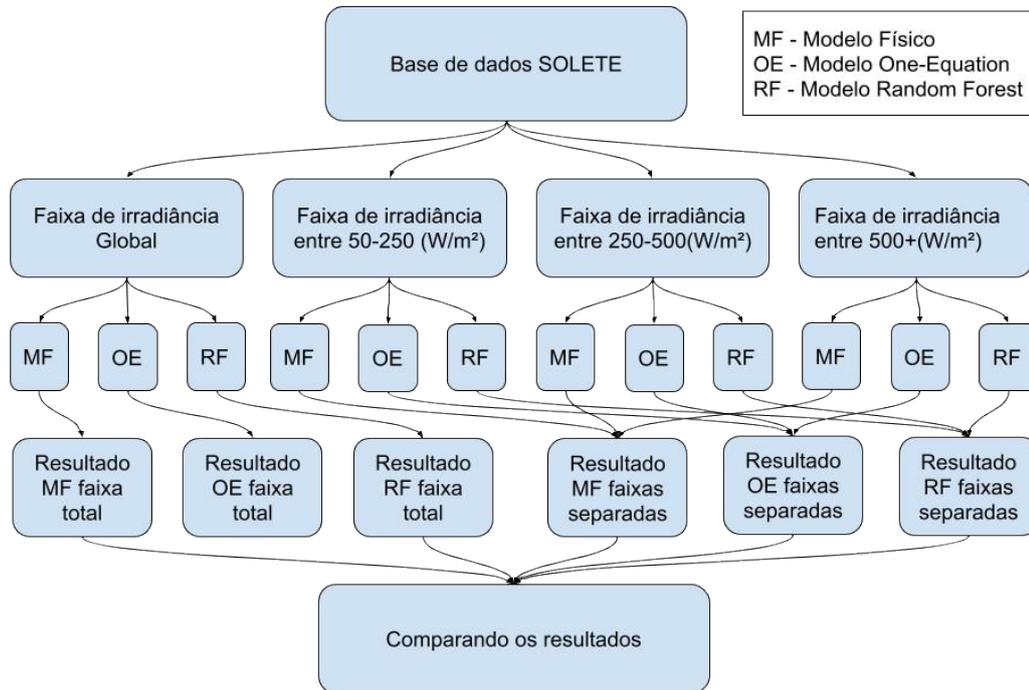
Nesta seção será tratada a metodologia dos modelos de predição implementados, o primeiro sendo o modelo físico, o segundo utilizando o algoritmo *random forest* de aprendizado de máquina e o último o modelo *one equation* (OE), todos os modelos foram treinados e testados para os dados com período de uma hora. Foi utilizado para ambos os modelos a mesma separação de dados para treinamento e para teste, onde, aproximadamente, o primeiro ano da série de dados foi utilizado para o treinamento dos modelos (70%) e os últimos 5 meses foram utilizados para o teste (30%). A seleção dos dados não foi feita de forma aleatória para permitir que o período de um ano seja completamente considerado no treinamento dos modelos.

Além disso, assim como em [3], para captar o comportamento do sistema fotovoltaico de acordo com os níveis de luz solar, modelos para diferentes intervalos de irradiância foram desenvolvidos. Podemos visualizar um diagrama esquemático geral da maneira que os modelos foram utilizados neste trabalho na Figura 8.

3.3.1 Modelo Físico

O modelo físico foi implementado utilizando as equações (2.2-2.5) presentes no Capítulo 2. A sua implementação foi simples, rápida e objetiva, além de obter o valor de potência CA predita, também foram obtidos e armazenados os valores das temperaturas do módulo e da célula de cada medida, para uso no treinamento do modelo RF.

Figura 8 – Diagrama esquemático geral dos modelos utilizados



Fonte: Autor

Além disso, nos 70% dos dados utilizados para treinamento, foi a partir da razão da potência AC gerada pela potência P_{AC} obtida nesta subseção que foram calculados os limites inferiores e superiores, como mostrados nas equações (2.6) e (2.7), respectivamente.

3.3.2 Modelo RF

O modelo RF foi implementado utilizando a biblioteca *sklearn*, uma das bibliotecas mais utilizadas em trabalhos com *machine learning* atualmente. No primeiro momento foi necessário obter os dados tratados e sem falhas com as medidas a cada hora, os mesmos dados normais obtidos na seção de análise dos dados. Após a obtenção dos dados, separamos os mesmos em dois *dataframes* diferentes, um contendo as características possíveis de serem utilizadas no treinamento e outro contendo a característica a ser predita, que no caso é a potência AC gerada.

Podemos visualizar na Figura 9 as características possíveis a serem utilizadas no treinamento, *dataframe* X e a característica a ser predita, *dataframe* y. Um fato interessante é que temos presente, o valor da predição de P_{AC} , gerado anteriormente pelo MF, incluso nestas características, pois segundo os resultados de [2], esse valor melhora o resultado do algoritmo de RF. Além de P_{AC} , também temos a temperatura ambiente, umidade, irradiância solar no plano do painel, temperatura do módulo, temperatura da célula e a

irradiância solar horizontal como características de treinamento.

Figura 9 – Características escolhidas para treinamento e teste do modelo RF

```
df2 = DATA.copy()
X = df2[['TEMPERATURE[degC]', 'HUMIDITY[%]', 'POA Irr[kW1m2]', 'Pac', 'TempModule', 'TempCell', 'GHI[kW1m2]']]
y = df2[['P_Solar[kW]']]]
```

Fonte: Autor

Após a separação das características, podemos realizar a separação dos *dataframes*, onde agora selecionamos o tamanho e o modo de como será feita a separação de dados para treinamento e dados para teste. O tamanho, como definido anteriormente, ficou 70% dos dados para treinamento e 30% para teste, foi escolhido um número de 100 de árvores ou *random states* para não gerar uma carga computacional muito grande, mas que da mesma maneira, gerou uma predição muito eficaz. Para essa separação foi utilizada a função *train_test_split()* da biblioteca *sklearn*.

De início foi realizado o treinamento para a faixa global de irradiância e salvo as predições e resultados de *R-squared* e RMSE, após isso foram separados os dados em três faixas de irradiância e realizado o mesmo processo descrito acima para cada uma destas faixas. Foram salvos os *dataframes* das potências AC preditas nos dados de teste em arquivos csv.

3.3.3 Modelo OE

O modelo OE foi feito a partir de [3], onde temos uma função com quatro coeficientes a_1 , a_2 , a_3 e a_4 , e dois valores do inversor, sendo eles a irradiação solar no painel e a temperatura do módulo. Após a separação dos *dataframes* de treinamento e teste, tem que ser realizado uma regressão no *dataframe* de treinamento, para que se obtenha os melhores coeficientes possíveis para realizar a predição da potência gerada.

Para a realização deste modelo no Python, foi utilizada a função *curve_fit* da biblioteca *Spicy*. Para a utilizar a *curve_fit* é necessário definir uma função que receba as variáveis independentes e retorne a partir destas a equação que queremos encontrar os coeficientes, esta é a mesma Equação (2.1) e pode ser visualizada na Figura 10. Após definir esta função, foram obtidas as variáveis independentes e dependentes do *dataframe* de treinamento e armazenadas em um *Array*. Com isso podemos chamar a função *curve_fit* passando os parâmetros na seguinte ordem: (i) A função retornando a equação pretendida de atingir; (ii) Uma lista de dois *Arrays*, um contendo a irradiação e outro a temperatura; (iii) Um *Array* contendo a potência gerada.

A função irá retornar uma lista de valores que são as variáveis a_1 , a_2 , a_3 e a_4 . Com essas variáveis é possível agora aplicar as mesmas em uma equação com os dados de teste para realizar a predição da potencia para o mesmo período. Após essa etapa basta dividir

Figura 10 – Função que retorna a expressão de Pac para o uso da função *curve_fit*

```
def func(data, a, b, c, d):  
    irr = data[0] * 1000 # Precisa ser em W/m²  
    tm = data[1]  
    return irr * (a + b * irr + c * np.log(irr)) * (1 + d * (tm - 25))
```

Fonte: Autor

a potência gerada pela potência predita e obter a razão que é utilizada para detectar as falhas.

4 Resultados

No presente capítulo serão demonstrados e discutidos os resultados obtidos no decorrer deste trabalho. Primeiro serão apresentados os resultados perante o valor RMSE de todos os modelos, por último teremos a comparação da detecção de falha de cada modelo tendo em vista os indicadores de classificação apresentados no Capítulo 2. Em todas as etapas serão feitas observações e conclusões sobre os resultados, além de sugestões para futuros trabalhos.

Temos alguns fatos importantes a serem observados, o primeiro é que para os dados de treinamento, foi realizada uma detecção de valores absurdos com base em limites estipulados a partir da *fitline*, que foi obtida pela relação incrementalmente linear entre potência gerada e irradiância solar, pois a correlação entre estas características é muito alta. Por outro lado, pela falta de falhas nos dados da base SOLETE, foi necessário realizar uma inserção de falhas artificiais para o teste dos modelos e uma pré detecção das mesmas. Estas falhas foram inseridas obtendo os valores de potência gerada no período de uma semana e os reduzindo em 30%, com a semana contida nos dados de testes para validar a detecção das mesmas.

4.1 Comparando os Valores de RMSE

Os primeiros valores obtidos e a serem comparados foram os valores de RMSE (%), na Tabela 4. Esses valores derivam da comparação dos valores gerados e preditos nos dados de testes, esses são os dados que não contém as falhas inseridas, obtendo assim o valor mais real possível para comparar os modelos de predição.

Tabela 4 – Valores de RMSE dos modelos de predição.

Irradiâncias (W/m ²)	Dados por faixa	Modelo RF - RMSE (%)	Modelo OE - RMSE (%)	Modelo Físico - RMSE (%)
Faixa Global	100%	3,368	3,805	4,989
Faixa 50-250	48,46%	0,426	0,390	1,167
Faixa 250-500	23,85%	3,347	4,048	3,854
Faixa 500+	27,68%	10,043	11,026	13,490

Fonte: Autor

Como podemos ver, no modelo RF temos a grande maioria dos menores valores de RMSE, justamente por ser o modelo mais complexo e conseqüentemente com mais acurácia na predição. Em seguida vem o modelo OE que, em comparação com o MF, no geral teve resultados melhores. Esses resultados, como esperado, seguem o padrão de que quanto mais complexo o modelo utilizado para fazer a predição, menor será o valor de RMSE.

Contudo, como os valores RMSE dos modelos estão consideravelmente próximos, podemos dizer que todos os modelos são bons preditores e que os modelos de baixa complexidade, MF e OE, são dois bons modelos para uma implementação online.

4.2 Detecção de Falha

Após a comparação dos valores de RMSE, foi realizado a detecção de falhas de acordo com o modelo de [3]. Os limites foram definidos a partir dos dados de treinamento e a predição foi realizada nos dados de teste, utilizando o método da média e dos desvios padrões. Primeiro serão tratados os resultados referentes a cada modelo de predição, dentro da faixa global e das faixas separadas de irradiação solar, e por último serão calculados e demonstrados os indicadores de classificação de cada modelo, com comparação e a conclusão sobre os resultados obtidos.

Os resultados de cada modelo podem ser visualizados por três gráficos que contém a razão da potência gerada por potência predita no eixo das ordenadas e a irradiância no eixo das abscissas, esses gráficos são: (i) Gráfico referente a detecção de falhas de todos as medidas dos dados testes incluindo a falha; (ii) Gráfico com apenas as falhas previamente detectadas e os limites do mesmo; (iii) Gráfico com apenas os dados normais previamente detectados. No gráfico (ii) é possível detectar os falsos positivos e no gráfico (iii) conseguimos detectar os falsos negativos.

Vale a pena ressaltar que em todos os gráficos gerados no trabalho, os pontos azuis (ou em tons azulados) são os pontos normais (não falhas) enquanto que os pontos vermelhos são as falhas, podendo ser as falhas inseridas ou as falhas detectadas pelos limites definidos, dependendo do gráfico observado.

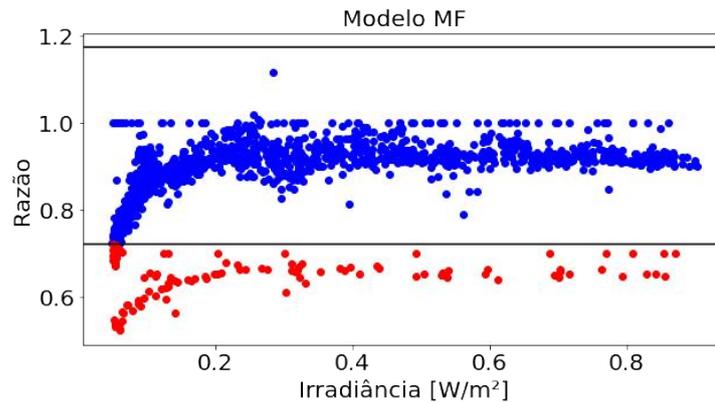
4.2.1 Modelo Físico

Nesta seção serão apresentados os gráficos referentes ao modelo físico de Pombo et al.[2], tanto da faixa global quanto das faixas separadas, serão obtidos os valores de TP, TN, FN e FP e será feita uma breve conclusão sobre os resultados. As conclusões serão rediscutidas na seção de comparação.

4.2.1.1 Faixa Global

Podemos visualizar na Figura 11, onde pontos azuis referem-se a dados considerados normais e os vermelhos aos dados considerados falhas, a detecção de falhas considerando toda a faixa de irradiância. Já é possível verificar um certo defeito no modelo físico, onde em irradiâncias baixas ele contém em sua fórmula uma eficiência maior do que a real, o que acaba gerando a detecção de muitas falhas que provavelmente são valores normais.

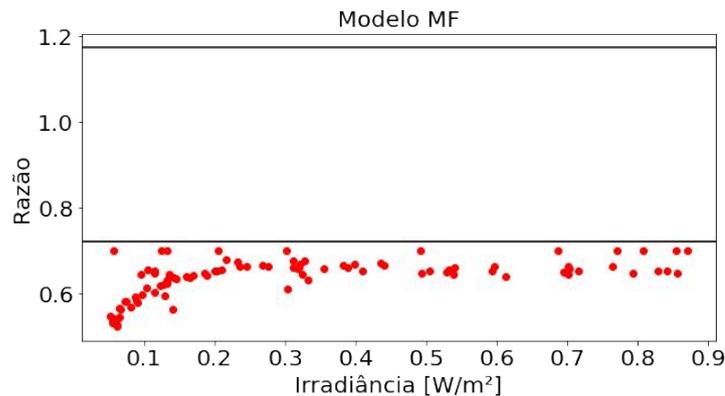
Figura 11 – Detecção de falhas dos pontos no modelo MF na faixa global



Fonte: Autor

Na Figura 12 é possível verificar a inexistência de falhas dentro dos limites normais de operação, ou seja, não há nenhum falso negativo. O significado é que esta detecção conseguiu identificar 100% das falhas utilizando a faixa global de irradiância, um resultado inesperado para o modelo de mais baixa complexidade.

Figura 12 – Falhas inseridas e visualizadas nos limites do modelo físico na faixa global

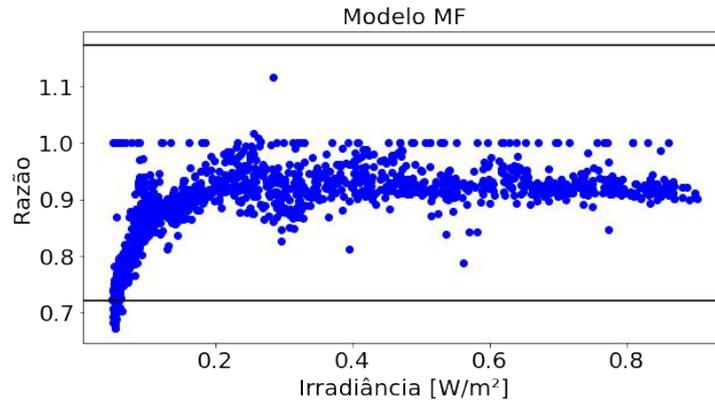


Fonte: Autor

Já na Figura 13 pode-se visualizar todos os pontos normais detectados anteriormente dentro dos limites, observando este gráfico concluiu-se que realmente foram detectados vários pontos de falsos positivos, que são valores normais (azuis) fora dos limites de detecção, na faixa de irradiação mais baixa, o que demonstra que o modelo acabou não funcionando muito bem nesta faixa.

Com base nos gráficos obtidos acima, no modelo MF na faixa global temos os seguintes indicadores de classificação: 25 FP, 0 FN, 1287 TP e 94 TN. Com esses valores, como visto no Capítulo 2, podemos calcular: (i) uma precisão de 98,09%; (ii) uma acurácia de 98,22%; (iii) uma especificidade de 78,99% e (iv) um *Recall* de 100,00%.

Figura 13 – Pontos normais nos limites do modelo físico na faixa global

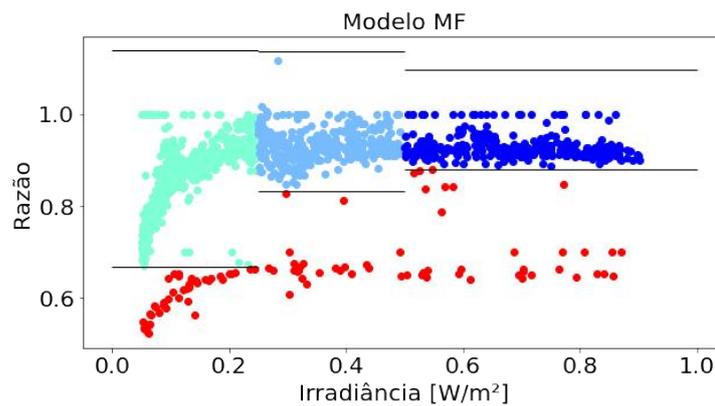


Fonte: Autor

4.2.1.2 Faixas Separadas

Agora tem-se os gráficos das faixas separadas, como visto anteriormente, os dados foram separados por faixas de irradiação, foi realizado o treinamento do modelo com uma parte destes dados e depois unidos os resultados em um *dataframe* único. Na Figura 14 pode-se visualizar a detecção de falhas pelo método da média e do desvio padrão em todas as medidas, onde as medidas normais estão em tons azulados e as falhas identificadas se encontram em vermelho. Observa-se que os limites tem uma tendência a serem menores para irradiação solar maiores, pois neste modelo quanto maior a irradiação menos espalhados estão as medidas da razão e, conseqüentemente, menor é o desvio padrão.

Figura 14 – Detecção pela média e desvio padrão de todos os pontos no modelo físico nas faixas separadas

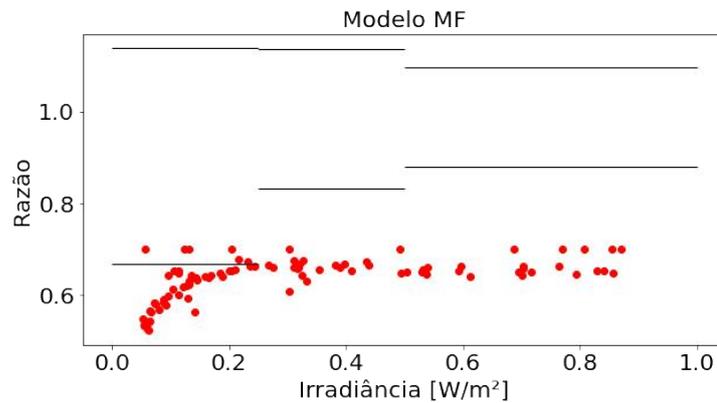


Fonte: Autor

Já na Figura 15, pode-se visualizar as falhas previamente detectadas dentro destes novos limites. Ao contrário da faixa global, é possível identificar-se alguns pontos de falsos negativos na faixa inicial, onde valores que são falhas foram considerados valores normais utilizando a detecção

Por fim, na Figura 16 pode-se identificar os falsos positivos e verdadeiros positivos do

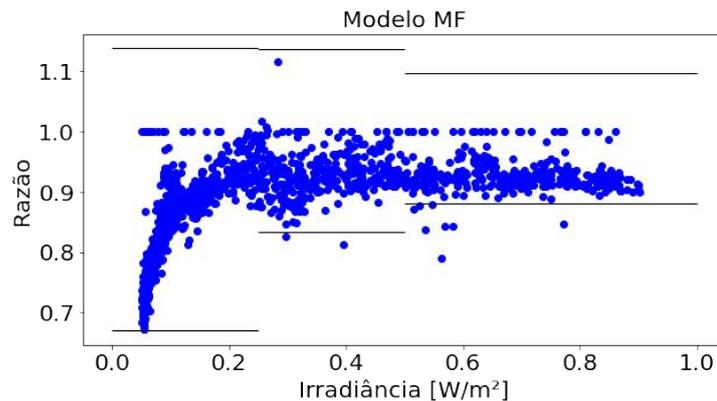
Figura 15 – Falhas inseridas e visualizadas nos limites do modelo físico nas faixas separadas



Fonte: Autor

modelo físico nas faixas separadas, pode-se ver que em comparação com o modelo na faixa global este teve uma melhoria significativa em termos de falsos positivos, principalmente nas baixas irradiações.

Figura 16 – Valores normais nos limites do modelo físico nas faixas separadas



Fonte: Autor

Com base nos gráficos obtidos acima, no modelo MF nas faixas separadas tem-se os seguintes indicadores de classificação: 10 FP, 6 FN, 1302 TP e 88 TN. Com esses valores, como visto no Capítulo 2, pode-se calcular: (i) uma precisão de 99,24%; (ii) uma acurácia de 98,86%; (iii) uma especificidade de 89,80% e (iv) um *Recall* de 99,54%.

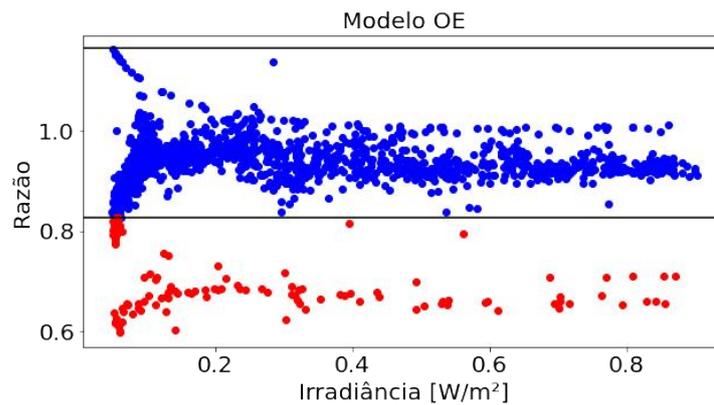
4.2.2 Modelo *One-equation*

Nesta seção serão apresentados os gráficos referentes ao modelo OE de [3], tanto da faixa global quanto das faixas separadas, serão obtidos os valores de TP, TN, FN e FP e será feita uma breve conclusão sobre os resultados.

4.2.2.1 Faixa Global

No primeiro momento tem-se os gráficos da faixa global, na Figura 17 pode-se verificar que o resultado é similar ao do modelo MF, apenas com uma diferença nos limites de detecção. Aparentemente, em irradiâncias baixas, tem-se o mesmo problema do modelo MF, onde o modelo em si não considera muito bem a eficiência em irradiâncias abaixo de 100 W/m^2 .

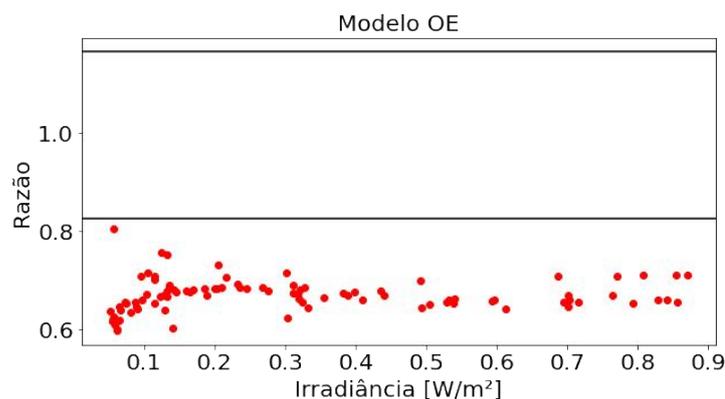
Figura 17 – Detecção de falhas dos pontos no modelo OE na faixa global



Fonte: Autor

Já na Figura 18, assim como na faixa global do modelo MF, verifica-se que o sistema conseguiu identificar 100% das falhas e, até o momento, parecido com o modelo MF. Considerando que as falhas estão com 70% do seu valor original este já é um resultado considerado bom para a complexidade do modelo de predição.

Figura 18 – Falhas inseridas e visualizadas nos limites do modelo OE na faixa global

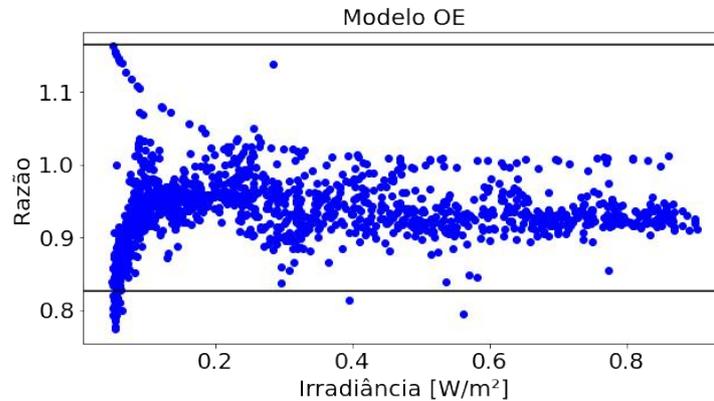


Fonte: Autor

Finalmente na Figura 19 pode-se visualizar a existência de valores falsos positivos, assim como no modelo MF, em valores de irradiação solar baixos e até alguns falsos positivos “perdidos” em irradiações maiores. Por ser um modelo que tem um RMSE maior do que o modelo MF, acaba-se por ter os limites um pouco menores, resultando

assim em um número um pouco maior de falsos positivos do que no modelo MF para a mesma faixa.

Figura 19 – Valores normais nos limites do modelo físico na faixa global



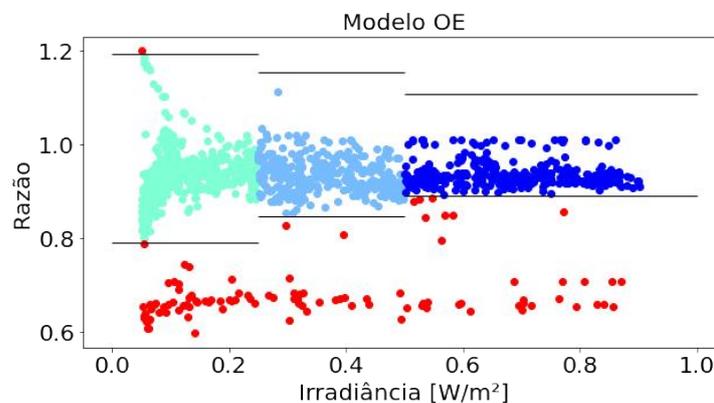
Fonte: Autor

Com base nos gráficos obtidos acima, no modelo OE na faixa global temos os seguintes indicadores de classificação: 27 FP, 0 FN, 1285 TP e 94 TN. Com esses valores, como visto no Capítulo 2, pode-se calcular: (i) uma precisão de 97,94%; (ii) uma acurácia de 98,08%; (iii) uma especificidade de 77,68% e (iv) um *Recall* de 100,00%.

4.2.2.2 Faixas Separadas

Com as faixas separadas tem-se os mesmos gráficos gerados, na Figura 20 pode-se visualizar a detecção de falhas pelo método da média e do desvio padrão em todas as medidas. Observa-se que neste modelo os limites continuam seguindo a tendência de serem menores para irradiação solar maiores, pois quanto maior a irradiação, menor é o desvio padrão da razão.

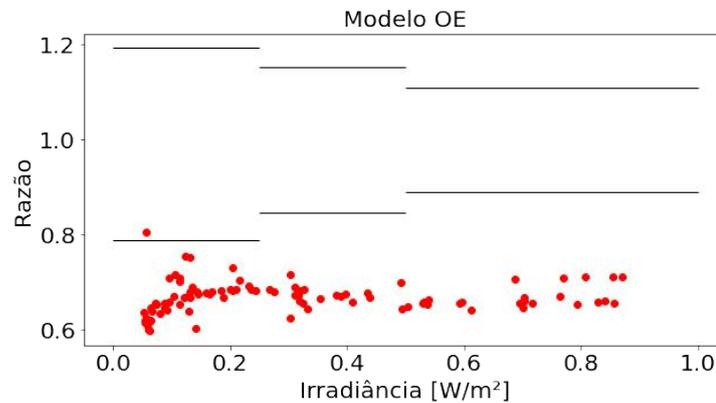
Figura 20 – Detecção pela média e desvio padrão de todos os pontos no modelo OE nas faixas separadas



Fonte: Autor

Em seguida tem-se a Figura 21, na qual verificamos uma acurácia na detecção de falhas de quase 100%, tendo apenas um valor detectado como valor normal. Um resultado menor em comparação com o modelo MF nas mesmas condições.

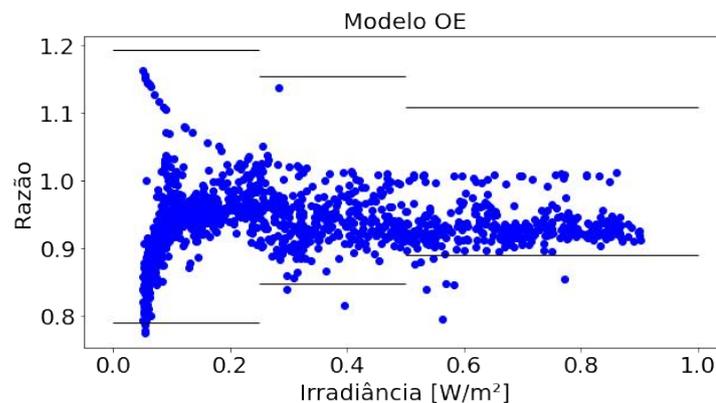
Figura 21 – Falhas inseridas e visualizadas nos limites do modelo OE nas faixas separadas



Fonte: Autor

Contudo, na Figura 22, é possível identificar um leve aumento nos pontos de valores normais considerados como falhas, em comparação com os gráficos do modelo MF.

Figura 22 – Valores normais nos limites do modelo físico nas faixas separadas



Fonte: Autor

Com base nos resultados obtidos acima, no modelo OE nas faixas separadas temos os seguintes indicadores de classificação: 12 FP, 1 FN, 1300 TP e 93 TN. Com esses valores, como visto no Capítulo 2, podemos calcular: (i) uma precisão de 99,08%; (ii) uma acurácia de 99,07%; (iii) uma especificidade de 88,57% e (iv) um *Recall* de 99,92%.

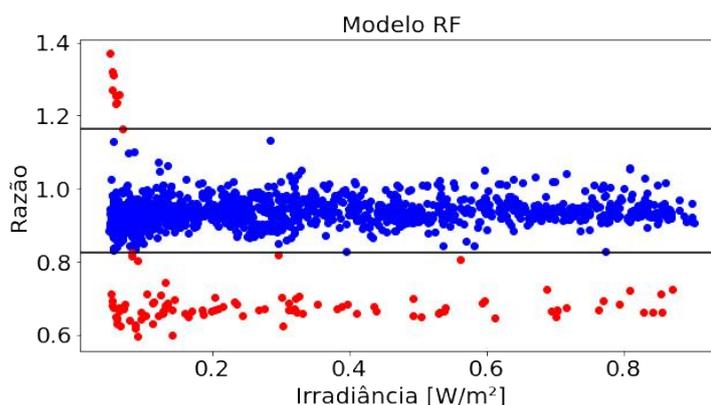
4.2.3 Modelo *Random Forest*

Nesta seção serão apresentados os gráficos referentes ao modelo RF, tanto da faixa global quanto das faixas separadas, serão obtidos os valores de TP, TN, FN e FP e será feita uma breve conclusão sobre os resultados.

4.2.3.1 Faixa Global

Como padrão, primeiro iremos tratar da faixa global, na Figura 23 podemos ver o gráfico dos dados das razões obtidas pela divisão da potência gerada pela potência prevista pelo modelo RF com os limites de detecção utilizando o método da média e dos desvios padrões. Observa-se que por ser um modelo mais complexo temos uma dispersão um pouco menor dos dados em comparação com os últimos dois modelos.

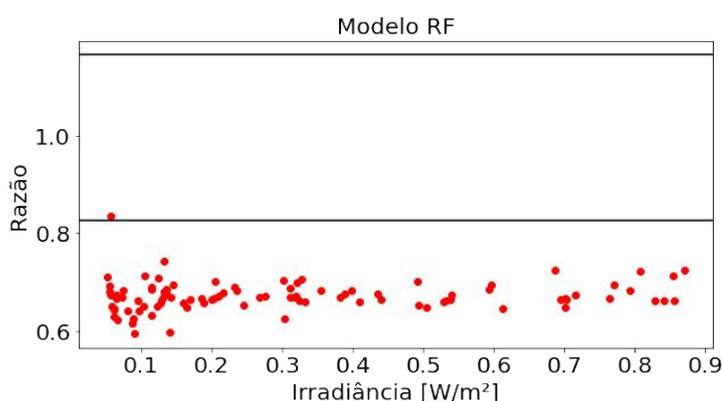
Figura 23 – Detecção de falhas dos pontos no modelo RF na faixa global



Fonte: Autor

Em seguida, na Figura 24, tem-se apenas as falhas detectadas anteriormente dentro dos limites, neste ponto percebe-se que existe apenas um ponto no qual uma das falhas não foi identificada, ao contrário dos outros dois modelos que foram identificadas 100% das mesmas.

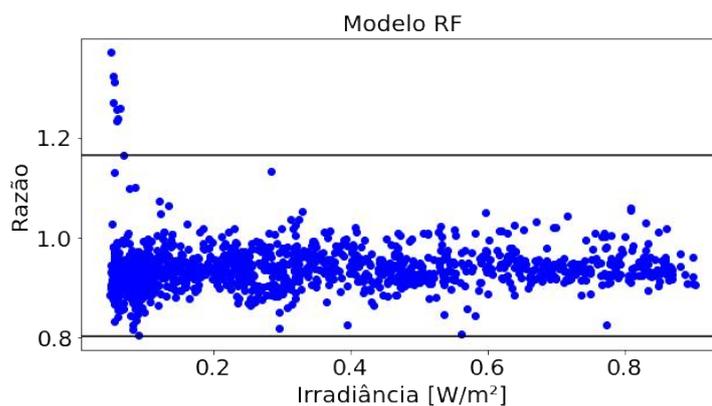
Figura 24 – Falhas inseridas e visualizadas nos limites do modelo RF na faixa global



Fonte: Autor

Contudo, na Figura 25, pode-se identificar um valor de falsos positivos muito menor em comparação com outros modelos, o que se justificaria pela menor dispersão dos dados. Ainda que a maioria desses falsos positivos foram obtidos acima do limite superior, o que os tornam dados sem muito significado, estes ainda foram contabilizados no decorrer do projeto.

Figura 25 – Valores normais nos limites do modelo físico nas faixas separadas



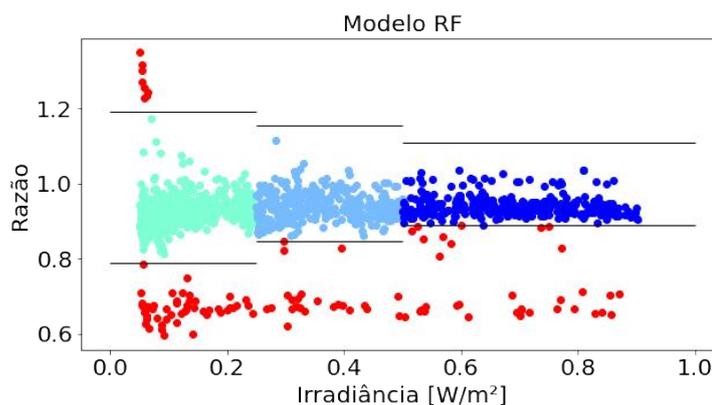
Fonte: Autor

Com base nos gráficos obtidos acima, no modelo RF na faixa global temos os seguintes indicadores de classificação: 9 FP, 1 FN, 1303 TP e 93 TN. Com esses valores, como visto no Capítulo 2, pode-se calcular: (i) uma precisão de 99,31%; (ii) uma acurácia de 99,29%; (iii) uma especificidade de 91,18% e (iv) um *Recall* de 99,92%.

4.2.3.2 Faixas Separadas

Nesta subseção serão tratados os resultados do modelo RF para as faixas separadas de irradiação solar. Começando com a Figura 26, na qual podemos verificar o modelo de detecção de falhas com as condições consideradas, nela pode-se ver que provavelmente existirão mais falsos positivos em comparação com o modelo RF na faixa global.

Figura 26 – Detecção pela média e desvio padrão de todos os pontos no modelo OE nas faixas separadas

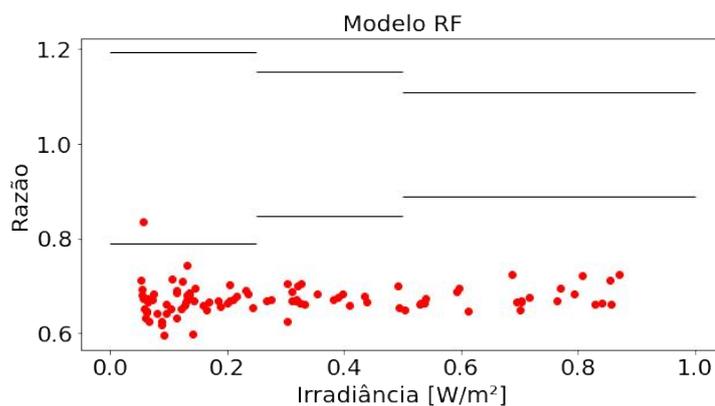


Fonte: Autor

Assim como no modelo OE, tem-se na Figura 27 a detecção de praticamente todas as falhas verificadas, apenas com uma delas ficando de fora.

Por último, na Figura 28, observa-se que foram obtidos mais falsos positivos em relação ao mesmo modelo com a faixa global. Pode-se dizer que alguns pontos ficaram mais

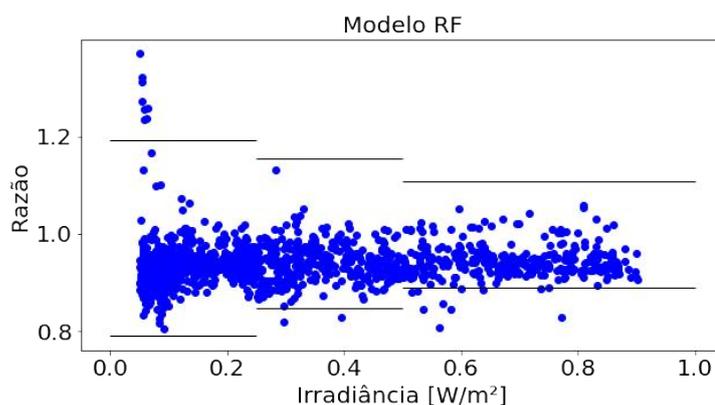
Figura 27 – Falhas inseridas e visualizadas nos limites do modelo RF nas faixas separadas



Fonte: Autor

espalhados do que o grande padrão que acabou resultando neste aumento.

Figura 28 – Valores normais nos limites do modelo RF nas faixas separadas



Fonte: Autor

Com base nos gráficos obtidos acima, no modelo RF nas faixas separadas tem-se os seguintes indicadores de classificação: 9 FP, 1 FN, 1303 TP e 93 TN. Com esses valores, como visto no Capítulo 2, pode-se calcular: (i) uma precisão de 98,32%; (ii) uma acurácia de 98,36%; (iii) uma especificidade de 80,87% e (iv) um *Recall* de 99,92%.

4.2.4 Comparando Resultados

Nesta seção serão reunidos os resultados anteriores de uma forma que fique melhor a visualização e comparação entre cada modelo, sendo os resultados apresentados em forma de diferentes tabelas. Ao mesmo tempo serão abordadas hipóteses e discussões sobre os resultados obtidos no decorrer do trabalho. Todos os modelos tiveram resultados positivos de modo geral, e conseqüentemente nenhum deles apresentou uma detecção ineficaz com relação às falhas. Os modelos acabaram obtendo resultados melhores em determinados aspectos e piores em outros e além disso foram encontrados pontos interessantes a serem

discutidos no sistema de detecção de falhas de [3] e também algumas considerações que foram implementadas.

A primeira comparação é feita entre os três modelos na faixa global de irradiação. Tem-se na Tabela 5 os resultados dos indicadores de classificação dos mesmos, apesar de todos os valores serem considerados bastantes elevados, um fato que chama atenção de início é o baixo valor da especificidade dos modelos MF e OE. Isto ocorre, pois em ambos modelos existe uma certa dificuldade de predição para valores de irradiância abaixo de 150 W/m^2 , onde os valores normais acabam ultrapassando os limites inferiores definidos pelos modelos, resultando em um número consideravelmente alto de falsos positivos.

Tabela 5 – Valores dos indicadores de classificação dos três modelos na faixa global de irradiação solar.

Modelo	Faixa Global		
	MF (%)	OE (%)	RF (%)
Precisão	98,09	97,94	99,31
Acurácia	98,22	98,08	99,29
Especificidade	78,99	77,68	91,18
<i>Recall</i>	100,00	100,00	99,92

Fonte: Autor

Em comparação com os resultados obtidos em [3] que obteve 17 falsos positivos em 1585 pontos de operação normal no qual foi utilizado uma base de dados própria, utilizando a base de dados SOLETE e com a inserção de falha realizada neste projeto, obteve-se 25 e 27 em 1312 pontos de operação normal, para os modelos MF e OE respectivamente. Isto demonstra um leve aumento percentual no número de falsos positivos, porém em compensação os modelos MF e OE conseguiram identificar 100% das falhas na faixa global, enquanto o artigo citado na mesma faixa apenas obteve 81,13% na taxa de detecção de falhas. A diferença entre falsos negativos pode vir a decorrer na diferença do *dataset* utilizado, por ser um sistema fotovoltaico com sensores mais recentes ou do modo de pré detecção de falhas que, como visto anteriormente, acabou sendo irreplicável a pré detecção “visual” de Platon et al.[3].

Como pode-se visualizar na Tabela 5, o modelo RF acabou tendo o melhor resultado médio, apesar de não ter reconhecido uma das 94 falhas. Além disso excluindo os resultados de especificidade, os outros dois modelos tiveram resultados bem atraentes em decorrência da sua baixa complexidade e facilidade de implementação.

A segunda comparação pode ser visualizada na Tabela 6, onde tem-se os resultados dos indicadores de classificação dos três modelos no método das faixas separadas de irradiação solar. A primeira discussão que pode-se ter é sobre o fato do modelo RF, nas condições especificadas, ter resultado em uma detecção de falhas pior em relação à média dos classificadores. Este foi um fato muito importante no trabalho que, a partir dele, foi realizada uma conclusão que pode ser utilizada para melhora futura.

Tabela 6 – Valores dos indicadores de classificação dos três modelos nas faixas separadas de irradiação solar.

Faixas Separadas			
Modelo	MF (%)	OE (%)	RF (%)
Precisão	99,24	99,08	98,32
Acurácia	98,86	99,07	98,36
Especificidade	89,80	88,57	80,87
<i>Recall</i>	99,54	99,92	99,92

Fonte: Autor

Após visualizar os gráficos das faixas separadas e tentar entender o motivo do fato que o modelo RF teve uma pior detecção do que os outros dois, percebeu-se que este fato decorre da definição dos limites de detecção, onde obtém-se a média e o desvio padrão para os *datasets* de treinamento. Como o modelo RF se trata de um modelo de alta complexidade e não se baseia em uma única fórmula como os modelos OE e MF, os limites do RF para os próprios dados de treinamento acabam por ser mais estreitos em comparação com os outros dois modelos. Assim, após utilizar a predição para os dados de testes, apesar de obter uma predição melhor, tem-se limites mais fechados que fazem os resultados do modelo RF ficar para trás dos resultados dos outros dois modelos no geral. Para resolver isto podemos considerar uma mudança na obtenção destes limites, aumentando o número que multiplica os desvios padrões ou obter esses limites dos dados de teste, nos quais mesmos os modelos mais complexos não terão limites tão apertados.

Apesar do fato observado acima, tem-se um incremento na detecção de falhas dos modelos MF e OE e, principalmente, no classificador de especificidade, pois com a separação de faixas temos limites mais amplos na faixa de irradiação entre 50 e 250 W/m², o que resultou em uma diminuição considerável de falsos positivos nestes dois modelos. Contudo, em decorrência do mesmo fato, acabou-se tendo um leve aumento nos falsos negativos de ambos, fato que foi mais ampliado no modelo MF, o que porém não afetou em demorado o resultado final dos classificadores.

Pode-se tirar a conclusão que a qualidade da detecção de falhas, em casos específicos, pode piorar em conta de um modelo mais complexo, fato que pode-se usar para considerar um estudo mais detalhado sobre os limites impostos pelo modelo de [3]. Outro fato foi que os modelos menos complexos, MF e OE, nas faixas de irradiação separadas, obtiveram resultados comparáveis ao modelo RF na faixa de irradiação global. Excluindo a irregularidade do modelo RF, no geral, a divisão entre faixas menores de irradiação resultou em uma melhoria na detecção de falhas, em comparação com os modelos globais. Por fim, mesmo com a diferença significativa entre as falhas presentes neste trabalho e as no artigo [3], onde as falhas deste trabalho têm uma diminuição de apenas 30% do valor de potência e no artigo de Platon et al.[3] as falhas eram valores praticamente nulos de potência, pode-se perceber uma detecção de falhas com uma alta precisão.

5 Conclusões

Este trabalho propôs o estudo e desenvolvimento sobre um sistema de detecção de estado de falhas em sistemas fotovoltaicos presente no artigo [3], com a base de dados presente em [2] e três modelos de predição de potência. Durante o período onde foram conduzidos os estudos, notou-se a importância do monitoramento e detecção de falhas em sistemas fotovoltaicos, em uma sociedade que está cada vez mais tendendo ao uso da energia limpa.

O enfoque principal do trabalho foi o desenvolvimento de seis abordagens diferentes, contendo três modelos de predição de potência gerada CA e dois modos de divisão das faixas de irradiação. Estas abordagens foram utilizadas no modelo de detecção de falhas advindo do artigo [3]. Tem-se o modelo de regressão para uma equação pré determinada, o modelo de *Random Forest* de *machine learning* e o modelo de equações com base em dados físicos do inversor. Estes modelos preditivos foram separados em dois modos, um deles prevendo a potência para a faixa de irradiação total nos dados e outro para diferentes faixas pré-determinadas de irradiação solar.

As seis abordagens descritas neste trabalho alcançaram resultados interessantes, porém os resultados obtidos nas abordagens: (i) Modelo RF no modo faixa global; (ii) Modelo OE no modo faixas separadas; (iii) Modelo MF no modo faixas separadas, mostram os melhores resultados no geral. Os resultados e análises desenvolvidas concluem que o método criado por Platon et al.[3], apesar de não ter uma taxa de acerto de 100%, acaba alcançando os objetivos pretendidos, mesmo utilizando outra base de dados e com uma falha inserida muito menos impactante e de mais difícil detecção do que a original.

Como trabalhos futuros, sugere-se a exploração de diferentes modelos de predição, como outros modelos de *machine learning* e a melhoria do método de obtenção dos limites de detecção de falhas, para que se obtenha resultados melhores em modelos de predição mais complexos. Além disso, pode-se evoluir o sistema de detecção de falhas para que o mesmo se torne um sistema de identificação de falhas, incluindo novos limites de detecção e até mesmo para realizar uma predição de futuras falhas no sistema.

Referências Bibliográficas

- 1 Marcelo Roubicek. *O crescimento da energia solar no Brasil. E o seu futuro*. Nexojornal, 2022. Acessado em 02 novembro 2022. Disponível em: <<https://www.nexojornal.com.br/expresso/2022/09/29/O-crescimento-da-energia-solar-no-Brasil.-E-o-seu-futuro>>. 9, 14
- 2 POMBO, D. V. et al. Benchmarking physics-informed machine learning-based short term pv-power forecasting tools. *Energy Reports*, Elsevier Ltd, v. 8, p. 6512–6520, 11 2022. ISSN 23524847. 14, 15, 16, 21, 22, 24, 27, 29, 31, 35, 47
- 3 PLATON, R. et al. Online fault detection in pv systems. *IEEE Transactions on Sustainable Energy*, Institute of Electrical and Electronics Engineers Inc., v. 6, p. 1200–1207, 10 2015. ISSN 19493029. 14, 15, 16, 19, 21, 24, 25, 29, 30, 32, 35, 38, 45, 46, 47
- 4 O Setor Elétrico. *2022: o melhor ano da energia solar no Brasil*. ABSOLAR, 2022. Acessado em 02 novembro 2022. Disponível em: <<https://www.absolar.org.br/noticia/2022-o-melhor-ano-da-energia-solar-no-brasil/>>. 14
- 5 MODELING and fault diagnosis of a photovoltaic system. *Electric Power Systems Research*, v. 78, p. 97–105, 1 2008. ISSN 03787796. 15
- 6 FIRTH, S.; LOMAS, K.; REES, S. A simple model of pv system performance and its use in fault detection. *Solar Energy*, v. 84, n. 4, p. 624–635, 2010. ISSN 0038-092X. International Conference CISBAT 2007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0038092X0900187X>>. 15
- 7 GAWRE, S. Pv module fault detection diagnosis. v. 5, 05 2018. 15
- 8 CHOUDER, A. et al. Modeling and simulation of a grid connected pv system based on the evaluation of main pv module parameters. *Simul. Model. Pract. Theory*, v. 20, p. 46–58, 2012. 15
- 9 BENEDETTI, M. et al. Anomaly detection and predictive maintenance for photovoltaic systems. *Neurocomputing*, v. 310, 05 2018. 15
- 10 POMBO, D. V. *The SOLETE dataset*. 2022. Retrieved from DTU-Data, <https://data.dtu.dk/articles/dataset/The_SOLETE_dataset/17040767>, DOI: 10.11583/DTU.17040767. Disponível em: <https://data.dtu.dk/articles/dataset/The_SOLETE_dataset/17040767>. 16, 27
- 11 MACHADO, C. T.; MIRANDA, F. S. Photovoltaic solar energy: A briefly review. *Revista Virtual de Química*, Sociedade Brasileira de Química, v. 7, p. 126–143, 1 2015. ISSN 19846835. 17
- 12 ALMEIDA, E. et al. Energia solar fotovoltaica: Revisão bibliográfica. 2015. 17, 18, 19
- 13 NASCIMENTO, C. A. do. Princípio de funcionamento da célula fotovoltaica. 2004. 18

- 14 MADETI, S. R.; SINGH, S. N. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Solar Energy*, Elsevier Ltd, v. 158, p. 161–185, 2017. ISSN 0038092X. 19
- 15 MUSTAFA, Z. et al. Fault identification for photovoltaic systems using a multi-output deep learning approach. *Expert Systems with Applications*, v. 211, p. 118551, 2023. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417422016207>>. 19
- 16 OVERVIEW and Importance of Data Quality for Machine Learning Tasks. In: . [S.l.]: Association for Computing Machinery, 2020. p. 3561–3562. ISBN 9781450379984. 19
- 17 WASKOM, M. seaborn: statistical data visualization. *Journal of Open Source Software*, The Open Journal, v. 6, p. 3021, 4 2021. 20
- 18 KING, D. L.; BOYSON, W. E.; KRATOCHVILL, J. A. Sandia report photovoltaic array performance model. 2004. Disponível em: <<http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>>. 22
- 19 KUHLMAN, D. A python book: Beginning python, advanced python, and python exercises. 2009. Disponível em: <<http://www.davekuhlman.org>>. 23
- 20 RASCHKA, S.; MIRJALILI, V. *Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2*. [S.l.: s.n.], 2015. 741 p. ISBN 9781789955750. 23
- 21 Deep AI. *Deep Learning*. 2022. Acessado em 06 out 2022. Disponível em: <<https://deepai.org/>>. 23
- 22 TING, K. M. *Encyclopedia of Machine Learning and Data Mining*. [S.l.]: Springer US, 2011. ISBN 978-0-387-30164-8. 25