



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO DE JOINVILLE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE SISTEMAS
ELETRÔNICOS

ANGELO ABDALLAH PELISSON

**APRENDIZADO DE MÁQUINA PARA PREVISÃO DE GERAÇÃO DE ENERGIA
FOTOVOLTAICA EM DADOS DE ESTAÇÕES SOLARIMÉTRICAS**

DISSERTAÇÃO DE MESTRADO

Joinville
2021

Angelo Abdallah Pelisson

**APRENDIZADO DE MÁQUINA PARA PREVISÃO DE GERAÇÃO DE ENERGIA
FOTOVOLTAICA EM DADOS DE ESTAÇÕES SOLARIMÉTRICAS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia de Sistemas Eletrônicos.

Orientador: Prof. Dr. Pablo Andretta Jaskowiak

Coorientador: Prof. Dr. Anderson Wedderhoff Spengler

Joinville

2021

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

Pelisson, Angelo

Aprendizado de máquina para previsão de geração de energia fotovoltaica em dados de estações solarimétricas / Angelo Pelisson ; orientador, Pablo Andretta Jaskowiak, coorientador, Anderson Wedderhoff Spengler, 2021.

110 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Campus Joinville, Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos, Joinville, 2021.

Inclui referências.

1. Engenharia de Sistemas Eletrônicos. 2. Energia Fotovoltaica. 3. Aprendizado de Máquina. 4. Estações Solarimétricas. 5. Previsão de Geração de Energia. I. Andretta Jaskowiak, Pablo . II. Wedderhoff Spengler, Anderson . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos. IV. Título.

Angelo Abdallah Pelisson

Aprendizado de máquina para previsão de geração de energia fotovoltaica em dados de estações solarimétricas

O presente trabalho em nível de Mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Dr. Diego Santos Greff
Universidade Federal de Santa Catarina

Prof. Dr. Thiago Ferreira Covões
Universidade Federal do ABC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Engenharia de Sistemas Eletrônicos.

Prof. Dr. Moisés Ferber de Vieira Lessa
Coordenador do Programa

Prof. Dr. Pablo Andretta Jaskowiak
Orientador

Joinville, 06 de Setembro de 2021.

Este trabalho é dedicado aos meus queridos pais e aos meus estimados avós, principalmente ao meu avô Osmar Antônio Pelisson que não está mais presente conosco.

AGRADECIMENTOS

Ao longo dos anos do mestrado e de muito estudo, em primeiro lugar gostaria de agradecer aos meus pais, Nohad Abdallah e Manoel Angelo Martins Pelisson, que sempre me apoiaram e estiveram ao meu lado nos momentos mais difíceis. Aos meus irmãos Gabriel e Lucas que sempre estiveram ao meu lado e me deram forças ao longo do caminho percorrido.

Aos meus professores, especialmente, o meu orientador Pablo Andretta Jaskowiak, pelos incentivos, amizade e dedicação ao meu trabalho, estando sempre disposto a me apoiar e ensinar.

Aos meus amigos da universidade Jesuíno, Thiago, Victor, Gabriel e Luís Fernando.

Aos meus amigos de Londrina do grupo Calas, Felipe, Vinícius, Daniel, Zufi, João, Kaio, Fausto, André, Hélder, Matheus, Pedro, Frederico e Leonardo.

Agradeço a minha namorada, Giovana Souza por sempre estar ao meu lado e me dar suporte nos momentos mais difíceis dessa caminhada, por estar comigo desde o começo e me apoiar em todas as minhas decisões.

Agradeço ao professor Ondrej Vaculin e seu aluno Franz Plashkies, pelo apoio e dedicação ao trabalho que foi desenvolvido no estágio de pesquisa realizado na *Technische Hochschule Ingolstadt* (THI), na Alemanha. A coordenadora do programa ENGLOBE, Lisa Hermsen pelo apoio e suporte dado do início ao final do estágio. Agradeço também ao *Deutscher Akademischer Austauschdienst* (DAAD) pelo suporte financeiro para que o estágio fosse realizado. Agradeço também professor Maurício de Campos Porath da Universidade Federal de Santa Catarina, pelo suporte dado com a documentação e os processos necessários para que o intercâmbio realizado fosse possível.

Agradeço também a Universidade Federal de Santa Catarina, por ter disponibilizado sua infraestrutura e apoio institucional. Ao Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por ter fornecido suporte financeiro para o desenvolvimento do trabalho.

RESUMO

O aumento da demanda por energia elétrica em âmbito mundial é resultado do avanço tecnológico e econômico, levando ao desenvolvimento e adoção de novas fontes de energia, sendo de particular interesse a energia fotovoltaica. Com o desenvolvimento dos sistemas fotovoltaicos de geração surgem alguns desafios, como, a sua produção intermitente, ligada a forte dependência de fatores ambientais, climáticos, e meteorológicos. A produção de energia inconstante destes sistemas pode resultar em problemas de confiabilidade, segurança, e planejamento para os produtores e concessionárias de distribuição de energia elétrica. Com o intuito de aumentar a confiabilidade do sistema como um todo, uma alternativa é a adoção de abordagens que realizam previsões acerca da energia produzida em sistemas fotovoltaicos. Este trabalho tem como objetivo estudar, aplicar e comparar estratégias de previsão de geração de energia baseadas em Aprendizado de Máquina para sistemas fotovoltaicos. Estas investigações são realizadas a partir de dados coletados por uma estação solarimétrica composta por sensores específicos. As estratégias investigadas podem ser divididas em duas grandes categorias: (i) estratégias baseadas em métodos individuais e (ii) estratégias baseadas em comitês de métodos. No primeiro caso, investigou-se o desempenho individual de diferentes métodos considerando variações em seus parâmetros e nos subconjuntos de atributos utilizados na descrição do problema. No segundo caso, investigou-se o uso de comitês de métodos com seu treinamento/aplicação baseado no tipo climático dos dias sob investigação, a fim de obter previsões mais direcionadas e precisas. A comparação de modelos individuais sugere melhores resultados com o uso de Redes Neurais Artificiais, especificamente com o uso da rede *Multilayer Perceptron* (MLP). A estratégia baseada em comitês apresentou melhores resultados quando comparada à estratégia de modelos individuais. Em particular, os melhores resultados foram observados quando os comitês foram empregados juntamente com conjuntos Fuzzy para a realização da categorização dos dias. O trabalho apresenta ainda uma proposta de estação solarimétrica de baixo custo que pode vir a ser utilizada para estudos futuros.

Palavras-chave: Sistemas Fotovoltaicos. Aprendizado de Máquina. Estação solarimétrica. Previsão de geração de energia.

ABSTRACT

The worldwide increase in electricity demand is a result of technological and economic advances, leading to the development and adoption of new energy sources. Among these, photovoltaic energy is of particular interest. Alongside the development of photovoltaic generation systems challenges emerge, noticeably related to their intermittent production which, in turn, is strongly dependent on environmental, climatic, and meteorological factors. The inconstant energy production of these systems can result in reliability, safety, and planning problems for power producers, distributors and consumers. In order to increase the reliability of the system as a whole, one alternative is to adopt approaches to forecast the expected amount of energy in photovoltaic systems. This work aims to study, employ and compare forecasting strategies based on Machine Learning for photovoltaic systems. These investigations are performed based on data collected by a solarimetric station composed of specific sensors. The strategies under investigation can be broadly divided into two categories: (i) individual method-based strategies and (ii) ensemble-based strategies. In the first case, the individual performance of different methods was investigated, considering variations in their parameters and in the subsets of attributes used in the problem description. In the second case, the use of ensembles was investigated. In this setting, methods were trained/applied based on the daily weather characteristics, in order to obtain targeted and accurate forecasts. The comparison of individual models suggests better results with the use of Artificial Neural Networks, specifically with the use of Multilayer Perceptron (MLP). Ensemble strategies showed better results when compared to the individual models. In particular, the best results were observed when ensembles were employed together with Fuzzy Sets to performing the categorization of days. The work also presents a proposal for a low-cost solarimetric station that can be used for future studies.

Keywords: Photovoltaic systems. Machine Learning. Solarimetric station. Power generation forecast

LISTA DE FIGURAS

Figura 2.1 – Painel fotovoltaico e o circuito elétrico equivalente de uma célula solar composta por um único diodo.	26
Figura 2.2 – Diagrama dos principais materiais de construção dos módulos fotovoltaicos.	27
Figura 2.3 – Componentes fundamentais da radiação solar.	29
Figura 2.4 – Diagrama esquemático do pireliômetro.	33
Figura 2.5 – Diagrama esquemático do piranômetro.	34
Figura 2.6 – Estação de medição das componentes de irradiação solar.	35
Figura 2.7 – All-Sky Imager montado em um rastreador solar.	36
Figura 2.8 – TSI(a) e CIR-7(b).	37
Figura 2.9 – Medidor de energia Solar TENMARS TM -201(a) e Termômetro Digital MS2101(b).	37
Figura 2.10 – Anemômetro Kestrel 3000(a) e Medidor de luz BK Model 615(b).	38
Figura 2.11 – Anemômetro WSD-100.	38
Figura 3.1 – Diagrama esquemático do piranômetro desenvolvido por Martínez, Andújar e Enrique (2009).	42
Figura 3.2 – Microcontrolador ESP32.	44
Figura 3.3 – Esquemático da placa eletrônica com o ESP32.	45
Figura 3.4 – Esquemático do circuito eletrônico montado na placa de ensaio.	46
Figura 3.5 – Teste de leitura do sinal gerado por um dos sensores.	47
Figura 3.6 – Envio de dados para base de dados hospedada na nuvem.	48
Figura 3.7 – Exemplo de estação solarimétrica de baixo custo.	48
Figura 3.8 – Corte da peça projetada(a) e peça impressa(b).	49
Figura 3.9 – Placa de acrílico responsável por fixar os sensores.	49
Figura 3.10 – Peça impressa e difusor de acrílico.	49
Figura 3.11 – Estrutura da estação com a placa de acrílico e os sensores fixados.	50
Figura 3.12 – Caixa de armazenamento das placas de circuito impresso.	50
Figura 3.13 – Estação solarimétrica de baixo custo.	51
Figura 3.14 – Imagem 3D das faces da placa eletrônica de aquisição de dados.	53
Figura 3.15 – Esquemático da placa eletrônica do sistema de aquisição de dados.	54
Figura 4.1 – Modelo de um neurônio artificial.	59
Figura 4.2 – Modelo de uma MLP com uma camada escondida.	60
Figura 4.3 – Arquitetura de um modelo <i>Stacking Ensemble</i>	64
Figura 4.4 – Função de pertinência forma trapezoidal.	66
Figura 4.5 – Função de pertinência forma triangular.	66
Figura 6.1 – Perfil médio diário de Janeiro e Junho de 2018.	78
Figura 6.2 – Perfil médio diário de Janeiro e Junho de 2019.	78

Figura 6.3 – Perfil médio diário de Janeiro e Junho de 2020.	79
Figura 6.4 – Método de validação por avanço não ancorado.	80
Figura 6.5 – Valores esperados e previstos da potência de saída para diferentes dias. No sentido horário, no canto superior esquerdo: 27/04/18, 11/07/2018, 12/12/2018 e 28/08/18.	83
Figura 6.6 – Valores dos atributos do subconjunto SA 1 para todos os registros de data e hora de 11/07/2018.	84
Figura 6.7 – Exemplo de perfis diários de irradiância dos tipos de dias considerados no trabalho.	85
Figura 6.8 – Método de validação por avanço ancorado.	86
Figura 6.9 – Processo de treinamento dos modelos individuais e <i>Stacking Ensemble</i>	87
Figura 6.10 – Processo de treinamento do modelo <i>Stacking Ensemble</i> de Intervalos Fuzzy Ponderados.	88
Figura 6.11 – Função de pertinência <i>fuzzy</i> na forma trapezoidal.	89
Figura 6.12 – Média das previsões semanais para os modelos que obtiveram o melhor desempenho.	94
Figura 6.13 – <i>Boxplot</i> descrevendo os erros individuais de cada modelo para os tipos climáticos de dias.	95
Figura 6.14 – Valores esperados e preditos para potência de saída e valores de irradiância para diferentes tipos dias. Da primeira linha para a última: 06/04/2020, 10/12/2018 e 27/04/2018.	96

LISTA DE TABELAS

Tabela 2.1 – Tabela de relação entre parâmetros e instrumentos.	40
Tabela 3.1 – Modelos dos dispositivos semicondutores utilizados como sensores.	43
Tabela 3.2 – Tabela de custo dos materiais utilizados na estação solarimétrica de baixo custo.	52
Tabela 6.1 – Parâmetros coletados pela estação solarimétrica e potência produzida pela planta PV.	76
Tabela 6.2 – Subconjuntos de entrada	81
Tabela 6.3 – Melhores resultados obtidos com cada modelo e conjunto de entrada.	82
Tabela 6.4 – Tabela para construção da função de pertinência fuzzy trapezoidal.	89
Tabela 6.5 – Resultados obtidos pelos modelos utilizando método de validação não ancorado.	90
Tabela 6.6 – Resultados obtidos pelos modelos utilizando método de validação ancorado.	90

LISTA DE ABREVIATURAS E SIGLAS

PV	<i>Photovoltaic</i>
GHI	<i>Global Horizontal Irradiance</i>
DNI	<i>Direct Normal Irradiance</i>
DHI	<i>Diffuse Horizontal Irradiance</i>
AT	<i>Ambient Temperature</i>
UV	<i>Ultraviolet Index</i>
WS	<i>Wind Speed</i>
PO	<i>Power Output</i>
MLP	<i>Multilayer Perceptron</i>
SVR	<i>Support Vector Regressor</i>
kNN	<i>k Nearest Neighbors</i>
GB	<i>GradientBoost</i>
LR	<i>Linear Regression</i>
AB	<i>AdaBoost</i>
DT	<i>Decision Tree</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
PCB	<i>Printed Circuit Board</i>
RMSE	<i>Root Mean Square Error</i>
MAE	<i>Mean Absolute Error</i>
RNA	<i>Redes Neurais Artificiais</i>
RBF	<i>Radial Basis Function</i>
SA	<i>Subconjunto de Atributos</i>
W	<i>Watts</i>

SUMÁRIO

1	INTRODUÇÃO	21
1.1	JUSTIFICATIVA	22
1.2	OBJETIVOS	22
1.2.1	Objetivos Específicos	22
1.3	ORGANIZAÇÃO DO TEXTO	23
2	GERAÇÃO DE ENERGIA FOTOVOLTAICA	25
2.1	MÓDULOS FOTOVOLTAICOS	26
2.2	PARÂMETROS METEOROLÓGICOS	28
2.2.1	Radiação Solar	29
2.2.1.1	Irradiância Normal Direta	30
2.2.1.2	Irradiância Horizontal Difusa	30
2.2.1.3	Irradiância Horizontal Global	31
2.2.2	Parâmetros Gerais	31
2.3	INSTRUMENTOS DE MEDIÇÃO	32
2.3.1	Irradiação Solar	33
2.3.1.1	Irradiância Normal Direta	33
2.3.1.2	Irradiância Horizontal Global	34
2.3.1.3	Irradiância Horizontal Difusa	34
2.3.2	Cobertura de nuvens	35
2.3.3	Parâmetros Climáticos	37
2.3.4	Sumarização dos instrumentos	39
3	ESTAÇÃO SOLARIMÉTRICA DE BAIXO CUSTO	41
3.1	SENSORES DE USO GERAL PARA A COLETA DE DADOS METEOROLÓGICOS	42
3.2	DESIGN DO SISTEMA DE AQUISIÇÃO DE DADOS	43
3.3	PROTOTIPAÇÃO DA ESTAÇÃO SOLARIMÉTRICA DE BAIXO CUSTO	45
3.3.1	Montagem do Circuito Eletrônico	46
3.3.2	Design Físico da Estação Solarimétrica	47
3.4	CONSIDERAÇÕES	52
4	CONCEITOS DE APRENDIZADO DE MÁQUINA	55
4.1	TRATAMENTO DE DADOS	56
4.1.1	Análise Exploratória de Dados	56
4.1.2	Pré-processamento de Dados	57
4.1.2.1	Imputação de Valores Ausentes	58
4.2	MÉTODOS DE APRENDIZADO DE MÁQUINA	58
4.2.1	Redes Neurais Artificiais	59
4.2.1.1	Perceptron Multicamadas	59

4.2.2	Support Vector Machines	61
4.2.3	k - Vizinhos Mais Próximos	61
4.2.4	Árvore de Decisão	61
4.2.5	Regressão Linear	62
4.3	COMITÊ DE MÉTODOS DE APRENDIZADO DE MÁQUINA	62
4.3.1	Bagging	63
4.3.2	Boosting	63
4.3.3	Stacking	64
4.4	TEORIA DE CONJUNTOS FUZZY	65
4.4.1	Função de Pertinência	65
4.5	MÉTRICAS DE ERRO	66
5	PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA	69
5.1	TRABALHOS RELACIONADOS	70
6	MODELOS DE PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA	75
6.1	ANÁLISE EXPLORATÓRIA DOS DADOS	75
6.1.1	Pré-processamento dos Dados	76
6.2	ESTRATÉGIA DE PREVISÃO BASEADA EM MODELOS INDIVIDUAIS	78
6.2.1	Materiais e Métodos	79
6.2.2	Resultados	81
6.3	ESTRATÉGIA DE PREVISÃO BASEADA EM COMITÊS DE MÉTODOS	83
6.3.1	Materiais e Métodos	83
6.3.2	Resultados	90
7	CONCLUSÕES	97
	REFERÊNCIAS	99

1 INTRODUÇÃO

O número de bens que necessitam da utilização de energia elétrica vem crescendo paralelamente ao desenvolvimento econômico, causando o aumento da demanda de energia e, conseqüentemente, da sua geração. O avanço tecnológico e a melhora dos padrões de vida da sociedade geraram um aumento no consumo industrial mundialmente, resultando no crescimento da demanda de energia elétrica. No ano de 2017 as indústrias mundiais foram responsáveis por um consumo de 35,6% da produção de energia elétrica (IEA, 2017). No Brasil, por exemplo, o uso de eletrodomésticos cresceu 1,8% entre 2018 e 2019, com uma previsão de 1,56% de aumento para 2020 (STATISTA, 2019).

O aumento da população e do consumo de bens e alimentos vem causando diversos tipos de impactos ambientais, como desmatamento e o aquecimento global gerados pela poluição industrial. Estes fatores fazem com que as buscas por novas fontes de geração de energia aumente, sendo elas energias limpas, como a eólica, hidroelétrica e principalmente a energia solar (RAZA; NADARAJAH; EKANAYAKE, 2016; INMAN; CHU; COIMBRA, 2016). Os sistemas de geração de energia solar que fazem uso de módulos fotovoltaicos têm aumentado nos últimos anos, visto que é uma fonte de energia limpa, confiável e simples de ser implementada. Esta fonte de energia, porém, carrega consigo alguns desafios.

A dificuldade encontrada na produção de energia por meio dos sistemas fotovoltaicos, baseia-se na sua geração inconstante, que depende de fatores geográficos, climáticos e meteorológicos (PEDRO; COIMBRA, 2012). Entre os fatores que influenciam na produção das plantas fotovoltaicas, estão a irradiância solar, temperatura ambiente, a direção e velocidade do vento, chuvas, nuvens, umidade do ar, pressão do ar e também a presença de aerossóis no ar (DAS et al., 2018). A variação destes parâmetros pode causar problemas de confiabilidade, segurança e planejamento para o produtor e consumidor do sistema, além das concessionárias responsáveis pela transmissão e distribuição de energia elétrica (MAHTTA; JOSHI; JINDAL, 2014; EHSAN; SIMON; VENKATESWARAN, 2017).

Os problemas causados pela intermitência na produção de energia podem ser minimizados via o uso de ferramentas de previsão de geração de energia. Uma estimativa precisa da quantidade de energia que será produzida por uma planta fotovoltaica, desempenha uma atividade importante para as concessionárias e produtores, evitando oscilações entre a energia esperada pelo planejamento e a produzida (ALMONACID et al., 2014; WANG et al., 2012). A previsão de energia fotovoltaica fornece ao sistema de distribuição de energia elétrica, operações de gerenciamento e planejamento. Além disso pode proporcionar aos sistemas de distribuição e produção de energia elétrica disponibilidade, confiabilidade e segurança, gerando uma redução de custos e au-

mento nos retornos econômicos (RAZA; NADARAJAH; EKANAYAKE, 2016; BRANO; CIULLA; FALCO, 2014; DAS et al., 2018).

O presente trabalho considera a utilização e a+ avaliação de duas estratégias de previsão de geração de energia fotovoltaica baseada em uma base de dados climáticos e meteorológicos históricos. Além disso apresenta uma investigação dos sensores utilizados para realizar a coleta dos dados e também a construção de uma estação solarimétrica de baixo custo, com sensores comerciais.

1.1 JUSTIFICATIVA

A geração de energia elétrica é um problema cotidiano e vem sendo alvo de estudos devido os impactos que podem ser causados aos sistemas de distribuição de energia. Portanto, a previsão de geração energia fotovoltaica pode proporcionar aos produtores de energia solar e as concessionárias de distribuição de energia elétrica mais segurança e um melhor planejamento da sua produção e de suas operações, gerando retornos econômicos e redução de gastos com problemas não previstos em seus sistemas. O trabalho propõe avaliar e desenvolver modelos e estratégias de previsão baseando-se em dados coletados por uma estação solarimétrica de uso específico, e demonstrar que é possível realizar a estimativa da produção de energia de plantas fotovoltaicas comuns por meio dados climáticos e de irradiação solar.

1.2 OBJETIVOS

O objetivo deste trabalho é empregar e avaliar métodos de Aprendizado de Máquina para a previsão de geração de energia fotovoltaica de uma planta fotovoltaica. Tal previsão é realizada com a coleta de dados climáticos e de irradiância por meio de uma estação solarimétrica composta por sensores específicos. O trabalho tem ainda como objetivo, a investigação e determinação de sensores adequados para a construção de uma estação solarimétrica de baixo custo. Abaixo são detalhados os objetivos específicos do trabalho.

1.2.1 Objetivos Específicos

- Estudar a geração de energia fotovoltaica e dos parâmetros que influenciam a sua geração;
- Estudar os sensores responsáveis pela coleta dos dados climáticos e meteorológicos;

- Construir uma estação solarimétrica de baixo custo, com base nos sensores identificados a partir do objetivo anterior¹;
- Investigar as aplicações de Aprendizado de Máquina neste cenário;
- Desenvolver e avaliar os modelos de previsão de geração de energia fotovoltaica.

1.3 ORGANIZAÇÃO DO TEXTO

Com os objetivos já apresentados, o trabalho é organizado da seguinte forma:

- Capítulo 2: Será apresentada uma revisão bibliográfica a respeito da geração de energia fotovoltaica e os parâmetros que a influenciam. Logo em seguida será abordada quais são os sensores que realizam a coleta de cada um dos parâmetros.
- Capítulo 3: Abordará o desenvolvimento da estação solarimétrica de baixo custo, mostrando os sensores de uso geral existentes e quais foram utilizados. Apresentará a construção da estação solarimétrica de baixo custo.
- Capítulo 4: Apresentará os conceitos de Aprendizado de Máquina utilizados no desenvolvimento do trabalho.
- Capítulo 5: Apresentará os conceitos de previsão de geração de energia, expondo alguns trabalhos relacionados ao assunto, quais métodos podem ser utilizados para realizar a tarefa, e métricas de avaliação.
- Capítulo 6: Serão apresentados os materiais e métodos empregados no desenvolvimento e avaliação das estratégias de previsão, além da análise dos resultados obtidos
- Capítulo 7: São apresentadas as considerações finais do trabalho e sugestões de trabalhos futuros.

¹ A estação solarimétrica proposta foi construída, porém devido aos problemas causados pela pandemia não foi possível realizar testes de validação.

2 GERAÇÃO DE ENERGIA FOTOVOLTAICA

Neste capítulo será feita uma introdução aos conceitos de geração de energia fotovoltaica. Primeiramente serão realizadas algumas considerações iniciais sobre o tema. Em seguida, serão apresentados os equipamentos responsáveis por desempenhar a conversão da energia solar em elétrica. O tópico seguinte abordará os parâmetros meteorológicos e climáticos que influenciam a geração de energia fotovoltaica. No final do capítulo serão retratados os sensores específicos, responsáveis por realizar a aquisição dos dados meteorológicos e climáticos.

Ainda que a energia solar seja uma grande fonte de energia renovável, a sua aplicação generalizada é dificultada pela sua sensibilidade às condições climáticas e meteorológicas. Além disso, a geração de energia solar é dependente de variações relacionadas a posição solar e o período entre a alvorada e o anoitecer. Um dos desafios mais críticos na transição para economia de energia baseada em fontes renováveis é superar problemas de capacidade e confiabilidade da produção de energia solar, eólica e hídrica. Problemas gerados pela natureza variável e às vezes intermitente dessas fontes de energia, implicam em desafios substanciais para o atual modo de operação dos produtores de energia, companhias públicas e operadores de serviço independente (PEDRO; COIMBRA, 2012).

As plantas de sistemas de energia solar são projetadas, dimensionadas e otimizadas com base nas estimativas dos recursos solares presentes no local da instalação. São considerados como recursos solares os fatores climáticos e meteorológicos. A avaliação destes recursos é efetuada por meio da medição de alguns parâmetros, que pode ser realizada em solo e também por meio de satélites (NONNENMACHER; KAUR; COIMBRA, 2014; ORTEGA et al., 2010).

Com a expansão da utilização de fontes de energias renováveis houve o aumento de estudos relacionados à métodos de predição de dados meteorológicos, como a velocidade do vento e a irradiação solar. Devido à natureza de tais variáveis, são necessários modelos precisos para prever com confiabilidade o desempenho de uma planta ou sistema fotovoltaico (MAHTTA; JOSHI; JINDAL, 2014; LORENZ et al., 2009).

A geração de energia fotovoltaica é considerada variável visto que a potência de saída é fortemente dependente de fatores meteorológicos, ambientais e de construção dos módulos fotovoltaicos. Portanto, a potência gerada varia dinamicamente com o tempo em virtude da constante oscilação desses fatores (ALZHRANI et al., 2017; KAPLANI; KAPLANIS, 2014; DAS et al., 2018). A seguir serão abordados o funcionamento e o tipo de construção dos módulos fotovoltaicos.

2.1 MÓDULOS FOTOVOLTAICOS

Um módulo fotovoltaico é formado por células¹ que são responsáveis por converter a luz solar em eletricidade. A luz que atinge uma célula, cria um campo elétrico entre suas camadas internas, que são constituídas de material semicondutor. Portanto, a energia fotovoltaica consiste na radiação solar convertida diretamente em eletricidade, sendo a intensidade da luz o fator principal que determina a quantidade de energia elétrica que cada célula gerará (TYAGI et al., 2013).

Na Figura 2.1 é apresentado um diagrama que exibe um painel fotovoltaico recebendo a radiação solar e, ao lado, um exemplo de circuito elétrico equivalente de uma célula solar composta apenas por um diodo. Nesta figura I_{ph} é a corrente gerada pela incidência de luz solar na célula, V_D é a queda de tensão sobre o diodo, R_p e R_s são respectivamente o resistor shunt e o resistor série da célula, por fim V e I (V_{pv} e I_{pv}) indicam, respectivamente, a tensão e corrente de saída do circuito elétrico (RAJESH; Carolin Mabel, 2015; MAHMOUD; MOHAMED, 2012).

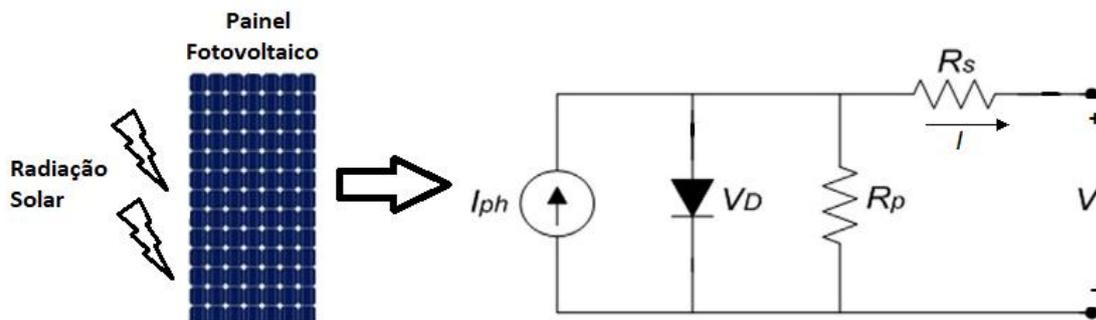


Figura 2.1 – Painel fotovoltaico e o circuito elétrico equivalente de uma célula solar composta por um único diodo.

Fonte: Adaptado de Rajesh e Carolin Mabel (2015).

A equação da corrente de saída do circuito equivalente é dada pela Equação (2.1), em que N_p e N_s são respectivamente os números de painéis conectados em paralelo e em série. A variável A é o fator de idealidade da junção P-N do semicondutor do qual a célula fotovoltaica é construída. A constante de Boltzmann é dada por k . O termo T é a temperatura em graus Kelvin, enquanto I_0 indica a corrente de saturação e q a carga de um elétron (RAJESH; Carolin Mabel, 2015). Os demais termos seguem a definição da figura acima.

$$I = N_p \left(I_{ph} - I_0 \left[\exp \left(\frac{q(V_{pv} + R_s I_{pv})}{N_s A k T} \right) - 1 \right] - \frac{(V_{pv} + R_s I_{pv})}{N_s R_p} \right) \quad (2.1)$$

As tecnologias fotovoltaicas referentes a geração de energia solar podem ser classificadas de acordo com o seu material de construção. Os módulos fotovoltaicos

¹ Definimos células como células solares de conversão de energia solar em elétrica.

são construídos principalmente a partir de materiais como silício cristalino, filme fino e semiconductor composto. A primeira geração de módulos fotovoltaicos teve o silício cristalino como material de construção. Essa tecnologia está em um processo contínuo de evolução, com o intuito de melhorar a sua eficiência e capacidade. Em comparação com os outros materiais de construção dos módulos, o silício detém o melhor desempenho, além de ser um material que pode ser facilmente encontrado no nosso planeta. As células de silício cristalino (c-Si) são divididas em, silício policristalino (p-Si), monocristalino (m-Si) e GaAs. (TYAGI et al., 2013; CHAAR; LAMONT; ZEIN, 2011; BRUTON, 2002; PARIDA; INIYAN; GOIC, 2011).

Um material que pode ser considerado como substituto do silício, é o filme fino. A produção de uma célula de filme fino necessita de menos material semiconductor comparada com a de silício, dessa forma reduzindo o custo de produção (MCCANN et al., 2001). As células de filme fino podem ser divididas em silício amorfo (a-Si); telureto de cádmio (CdTe) e sulfureto de cádmio (CdS); e também disseleneto de cobre e índio (CIS) ou disseleneto de cobre, índio e gálio (CIGS) (GREEN, 2007). A Figura 2.2 mostra um diagrama que separa os tipos de tecnologia de materiais dos quais os painéis são constituídos.

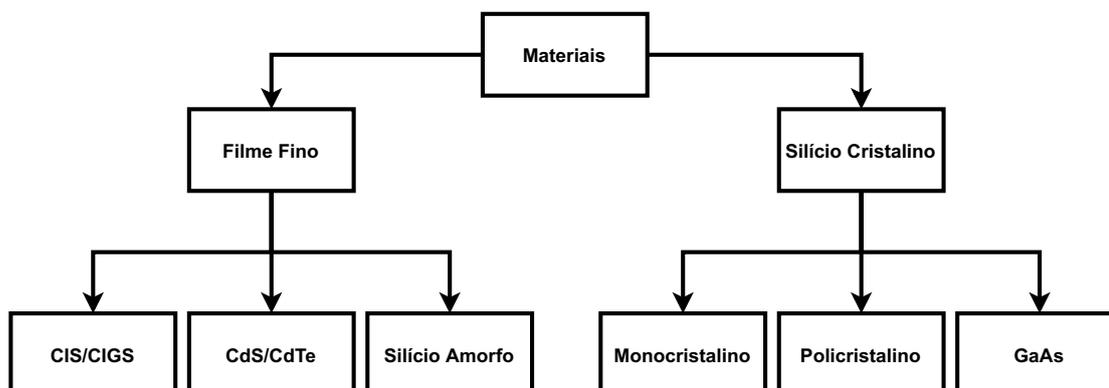


Figura 2.2 – Diagrama dos principais materiais de construção dos módulos fotovoltaicos.

Fonte: Adaptado de Tyagi et al. (2013).

Os fabricantes responsáveis pela produção dos módulos fotovoltaicos desempenham testes de operação sob condições padronizadas, nos quais fazem a aquisição de informações a respeito do funcionamento dos módulos. Com base nestes testes são disponibilizados alguns parâmetros de performance dos painéis, como a sua eficiência. A partir das informações colhidas em testes não é possível determinar quanto um painel específico pode gerar de energia em uma situação real de operação. Em uma situação particular de operação, o painel está exposto a variabilidade de todos os fatores que influenciam na geração de energia fotovoltaica e não a uma situação de teste com condições padronizadas (CAÑETE; CARRETERO; CARDONA, 2014).

Alguns autores realizaram a comparação entre o desempenho de módulos fo-

tovoltaicos fabricados a partir de diferentes materiais. Além disso avaliaram o desempenho dos módulos em diferentes cenários de operação, como inverno e verão. Os resultados indicam que cada material é afetado de uma forma diferente, de acordo com o cenário de operação, conforme discutido a seguir.

Başolu et al. (2015) compararam o desempenho de três módulos compostos por diferentes materiais, instalados nas mesmas posições geográficas e condições climáticas. Os autores realizaram as suas avaliações de desempenho dos módulos no verão e inverno. Os resultados obtidos sugerem que o módulo construído a partir de telureto de cádmio (CdTe) obteve a melhor performance nos dois cenários enquanto que no inverno o modelo constituído de silício policristalino (p-Si) atingiu uma eficiência mais alta que o de o silício cristalino (c-Si).

Já a pesquisa desenvolvida por Cañete, Carretero e Cardona (2014) concluiu que os módulos fotovoltaicos de CdTe e p-Si possuem uma eficiência maior no inverno. As outras duas tecnologias de c-Si que foram avaliadas, uma delas composta por uma estrutura de silício microcristalino e outra de a-Si, obtiveram um desempenho superior no verão. Os autores avaliaram também um módulo construído a partir do material de filme fino, para o qual foi observada uma variação de 15% em sua eficiência ao longo das estações do ano.

Os estudos apresentados anteriormente ilustram que além dos parâmetros meteorológicos, o desempenho das plantas fotovoltaicas é influenciado também pelo material do qual são construídos os módulos fotovoltaicos. Existem ainda alguns outros parâmetros que podem interferir na geração de energia solar, como a fixação e orientação dos módulos. A fixação e orientação dos módulos definem o seu posicionamento, podendo influenciar na formação de sombreamentos causados por agentes externos. Além do posicionamento, o acúmulo de sujeira na superfície dos módulos, perdas no inversor e o desgaste desses equipamentos, afetam a sua eficiência (KAPLANI; KAPLANIS, 2014; ROUHOLAMINI et al., 2016; HICKEL, 2017). Estes fatores não serão abordados, uma vez que o objetivo do presente trabalho é prever a potência de saída gerada em uma planta PV já existente e instalada. Tal previsão será realizada a partir de dados climáticos e meteorológicos coletados por uma estação solarimétrica instalada no mesmo local.

2.2 PARÂMETROS METEOROLÓGICOS

Esta seção abordará as grandezas meteorológicas que variam no espaço e no tempo, influenciando na produção de energia das plantas fotovoltaicas. A seção será dividida em duas partes. Primeiramente será abordado o parâmetro meteorológico que possui maior influência na geração de energia solar, a radiação solar. Em seguida, serão apresentados os demais parâmetros meteorológicos (os quais serão chamados de parâmetros gerais). Dentre destacam-se: a cobertura de nuvens, umidade relativa

do ar, velocidade do vento e temperatura ambiente.

2.2.1 Radiação Solar

A radiação solar é a energia gerada pelo sol que chega na superfície da terra, enquanto a irradiância corresponde a densidade de fluxo de radiação incidente sobre a superfície. A irradiância é expressa por uma razão da potência pela área (W/m^2) (MARTINS; PEREIRA; ECHER, 2004).

As interações complexas da atmosfera da Terra com a radiação solar resultam em três componentes (Figura 2.3) fundamentais de grande interesse para as tecnologias de conversão de energia solar. São elas Irradiância Horizontal Global (*Global Horizontal Irradiance* - GHI), Irradiância Normal Direta (*Direct Normal Irradiance* - DNI) e Irradiância Horizontal Difusa (*Diffuse Horizontal Irradiance* - DHI) (STOFFEL et al., 2010).

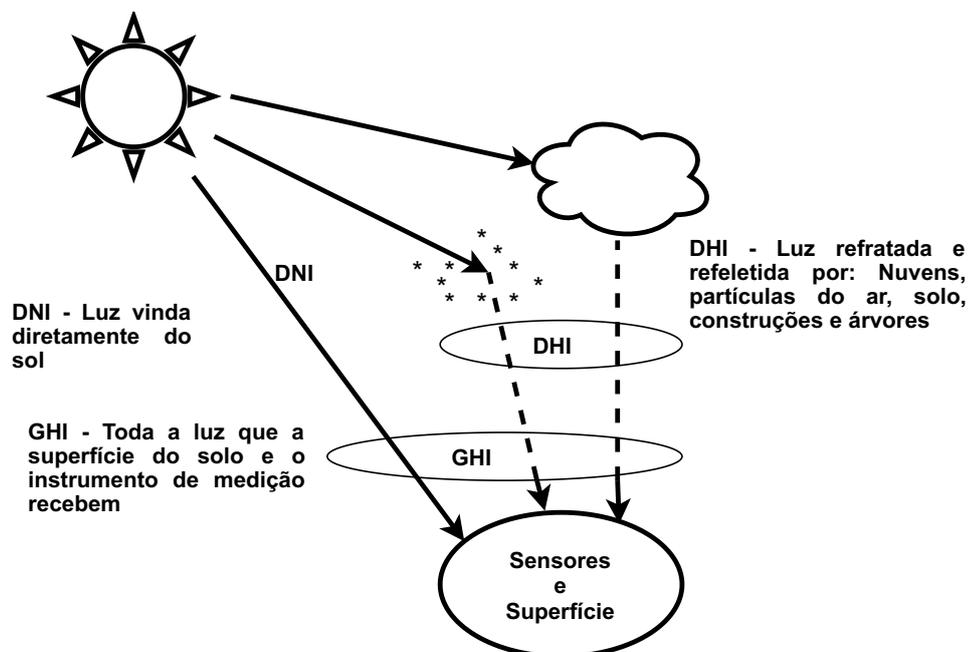


Figura 2.3 – Componentes fundamentais da radiação solar.

Fonte: Autoria própria.

Mellit, Massi Pavan e Lughì (2014) afirmam que a irradiância solar é o parâmetro que possui a maior influência na energia produzida por uma planta PV. Qualquer projeto que vise explorar a radiação solar requer uma avaliação precisa da disponibilidade deste recurso. Em particular, as medições de GHI possuem um importante papel no início do desenvolvimento do design de sistemas fotovoltaicos (BACHOUR; PEREZ-ASTUDILLO, 2014). Embora a radiação que atinge o topo das camadas da atmosfera seja bem definida e possa ser facilmente calculada, a irradiância solar que atinge o nível do solo, onde os coletores solares (térmicos e fotovoltaicos) operam, depende fortemente das condições atmosféricas e climáticas (MARQUEZ; COIMBRA, 2011).

2.2.1.1 Irradiância Normal Direta

A Irradiância Normal Direta (*Direct Normal Irradiance* - DNI) é a porção da radiação solar que atinge o topo da atmosfera e chega à superfície da Terra em um plano normal ou perpendicular à linha que conecta o observador ao centro do disco solar (MYERS, 2017). De acordo com Blanc et al. (2014), o DNI é a irradiância recebida diretamente por uma superfície plana normal, desempenhando uma função essencial para produção de energia em dias que estão sob condições de céu limpo. O DNI representa uma das componentes da radiação solar que pode ser usada por várias tecnologias de geração de energia solar, como os sistemas térmicos solares, a torre solar e os sistemas fotovoltaicos.

A aquisição de informações relacionadas ao DNI são de obtenção mais complexa de obter do que informações a respeito das outras componentes de irradiância. Os dados sobre o DNI são de grande importância para o desenvolvimento de modelos de previsão de geração de energia e também para modelos que tem o objetivo de prever as componentes de irradiância (STOFFEL et al., 2010).

O ângulo de separação entre os feixes de luz do sol é muito pequeno, então são considerados paralelos na maioria das aplicações, portanto o DNI é considerado a quantidade de radiação emitida pelo sol em um anel estreito (PARMAR, 2011). O DNI é medido por meio de pireliômetros projetados com um ângulo de campo de visão de 5 graus, cujas superfícies receptoras estão dispostas de maneira normal à direção do Sol (WMO, 2018). Os instrumentos serão abordados com mais detalhes na Seção 2.3 do trabalho.

2.2.1.2 Irradiância Horizontal Difusa

Em um dia de céu limpo em que não há presença de nuvens, parte do DNI que chega ao solo é absorvida e refletida em outras direções. A partir desta reflexão surge a radiação espalhada, que é vista como a radiação do céu no hemisfério acima do solo. Alguns cientistas dos Séculos e desenvolveram teorias que explicam o motivo da radiação do céu parecer azul e do disco solar parecer amarelo e vermelho no nascer e no pôr do sol. Respectivamente o céu é azul devido ao comprimento de onda da luz azul ser curto, o que faz que sua dispersão ocorra de forma mais eficiente pelos gases atmosféricos. O sol é visto na cor vermelha pelo motivo contrário: o comprimento de onda da luz vermelha do disco solar é mais longo e acaba não sofrendo dispersão (STOFFEL et al., 2010).

A Irradiância Horizontal Difusa (*Diffuse Horizontal Irradiance* - DHI) é então a radiação solar dispersa na cúpula do céu, ou seja, é a radiação que foi espalhada e refletida por nuvens, aerossóis e outros elementos atmosféricos. O DHI representa a radiação vista do solo sem a presença do DNI (PARMAR, 2011; STOFFEL et al.,

2010). Segundo Zell et al. (2015) os níveis de DHI são altos na presença de nuvens e baixos quando o céu está limpo.

O DHI é mensurado por radiômetros sombreados, que possuem um campo de visão de 180° e um disco ou esfera responsável pelo sombreamento do disco solar de 5°. O sombreamento é necessário para bloquear o feixe de luz solar direto (DNI). Na próxima seção esse instrumento vai ser tratado com maior destaque. Os dados de DHI são úteis para estimar a irradiância do plano onde estaria localizada uma planta PV, além de auxiliar nas avaliações de recursos solares e projetos arquitetônicos que envolvam a iluminação natural (ZELL et al., 2015).

2.2.1.3 Irradiância Horizontal Global

A Irradiância Horizontal Global (*Global Horizontal Irradiance* - GHI) é a quantidade de irradiância terrestre que atinge uma superfície horizontal, sendo ela a soma total dos componentes geométricos de radiação solar, DNI e DHI (ZELL et al., 2015). Segundo Myers (2017), a GHI é a irradiância direta multiplicada pelo cosseno do ângulo zênite (Z - ângulo entre o ponto verticalmente acima da posição em questão e a declinação do sol no céu) mais a irradiância difusa, conforme a Equação (2.2).

$$GHI = (DNI * \cos(Z)) + DHI \quad (2.2)$$

As medições de GHI podem ser feitas por meio de diversos instrumentos. O instrumento mais utilizado é o piranômetro sem sombreamento, que possui um ângulo de visão hemisférico, ou seja, de 180° (ZELL et al., 2015).

2.2.2 Parâmetros Gerais

Nesta seção serão abordados primeiramente, os papéis de cada uma das grandezas meteorológicas na geração de energia solar. Em seguida, serão discutidos trabalhos que desenvolveram estudos sobre o efeito individual desses parâmetros na produção de energia fotovoltaica.

As nuvens desempenham um papel relevante na transferência de radiação da atmosfera à superfície da terra. Possuem uma estrutura tridimensional e macroscópica, tipos e padrões variáveis, e características de composição microfísicas que determinam as suas particularidades ópticas (FEISTER et al., 2010). A cobertura de nuvens tem um papel de extrema importância na produção de energia solar, por estar diretamente ligada a quantidade de radiação solar que atinge o solo. Portanto, informações sobre a quantidade de nuvens que cobrem o céu acabam sendo necessárias para a avaliação e utilização de energias renováveis como a solar (SOUZA-ECHER et al., 2006).

Em uma planta PV que opera ao ar livre, a velocidade e direção do vento podem afetar o desempenho dos módulos fotovoltaicos. Projetos que visam prever

com precisão o potencial de uma planta fotovoltaica em locais nos quais existe um alto índice de velocidade do vento, devem levá-lo em consideração. Geralmente o vento pode ocasionar um efeito de refrigeração nos módulos fotovoltaicos, afetando a sua eficiência (KAPLANI; KAPLANIS, 2014; GÖKMEN et al., 2016).

A umidade relativa do ar tem efeito sobre a irradiância solar, afetando a quantidade de irradiância que atinge os módulos fotovoltaicos. De acordo com Rouholamini et al. (2016), um índice alto de umidade relativa do ar diminui a intensidade da radiação solar e pode causar degradação das células solares, induzindo a redução da eficiência dos módulos fotovoltaicos. O feixe de luz do sol, que atinge a atmosfera atinge também partículas de água que fazem com que a luz seja refletida, refratada ou difratada. Desta forma, a umidade afeta de forma não linear a componente de Irradiância Normal Direta (*Direct Normal Irradiance* - DNI) da radiação solar devido aos diferentes tamanhos das partículas (MEKHILEF; SAIDUR; KAMALISARVESTANI, 2012).

Alguns trabalhos visam analisar certas localidades para a sua utilização na geração de energia através de sistemas fotovoltaicos. Esses estudos podem ser considerados como uma avaliação dos recursos solares de uma região. Kruakaew, Banjerdpongchai e Hoonchareon (2019), Roy, Hayman e Julian (2006) afirmam que as informações sobre nuvens exercem uma função importante no desempenho dos algoritmos de Aprendizado de Máquina, que têm a finalidade de prever a irradiação. Em um estudo feito por Escobar et al. (2014), é analisada a intensidade do GHI em regiões com o céu limpo (onde ele chega a 1200 W/m^2) e em locais com cobertura de nuvens no céu (onde ele decai para 700 W/m^2).

A eficiência dos módulos fotovoltaicos também é afetada pela variação de temperatura dos módulos (ZDRAVKOVIC et al., 2009). De modo geral a temperatura do módulo depende de parâmetros como a radiação solar que envolve a ocorrência de nuvens no céu, a temperatura ambiente, a umidade, e também a velocidade e direção do vento (ROUHOLAMINI et al., 2016).

Segundo Malvoni, De Giorgi e Congedo (2017), a temperatura ambiente pode melhorar o acerto da previsão da potência de saída de um PV quando é utilizada como entrada em modelos de previsão. Em um estudo realizado por Rouholamini et al. (2016), demonstrou-se que a relação entre a temperatura ambiente e a temperatura dos módulos fotovoltaicos é linear, sendo a temperatura dos módulos mais alta que a do ambiente.

2.3 INSTRUMENTOS DE MEDIÇÃO

O objetivo desta seção é abordar os instrumentos utilizados para a realização de monitoramento e aquisição dos dados em solo dos parâmetros meteorológicos expostos na seção anterior, em solo. Serão apresentados alguns trabalhos com foco nos sensores e parâmetros meteorológicos considerados.

2.3.1 Irradiação Solar

Informações sobre radiação solar adquiridas através de satélites e estações de medição terrestre, possuem grande importância para os projetos que visam a construção de um sistema PV (PDUGARIA et al., 2015). Em trabalhos onde são necessários os dados de radiação, são utilizados sensores como piranômetro, pireliômetro, pirgeômetro, radiômetro com sombreado giratório e ceilômetro (STOFFEL et al., 2010; WMO, 2018). Estes instrumentos são discutidos a seguir.

2.3.1.1 Irradiância Normal Direta

A componente DNI é medida com um pireliômetro que é montado em conjunto com um rastreador ou seguidor solar, o que torna o custo da instrumentação do sistema mais alta (PADOVAN et al., 2014; MADETI; SINGH, 2017). O rastreador solar é responsável por fazer com que o sensor tenha um ajuste rápido para a posição correta do sol, ou seja, ajusta o pireliômetro aos ângulos solares para que haja uma gravação em tempo real mais confiável e sem intervalos (ZELL et al., 2015; DEKKER et al., 2012; WMO, 2018).

Esses equipamentos são expostos a diversas condições climáticas e físicas, por isso precisam possuir uma construção robusta. A qualidade das medições pode ser prejudicada, devido a sensibilidade, tempo de resposta, compensação térmica e estabilidade do instrumento (WMO, 2018; PADOVAN et al., 2014).

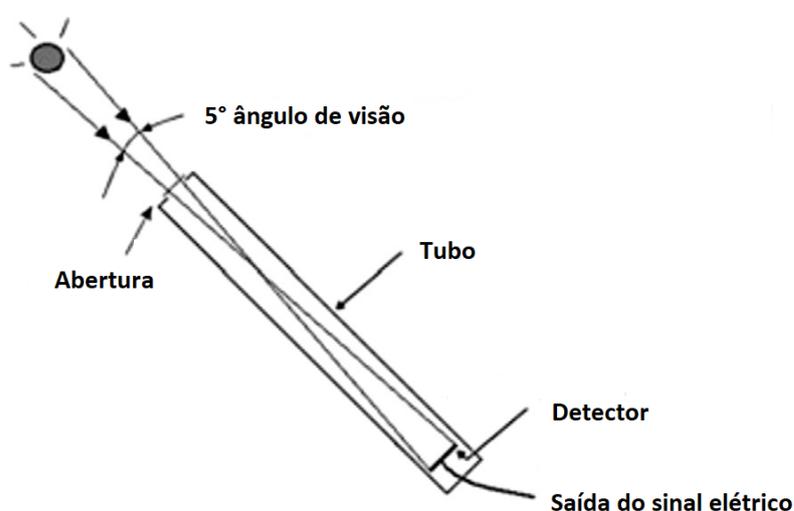


Figura 2.4 – Diagrama esquemático do pireliômetro.

Fonte: Adaptado de Madeti e Singh (2017).

A superfície receptora do pireliômetro (Figura 2.4) possui uma abertura estreita com um ângulo de visão de 5° a 6°, portanto o DNI é medido dentro de uma região circunsolar limitada, possuindo sua posição sempre normal ao feixe direto da irradiação

solar, omitindo o DHI fornecido pelo céu (GUEYMARD; MYERS, 2008; STOFFEL et al., 2010; MADETI; SINGH, 2017).

2.3.1.2 Irradiância Horizontal Global

As informações a respeito do GHI a partir de piranômetros não sombreados. Este instrumento mede toda a radiação que incide sobre a cúpula do sensor em um ângulo de 180° . O piranômetro deve ser montado em um local sem obstruções naturais ou artificiais no horizonte, em uma plataforma horizontal, para que tenha acesso completo à cúpula do céu (STOFFEL et al., 2010; DEKKER et al., 2012).

O receptor ou detector do piranômetro (Figura 2.5) é instalado abaixo de uma cúpula (construída com quartzo, vidro ou acrílico) que desempenha a função de proteção, visto que o detector pode ser danificado se exposto à intemperes. Uma solução para evitar danos ao instrumento é a instalação de ventiladores que emitem ar constantemente sobre a cúpula e o instrumento, além disso para que não ocorra danos através de umidade, é utilizado um dessecante responsável por regular o nível de umidade dentro do instrumento (MADETI; SINGH, 2017; STOFFEL et al., 2010; WMO, 2018).

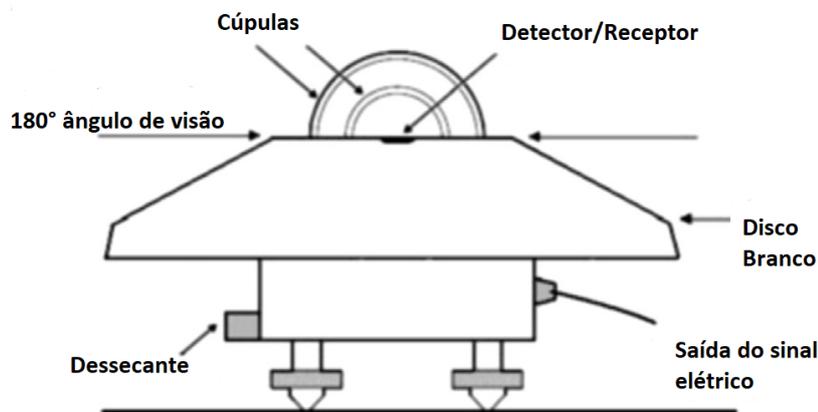


Figura 2.5 – Diagrama esquemático do piranômetro.
Fonte: Adaptado de Madeti e Singh (2017).

2.3.1.3 Irradiância Horizontal Difusa

De acordo com Siuzdak (1996) medições de alta qualidade de radiação solar são feitas com piranômetro instalado em conjunto com um disco ou esfera de sombra e um rastreador solar. O piranômetro é sombreado, utilizando uma esfera para não receber a radiação direta do sol. Uma alternativa que torna o sistema mais barato é a utilização de uma faixa de sombra, necessitando apenas de um ajuste manual no decorrer dos dias conforme a angulação do sol varia. Neste último caso, as medições

obtidas devem ser corrigidas, pois a faixa de sombra inibe uma parte do próprio DHI, tornando o método menos preciso (PDUGARIA et al., 2015; PADOVAN; Del Col, 2010).

Na Figura 2.6, temos uma estação de medição composta por um rastreador solar Solys 2 em conjunto com piranômetro e pireliômetros fabricados por Kipp & Zonnen. Indicado como 1 na imagem tem-se um pireliômetro do modelo CHP1 responsável por realizar a medida do DNI, já os instrumentos 2 e 3 são os piranômetros SMP11 responsáveis por medir respectivamente GHI e DHI e por último o instrumento 4, que corresponde ao rastreador solar equipado com a esfera de sombreamento (DEKKER et al., 2012).



Figura 2.6 – Estação de medição das componentes de irradiação solar.

Fonte: Adaptado de Dekker et al. (2012)

O piranômetro e o rastreador solar são classificados de acordo com a incerteza das suas medições (WMO, 2018; ISO, 1990). A classificação dos instrumentos é importante para que o utilizador tenha conhecimento da qualidade dos seus dados e do instrumento que está utilizando.

2.3.2 Cobertura de nuvens

A instrumentação para a detecção e determinação do grau de cobertura de nuvens, pode ser feita através de satélites, câmeras, sensores infravermelhos e também pela visualização do céu feita por técnicos meteorológicos treinados. No procedimento de avaliação das nuvens que surgem no céu, tem-se como objetivo a definição da fração de nuvens que cobrem o céu e também a sua posição. Uma das dificuldades encontradas na detecção de nuvens, é o brilho da região que está em torno do sol.

O processo que possui alta imprecisão devido a sua taxa oscilação espacial e temporal ser frequente (KRUAKAEW; BANJERDPONGCHAI; HOONCHAREON, 2019; ROY; HAYMAN; JULIAN, 2006).

O trabalho realizado por Cazorla, Olmo e Alados-Arboledas (2008) apresenta, baseado em uma câmera digital CCD (Dispositivo de Carga Acoplada), um sistema que produz imagens do céu com o objetivo de estimar e caracterizar a cobertura de nuvens. As imagens obtidas são classificadas por métodos de aprendizado de máquina. O sensor utilizado é chamado de “All-Sky Imager” ou “Total Sky Imager” (ASI/TSI - Figura 2.7), que faz a aquisição de imagens coloridas da cúpula do céu em um ângulo de 180° por meio de uma câmera CCD, no formato RGB (*Red, Green and Blue*). O autor obteve bons resultados com o uso de Aprendizado de Máquina, fornecendo também uma classificação das nuvens em duas classes diferentes, nuvens finas e opacas.

Berger (2005) apresenta uma comparação de imagens produzidas por dois instrumentos, um baseado em uma câmera CCD e outro em sensores infravermelhos. O primeiro é o TSI (Figura 2.8), o qual calcula a cobertura de nuvens no céu pela razão de intensidade das cores nos canais RGB. Já o sensor CIR-7 é um instrumento composto por sete sensores infravermelhos que estão dispostos em ângulos de visão de 12° , montados nos ângulos zenitais 0° , 12° , 24° , 36° , 48° , 60° e 72° em uma torre azimutal de semicírculo (Figura 2.8). Foi observado nesse estudo, que o sensor CIR-7 possui um tempo de coleta das imagens mais longo que o TSI, porém consegue fazer a aquisição de dados durante o dia e noite.

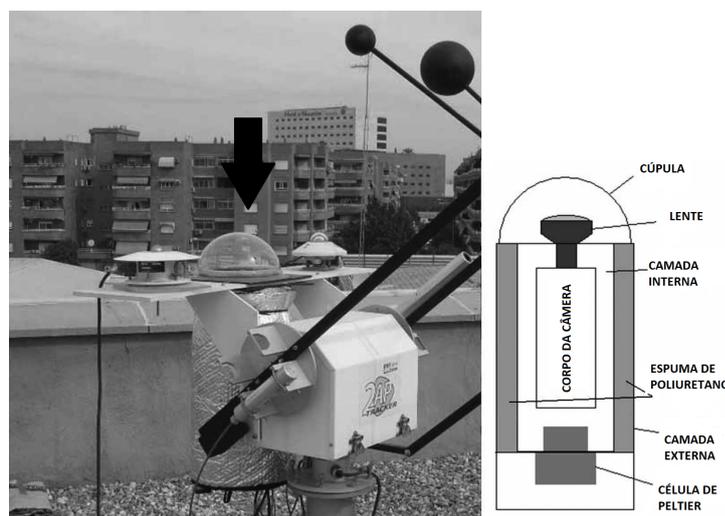


Figura 2.7 – All-Sky Imager montado em um rastreador solar.
Fonte: Adaptado de Cazorla, Olmo e Alados-Arboledas (2008)

Em seguida serão apresentados os sensores responsáveis pela aquisição de dados como velocidade e direção do vento, umidade e temperatura ambiente. Serão também abordados alguns estudos relacionados a geração de energia fotovoltaica, que fizeram a utilização de informações desses parâmetros.



(a)



(b)

Figura 2.8 – TSI(a) e CIR-7(b).
Fonte: Adaptado de Berger (2005)

2.3.3 Parâmetros Climáticos

Bhattacharya, Chakraborty e Pal (2014) realizaram um estudo sobre a variação da eficiência dos módulos fotovoltaicos de acordo com a mudança de alguns fatores climáticos. A temperatura ambiente foi medida a partir de um termômetro digital e a intensidade de radiação solar por um medidor de energia solar (Figura 2.9).



(a)



(b)

Figura 2.9 – Medidor de energia Solar TENMARS TM -201(a) e Termômetro Digital MS2101(b).

Fonte: Adaptado de TENMARS (1997), MASTECH (2019)

O trabalho de Omubo-Pepple, Tamunobereton-ari. e Briggs-Kamara. (2013) visou avaliar a eficiência da conversão de energia solar em eletricidade dos módulos

fotovoltaicos, com base em informações sobre temperatura, fluxo solar e umidade relativa. Para realizar a medição desses fatores, utilizou-se um sensor medidor de luz para o fluxo solar, um termômetro digital para a temperatura e um anemômetro digital que fornece informações do vento, da umidade relativa do ambiente e também da temperatura (Figura 2.10).



Figura 2.10 – Anemômetro Kestrel 3000(a) e Medidor de luz BK Model 615(b).
Fonte: Adaptado de Kestrel (2009), PRECISION (2000)

Rouholamini et al. (2016) avaliou como a temperatura dos módulos fotovoltaicos afeta a sua eficiência. Para essa avaliação de eficiência dos módulos, foi feita a aquisição de dados de irradiância por um piranômetro, da temperatura ambiente e dos módulos fotovoltaicos por um sensor de temperatura, e da velocidade e direção do vento por um anemômetro (Figura 2.11).



Figura 2.11 – Anemômetro WSD-100.
Fonte: (SCIENTIFIC, 2011)

2.3.4 Sumarização dos instrumentos

Neste capítulo diferentes trabalhos sobre a geração de energia fotovoltaica e os fatores que podem interferir na geração foram discutidos. Esses trabalhos realizaram a previsão e avaliação destes parâmetros em localizações determinadas, onde estavam instaladas plantas PV e estações solarimétricas. Com a Tabela 2.1, que apresenta um resumo destes trabalhos, podemos observar que para os fatores de irradiação solar os três instrumentos mais utilizados e comuns entre as pesquisas expostas são: piranômetro, pireliômetro e rastreador solar. Já para a detecção de nuvens o TSI é o instrumento mais comum, no qual as imagens são adquiridas através de câmeras digitais e avaliadas através do processamento de imagens.

As colunas da Tabela 2.1 representam os autores do trabalho; data de publicação; a data publicada do trabalho; a localização dos equipamentos de medição onde o estudo foi realizado; quais os parâmetros utilizados; os instrumentos utilizados no trabalho.

No próximo capítulo será apresentado a estação solarimétrica de baixo custo desenvolvida neste trabalho. Serão abordados os sensores de baixo custo utilizados e também os processos de desenvolvimento da estação solarimétrica proposta no trabalho.

Tabela 2.1 – Tabela de relação entre parâmetros e instrumentos.

Referências	Ano	Localização	Parâmetros	Instrumentos
PDUGARIA et al.	2015	Itália/Padova e Tri- saia		
RONZIO; COLLINO; BONELLI	2013	Itália/Milano, Casac- cia e Catania		Piranômetro, Pireliômetro e Rastreador Solar
PEREZ-ASTUDILLO; BACHOUR	2014	Catar, Doha		
RELVA et al.	2014	Brasil, São Paulo		
DEKKER et al.	2012	África do Sul		
GUEYMARD; MYERS	2009	EUA		Piranômetro, Pireliômetro e Pirgeômetro
STÖKLER; SCHILLINGS; KRAAS	2016	Paquistão	GHI, DNI e DHI	Piranômetro, Pireliômetro e Radiômetro de faixa de sombra rotativa
GUEYMARD; MYERS	2008	EUA		Radiômetro de faixa de sombra rotativa
ZELL et al.	2015	Arábia Saudita		Piranômetro, Pireliômetro, Radiômetro de faixa de sombra rotativa, TSI, Sensores Meteorológicos, Fotometro, Pirgeômetro
GHEDIRA; EISSA	2014	Emirados Árabes		Radiômetro de faixa de sombra rotativa e Piranô- metro
MARTINS; SOUZA; PEREIRA	2003	Brasil/Florianópolis		TSI equipado com Disco de Sombra
KRUAKAEW; BANJERDPONGCHAI; HOON- CHAREON	2019	EUA/Califórnia		TSI
FEISTER et al.	2010	Alemanha/Lindenberg	Cobertura de Nuvens	Pirômetro/Nubiscope, Radar de nuvens e Celio- grafo
BERGER	2005	França e EUA		TSI e CIR-7
AGUIAR et al.	2016	Espanha		
BACHOUR; PEREZ-ASTUDILLO	2014	Catar	GHI	Piranômetro
MATHIESEN; KLEISSL	2011	EUA		
INMAN; CHU; COIMBRA	2016	EUA	GHI e DNI	Piranômetro, Pireliômetro, Rastreador Solar e Ra- diômetro de faixa de sombra rotativa
ORTEGA et al.	2010	Chile	GHI e DNI	Piranômetro e Rastreador Solar
NONNENMACHER; KAUR; COIMBRA	2014	EUA	DNI	Piranômetro, Pireliômetro, Rastreador Solar e Ra- diômetro de faixa de sombra rotativa
ROUHOLAMINI et al.	2016	Irã	GHI, Temperatura e Vento	Anemômetro, Termômetro e Piranômetro
KAPLANI; KAPLANIS	2014	Grécia	GHI, Temperatura e Vento	Anemômetro, Termômetro e Piranômetro
OMUBO-PEPPLE; TAMUNOBERETON-ARI.;	2013	Nigéria	Fluxo solar, Umidade relativa e	Termômetro, Luxímetro e Medidor climático
BRIGGS-KAMARA.; OMUBO-PEPPLE; ISRAEL- COOKEY; ALAMINOKUMA	2009		Temperatura ambiente	

Fonte: Autoria própria.

3 ESTAÇÃO SOLARIMÉTRICA DE BAIXO CUSTO

Este capítulo apresentará os sensores de baixo custo que podem ser utilizados de forma alternativa para a coleta de dados de irradiância, temperatura ambiente e umidade relativa do ar. A identificação dos possíveis sensores que irão compor a estação solarimétrica de baixo custo e o seu desenvolvimento também serão descritos.

A estação solarimétrica de baixo custo proposta neste trabalho é responsável por realizar a aquisição de dados meteorológicos. A estação é constituída por sensores de uso geral, que realizam a medição de temperatura ambiente, umidade relativa do ar e das componentes de irradiância solar. A utilização desses sensores é justificado pelo alto custo dos sensores específicos, como o piranômetro e o pireliômetro. Espera-se, dessa forma, um menor custo nas aplicações que envolvem o estudo da previsão de geração de energia fotovoltaica e a aquisição de dados meteorológicos utilizáveis.

Alguns autores propuseram em seus trabalhos o uso de módulo e células fotovoltaicas para a estimação da irradiância. No trabalho desenvolvido por Tan, Tai e Mok (2013), a irradiância foi estimada através da corrente de curto circuito de um módulo fotovoltaico. Os autores concluíram que foi possível obter uma precisão aceitável, validando as medições com um piranômetro comercial e demonstram que o uso de um sensor mais caro pode ser substituído por um módulo fotovoltaico. Tan, Tai e Mok (2013) obtiveram bons resultados, porém com erros obtidos na faixa de 50%. Os erros foram justificados por sombras geradas pela árvore de instalação das células. Para avaliar os erros obtidos os autores utilizaram um sensor de temperatura, com o objetivo de calcular o coeficiente de temperatura de trabalho da célula. Carrasco et al. (2017) utilizou a temperatura do módulo em conjunto com a corrente e a tensão de trabalho para estimar a irradiância, mostrando que os custos da implementação do sistema proposto são baixos.

Outro tipo de sensor de uso geral utilizado para estimar a irradiância, é o fotodiodo. O fotodiodo gera uma pequena corrente elétrica quando exposto a luz, e a partir dessa corrente é estimada a irradiância. Martínez, Andújar e Enrique (2009) desenvolveram um piranômetro de baixo custo a partir de um fotodiodo, demonstrando que o modelo construído possuiu um custo muito baixo em relação aos piranômetro comerciais (Figura 3.1). O autor obteve resultados confiáveis, que foram validados com o piranômetro CM21. Além disso observaram algumas características do equipamento, como baixa sensibilidade a mudanças de temperatura e boas respostas a variações angulares do sol.

Medugu, Burari e Abdulazeez (2010) construíram um piranômetro visando a medição da irradiância em situações nas quais o baixo custo é um fator importante para a construção de uma estação solarimétrica. Os resultados obtidos foram validados por meio de medições realizadas com o piranômetro comercial. Lave et al. (2015) concluiu

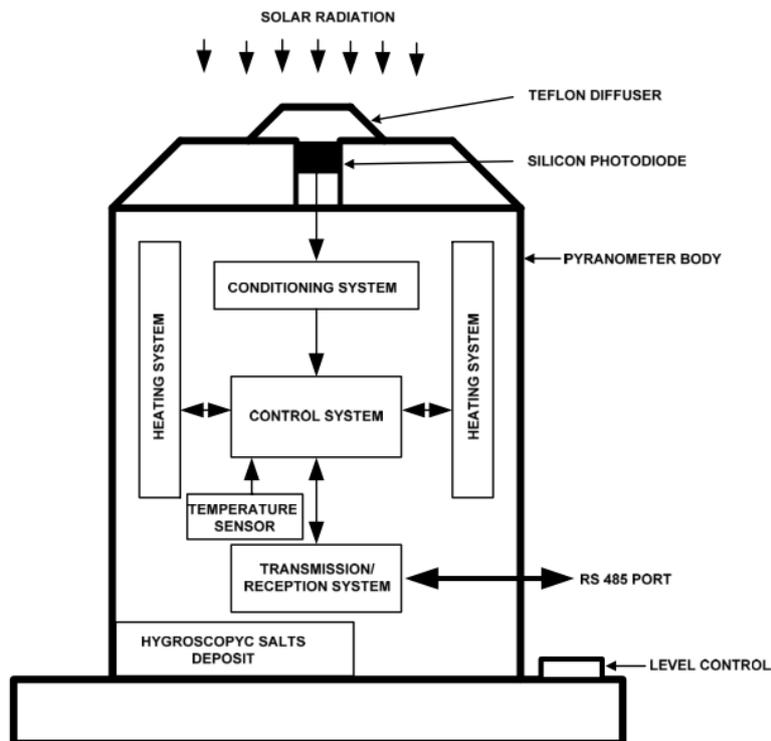


Figura 3.1 – Diagrama esquemático do piranômetro desenvolvido por Martínez, Andújar e Enrique (2009).

Fonte: (MARTÍNEZ; ANDÚJAR; ENRIQUE, 2009)

que o uso de sensores de uso geral para realizar a medição de irradiância solar, como fotodiodos e células fotovoltaicas, podem levar a obtenção de dados com precisão e qualidade quando comparados com sensores de uso específico.

A seção seguinte abordará os sensores escolhidos para o desenvolvimento da estação solarimétrica de baixo custo proposta neste trabalho.

3.1 SENSORES DE USO GERAL PARA A COLETA DE DADOS METEOROLÓGICOS

Entre os sensores selecionados para o desenvolvimento do trabalho estão: fotodiodos, fototransistores e células fotovoltaicas. Alguns dos sensores foram selecionados baseando-se nos trabalhos citados na seção anterior. Para que houvesse uma comparação das medições realizadas, outros modelos também foram considerados.

Os fotodiodos são dispositivos semicondutores fotoelétricos que convertem luz em energia elétrica gerando uma corrente proporcional a quantidade de radiação incidente. Cada fotodiodo possui uma resposta diferente para cada faixa de comprimento de onda de luz, variando assim a corrente gerada. Dessa forma é possível medir as variações da radiação solar ao longo do dia (MEDUGU; BURARI; ABDULAZEEZ, 2010). O fototransistor funciona de forma semelhante ao fotodiodo, possuindo uma base sen-

sível a luz que gera uma corrente entre o coletor e o emissor proporcional a luz que incide no componente (HIDALGO; MARTINEZ; VIDAL, 2013).

O LDR (*Light Dependent Resistor*) é um resistor dependente de luz que varia a sua resistência de acordo com a intensidade de luz que atinge a sua superfície. O resistor LDR possui um aumento logarítmico na sua resistência e, de acordo com De Barros et al. (2019), para que respostas precisas a respeito da irradiância sejam obtidas é necessário validar a medição com diferentes modelos que realizam a aquisição em vários espectros.

O último sensor utilizado no desenvolvimento do trabalho é a célula fotovoltaica, que funciona como um painel solar comum, no qual é possível obter a corrente de curto circuito gerada pela célula quando atingida pela luz solar (LAVE et al., 2015).

Na Tabela 3.1 são sumarizados os sensores utilizados no desenvolvimento do trabalho e as referências que levaram a sua adoção (quando existente). Alguns dos sensores não possuem referências pois a ideia é que seja realizada uma comparação entre as medidas dos diferentes sensores, que possuem aspectos diferentes de construção e uma faixa de espectro de leitura dos feixes de luz variada.

Tabela 3.1 – Modelos dos dispositivos semicondutores utilizados como sensores.

Modelo	Tipo do Sensor	Referências
BPW21	Fotodiodo	(MEDUGU; BURARI; ABDULAZEEZ, 2010)
BPW34	Fotodiodo	(SYSTEMS; SCIENCE, 2019)
TPD5700	Fototransistor	-
OPA505B	Fototransistor	-
TEFD4300F	Fotodiodo	-
MTD5010W	Fotodiodo	-
MTD3910N	Fotodiodo	-
PT333-3C	Fototransistor	-
SM111K09L	Célula PV	-
AM-5610CAR	Célula PV	(LAVE et al., 2015)
DHT11	Sensor de Umidade e Temperatura	-

3.2 DESIGN DO SISTEMA DE AQUISIÇÃO DE DADOS

Para a construção da estação solarimétrica de baixo custo foi desenvolvido um sistema de aquisição de dados. Esse sistema é composto por duas placas de circuito idênticas, e que possuem o mesmo objetivo, realizar a leitura dos dados e o envio via Wi-Fi para uma plataforma de base de dados *online*. A existência de duas placas de aquisição se deve a quantidade de sensores que foram utilizados no desenvolvimento do projeto, sendo assim, foram divididos entre duas placas. O projeto foi desenvolvido visando a utilização de 11 sensores diferentes (Tabela 3.1).

O projeto da placa de circuito impresso foi baseado no Microcontrolador ESP32 (Figura 3.2) que possui módulos Wi-Fi e Bluetooth integrados. Este microcontrolador possui dois núcleos de CPU (*Central Process Unit* - Unidade Central de Processamento). Além de integrar um conjunto de periféricos, desde sensores capacitivos de toque, sensores Hall, Ethernet, SPI (*Serial Peripheral Interface*) de alta velocidade, UART (*Universal Asynchronous Receiver/Transmitter*), I²C e um conversor sigma-delta integrado de 12 bits, com 10 canais.



Figura 3.2 – Microcontrolador ESP32.

Fonte: Adaptado de ESPRESSIF (2019)

A Figura 3.3 representa o esquemático elétrico de uma das placas de circuito responsável por realizar a leitura dos dados. Nessa figura estão sendo representados alguns dos sensores utilizados que foram abordados anteriormente, como, fotodiodos, a célula fotovoltaica e o sensor de temperatura e umidade relativa do ar (DHT11). A alimentação das placas é feita pelo próprio ESP32, que é alimentado por um cabo USB (*Universal Serial Bus* - Porta Serial Universal), sendo então capaz de fornecer uma tensão de 3.3V para o circuito eletrônico.

Esse sistema possui um circuito auxiliar responsável pelo condicionamento do sinal gerado pelos fotodiodos. O circuito auxiliar é composto por um circuito integrado de amplificadores operacionais (LM124), que junto com os resistores divisores de tensão, exerce o papel de elevar a tensão dissipada pelos fotosensores. O circuito é necessário devido ao baixo nível de tensão gerado pelos fotosensores, fazendo com que o conversor, que possui uma resolução de 12 bits, não seja capaz de realizar a leitura adequada desses valores.

Após o condicionamento do sinal que foi gerado pelos sensores, eles estão prontos para serem entregues ao conversor sigma-delta do ESP32. O sinal gerado pelo sensor DHT11 não necessita de condicionamento, pois o ESP32 possui uma biblioteca própria responsável pela leitura dos dados provenientes deste sensor. As células fotovoltaicas utilizadas podem gerar tensões em uma faixa de 0.1 V até 3V, um nível de tensão aceitável para a resolução do conversor integrado do ESP32. O mesmo

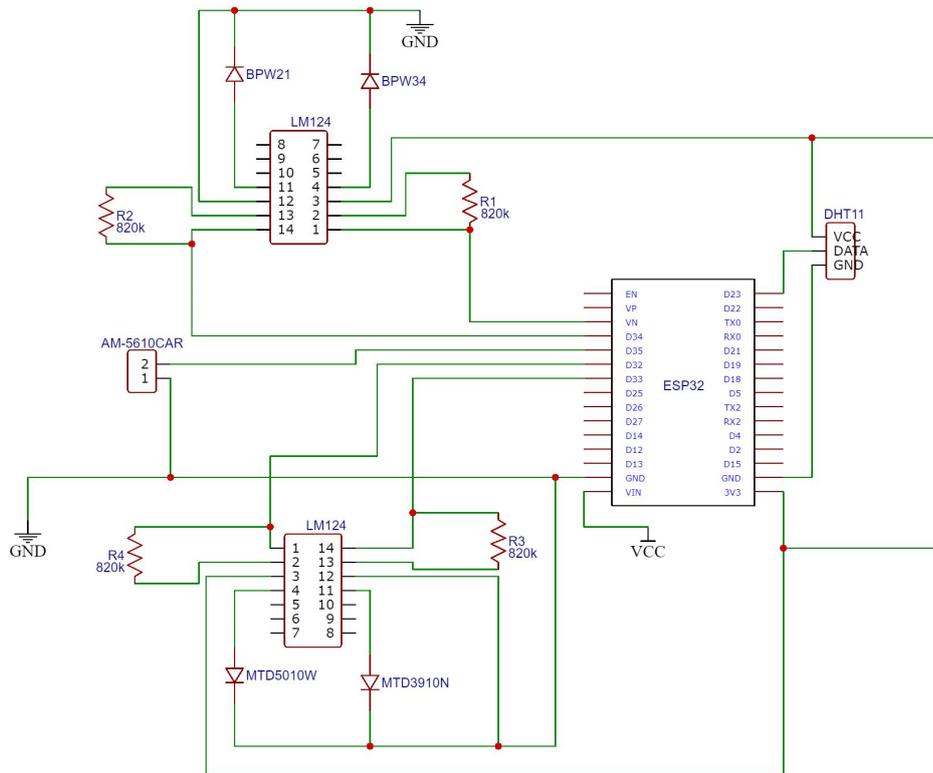


Figura 3.3 – Esquemático da placa eletrônica com o ESP32.
 Fonte: Autoria própria.

esquemático se aplica para a outra placa, sendo ela composta por fototransistores, fotodiodos e uma célula fotovoltaica.

Com o sistema eletrônico pronto para ser implementado, o passo seguinte foi definir como os dados coletados seriam armazenados. Para realizar o armazenamento dos dados foi utilizada a plataforma *Google, Firebase*. O *Firebase* é uma plataforma que tem como objetivo principal o desenvolvimento de aplicativos móveis e web. Além disso, essa plataforma possui uma ferramenta chamada *Firebase RealTime Database*, que permite o envio de dados via internet para um banco de dados hospedado na nuvem, em tempo real.

3.3 PROTOTIPAÇÃO DA ESTAÇÃO SOLARIMÉTRICA DE BAIXO CUSTO

A prototipação da estação foi dividida em duas etapas. Na primeira etapa foi realizada a montagem do esquemático do circuito eletrônico (Figura 3.3) em *pront-o-board* (Figura 3.4). Em seguida foi desenvolvido o design físico da estação solarimétrica e por fim a sua montagem.

3.3.1 Montagem do Circuito Eletrônico

Com a montagem do circuito eletrônico na placa de ensaio, foi dado início ao desenvolvimento do programa a ser aplicado no microcontrolador. A plataforma utilizada para a compilação e carregamento dos programas desenvolvidos para o sistema, é a Arduino Software IDE (*Arduino Integrated Development Environment*). O desenvolvimento foi realizado na linguagem C/C++. O programa é responsável por realizar a leitura dos dados gerados pelos sensores e enviá-los para o *Firebase* via comunicação Wi-Fi, com uma rede de internet.

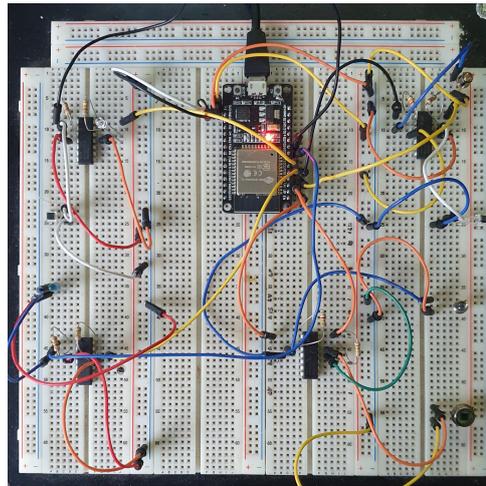


Figura 3.4 – Esquemático do circuito eletrônico montado na placa de ensaio.
Fonte: Autoria própria.

O primeiro teste com a placa de ensaio foi realizado com objetivo de estabelecer a comunicação entre o circuito eletrônico e a IDE utilizada. Em seguida, foi verificado se o microcontrolador estava realizando as leituras e conversões dos sinais gerados pelos sensores, sendo essa conversão de um sinal analógico para digital. O teste de leitura foi realizado com um dos sinais, apenas para que fosse validado o circuito eletrônico, e pode ser observado na Figura 3.5, no qual as leituras estão sendo realizadas e impressas na variável valor.

Estabelecida a comunicação entre os sensores e o microcontrolador, o próximo passo foi realizar a comunicação entre o microcontrolador e o *Firebase*. Primeiramente para que houvesse acesso a plataforma de base dados na nuvem, foi necessário estabelecer a comunicação entre o ESP32 e uma rede de internet Wi-Fi. Após configurar a ligação Wi-Fi, foi preciso setar as configurações necessárias diretamente no *Firebase*, no qual foi criado uma base de dados de teste chamada "*solar*" no *Realtime Database*. Em seguida, foi utilizada a biblioteca "*IOXhopFirebaseESP32*" para realizar a comunicação entre os ESP32 e o banco de dados "*solar*" hospedado na nuvem.

O último passo antes de realizar a solda dos componentes eletrônicos na placa de circuito impresso, é dado pelo teste de comunicação com o banco de dados. O teste foi desenvolvido realizando o envio dos dados coletados de alguns sensores.

```

sketch_jun17a | Arduino 1.8.13 (Windows Store)
Arquivo Editar Sketch Ferramentas Ajuda

sketch_jun17a

void setup() {
  Serial.begin(9600);

  // configura o preescaler do ADC
  ADCSRA &= ~PS_128; //limpa config

  // valores possiveis de prescaler
  // PS_16, PS_32, PS_64 or PS_128
  //ADCSRA |= PS_128; // 64 prescale
  ADCSRA |= PS_128; // 64 prescal
  // ADCSRA |= PS_32; // 32 prescal
  // ADCSRA |= PS_16; // 16 prescale
}

void loop() {
  // leitura
  tempo_inicio = micros(); //marca
  valor = analogRead(0); //le val
  tempo_fim = micros(); //le tem

  //exibe valor lido e tempo de co
  Serial.print("Valor = ");
  Serial.print(valor);
  Serial.print(" -- Tempo leitura
  Serial.print(tempo_fim - tempo_i
  Serial.println(" us");
  delay(500);
}

Carregado
o sketch usa 2080 bytes (6%) de espa
Variáveis globais usam 234 bytes (11
Valor = 63 -- Tempo leitura = 116 us
Valor = 62 -- Tempo leitura = 116 us
Valor = 62 -- Tempo leitura = 116 us
Valor = 62 -- Tempo leitura = 112 us
Valor = 62 -- Tempo leitura = 116 us
Valor = 63 -- Tempo leitura = 112 us
Valor = 62 -- Tempo leitura = 116 us
Valor = 63 -- Tempo leitura = 116 us
Valor = 70 -- Tempo leitura = 116 us
Valor = 77 -- Tempo leitura = 116 us
Valor = 77 -- Tempo leitura = 116 us
Valor = 75 -- Tempo leitura = 116 us
Valor = 80 -- Tempo leitura = 116 us
Valor = 79 -- Tempo leitura = 116 us
Valor = 80 -- Tempo leitura = 112 us
Valor = 80 -- Tempo leitura = 116 us
Valor = 83 -- Tempo leitura = 116 us
Valor = 85 -- Tempo leitura = 116 us
Valor = 84 -- Tempo leitura = 116 us
Valor = 81 -- Tempo leitura = 116 us
Valor = 79 -- Tempo leitura = 116 us
Valor = 81 -- Tempo leitura = 116 us
Valor = 78 -- Tempo leitura = 116 us
Valor = 73 -- Tempo leitura = 116 us
Valor = 105 -- Tempo leitura = 116 us
Valor = 109 -- Tempo leitura = 112 us
Valor = 108 -- Tempo leitura = 112 us
Valor = 106 -- Tempo leitura = 112 us
Valor = 107 -- Tempo leitura = 112 us
Valor = 107 -- Tempo leitura = 112 us
Valor = 104 -- Tempo leitura = 108 us
Valor = 105 -- Tempo leitura = 108 us
Valor = 98 -- Tempo leitura = 112 us
Valor = 86 -- Tempo leitura = 116 us
Valor = 83 -- Tempo leitura = 116 us
Valor = 80 -- Tempo leitura = 116 us
Valor = 79 -- Tempo leitura = 116 us
Valor = 76 -- Tempo leitura = 116 us

```

Figura 3.5 – Teste de leitura do sinal gerado por um dos sensores.
Fonte: Autoria própria.

Além do envio dos dados dos sensores, a data e hora também foram encaminhados. Um exemplo da comunicação estabelecida entre o sistema de aquisição de dados e o banco de dados criado é demonstrado na Figura 3.6. Nesta imagem é possível observar a data e hora do envio, como também os sensores.

Finalizado os testes com o circuito eletrônico montado na placa de ensaio, os componentes foram soldados em uma placa de circuito impresso perfurada. Após a soldagem foi dado início ao projeto do design físico da estação solarimétrica.

3.3.2 Design Físico da Estação Solarimétrica

A última fase do processo de construção da estação solarimétrica de baixo custo foi o projeto físico do equipamento responsável pelo alojamento dos sensores e da placa de circuito impresso (*Printed Circuit Board* - PCB) (exemplo de instalação de uma estação solarimétrica de baixo custo na Figura 3.7).

O primeiro passo foi projetar as peças encarregadas pelo alojamento dos foto-

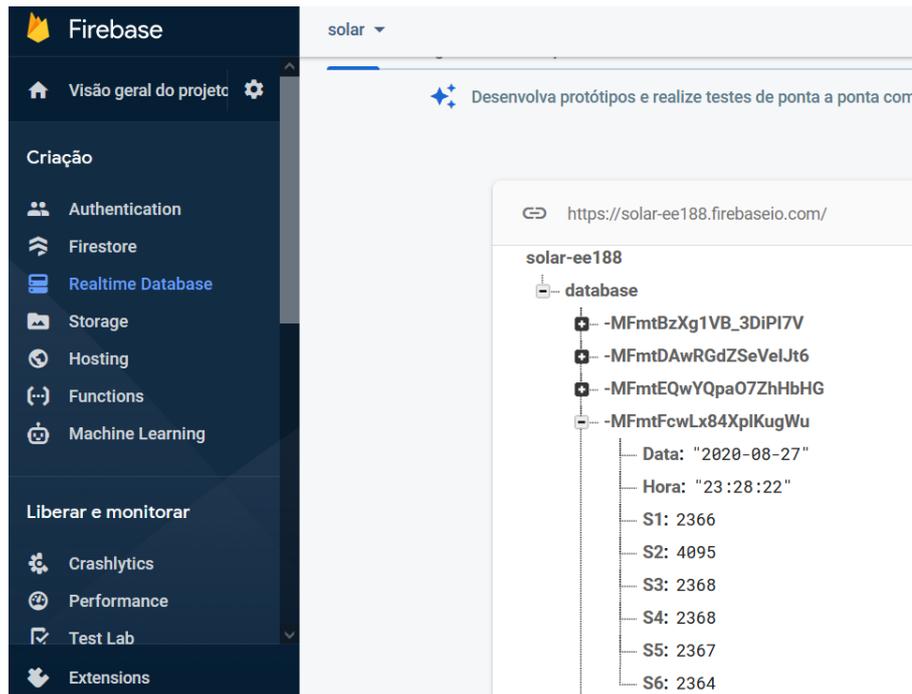


Figura 3.6 – Envio de dados para base de dados hospedada na nuvem.
Fonte: Autoria própria.

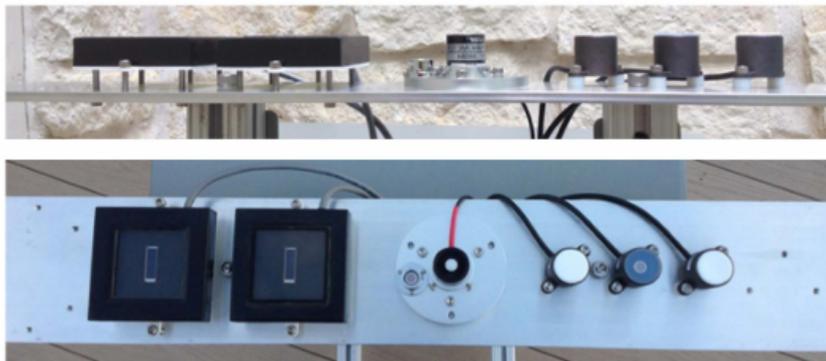
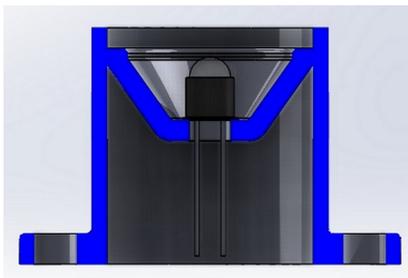


Figura 3.7 – Exemplo de estação solarimétrica de baixo custo.
Fonte: (LAVE et al., 2015).

sensores. O projeto e a peça original impressa por uma impressora 3D, estão expostos respectivamente na Figura 3.8a e 3.8b.

A fixação das peças foi feita em uma placa de acrílico, projetada com 40 centímetros de comprimento e 20 centímetros de largura (Figura 3.9). Além da placa, foi projetado um difusor de acrílico (Figura 3.10), para ser acoplado nas peças responsáveis pelo armazenamento dos fotosensores. Em seguida, foi feita a montagem da placa de acrílico em cima de quatro hastes de metal que são fixadas com outras quatro hastes de madeira na parte inferior da estrutura, como mostrado na Figura 3.11.

Após a montagem da estrutura física da estação solarimétrica, foi realizada a fixação dos sensores já armazenados nas peças projetadas (Figura 3.11). Em seguida foi feito o armazenamento das placas de circuito impresso soldadas dentro de uma



(a)



(b)

Figura 3.8 – Corte da peça projetada(a) e peça impressa(b).
Fonte: Autoria própria

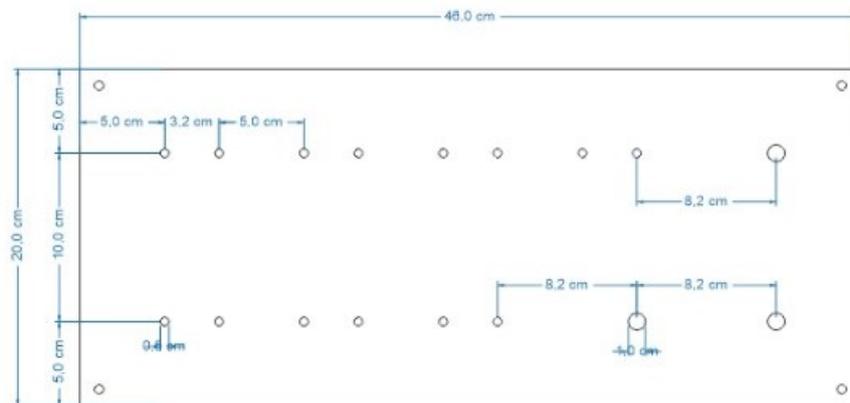


Figura 3.9 – Placa de acrílico responsável por fixar os sensores.
Fonte: Autoria própria.



Figura 3.10 – Peça impressa e difusor de acrílico.
Fonte: Autoria própria.

caixa de plástico preta (Figura 3.12), com intuito de proteger o circuito eletrônico de intemperes provenientes de um ambiente aberto, como a chuva.

No último passo dessa etapa da construção da estação solarimétrica proposta, foi realizar a ligação dos cabos dos sensores com as placas de circuito armazenadas na caixa (Figura 3.12). O resultado final da estação solarimétrica construída é exposto na Figura 3.13.



Figura 3.11 – Estrutura da estação com a placa de acrílico e os sensores fixados.
Fonte: Autoria própria.

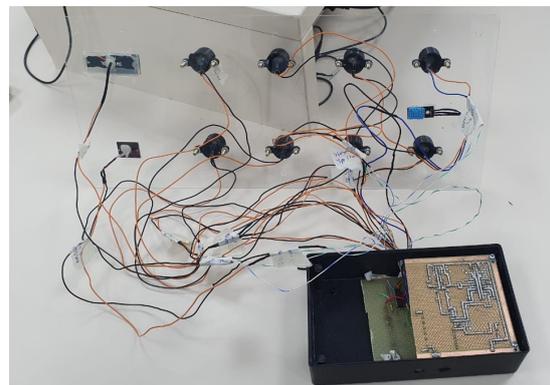
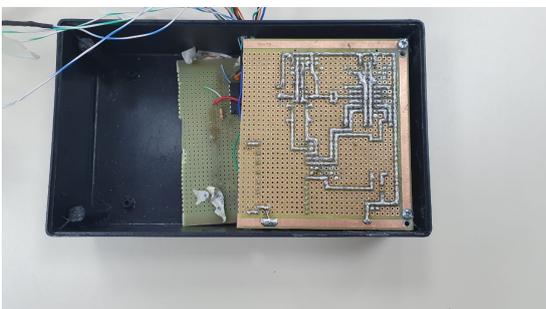


Figura 3.12 – Caixa de armazenamento das placas de circuito impresso.
Fonte: Autoria própria.

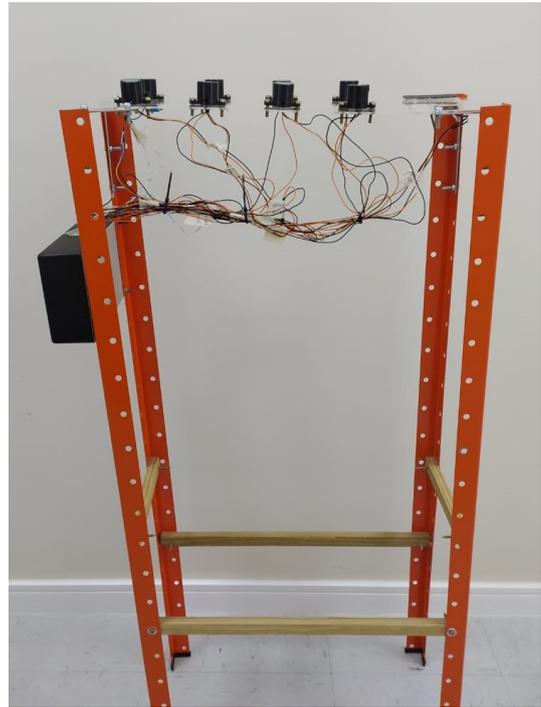


Figura 3.13 – Estação solarimétrica de baixo custo.
Fonte: Autoria própria.

3.4 CONSIDERAÇÕES

O objetivo almejado com a estação solarimétrica de baixo custo, era desenvolver um sistema de aquisição de dados de irradiância e climáticos utilizando sensores de uso geral. A Tabela 3.2 demonstra a quantidade e o custo de cada um dos materiais utilizados para a construção da estação, com um custo final de aproximadamente de R\$ 500,00. Um piranômetro de uso comercial e padrão secundário custa em média R\$ 10.000,00 de acordo com um orçamento feito pela empresa especializada, chamada HukseFlux. Portanto, um preço 20 vezes maior que o valor gasto para a construção da estação solarimétrica de baixo custo.

Tabela 3.2 – Tabela de custo dos materiais utilizados na estação solarimétrica de baixo custo.

Material	Quantidade	Preço
Hastes de ferro	4	R\$ 15.00
Placa de acrílico	1	R\$ 40.00
Difusor	8	R\$ 0.50
Peça sensor	8	R\$ 8.00
BPW21	1	R\$ 49.28
BPW34	1	R\$ 4.20
TPD5700	1	R\$ 3.47
OPA505B	1	R\$ 2.48
TEFD4300F	1	R\$ 2.96
MTD5010W	1	R\$ 21.48
MTD3910N	1	R\$ 21.48
PT333-3C	1	R\$ 1.20
SM111K09L	1	R\$ 28.12
AM-5610CAR	1	R\$ 19.20
DHT11	1	R\$ 14.90
ESP32	2	R\$ 50.00
Total		R\$ 436.77

Fonte: Autoria própria.

Inicialmente foi desenvolvido o projeto de um sistema de aquisição, no qual foi desenvolvida uma placa de circuito impresso (*Printed Circuit Board* - PCB) que tem o objetivo de realizar a coleta dos dados adquiridos com os sensores e enviá-los via comunicação serial para um computador, onde os dados serão tratados e analisados. O projeto da placa foi baseado no microcontrolador TM4C123GH6PM, embarcado na plataforma Tiva LaunchPad Series C, da empresa Texas Instruments, que será responsável por controlar o funcionamento do sistema de aquisição. Em seguida foi selecionado um conversor sigma-delta ADS131E08 para a conversão dos dados analógicos de saída dos sensores para valores digitais, sendo um conversor de 24bits que possui 8 canais. Foram projetados circuitos condicionadores de sinal entre a saída dos sensores e a entrada dos conversores analógico digital, para melhor utilizar

a faixa de entrada do conversor. A placa (Figura 3.14) foi desenvolvida visando os testes simultâneos de até 24 sensores diferentes, fotodiodos, fototransistores, células fotovoltaicas, foto resistores e sensores de temperatura. A Figura 3.15 representa o esquemático elétrico da placa responsável pela interpretação dos sinais gerados pelos sensores.

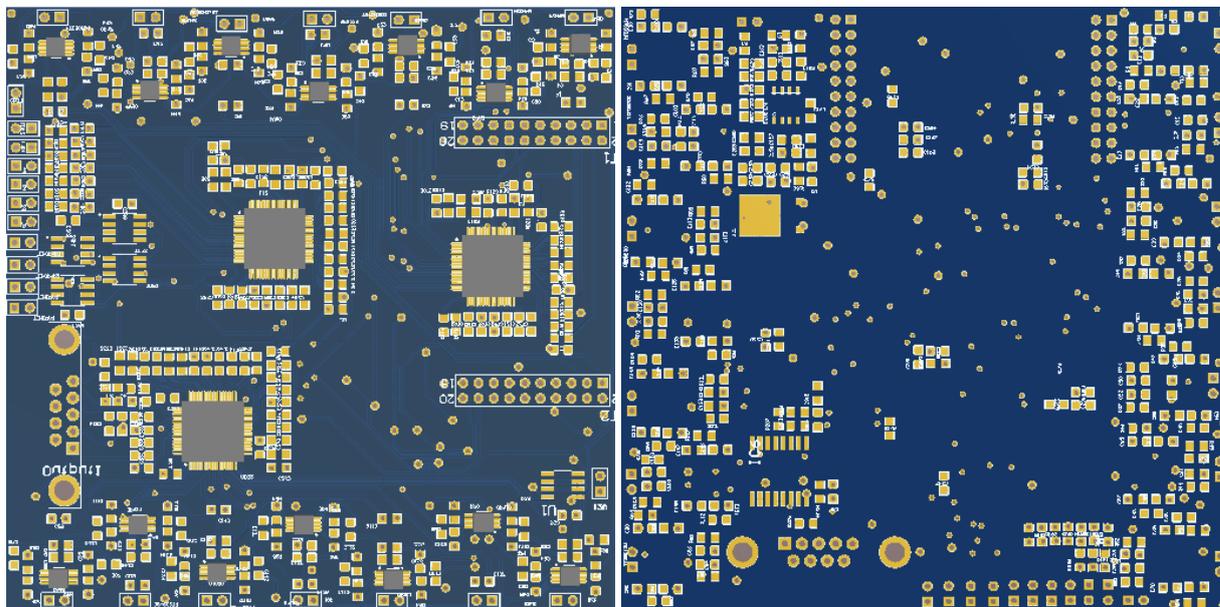


Figura 3.14 – Imagem 3D das faces da placa eletrônica de aquisição de dados.

Fonte: Autoria própria

Esta frente de pesquisa foi interrompida no estado descrito devido ao início da pandemia declarada pela Organização Mundial da Saúde em Março de 2020, causada pelo *SARS-CoV-2*. A pandemia causou problemas com a fabricação das placas de circuito impresso, que eram desenvolvidas por uma empresa situada na China. Além disso, o fechamento de ambientes de trabalho nas universidades ou horário reduzido, dificultando a execução de algumas tarefas deste trabalho. Apesar disso, a construção da estação solarimétrica foi concluída com sucesso. Não foi possível, porém, a execução de testes de validação, frente a estação solarimétrica comercial presente no Laboratório Fotovoltaica localizada na Universidade Federal de Santa Catarina na cidade de Florianópolis.

O próximo capítulo abordará os conceitos de Aprendizado de Máquina utilizados ao longo do desenvolvimento do trabalho.

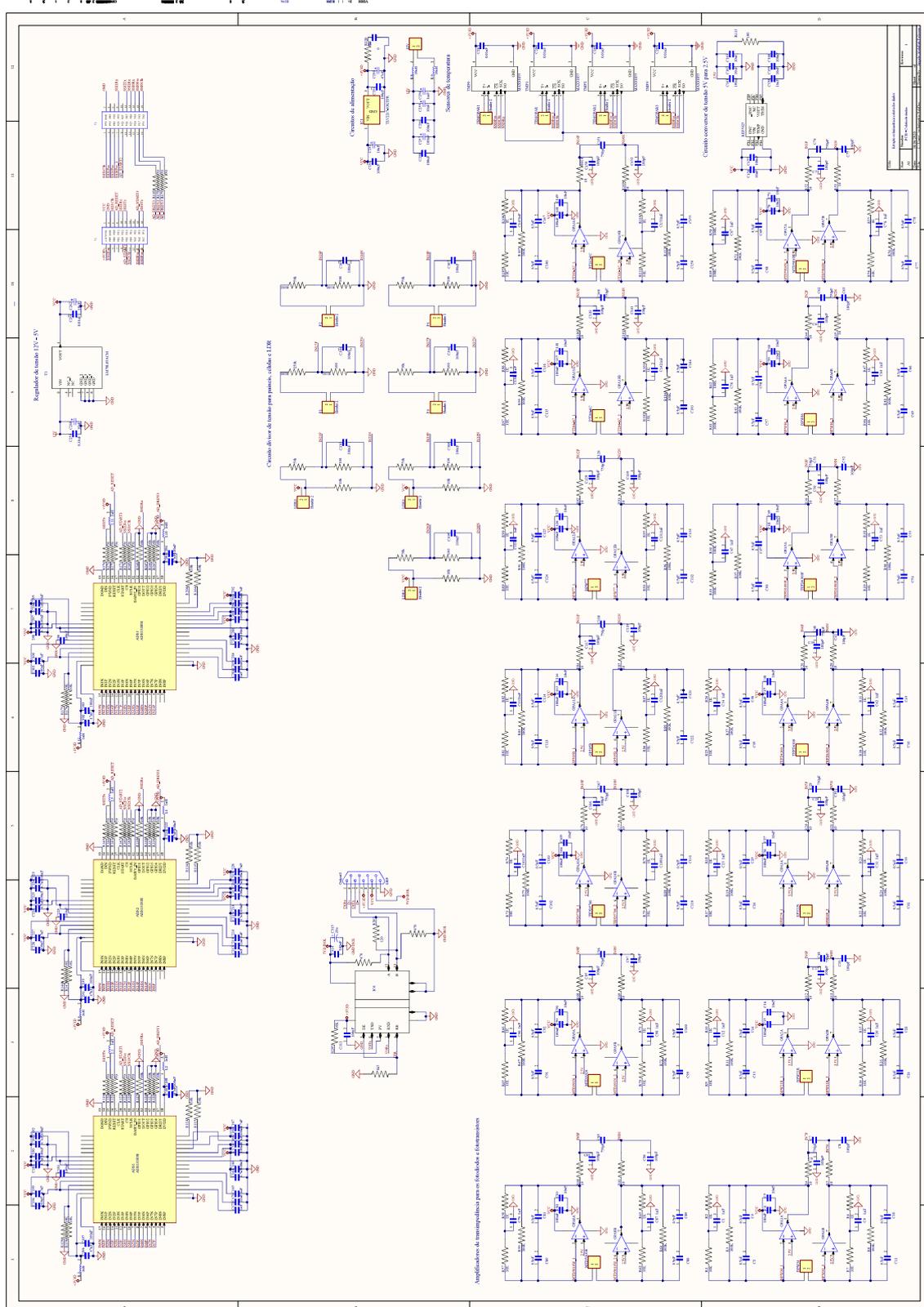


Figura 3.15 – Esquemático da placa eletrônica do sistema de aquisição de dados.
Fonte: Autoria própria

4 CONCEITOS DE APRENDIZADO DE MÁQUINA

Os conceitos de aprendizado de máquina são utilizados desde 1950 e fazem parte de um dos campos de pesquisa mais abordados na engenharia de computação moderna, a inteligência artificial (WANG; MA; ZHOU, 2009; SHINDE; SHAH, 2018). O aprendizado de máquina envolve diferentes processos, podendo ser brevemente definido como um processo de aprendizado computacional que a partir da experiência adquirida por meio da execução de tarefas e uma métrica de desempenho, consegue aperfeiçoar a sua habilidade de realizar determinada tarefa (MITCHELL, 1997).

Os métodos de aprendizado de máquina são algoritmos, nos quais são aplicados a uma base de dados de um determinado problema. Os algoritmos tem a capacidade de melhorar o seu desempenho adaptando a sua estrutura por meio do processo de aprendizado. O processo de aprendizado pode ser dividido em dois paradigmas de acordo com a base histórica de dados fornecida por um determinado problema, sendo eles o aprendizado supervisionado e não supervisionado (NAQA; MURPHY, 2015).

O aprendizado supervisionado é utilizado quando a base de dados entregue para a solução de uma determinada tarefa é composta por um conjunto de dados com instâncias rotuladas. As instâncias são consideradas como o conjunto de registros dos atributos/variáveis presentes no conjunto de dados. O rótulo de cada instância é responsável por representar o fenômeno desejado. Os atributos que constituem as instâncias e os rótulos são considerados respectivamente variáveis independentes e dependentes (ROKACH; MAIMON, 2010).

Os métodos de aprendizado supervisionado têm o objetivo de descrever a relação entre as variáveis independentes de entrada e o atributo alvo (variável dependente). Os modelos de classificação e regressão são os dois principais métodos de aprendizado supervisionado. Problemas de classificação consistem em métodos classificadores aplicados a uma base de dados rotulada em classes, nos quais os métodos realizam a identificação das classes de cada instância após serem treinados na base de dados histórica rotulada. Os métodos de regressão tem o objetivo de mapear a relação entre as variáveis de entrada e o atributo alvo por meio do treinamento do modelo em uma base de dados histórica, gerando previsões para novas instâncias de entrada (HAYKIN, 2008; SU; YAN; TSAI, 2012; ROKACH; MAIMON, 2010).

O aprendizado não supervisionado se refere a métodos aplicados a uma base de dados no qual os seus objetos não possuem rótulos. Os métodos de aprendizado não supervisionado possuem o objetivo de extrair características a partir dos dados brutos de uma base de dados. A sua utilização pode ajudar na identificação e revelação de padrões e grupos de dados caracterizados por causas físicas, que não são claramente observadas em uma base de dados brutos (HINTON; SEJNOWSKI et al., 1999).

Para o desenvolvimento deste trabalho, foram utilizados métodos de aprendizado supervisionado, com o objetivo de resolver um problema de regressão. Neste capítulo serão descritas as técnicas e conceitos de aprendizado de máquina utilizados no desenvolvimento do trabalho. Entre elas estão, a análise exploratória e pré-processamento dos dados, métodos de aprendizado de máquina e métricas de avaliação de desempenho.

4.1 TRATAMENTO DE DADOS

Esta seção abordará os conceitos de análise exploratória de dados utilizados para o desenvolvimento do trabalho. A investigação dos dados está dividida em duas etapas, a análise e o pré-processamento, que consistem em uma das etapas mais importantes de um trabalho que envolve a utilização de métodos de aprendizado de máquina. Para que os métodos de aprendizado de máquina consigam alcançar o seu objetivo, é necessário que a base de dados em que os métodos são aplicados tenha uma boa qualidade.

4.1.1 Análise Exploratória de Dados

A análise exploratória dos dados consiste no estudo e investigação da base de dados. Com isso, é realizada a identificação dos atributos da base de dados e como eles são distribuídos. O conjunto de atributos ou variáveis, é responsável por descrever as características da base de dados. Essas características podem ser, por exemplo, a temperatura ambiente e a umidade relativa do ar. Cada atributo então é representado por um conjunto de objetos, em que cada objeto corresponde a uma observação (por exemplo, a temperatura ambiente) em um instante tempo t .

Em seguida foi feita a análise de correlação de *Pearson*. A análise de correlação é comumente usada na análise de dados em trabalhos que tem como finalidade a previsão de geração de energia fotovoltaica (BUGALA et al., 2018; DOBREVA; DYK; VORSTER, 2020). O coeficiente de correlação tem o objetivo de avaliar a relação linear entre pares de atributos de uma base de dados. O cálculo do coeficiente de correlação entre dois atributos, x e y , é dado pela Equação (4.1), no qual n é o número de objetos (ZHONG et al., 2018).

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.1)$$

Em que:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

A interpretação da correlação é baseada no valor de r calculado, podendo ela ser de valor negativo ou positivo. Quando o valor for negativo, significa que o par de variáveis analisado possui uma correlação negativa, que acontece devido as variáveis serem inversamente proporcionais. A correlação positiva demonstra que as variáveis possuem uma relação diretamente proporcional, ou seja, quando o valor de x aumenta o de y também aumenta. O grau de correlação entre os pares de variáveis pode ser classificado, como: muito fraca ($0.00 < r < 0.19$), fraca ($0.20 < r < 0.39$), moderada ($0.40 < r < 0.69$), forte ($0.70 < r < 0.89$) e bem forte ($0.90 < r < 1.00$).

A etapa de análise de dados envolve também a seleção dos atributos. A seleção dos atributos realizada no trabalho não foi feita baseada em nenhum método de seleção de atributos específico. A seleção dos atributos foi feita baseada no coeficiente de correlação entre os conjuntos de atributos disponíveis e o conjunto de observações dos dados de potência de saída do sistema fotovoltaico. Esse método de seleção dos atributos foi baseada nos trabalhos de Zhong et al. (2018), Bugala et al. (2018), Khan et al. (2017), que tinham como um dos objetivos de seus trabalhos realizar a previsão de geração de energia fotovoltaica.

4.1.2 Pré-processamento de Dados

O pré-processamento dos dados possui como objetivo realizar a limpeza da base dados e prepará-la para a sua utilização como entrada dos métodos de aprendizado de máquina. Nesta etapa um dos objetivos é encontrar problemas contidos na base de dados. Esses problemas podem ser fruto de falhas em sensores, valores ausentes ou inconsistentes, e valores anormais em relação a distribuição dos dados, chamados de *outliers*.

Para a limpeza dos dados foi necessário realizar uma análise de *outliers* no conjunto de atributos selecionado na etapa anterior. Essa análise pode ser realizada por meio de um tipo específico de gráfico, o *boxplot*. Os gráficos de *boxplot* foram utilizados para auxiliar na visualização e identificação de valores anormais, que se dispersam em relação a distribuição normal dos dados (KWAK; KIM, 2017).

Além da identificação de possíveis *outliers*, outras abordagens para a limpeza dos dados foram utilizadas neste trabalho. Foi utilizado um filtro para remoção de observações com valores consecutivos idênticos, justificados por falhas de sensores e do equipamento responsável por realizar a gravação dos dados. Para lidar com os valores ausentes da base de dados, foi utilizado um algoritmo de aprendizado de máquina (kNN *Imputer*) para a imputação destes valores.

4.1.2.1 Imputação de Valores Ausentes

A imputação de dados é um tópico importante quando estamos tratando valores ausentes. Existem diversas técnicas de imputação de dados e alguns tipos de valores ausentes, dentre eles *Missing At Random* (MAR), *Missing Completely At Random* (MCAR) e *Missing Not At Random* (MNAR). Os valores do tipo MAR são dados que estão aleatoriamente ausentes na base de dados, a sua falta é dependente de um fator externo. Por exemplo, um formulário sobre depressão em que homens e mulheres respondem diariamente, porém a porcentagem de homens que respondem corretamente o formulário é baixa, deixando algumas perguntas em branco devido a fatores externos, sendo assim gerando uma variável ausente no banco de dados de respostas (DONDEERS et al., 2006).

Os dados do tipo MCAR são considerados totalmente aleatórios, no qual a falta das observações não estão relacionadas a um motivo específico. Os valores MCAR podem ser gerados por falhas de leituras de sensores e de equipamentos de gravação. Os valores ausentes do tipo MNAR são valores que existem porém foram omitidos da base dados, por exemplo, uma empresa que não declara os seus impostos. No caso do trabalho proposto, os valores faltantes são do tipo MCAR, isso devido a sua ausência ser justificada por falhas de gravação e dos sensores (ROYSTON, 2005; DONDEERS et al., 2006).

Para a imputação de valores ausentes neste trabalho, foi utilizado o método do k vizinho mais próximo (kNN - *k Nearest Neighbor*), que quando usado no contexto de valores ausentes é chamado de kNN *Imputer* (ZHANG, 2012). Dados valores de uma determinada instância, esse método tem como objetivo encontrar os vizinhos mais próximos a partir das instâncias completas da base dados utilizada. O kNN precisa desempenhar uma busca entre todas instâncias, exceto a que possui o valor faltante, e selecionar as instâncias com valores mais próximos, que são calculados por meio de uma distância. O cálculo das distâncias pode ser feito, por exemplo, com a distância Euclidiana. Assim, uma opção é observar o valor faltante nos vizinhos mais próximos e calcular a média para realizar a imputação desse valor (ZHANG, 2012; KIM; KIM; YI, 2004).

4.2 MÉTODOS DE APRENDIZADO DE MÁQUINA

Esta seção abordará os métodos de aprendizado de máquina utilizados no desenvolvimento do trabalho. Todos os métodos foram treinados e validados utilizando *scikit-learn*, exceto o modelo MLP utilizado na Estratégia de Previsão Baseada em Modelos Individuais, no qual foi usado o *keras* (PEDREGOSA et al., 2011; CHOLLET et al., 2015). Além dos métodos de aprendizado de máquina, será apresentada a técnica de combinação de métodos utilizada no desenvolvimento do trabalho.

4.2.1 Redes Neurais Artificiais

A primeira ideia de uma Rede Neural como uma máquina de computação foi introduzida por McCulloch e Pitts (1943), e na década de 1950, Rosenblatt (1958) propôs o primeiro modelo de Redes Neurais Artificiais (RNA), chamado *Perceptron*. De acordo com Haykin (2008), uma Rede Neural é uma máquina desenvolvida para modelar a forma em que o cérebro humano desempenha tarefas específicas. O modelo de neurônio proposto por Rosenblatt (1958), é a base para a construção de uma RNA. O neurônio é um modelo matemático linear que possui um limitador (HAYKIN, 2008). Um modelo de neurônio é representado pela Figura 4.1, para um neurônio k .

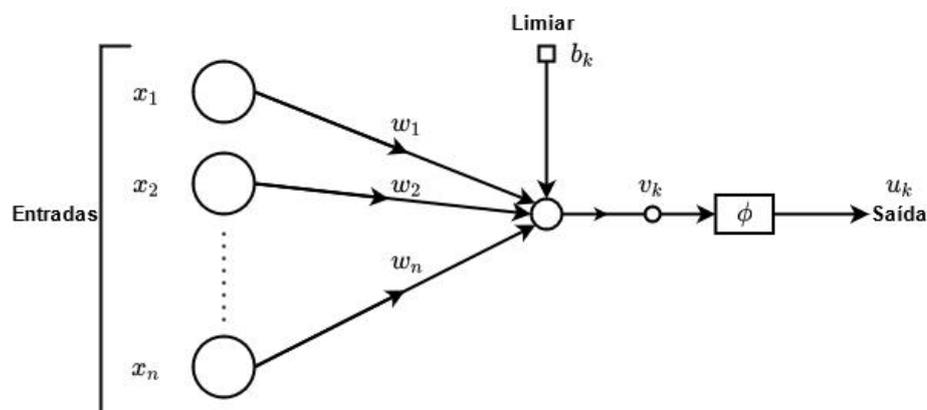


Figura 4.1 – Modelo de um neurônio artificial.

Fonte: Adaptado de Haykin (2008).

As entradas do neurônio são representadas por x_1, x_2, \dots, x_n , e os pesos sinápticos por w_1, w_2, \dots, w_n . O limiar externo b_k é aplicado na saída dos neurônios, gerando o somatório v_k que é aplicado a função de ativação ϕ . O somatório v_k é dado pela Equação (4.2).

$$v_k = \sum_{i=1}^n w_i x_i + b_k \quad (4.2)$$

A função de ativação é responsável por definir a saída do neurônio (HAYKIN, 2008). De acordo com Haykin et al. (2009), um exemplo de função de ativação muito comum é a função sigmoideal logística, dada pela Equação (4.3). Essa função vai gerar uma saída limitada entre 0 e 1, que é determinada por meio da variação do parâmetro de inclinação a , que deve ser maior que 0.

$$\phi(v_k) = \frac{1}{1 + \exp(-a * v_k)}, a > 0 \quad (4.3)$$

4.2.1.1 Perceptron Multicamadas

A Rede Neural Perceptron Multicamadas (*Multilayer Perceptron* - MLP) é utilizada tanto para problemas de regressão como para classificação (HAYKIN, 2008). Possui a capacidade de mapear a relação não linear entre as variáveis de entrada e

saída. A estrutura de uma Rede Neural MLP é composta por três camadas, sendo elas a camada de entrada, a camada oculta e a de saída. O número de camadas ocultas pode ser variado, permitindo que a MLP possua mais de três camadas. Cada camada possui um número determinado de neurônios. A Figura 4.2 representa a estrutura de uma MLP constituída por três camadas.

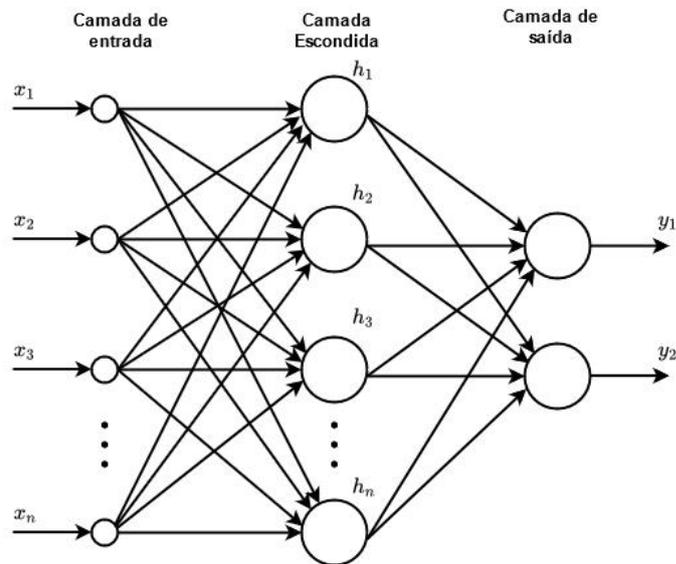


Figura 4.2 – Modelo de uma MLP com uma camada escondida.

Fonte: Adaptado de Haykin (2008).

Esse modelo de Rede Neural possui um alto grau de conectividade entre os neurônios, no qual todos os neurônios de uma determinada camada estão conectados com os neurônios da camada anterior e da seguinte. O sinal de entrada x_n aplicado a MLP é propagado para frente, camada por camada, ou seja, o sinal de entrada de cada um dos neurônios que compõem cada uma das camadas são as saídas dos neurônios da camada anterior (HAYKIN, 2008).

As MLPs têm sido aplicadas com sucesso em diversos campos de estudos. Os resultados obtidos por meio dessas aplicações são gerados a partir de um processo chamado de treinamento. O algoritmo de treinamento mais comumente utilizado na aplicação da MLP, é o *Back-Propagation* (HAYKIN et al., 2009).

O processo de treinamento *Back-Propagation* é dividido em duas etapas. Na primeira etapa, os sinais de entrada são propagados através das camadas da rede até alcançar a saída. A camada de entrada e oculta possuem pesos sinápticos em todas as conexões entre os neurônios, e cada um dos neurônios possuem uma função de ativação. Portanto, a qualidade do resultado final da primeira etapa do processo depende majoritariamente dos pesos e da entrada de cada neurônio. Na segunda etapa do algoritmo *Back-Propagation*, é executado o caminho inverso realizado na primeira etapa. O resultado obtido na primeira etapa do processo é comparado com o resultado desejado, e é calculado um erro. O sinal de erro é propagado da camada de

saída para a camada de entrada, e em cada uma das camadas é realizado um cálculo de reajuste dos pesos. O processo é repetido até o final do treinamento da MLP, com objetivo de alcançar um desempenho melhor através do reajuste dos pesos (ZHONG et al., 2018; HAYKIN, 2008).

4.2.2 *Support Vector Machines*

O modelo de máquinas de vetor de suporte (*Support Vector Machines* - SVM) é um método estatístico desenvolvido por VAPNIK (1995) e utilizado para resolver problemas de classificação. Esse método também pode ser utilizado para resolver problemas de regressão, sendo chamado de *Support Vector Regressor* (SVR) (LIU et al., 2011).

O objetivo do SVM é encontrar o melhor ajuste dos dados para um hiperplano. O SVR utiliza do mesmo princípio, porém no caso da regressão é estipulado um valor limite ou margem máxima de erro permitido. O objetivo é minimizar o erro de uma forma que o hiperplano se aproxime do valor real no treinamento. Então, o ajuste do hiperplano é calculado com intuito de ajustá-lo entre os valores limites em que irá existir o maior número de dados. Sendo assim, a predição final será o valor do hiperplano (ARIZE; RIOS, 2019; AWAD; KHANNA, 2015).

4.2.3 *k - Vizinhos Mais Próximos*

O algoritmo *k*-Vizinhos Mais Próximos (*k-Nearest Neighbors* - kNN) é um dos métodos de aprendizado de máquina mais simples para resolver problemas de classificação e regressão. O algoritmo kNN têm sido muito utilizado devido a sua simplicidade e baixo custo computacional (ZHANG, 2012).

O funcionamento do kNN é baseado no cálculo da distância entre os dados armazenados no treinamento e o dado para o qual se deseja realizar uma predição. O cálculo da distância pode ser feito, por exemplo, com a distância Euclidiana. Após o cálculo da distância entre os objetos presentes no treinamento e objeto a ser predito, o número de vizinhos k dever ser definido. O número de vizinhos irá determinar o número de objetos que serão utilizados para realizar predição. Portanto, a predição final será definida, por exemplo, pelo cálculo da média dos objetos de menor distância do objeto a ser predito (MYLES et al., 2004; PAOLI et al., 2010).

4.2.4 *Árvore de Decisão*

A primeira árvore de decisão ou *Decision Tree* (DT) foi proposta por Breiman (1984). É um método que pode ser usado na extração de padrões em uma base dados e resolver problemas de regressão e classificação. O método gera uma árvore que se inicia a partir de um nó ramificando-se em possíveis resultados. Os possíveis

resultados tornam-se nós, que se ramificam em mais nós até chegar a uma decisão final (MITCHELL, 1997).

O resultado final é dado por um nó folha, que é definido por meio do critério de parada. O critério de parada pode ser, por exemplo, um número mínimo de nós folhas gerados por um nó de decisão. A partir das regras induzidas pelas árvores geradas é definida a predição final (MYLES et al., 2004; MITCHELL, 1997).

4.2.5 Regressão Linear

A regressão linear (*Linear Regression* - LR) é o modelo de aprendizado de máquina mais utilizado para problemas de regressão, por ser simples, fácil de se aplicar e garante bons resultados. O método LR é definido por uma equação matemática linear, permitindo uma análise estatística simples (HAYKIN, 2008).

O método tem o objetivo de mapear a relação entre pares de variáveis por meio do cálculo de ajuste de uma equação linear. São essas as variáveis dependentes e independentes já abordadas anteriormente no início do capítulo. O cálculo da equação linear realizará o ajuste de uma linha reta aos dados de treinamento, modelando a relação linear entre os objetos de treinamento e saída do modelo. Portanto, a partir da relação linear entre as variáveis de entrada e saída o modelo gerará suas predições (SU; YAN; TSAI, 2012).

4.3 COMITÊ DE MÉTODOS DE APRENDIZADO DE MÁQUINA

Um modelo *Ensemble* ou de comitê de métodos é definido pela combinação dos resultados de diferentes métodos de aprendizado de máquina (GANAIÉ et al., 2021). As abordagens que fazem a utilização de *Ensembles*, tem como objetivo melhorar o desempenho de uma predição a partir da combinação das predições de vários métodos (VALENTINI; MASULLI, 2002).

Os modelos *Ensemble* utilizados no desenvolvimento do presente trabalho, foram aplicados em um problema de regressão. De acordo com Allende e Valle (2017), o uso de modelos *Ensemble* neste tipo de problema possui suas vantagens, podendo melhorar a precisão e reduzir os erros de previsão obtidos por um determinado método. O uso de *Ensembles* em problemas de regressão, pode também auxiliar na generalização dos resultados e reduzir variações do modelo de previsão (RIBEIRO; COELHO, 2020; MENDES-MOREIRA et al., 2012). Nas próximas seções serão abordados os três tipos de *Ensemble* mais utilizados na literatura: *Bagging*, *Boosting* e *Stacking* (DIVINA et al., 2018).

4.3.1 *Bagging*

O *Bagging* ou *Bootstrap Aggregating* é um modelo de comitê de métodos proposto primeiramente na década de 1990 por Breiman (1996). Esse modelo faz o uso de múltiplos conjuntos de dados, construídos a partir de uma única base de dados. Os conjuntos de dados são criados a partir do método de amostragem *Bootstrapping*, que realiza a seleção dos objetos gerando novos conjuntos de dados aleatórios com o mesmo tamanho do conjunto original e com a reposição de objetos. Esses conjuntos são, então, aplicados individualmente em métodos distintos. Os resultados de cada um dos métodos são agregados para se obter um resultado final (DIVINA et al., 2018; VALENTINI; MASULLI, 2002).

O resultado ou previsão final, pode ser obtido por meio de votação se estiver sendo aplicado a um problema de classificação, ou por meio da média final das previsões caso o problema abordado seja de regressão (RIBEIRO; COELHO, 2020). Um modelo *Bagging* é composto por um algoritmo de aprendizado de máquina base e o número de modelos que vão ser gerados a partir do algoritmo base (MENDES-MOREIRA et al., 2012).

No desenvolvimento deste trabalho foi utilizado um método de *Bagging*, chamado de *Random Forest* (RF). O algoritmo RF foi proposto por Breiman (2001), sendo este algoritmo um método de aprendizado de máquina que aplica o conceito de *Bagging*, porém com características adicionais, como a seleção aleatória de atributos. O RF realiza a construção de múltiplas árvores de decisão, e a partir da predição de cada uma das árvores construídas, a predição final é obtida.

4.3.2 *Boosting*

Os modelos *Boosting* foram introduzidos por Schapire (1990). Esses são treinados em diferentes conjuntos de dados, assim como para os modelos *Bagging*. A diferença entre os modelos *Bagging* e *Boosting* é que para os modelos *Boosting*, os métodos são aplicados sequencialmente. Para problemas de regressão e classificação, isso significa que o erro de predição do primeiro método influenciará nos pesos de treinamento do método seguinte, conseqüentemente influenciando nos seus resultados. A predição final é dada quando um método não for capaz de reduzir o erro de predição do método anterior (DIVINA et al., 2018; VALENTINI; MASULLI, 2002).

Modelos *Boosting* são responsáveis pela combinação de algoritmos de aprendizado de máquina, capazes de obter um desempenho melhor do que a utilização de apenas um algoritmo (GANAIE et al., 2021). Entre os modelos *Boosting* utilizados neste trabalho, estão os algoritmos de aprendizado de máquina, *GradientBoost* (GB) e *AdaBoost* (AB) (RIBEIRO; COELHO, 2020; FREUND; SCHAPIRE, 1997; FRIEDMAN, 2001).

4.3.3 *Stacking*

O primeiro modelo *Stacking Ensemble*, foi introduzido na década de 1990 por Wolpert (1992). A construção de um modelo *Stacking* é feita por métodos de aprendizado de máquina base e um *meta-learner*. O *meta-learner* é um método de generalização ou meta-modelo, que vai combinar as previsões obtidas por meio dos métodos base (DIVINA et al., 2018).

A arquitetura de um modelo *Stacking*, é composta por duas camadas (Figura 4.3). Na primeira camada, estão os métodos base, no qual serão aplicados os dados de entrada e serão geradas as previsões individuais dos métodos. A segunda camada existente é integrada pelo meta-modelo (*meta-learner*), que vai receber como entrada as previsões individuais, e gerar a previsão final do modelo *Stacking* (GANAIÉ et al., 2021; RIBEIRO; COELHO, 2020).

O meta-modelo utilizado nesse trabalho, foi o *Least Absolute Shrinkage and Selection Operator* (LASSO). O LASSO é um método de aprendizado de máquina aplicado em problemas de regressão, que tem como objetivo minimizar os erros de previsão por meio da redução dos coeficientes de regressão das amostras que não possuem uma contribuição relevante para a previsão final do modelo (RANSTAM; COOK, 2018).

No desenvolvimento deste trabalho foram utilizados dois modelos *Stacking*. Um *Stacking Ensemble* aplicando o conceito abordado e um modelo *Weighted Stacking Ensemble*, no qual a previsão dos modelos base são ponderadas. Para o desenvolvimento do modelo de comitê de métodos ponderados foi utilizado o conceito de função de pertinência da teoria de conjuntos *fuzzy*. Em seguida, serão abordados os conceitos de teoria de conjuntos *fuzzy* utilizados no desenvolvimento do trabalho.

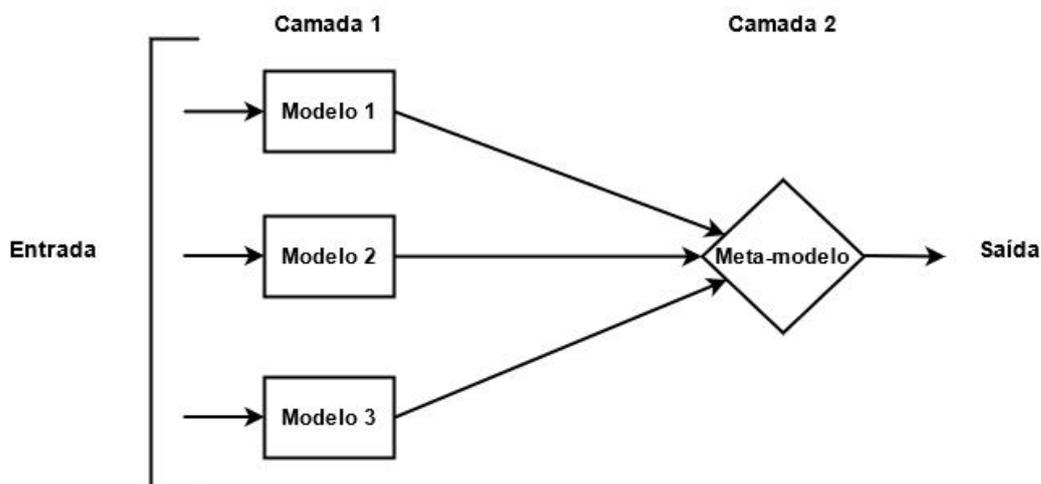


Figura 4.3 – Arquitetura de um modelo *Stacking Ensemble*.

Fonte: Autoria própria.

4.4 TEORIA DE CONJUNTOS FUZZY

A Teoria de Conjuntos *Fuzzy* foi introduzida por Zadeh (1965). Os conceitos dessa teoria foram construídos a partir de conceitos estabelecidos pela lógica clássica. Essa teoria tem como objetivo tratar de forma ordenada objetos que não possuem suas classes explicitamente definidas (DUBOIS D.; YAGER, 1993). Um conjunto *Fuzzy* pode ser considerado como uma classe em um conjunto de dados. O conjunto de dados pode ser composto por mais de uma classe, sendo que um objeto nesse conjunto pode pertencer a mais de uma classe. A Teoria de Conjuntos *Fuzzy* tem o objetivo de dimensionar o grau de pertinência de um objeto em uma classe.

Um exemplo dado por Zadeh (1965) como um problema que poderia ser resolvido pela Teoria de Conjuntos *Fuzzy*, seria de uma dona de casa que ao ir no mercado precisa selecionar os produtos ordenadamente. Os produtos devem ser escolhidos de uma forma que atenda os requisitos da dona de casa da melhor maneira possível, atendendo critérios conflitantes como preço, qualidade e variedade. Nesse problema seriam considerados como classes os critérios definidos e os produtos que atenderiam as necessidades da dona de casa os objetos. Assim, os produtos selecionados deveriam pertencer a mais de uma das classes.

4.4.1 Função de Pertinência

Um conjunto *Fuzzy* é definido por meio de uma função de pertinência (*membership function*). Portanto, vamos considerar que um conjunto *Fuzzy* A em um conjunto de dados X é descrito por uma função de pertinência $f_A(x)$, onde x é um valor discreto do objeto. Então, a função de pertinência relaciona os pontos presentes em X com um valor entre 0 e 1, assim representando o grau de pertinência de x no conjunto A (ZADEH, 1965).

No desenvolvimento da estratégia de previsão baseada nos tipos climáticos proposto por este trabalho que será discutida nos capítulos seguintes, foram utilizadas funções de pertinência em duas formas. As funções de pertinência de forma trapezoidal e triangular, foram utilizadas para definir os pesos de ponderação do modelo *Weighted Stacking Ensemble*.

As funções de pertinência nas formas trapezoidal e triangular, são funções *piecewise-linear* (conjuntos de retas). Estas funções podem ter formas simétricas ou assimétricas, representadas por trapézios e triângulos (ZHAO; BOSE, 2002). Na Figura 4.4, o trapézio é descrito pelos parâmetros k_a , k_b , k_c e k_d . Os parâmetros k_a e k_d representam os pontos na base do trapézio, enquanto k_b e k_c os ombros do trapézio. Enquanto no triângulo (Figura 4.5), k_b representa o topo e os parâmetros k_a e k_c a base.

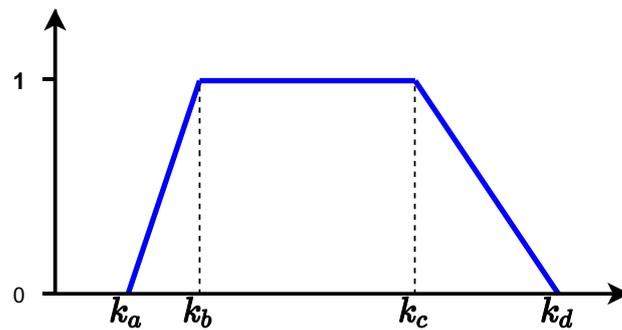


Figura 4.4 – Função de pertinência forma trapezoidal.
Fonte: Autoria própria.

A função de pertinência na forma trapezoidal é dada por:

$$f(x, k_a, k_b, k_c, k_d) = \max \left\{ \min \left(\frac{x - k_a}{k_b - k_a}, 1, \frac{k_d - x}{k_d - k_c} \right), 0 \right\} \quad (4.4)$$

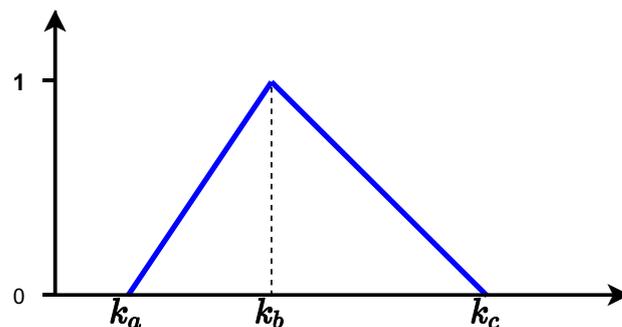


Figura 4.5 – Função de pertinência forma triangular.
Fonte: Autoria própria.

A função de pertinência na forma triangular é dada por:

$$f(x, k_a, k_b, k_c) = \max \left\{ \min \left(\frac{x - k_a}{k_b - k_a}, \frac{k_c - x}{k_c - k_b} \right), 0 \right\} \quad (4.5)$$

Na próxima seção serão discutidas as métricas de erro utilizadas para a avaliação dos modelos de previsão propostos no presente trabalho.

4.5 MÉTRICAS DE ERRO

A avaliação de desempenho dos modelos de previsão desenvolvidos neste trabalho é baseada em algumas métricas de erro. De acordo com Raza, Nadarajah e Ekanayake (2016), é necessário o uso métricas que indiquem a qualidade das previsões, devido ao alto índice de variação na geração de energia fotovoltaica ocasionado pelos fatores abordados Capítulo 2.

Uma das métricas utilizadas no desenvolvimento do projeto é o coeficiente de determinação R^2 calculado com base na correlação de *Pearson*, representado por

uma grandeza entre 0 e 1. Quanto mais próximo de 1 estiver o valor melhor é o resultado alcançado (PAK; OH, 2010). A Equação (4.6) define o cálculo do R^2 , em que y_i representa o valor real da i -ésima observação, \hat{y}_i é o resultado obtido com o modelo de previsão para i -ésima observação e \bar{y} é a média de todos os valores observados (COIMBRA; KLEISSL; MARQUEZ, 2013).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.6)$$

Outra métrica utilizada é Raiz Quadrada do Erro-Médio Quadrático (*Root Mean Square Error - RMSE*), o erro é dado pela raiz quadrada da soma dos quadrados entre a diferença dos valores reais e previstos, divididos pela quantidade de valores considerados. O RMSE revela o tamanho do desvio dos erros, sendo importante para que possamos analisar a grandeza dos erros e o quão preciso é o modelo de previsão desenvolvido (HOFF; PEREZ, 2010; ZHANG et al., 2015). A Equação (4.7) define o cálculo do RMSE, em que y_i representa o valor real da i -ésima observação, \hat{y}_i é o resultado obtido com um modelo de previsão e n a quantidade valores considerados (COIMBRA; KLEISSL; MARQUEZ, 2013; RAZA; NADARAJAH; EKANAYAKE, 2016).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.7)$$

O Erro Médio Absoluto (*Mean Absolute Error - MAE*) é mais uma métrica comumente utilizada com o RMSE para avaliar problemas de previsão. O MAE tem o objetivo de estimar valores de produção de energia e até parâmetros que indicam recursos solares, como a irradiância solar. O cálculo do MAE é dado a partir da soma das diferenças absolutas no intervalo de tempo que houve a aquisição dos dados dividido pelos números de observações (Equação (4.8)), indicando os erros de previsão uniformemente (COIMBRA; KLEISSL; MARQUEZ, 2013; RAZA; NADARAJAH; EKANAYAKE, 2016; HOFF; PEREZ, 2010).

$$MAE = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n} \quad (4.8)$$

O próximo capítulo abordará os conceitos relacionados a previsão de geração de energia fotovoltaica. Além disso serão abordados alguns trabalhos relacionados que auxiliaram no desenvolvimento do presente trabalho.

5 PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA

Entre os anos de 2018 e 2020 a utilização dos sistemas de geração de energia solar cresceu aproximadamente 223% no Brasil e 46% entre todos os países do mundo (IRENA, 2020a; IRENA, 2020b). A energia gerada pela luz solar é extremamente sensível a variações climáticas e meteorológicas. Essa sensibilidade impõe alguns desafios para os produtores, empresas públicas, operadoras independentes de serviços e principalmente as concessionárias responsáveis pela distribuição de energia, que necessitam de segurança, disponibilidade e confiabilidade (PEDRO; COIMBRA, 2012; GIELEN et al., 2019).

Para melhorar o desempenho da resposta da rede de distribuição, a geração de energia fotovoltaica deve ser prevista com a maior precisão possível, provendo aos gerentes da rede, informações sobre a quantidade de energia que será disponibilizada em determinado momento (KUDO et al., 2009; LORENZ et al., 2009; MAHTTA; JOSHI; JINDAL, 2014). Dessa forma, a estimativa precisa da produção de um sistema PV fornece segurança para todo sistema energético, melhorando a estabilidade e permitindo um melhor planejamento da geração e distribuição de energia, gerando também um retorno econômico (DAS et al., 2018; BRANO; CIULLA; FALCO, 2014; EHSAN; SIMON; VENKATESWARAN, 2017; RAZA; NADARAJAH; EKANAYAKE, 2016).

A previsão de produção de energia fotovoltaica pode ser classificada de acordo com seu horizonte de previsão (DAS et al., 2018). A previsão pode ser considerada de longo prazo, buscando prever a geração de 1 a 10 anos à frente. A estimativa de médio prazo é dada por meio de uma previsão de 1 mês a 1 ano à frente. Existem também as previsões de curto prazo, que buscam prever a geração na próxima hora, dia ou semana. Por último, o horizonte de previsão de prazo muito curto que tem o objetivo de prever os dados minuto a minuto, ou seja, uma previsão quase em tempo real (RAZA; NADARAJAH; EKANAYAKE, 2016).

De acordo com Das et al. (2018) e Raza, Nadarajah e Ekanayake (2016) as ferramentas de previsão podem ser categorizadas em:

- **Métodos Estatísticos:** os métodos estatísticos são baseados em bases históricas de medições de potência e características meteorológicas do local onde está instalada planta fotovoltaica. Sendo assim, quanto melhor a qualidade desses dados e maior a base de dados históricos coletados sobre os funcionamento da planta, mais precisos serão os resultados de previsão obtidos (ANTONANZAS et al., 2016). De acordo com Ehsan, Simon e Venkateswaran (2017) os métodos estatísticos podem ser classificados em modelos de séries temporais e em métodos de aprendizado de máquina. Os modelos estatísticos baseados em métodos de aprendizado de máquina são, por exemplo, RNAs, SVM, LR, kNN e DT. Dentre os modelos de séries temporais são: Método Auto Regressivo de Média Móvel

(*Autoregressive Moving Average* - ARMA), Método Auto Regressivo Integrado de Médias Móveis (*Autoregressive Integrated Moving Average* - ARIMA) (RAZA; NADARAJAH; EKANAYAKE, 2016).

- **Métodos Físicos:** os métodos físicos são baseados na previsão das variáveis meteorológicas, que é desempenhada por modelos de previsão numérica do tempo (*Numeric Weather Predictor* - NWP). A precisão dos modelos é afetada por variações nos atributos usados para realizar previsão, e devem ser desenvolvidos exclusivamente para um determinado sistema PV e localização (DAS et al., 2018; De Giorgi; CONGEDO; MALVONI, 2014).
- **Métodos Híbridos:** o desempenho das outras técnicas de previsão pode ser melhorado com a hibridização dos modelos, que acontece quando mais de uma técnica é combinada para alcançar um determinado resultado. O desenvolvimento de um modelo híbrido tem a intenção de minimizar os erros individuais de cada modelo, fazendo com que os pontos fortes de cada técnica seja explorado, assim buscando melhorar a precisão da previsão gerada individualmente (ANTONANZAS et al., 2016; VOYANT et al., 2017; DAS et al., 2018; RAZA; NADARAJAH; EKANAYAKE, 2016).
- **Métodos Persistentes:** os modelos de previsão baseados em métodos persistentes são simples, e na maioria das vezes usados como referência para os outros modelos mais precisos e que possuem um custo computacional mais alto. Estes modelos fornecem a previsão de acordo com o valor que foi encontrado no dia ou hora anterior, o que determina que não é um modelo com boa precisão, pois é facilmente afetado por variações climáticas, como temperatura, surgimento de cobertura de nuvens, umidade e velocidade do vento (VOYANT et al., 2017; FERNANDEZ-JIMENEZ et al., 2012; ANTONANZAS et al., 2016).

Na próxima seção serão abordados alguns trabalhos relacionados à previsão de geração de energia fotovoltaica com métodos estatísticos. Serão apresentados estudos que foram utilizados como referência para o desenvolvimento do trabalho.

5.1 TRABALHOS RELACIONADOS

Diversos trabalhos possuem foco principal na previsão da potência de saída em sistemas fotovoltaicos, o que envolve diferentes variáveis, tipos de modelos, horizontes de previsão e as características solares e meteorológicas que influenciam na produção de energia solar. Raza, Nadarajah e Ekanayake (2016) e Das et al. (2018), desenvolveram um estudo a respeito de todo o processo, e das ferramentas que são necessárias para a construção de uma técnica de previsão de geração de energia que tenha uma alta precisão.

Os trabalhos que serão apresentados a seguir utilizaram diferentes parâmetros meteorológicos para obter a previsão de geração de energia de determinada planta PV. Foi realizada a avaliação de comportamento dos modelos de previsão de acordo com a variação de seus conjuntos de parâmetros meteorológicos de entrada. Assim, analisando quais os parâmetros que influenciam de forma positiva no desempenho dos modelos de previsão propostos.

Os trabalhos desenvolvidos por De Giorgi, Congedo e Malvoni (2014) e De Giorgi et al. (2013), realizaram a análise do comportamento de diferentes combinações de conjuntos de entradas envolvendo dados históricos de um ano. Entre os dados utilizados, estavam a temperatura ambiente e do módulo fotovoltaico, a irradiância solar medida em dois planos posicionados em ângulos diferentes e a potência de saída do sistema. Com os testes realizados foi observado que o modelo que utilizou como conjunto de entrada todos os parâmetros propostos, obteve os melhores resultados. Já Huang et al. (2010), fizeram a comparação entre métodos físicos e estatísticos utilizando como entrada em seus métodos, dados de umidade relativa do ar, temperatura ambiente, posição do sol, irradiância e coeficiente de cobertura do céu. O autor obteve os melhores resultados com o modelo de previsão composto por métodos estatísticos. Chow, Lee e Li (2012) afirmam em sua pesquisa que apenas os dados solares como ângulo azimute, ângulo de elevação solar, irradiância e a temperatura de bulbo seco, são capazes de proporcionar aos modelos de previsão uma estimativa precisa da potência de saída de um sistema PV.

Raza, Nadarajah e Ekanayake (2016) observaram que os modelos desenvolvidos para tipos de dias específicos como chuvosos, nublados e de céu limpo, possuem um desempenho acima dos modelos que não levam em consideração os tipos climáticos dos dias, realizando previsões mais precisas. Shi et al. (2012) desenvolveram um modelo de previsão fundamentado na classificação dos dias em quatro tipos de dias, ensolarados, nublado, chuvoso e nebuloso (presença de muita neblina). Foram utilizados modelos SVR pra realizar a previsão, nos quais o autor obteve os melhores resultados para os modelos utilizados em dias ensolarados e nebuloso. Outros autores como Leva et al. (2017) e Ding, Wang e Bi (2011), utilizaram além das classificações de tipos climáticos outros parâmetros de entrada, como temperatura ambiente e irradiância. Os autores observaram resultados melhores quando os modelos de previsão estavam atuando em dias de céu limpo, sem a presença de nuvens ou chuva.

Mellit, Massi Pavan e Lughì (2014) avaliaram modelos específicos para dias de céu limpo, parcialmente nublados e nublados. Além disso realizaram a comparação da aplicação dos modelos para um conjunto de dados no qual os tipos de dias estavam estratificados e para um conjunto sem a estratificação. Os autores concluíram por meio dos resultados obtidos, que o modelo que alcançou os melhores resultados para todos os tipos de dias, foi o modelo desenvolvido para atuar em dias nublados. No trabalho

de Kardakos et al. (2013) foram comparados diferentes modelos de previsão, entre eles um método persistente, uma RNA e um modelo de séries temporais, sendo cada um deles testado em todas as estações do ano. Os melhores resultados individuais, foram obtidos no verão pelo método persistente, no entanto o modelo com a menor média anual de erro foi o de séries temporais sem a inclusão de fatores exógenos (dados coletados em solo através de estações solarimétricas).

Além do clima, a localização geográfica de uma usina fotovoltaica é frequentemente considerada um fator-chave na produção observada do sistema. Dado que as correlações entre os parâmetros meteorológicos podem mudar de acordo com a localização, há recomendações para sua investigação em diferentes cenários de aplicação (DAS et al., 2018).

Vários trabalhos que possuem como objetivo prever a geração de energia fotovoltaica realizam a comparação de alguns métodos. Dentre eles, os que apresentam os melhores resultados quando comparados, são os modelos baseados em RNAs, principalmente a rede MLP (DAS et al., 2018; RAZA; NADARAJAH; EKANAYAKE, 2016). Graditi, Ferlito e Adinolfi (2016), Ehsan, Simon e Venkateswaran (2017) e Brano, Ciulla e Falco (2014) implementaram um modelo de previsão com uma MLP, mostrando que é uma técnica simples que alcança resultados aceitáveis com erros baixos, custo computacional baixo e podendo ser executado em qualquer planta PV.

Fernandez-Jimenez et al. (2012) desenvolveram modelos de previsão a curto prazo de um sistema conectado à rede de distribuição de energia elétrica. Foram validados e comparados oito métodos diferentes, dentre eles, ARIMA, MLP, persistente, Função de Base Radial (*Radial Bases Function* - RBF), Rede de Elman, Sistema de Inferência Neuro-Fuzzy Adaptável (ANFIS - *Adaptive Neuro Fuzzy Inference System* - ANFIS), Rede Neural com Retardo de Tempo (*Time Delay Neuro Network* - TDNN) e kNN, obtendo melhores resultados com a MLP.

Arize e Rios (2019) realizaram a comparação entre alguns modelos com objetivo de prever séries temporais de consumo de energia para controladores de redes inteligentes, usando Redes Neurais Recorrentes (*Recurrent Neural Network* - RNN), SVR, *Random Forest*, SARIMAX e ANFIS, obtendo por meio do modelo ANFIS os melhores resultados em comparação com os outros modelos. Já Hossain et al. (2017), desenvolveram um modelo de aprendizado de máquina extrema (*Extreme Learning Machine* - ELM) para realizar a estimação precisa da potência de saída de uma planta PV com prazos de uma 1 hora e um dia e, para validar a técnica proposta, usou um modelo SVR e uma RNA. A partir da avaliação dos modelos através de algumas métricas de desempenho como o R^2 e o RMSE, concluíram que o modelo ELM obteve os melhores resultados e maior eficiência computacional. Das et al. (2018) recomendaram a partir do seu estudo sobre a literatura, que as ferramentas baseadas em RNAs e SVR apresentam os melhores desempenhos para a previsão de potência de saída PV.

Além dos métodos de aprendizado de máquina utilizados para realizar a previsão de geração de energia fotovoltaica já abordados nesta seção, existem outros trabalhos que utilizam modelos de conjuntos de métodos para obter uma previsão de geração mais precisa. Rana, Koprinska e Agelidis (2016) desenvolveram um trabalho utilizando métodos de agrupamento para separar os tipos de dias e modelos de conjuntos de métodos para previsão de energia fotovoltaica. Propuseram um modelo de conjunto de RNAs de previsão específico para cada tipo de dia, comparando com um modelo composto por apenas uma RNA. Concluíram com essa comparação, que o modelo de comitê de métodos obteve uma precisão de 3.60% melhor do que o modelo de uma RNA.

Zhu, Guo e Gong (2019) propuseram um modelo de previsão de geração PV a curto prazo utilizando um modelo de comitê de métodos. O comitê de métodos combinou três algoritmos de aprendizado de máquina, uma MLP, *Extreme Gradient Boosting* e uma RNN. Em seus resultados, concluíram que os erros médios de previsão do modelo de conjuntos de métodos proposto foram menores do que os erros obtidos por métodos individuais. Um modelo de comitê de métodos *Stacking* proposto por Divina et al. (2018), tinha como objetivo prever o consumo de energia elétrica a curto prazo. Realizaram a comparação deste modelo com métodos individuais, como regressão linear e árvore de decisão. Utilizaram para validar os modelos diferentes tamanhos de janelas de treinamento e concluíram que janelas com uma quantidade menor de dados históricos, quando aplicadas nos dois tipos de modelo, possuem uma performance melhor para o modelo *Stacking*. Os autores discutiram que os modelos de métodos individuais necessitam de um maior número de dados para alcançar bons resultados, devido a sua degradação de desempenho.

Por meio dos trabalhos discutidos neste Capítulo foi possível observar diferentes estudos e abordagens a respeito da previsão de geração fotovoltaica. De forma geral, os modelos baseados em Redes Neurais Artificiais possuem um desempenho melhor que os outros modelos. A previsão de geração fotovoltaica baseada nos tipos climáticos de dias pode aperfeiçoar a performance dos modelos, gerando previsões mais assertivas. Outra abordagem demonstrada por alguns autores é relacionada ao uso de comitê de métodos, que pode melhorar o desempenho de modelos individuais e diminuir a quantidade de erros de previsão.

O próximo capítulo do trabalho abordará as estratégias de previsão propostas no trabalho. Assim como uma análise dos resultados obtidos a partir das estratégias desenvolvidas.

6 MODELOS DE PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA

Os modelos de previsão deste trabalho foram empregados por meio de duas estratégias diferentes. As duas abordagens possuem como objetivo realizar a previsão de geração de energia fotovoltaica de uma planta PV a curto prazo. Primeiramente foi realizada uma comparação entre cinco métodos de aprendizado de máquina distintos. Além dos métodos foram utilizados diferentes conjuntos de atributos como vetores de entrada, e também múltiplas configurações dos parâmetros para cada método. Os resultados obtidos com a primeira abordagem utilizada, foram publicados no XVI Encontro Nacional de Inteligência Artificial e Computacional no ano de 2020 (PELISSON et al., 2020).

Em seguida foi realizada a investigação da segunda estratégia considerada no trabalho, no qual além dos métodos de Aprendizado de Máquina do primeiro procedimento, foi utilizado também modelos de comitês de métodos. Essa estratégia de previsão foi baseada no tipo climático dos dias.

Este Capítulo está dividido em três seções. Na primeira seção será apresentada a análise exploratória e pré-processamento da base de dados utilizada na elaboração do trabalho. Em seguida, serão apresentadas as duas estratégias de previsão utilizadas e os resultados obtidos.

6.1 ANÁLISE EXPLORATÓRIA DOS DADOS

Para que as estratégias de previsão consideradas neste trabalho possam ser implementadas, algumas etapas prévias devem ser realizadas com o objetivo de fornecer uma base de dados cuja qualidade e estrutura sejam satisfatórias para a aplicação das estratégias. As etapas preliminares envolvem a análise exploratória dos dados e em seguida o seu pré-processamento.

A base de dados adotada para o desenvolvimento do trabalho foi obtida a partir de medidas realizadas em solo por uma estação solarimétrica localizada no Laboratório Fotovoltaica da Universidade Federal de Santa Catarina (UFSC), situado na cidade de Florianópolis (MANTELLI; MARTINS; RUTHER, 2020). A planta fotovoltaica utilizada como base para a investigação da previsão da potência de saída está localizada no mesmo local que a estação solarimétrica responsável pela coleta dos dados.

A obtenção dos dados de geração de energia e os dados da estação solarimétrica, foi efetuada no decorrer dos anos de 2018, 2019 e 2020. O total de observações colhidas para a construção da base de dados é de 1.452.203 com uma resolução de um minuto entre cada leitura. A coleta dos dados foi realizada por um *datalogger* (registrador de dados) CR6 da *Campbell* (CAMPBELL, 2021). Os atributos considerados para o desdobramento do estudo são na Tabela 6.1.

Foram utilizadas para a primeira abordagem, apenas os dados referentes ao

ano de 2018, com um total de 545.806 observações. Já na segunda parte do desenvolvimento do trabalho, utilizou-se os dados pertencente aos anos de 2018, 2019 e 2020. Essas duas bases de dados serão identificadas daqui a diante, respectivamente como Base de Dados 1 e Base de Dado 2.

Tabela 6.1 – Parâmetros coletados pela estação solarimétrica e potência produzida pela planta PV.

Parâmetro	Abreviação	Sensor	Unidade
Direct Normal Irradiance	DNI	SHP1	W/m ²
Global Horizontal Irradiance	GHI	SMP22	W/m ²
Diffuse Horizontal Irradiance	DHI	SMP11	W/m ²
Ultraviolet Index	UV	CUV 5	W/m ²
Ambient Temperature	AT	HMP155	°C
Wind Speed	WS	WindSonic1	m/s
Relative Humidity	RH	HMP155	%
Power Output	PO	ABB Uno 2.0 Inverter	W

6.1.1 Pré-processamento dos Dados

Antes de elaborar qualquer tipo de análise foi necessário definir os horários de atuação, ou seja, qual seria o espaço de tempo em que as predições seriam realizadas. Foram definidos dois intervalos, para a Base de Dados 1 foi implementado um filtro para selecionar as observações entre 07:30 às 17:00 baseado no trabalho de Ehsan, Simon e Venkateswaran (2017). Em seguida, para a Base de Dados 2, foram filtradas as observações em um intervalo de horário maior, situado entre às 6:30 e 19:30. Os intervalos de horário definidos possuem a maior parte dos dados de geração da planta PV e também de maior concentração de produção de energia fotovoltaica.

Após definido o horário seguiu-se para a limpeza das bases de dados. Primeiramente foram removidas as observações que possuíam *timestamps* (hora e data em que a observação foi feita) duplicados. Foi verificada a presença de falhas de gravação do *datalogger*, geradas por travamento e desligamento do equipamento, o que gerou intervalos entre algumas observações. Investigou-se também a presença de valores nulos, valores consecutivos iguais (constantes) e os horários em que a potência de saída do sistema era igual a zero. Sendo assim, essas medidas foram removidas da Base de Dados 1.

A planta PV e a estação solarimétrica estão presentes em um laboratório que ocasionalmente precisa de manutenção nos inversores, nos módulos PV e também na estação de aquisição de dados. Além dessas manutenções, eventualmente, é efetuado no local treinamento de pessoal, causando o desligamento dos inversores. O desligamento dos inversores faz com que não haja potência gerada, assim não existindo

também medidas de potência nesses momentos, o que ocasionou em observações com valores iguais a zero na base de dados (valores que foram removidos).

O processo de filtragem realizado para a Base de Dados 1 também foi repetido para a Base de Dados 2. No entanto, valores consecutivos iguais, valores ausentes e dados de potência igual a zero, foram substituídos por valores não numéricos, chamados de *not a number (NaN)*. Por meio da análise de *outliers*, foram encontradas medidas de temperatura ambiente no ano de 2019, abaixo do limite inferior da análise. Esses valores foram investigados, pois apresentavam medidas de temperatura baixas para a localização da estação solarimétrica. Foi constatado por meio de um banco de dados histórico INMET, Instituto Nacional de Meteorologia (2021), que a temperatura mais baixa registrada em Florianópolis no ano de 2020, foi de 8,4° graus *Celsius*. Portanto, os valores de temperatura ambiente abaixo de 8,4°, foram substituídos por valores *NaN*.

Os valores da Base de Dados 2 substituídos por valores *NaN* foram tratados como valores ausentes. Com isso, através de um filtro foram identificados os dias que possuíam menos de 30% de valores ausentes em seu total de observações. Os dias com mais de 30% das observações ausentes foram removidos da base de dados. Portanto, foram utilizados para a criação da versão final da Base de Dados 2, apenas os dias com menos de 30% do seu número total de observações como valores ausentes. Sendo assim, a Base de Dados 2 passou por um processo de imputação de valores ausentes. Esse processo foi realizado por meio do algoritmo kNN - *Imputer*, que foi abordado anteriormente na Seção 4.1.2.1.

Com a intenção de diminuir a quantidade de observações, entregando bases de dados mais consistente e robusta para o treinamento dos modelos propostos, agregou-se os registros de data e hora em uma resolução de 30 minutos. Portanto, para a Base de Dados 1, foram geradas 20 leituras diárias intercaladas entre 30 minutos {07:30, 08:00, ..., 16:30, 17:00}. Para a Base de Dados 2, foram geradas 27 observações diárias intercaladas entre 30 minutos {06:30, 07:00, ..., 19:00, 19:30}. Para tanto, obteve-se a média das leituras dos 15 minutos que antecedem e sucedem os horários das observações geradas.

Nas Figuras 6.1, 6.2 e 6.3 são apresentados exemplos dos perfis médios diários dos meses de Janeiro e Junho, respectivamente dos anos de 2018, 2019 e 2020. Os meses de Janeiro e Junho foram escolhidos pelo motivo de estarem em diferentes estações do ano, podendo demonstrar diferentes variações dos valores de irradiância e conseqüentemente de potência gerada pelo sistema PV. O mês de Janeiro é considerado um mês de verão, no qual os índices de irradiância são maiores que em meses de inverno como Junho. É possível observar a diferença entre os índices de irradiância nestes meses quando olhamos para o ponto na curva as 12:00 horas nas Figuras 6.1b e 6.1c para o ano de 2018, Figuras 6.2b e 6.2c para o ano de 2019

e Figuras 6.3b e 6.3c para o ano de 2020. Os perfis médios diários de 2018 para a Base de Dados 2, possuem o mesmo comportamento que os da Base de Dados 1, se diferenciando apenas pela faixa de horário entre as 6:30 e as 19:30. A Base de Dados 2 compreende um total de 630 dias após o pré-processamento dos dados, totalizando 17.010 observações.

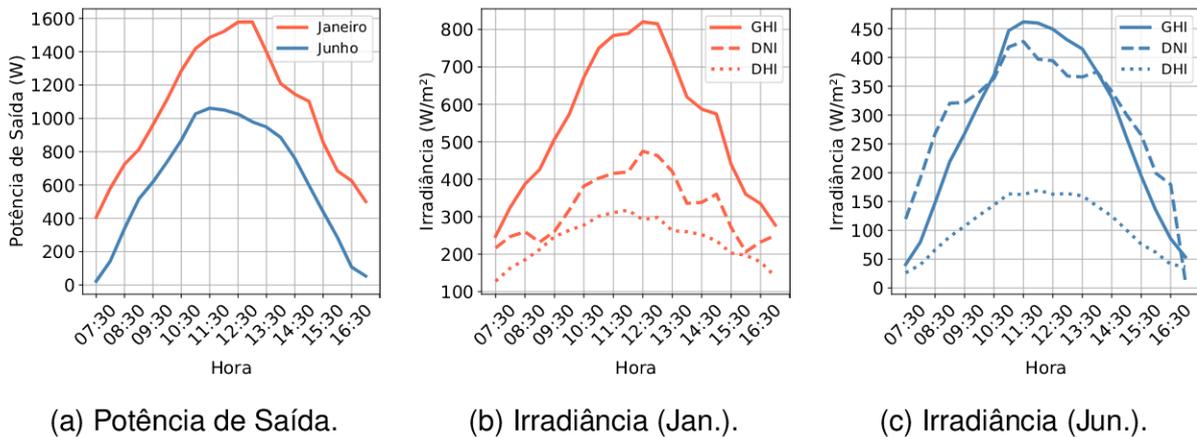


Figura 6.1 – Perfil médio diário de Janeiro e Junho de 2018.

Fonte: Autoria própria

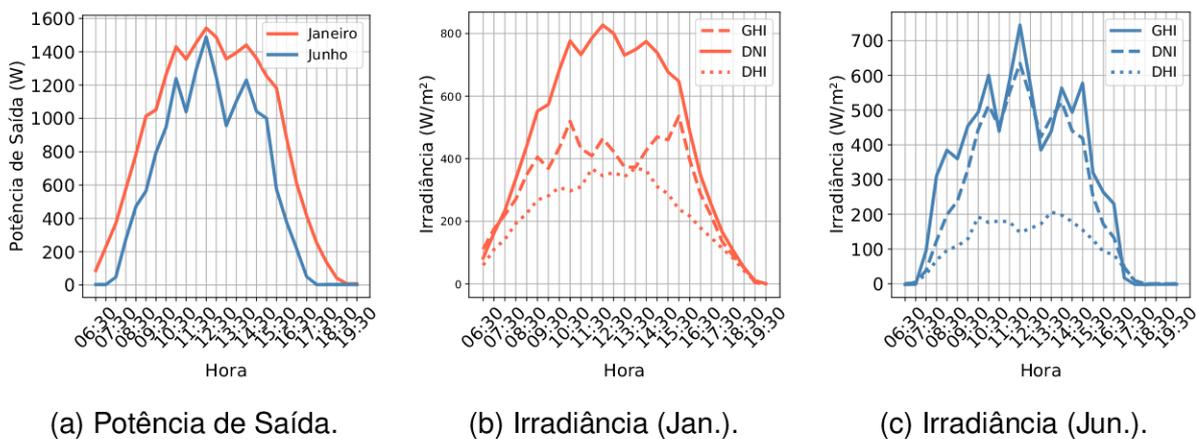


Figura 6.2 – Perfil médio diário de Janeiro e Junho de 2019.

Fonte: Autoria própria

O próximo passo do estudo foi marcado pela definição de como seriam feitos os treinamentos e a validação das duas estratégias de previsão comparadas. Para o treinamento e teste dos modelos foram utilizados os métodos de validação de avanço não ancorado e ancorado, que serão detalhados nas próximas seções. Nas seções seguintes serão abordadas as estratégias de previsão propostas neste trabalho.

6.2 ESTRATÉGIA DE PREVISÃO BASEADA EM MODELOS INDIVIDUAIS

Esta seção foi dividida em duas subseções, na qual primeiramente será apresentado os materiais e métodos utilizados no desenvolvimento da estratégia de previsão

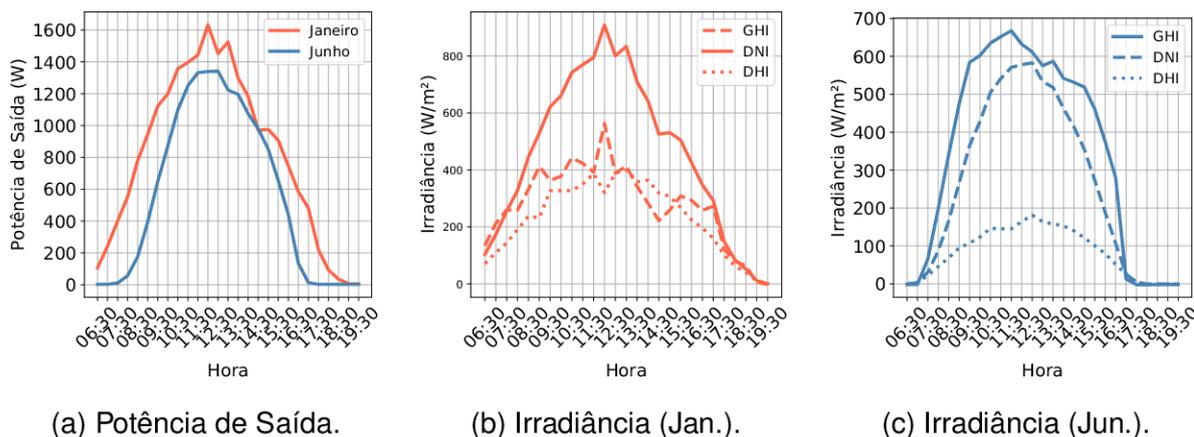


Figura 6.3 – Perfil médio diário de Janeiro e Junho de 2020.

Fonte: Autoria própria

de potência de saída baseada em modelos individuais. A segunda subseção apresentará os resultados obtidos por meio dessa estratégia.

6.2.1 Materiais e Métodos

Para o desenvolvimento da estratégia de previsão baseada em diferentes subconjuntos de entrada, utilizou-se os dados da Base de Dados 1. Os métodos de aprendizado de máquina utilizados para o desenvolvimento desta estratégia, foram a MLP, SVR, kNN, Árvore de Decisão e Regressão Linear. Os métodos foram treinados e validados por meio de diferentes configurações dos seus parâmetros e diferentes subconjuntos de atributos de entrada, conforme discutido a seguir.

Para a MLP foi definido a utilização de apenas uma camada escondida, na qual consideramos $\{3, 5, 7, 9, 11\}$ neurônios na camada oculta e duas funções de ativação, Unidade Linear Retificada (*Rectified Linear Unit - ReLU*) e a Tangente Hiperbólica (*Hyperbolic Tangent - tanh*). Foram obtidas 10 combinações diferentes de configurações dos modelos de MLP para avaliação e validação. Os modelos foram treinados com o otimizador Adam (KINGMA; BA, 2014) com os parâmetros padrões do Keras (CHOLLET et al., 2015).

Para o método SVR foram consideradas variações nos parâmetros de configuração C e $kernel$, adotamos valores de C iguais a $\{1, 10, 50, 100, 1000\}$ e $kernel$ s, função de base radial (*Radial Basis Function - RBF*), linear e polinomial, usando um valor fixo de ϵ (0.01) e obtendo então 15 configurações diferentes do modelo para análise.

Para o kNN, foram considerados valores de k (número de vizinhos): $\{3, 5, 7, 9, 11\}$, distância Euclidiana e atributos de entrada normalizados com média zero e desvio padrão igual a 1 (normalização z-score). Foram obtidas cinco configurações diferentes do método kNN e consideramos uma média aritmética não ponderada do valor alvo dos vizinhos.

Na utilização da Árvore de Decisão (DT), foi adotado o erro quadrático médio

como critério de divisão. Consideramos para a configuração do modelo, o número mínimo de amostras por nó folha de 5 e 10, controlando indiretamente a profundidade da árvore final.

O algoritmo de regressão linear (LR) foi utilizado na sua forma mais simples, sem regularização e expansão da base. Como o conjunto de dados utilizado é composto por um pequeno número de atributos, a maioria dos problemas comuns encontrados na regressão linear são evitados

O treinamento e teste foi realizado por meio do método de avanço não ancorado de uma janela deslizante (Figura 6.4), no qual foi estabelecida uma janela fixa de 30 dias de observações. Como cada dia possui 20 observações, tem-se um total 600 observações por janela. As previsões de potência de saída foram realizadas para as 20 observações do dia seguinte. A janela de treinamento de tamanho constante anda um dia a cada processo de treinamento, esse processo se repete até o final da Base de Dados 1. Essa abordagem tem a intenção de realizar o treinamento de um novo modelo para cada janela, eliminando a sazonalidade presente na base de dados e evitando a utilização de dados passados e comportamentos diferentes a cada mês e estação do ano.

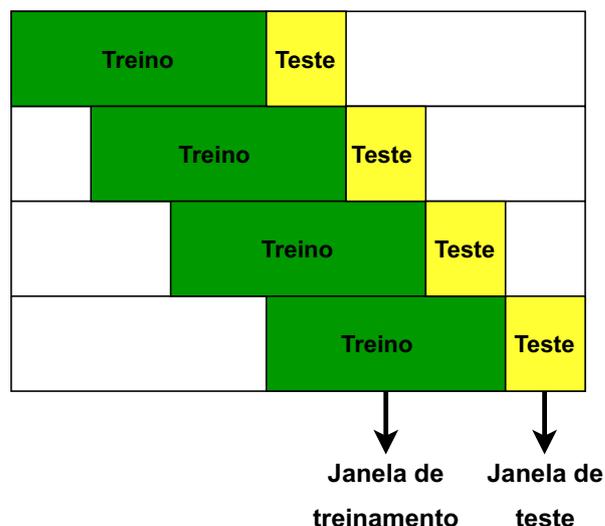


Figura 6.4 – Método de validação por avanço não ancorado.

Fonte: Autoria própria

Foram estabelecidos cinco conjuntos de atributos de entrada em diferentes arranjos. A correlação de *Pearson R* nos mostra qual a intensidade da relação entre os parâmetros de entrada e a potência de saída. De acordo com a Tabela 6.2, podemos observar que O GHI, UV e DNI são as características mais bem correlacionadas com a potência de saída, em seguida com uma correlação moderada temos AT, RH e WS, e o DHI com a correlação mais fraca entre todos os parâmetros.

O primeiro subconjunto de atributos (SA 1) é composto por todos os parâme-

tros, o segundo (SA 2) é composto pelas variáveis que compõe a irradiação solar, o terceiro (SA 3) é formado apenas pelos parâmetros climáticos e o quarto (SA 4) é uma mistura entre as variáveis climáticas e as componentes de irradiância. O último subconjunto (SA 5) é composto apenas pelo GHI devido a sua alta correlação com a potência de saída, e foi utilizado apenas para o treinamento do modelo de regressão linear.

Tabela 6.2 – Subconjuntos de entrada

Subconjunto de Atributos	GHI	DNI	DHI	UV	RH	AT	WS
SA 1	✓	✓	✓	✓	✓	✓	✓
SA 2	✓	✓	✓	✓	✓		
SA 3				✓	✓	✓	✓
SA 4	✓	✓	✓				
SA 5	✓						
R	0.98	0.77	0.23	0.96	0.46	-0.55	0.43

6.2.2 Resultados

Os resultados foram obtidos a partir de um total de 33 configurações (10 MLPs, 15 SVRs, 5 kNNs, 2 Árvores de Decisão, e 1 Regressão Linear) entre todos os métodos considerados. Cada método foi treinado e testado com os cinco subconjuntos de atributos de entrada e assim obtivemos os melhores resultados considerando as melhores configurações dos métodos expostos na Tabela 6.3. Os métodos foram avaliados de acordo com as métricas de erro RMSE e R^2 .

Em todos os diferentes casos de subconjuntos de atributos, a MLP se comportou melhor que os demais modelos, com uma configuração de 9 neurônios na camada escondida e função de ativação *tanh*. O modelo SVR testado com diferentes números de parâmetros de regularização obteve o melhor resultado em todos os casos com C igual a 1000, porém não houve um tipo específico de *kernel* que se destacou. O kNN teve os melhores resultados com o número de vizinhos igual a 7, 5, 9, 5 e 11 para os subconjuntos de entrada respectivamente de 1 a 5. Já a DT (Árvore de Decisão) demonstrou para os SA 3 e 5, que o melhor número de amostras mínimas por nó foi 10 e para os outros 5. A Regressão Linear resultou em um desempenho melhor que os modelos DT e kNN em todos os vetores de entrada, sendo superado apenas pela MLP e o SVR.

Foi realizada a análise de alguns dias individualmente, para que conseguíssemos observar o comportamento de cada modelo em dias específicos. Para uma melhor investigação foram selecionados para comparação dias com erros mais significativos a partir dos maiores valores de RMSE observados e também dias em que os modelos tiveram um bom desempenho. Os maiores erros de previsão, podem ser causados devido a existência de cobertura de nuvens, chuvas ou sombreamento dos sensores

Tabela 6.3 – Melhores resultados obtidos com cada modelo e conjunto de entrada.

Modelo	SA 1		SA 2		SA 3		SA 4		SA 5	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
MLP	45.92	0.99	49.43	0.99	107.68	0.96	53.19	0.99	66.17	0.99
SVR	52.20	0.99	52.63	0.99	112.69	0.96	56.78	0.99	88.61	0.98
kNN	94.95	0.97	76.65	0.98	155.42	0.92	71.68	0.98	87.33	0.98
DT	65.47	0.99	64.81	0.99	122.47	0.95	63.31	0.99	89.93	0.98
LR	59.03	0.99	60.42	0.99	112.35	0.96	63.33	0.99	87.82	0.98

de irradiância. Logo foram verificados dias dos meses de Abril, Julho, Agosto e Dezembro, no qual consideramos o melhor subconjunto de atributos definido a partir dos resultados na Tabela 6.3 como sendo o SA 1.

É possível observar o comportamento das curvas de cada modelo em cada um dos dias selecionados na Figura 6.5, em que são apresentados os valores preditos e os esperados de potência de saída do sistema PV. Conseguimos constatar que por meio das figuras abaixo apresentadas que na parte da manhã e da noite, os valores de potência gerada são mais baixos. Em um primeiro momento verificamos que esses valores são maiores no mês de Dezembro que pertence a estação de verão do ano, em que o sol nasce mais cedo e se põe mais tarde, comparado com os valores de potência do mês de Abril, no qual o sol nasce mais tarde devido ser outono.

Os melhores resultados foram obtidos nos meses de Abril e Dezembro, nos quais as previsões foram mais próximas dos valores esperados. Julho e Agosto foram os meses com os piores resultados, nos quais os modelos tiveram uma maior dificuldade em prever a geração de energia ao meio dia. É possível observar que o modelo de Regressão Linear no mês de Julho realizou a previsão de valores corretos ao meio dia, porém no resto do dia obteve resultados piores. As previsões ruins podem ser justificadas pela ocorrência de nuvens e chuvas, que afetam a potência de saída e também as medições de irradiância. O kNN foi o método que apresentou os piores erros de previsão para o dia 11 de Julho. Esse padrão é frequente para o kNN, o pior método, em geral.

A Figura 6.5 apresenta os resultados de previsão obtidos para os modelos no dia 11 de Julho e a Figura 6.6 apresenta os valores de cada atributo utilizado no treinamento dos modelos, para que possamos ter um melhor entendimento a respeito dos erros obtidos neste dia. É possível notar que próximo às 11:30, os valores de DNI foram zerados devido a um possível problema de medição do sensor, que só é restabelecido após as 15:30. Assim, todos os modelos que tiveram suas previsões influenciadas pelo DNI tiveram pior desempenho nesse período, sendo o kNN e o SVR os mais afetados. Esse exemplo mostra a importância de métodos robustos para previsão de produção de energia em sistemas fotovoltaicos.

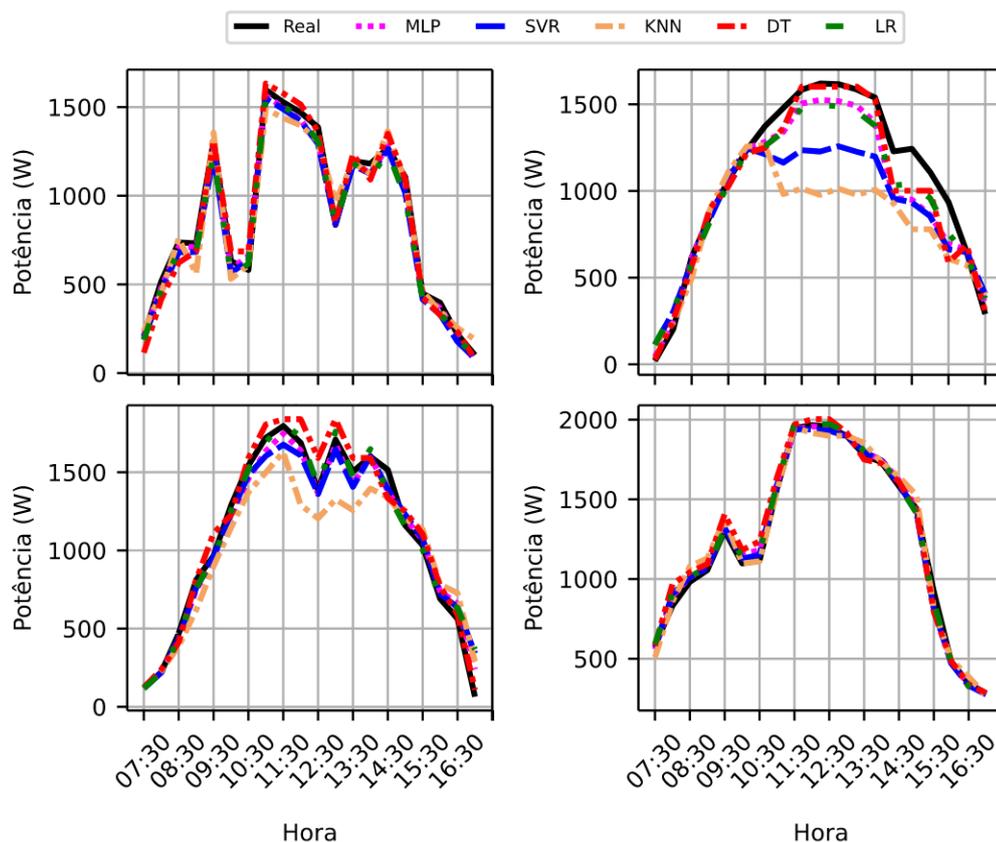


Figura 6.5 – Valores esperados e previstos da potência de saída para diferentes dias. No sentido horário, no canto superior esquerdo: 27/04/18, 11/07/2018, 12/12/2018 e 28/08/18.

Fonte: Autoria própria

6.3 ESTRATÉGIA DE PREVISÃO BASEADA EM COMITÊS DE MÉTODOS

Esta seção foi dividida em duas subseções. Primeiramente serão apresentados materiais e métodos utilizados para o desenvolvimento da estratégia de previsão de potência de saída baseada comitês de métodos. A segunda subseção apresentará os resultados obtidos por meio dessa estratégia.

6.3.1 Materiais e Métodos

A estratégia de previsão baseada em modelos de comitês de métodos, envolve a utilização, em separado, de modelos individuais e de comitês. Para o desenvolvimento desta estratégia foram utilizados os dados referentes a Base de Dados 2. O conjunto de atributos utilizado como entrada para os modelos, é o subconjunto de atributos 1 utilizado na Estratégia de Previsão Baseada em Modelos Individuais. Esse subconjunto foi escolhido, pois na estratégia apresentada na Seção 6.2 os modelos individuais obtiveram seus melhores resultados a partir do uso deste subconjunto (SA 1).

Os tipos climáticos dos dias foram categorizados em dias ensolarados (*sunny*), dias parcialmente nublados (*cloudy*) e dias nublados (*overcast*). Os dias foram clas-

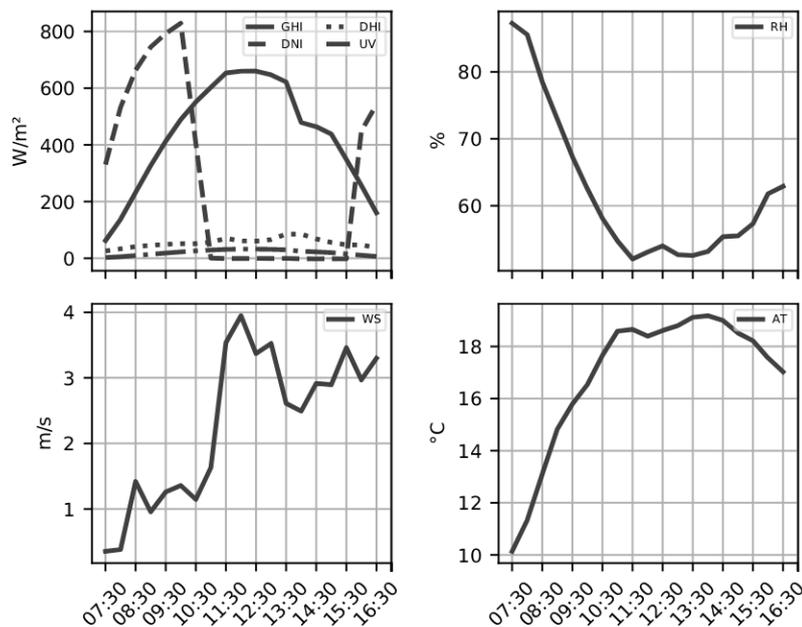


Figura 6.6 – Valores dos atributos do subconjunto SA 1 para todos os registros de data e hora de 11/07/2018.

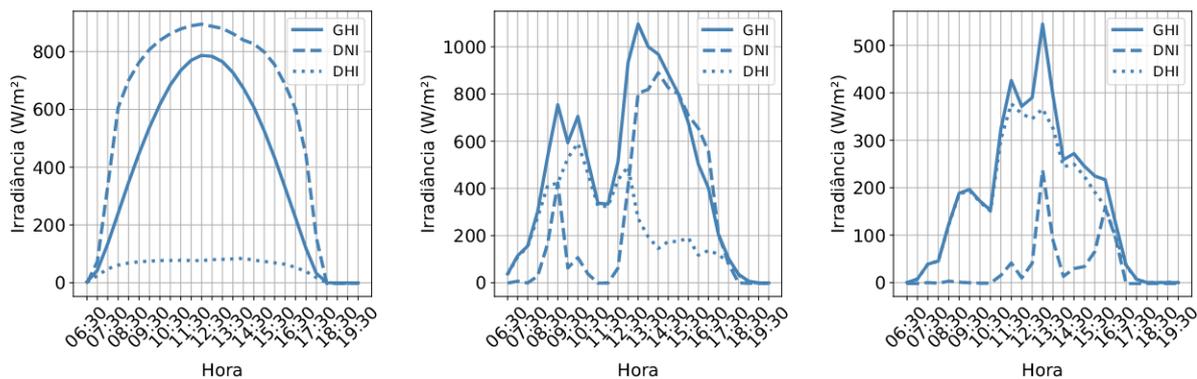
Fonte: Autoria própria

sificados com base no trabalho de Mellit, Massi Pavan e Lughi (2014). Os autores propuseram a classificação do tipo climático dos dias de acordo a média da irradiância diária. Os dias de sol, possuem um valor de GHI médio maior que $500W/m^2$. Os dias considerados como parcialmente nublados possuem um GHI médio entre 200 e $500W/m^2$. Para os dias totalmente nublados a faixa de GHI médio é menor que $200W/m^2$.

Além da categorização dos tipos climáticos de dias foram realizados testes com métodos de agrupamentos com o intuito de extrair mais informações da base de dados. Foram utilizados métodos de aprendizado de máquina não supervisionados, como *k- Means* e *Fuzzy c-Means* (CHEN et al., 2011; BEZDEK, 2013). Os métodos foram respectivamente avaliados por meio das métricas de desempenho *Silhouette* e *Fuzzy Partition Coefficient* (FPC). O resultado esperado quando utilizado os algoritmos de agrupamento, era de um valor de *Silhouette* e FPC maior que 0.5 indicando a existência de no mínimo mais de dois grupos distintos no conjunto de dados, um número maior do que o definido anteriormente por meio dos valores de GHI médio. A partir da avaliação dos métodos por meio das métricas de desempenho citadas, não foram encontrados resultados significativos. O objetivo dessa abordagem era encontrar mais grupos de tipos de dias na base de dados além dos citados a cima, porém os resultados obtidos não foram como os esperados, sendo assim, a abordagem de agrupamento foi interrompida.

A Figura 6.7 apresenta exemplos dos perfis diários de irradiância para os tipos

de dias abordados no trabalho. É possível observarmos na Figura 6.7a o comportamento das curvas de irradiância nos dias ensolarados, em que elas se mantêm estáveis ao longo do dia. No entanto, para os dias parcialmente nublados (Figura 6.7b) é observado uma variação nas curvas de irradiância ao longo do dia, e o aumento dos valores de *DHI* entre às 6:30 e 12:30 indicando a presença de nuvens. Já na Figura 6.7c é apresentado um exemplo de um dia nublado, em que é facilmente identificado valores de irradiância significativamente mais baixos que nos outros tipos de dias com um pico de $500W/m^2$ às 12:00 horas.



(a) 24/04/2018 - Ensolarado. (b) 01/11/2020 - P. Nublado. (c) 26/07/2019 - Nublado.

Figura 6.7 – Exemplo de perfis diários de irradiância dos tipos de dias considerados no trabalho.

Fonte: Autoria própria

O processo de treinamento e teste dos métodos foi realizado por meio do método de avanço não ancorado (Figura 6.4) e também pelo método de avanço ancorado (Figura 6.8). Para o método de avanço não ancorado, foi estabelecida uma janela fixa de 30 dias de observações, em que cada dia era composto 27 observações, assim totalizando um total 810 observações por janela de treinamento. Já o método de avanço ancorado, não possui uma janela de treinamento de tamanho fixo. Portanto, a cada processo de treinamento e teste, a janela de teste anterior é adicionada ao conjunto de dados que compõem a janela de treinamento. Como o objetivo do estudo é realizar a estimação da produção de energia no próximo dia, a validação foi feita através de 27 observações. Essa abordagem tem a intenção de comparar o desempenho dos modelos por meio dos métodos de validação, com objetivo de identificar se o número de observações por janela e também a sazonalidade presente na base de dados entregues ao modelo no treinamento afetará o seu desempenho quando utilizado o método de validação por avanço ancorado.

Os modelos individuais utilizados foram: MLP, kNN, Árvore de Decisão, SVR, Regressão Linear. Como base foram utilizados os mesmos modelos individuais utilizados no experimento apresentado na Seção 6.2. Para que os experimentos não fossem repetidos, foram utilizados os modelos individuais com as melhores configura-

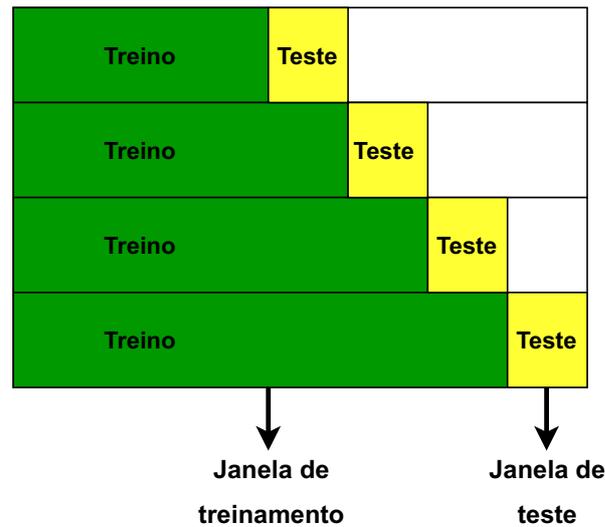


Figura 6.8 – Método de validação por avanço ancorado.

Fonte: Autoria própria

ções encontradas no experimento anterior. Além disso, foram utilizados como métodos individuais: *GradientBoost* (GB), *Adaboost* (AB) e *Random Forest* (RF), que aplicam os conceitos de comitês de métodos abordados anteriormente.

Basicamente foram utilizados dois modelos *Stacking Ensemble* nesta estratégia de previsão, um para cada tipo de dia e um ponderado. A primeira abordagem de comitês utilizou o modelo *Stacking Ensemble* de Intervalos Rígidos (SEIR). O modelo utilizado na segunda abordagem é chamado de *Stacking Ensemble* de Intervalos Fuzzy Ponderados (SEIFP).

O modelo SEIR é composto por três modelos *Stacking Ensemble*, cada um deles responsável por realizar respectivamente a previsão dos tipos de dias ensolarados, parcialmente nublados e nublados. Os modelos *Stacking Ensemble* serão constituídos pelos três modelos individuais que obtiverem os melhores resultados de previsão de acordo com as métricas de desempenho RMSE e R^2 , estes resultados serão expostos na próxima seção. O algoritmo de aprendizado de máquina LASSO foi utilizado como meta-modelo (RANA; KOPRINSKA; AGELIDIS, 2016). O processo de treinamento e teste dos modelos *Stacking Ensemble* e dos modelos individuais é apresentado por um diagrama na Figura 6.9.

O diagrama apresentado na Figura 6.9 tem como objetivo exemplificar como foi realizado o treinamento e teste dos modelos. Na Etapa 1 é realizado o cálculo do GHI médio (GHI_{avg}) diário dos dias que constituem a base dados, e então é feita a construção dos subconjuntos referentes a cada tipo climático abordado de acordo com as faixas de GHI. A Etapa 2 é responsável pelo treinamento e teste dos modelos por meio dos métodos de validação de avanço ancorado e não ancorado, utilizando os respectivos subconjuntos de dados, gerando as predições finais.

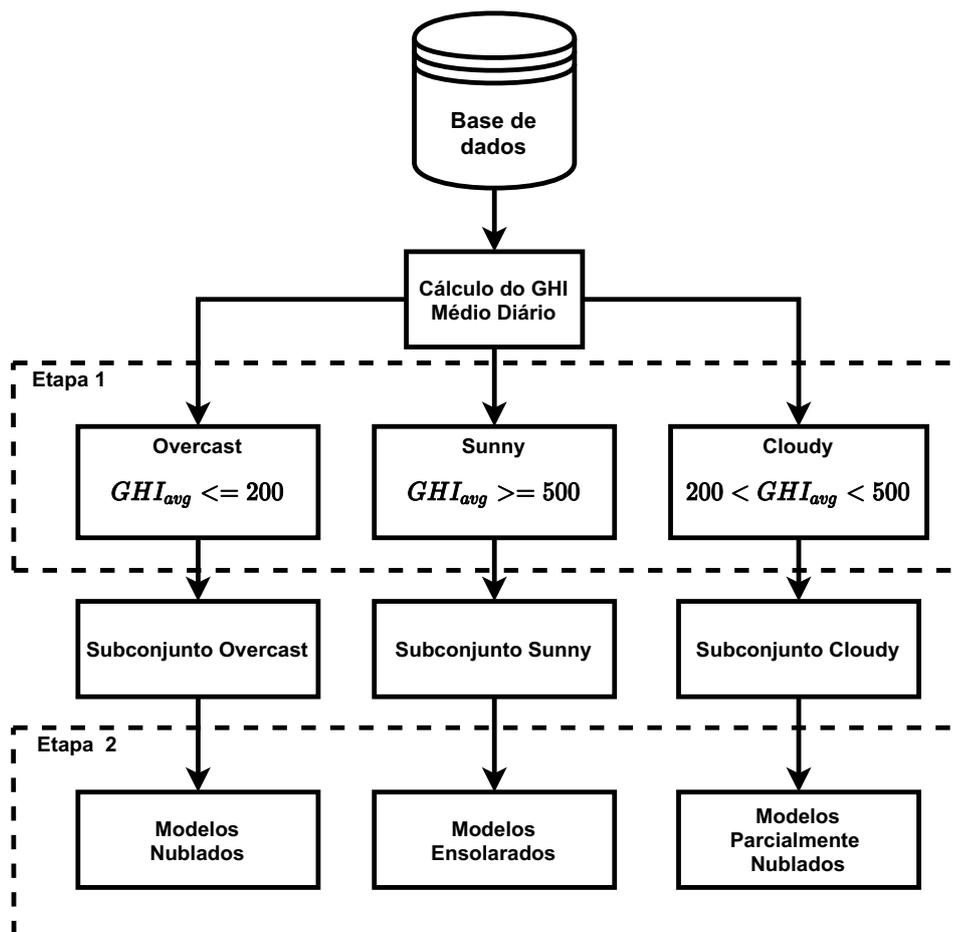


Figura 6.9 – Processo de treinamento dos modelos individuais e *Stacking Ensemble*.
Fonte: Autoria própria

O segundo modelo de comitês utilizado no desenvolvimento do trabalho, foi o *Stacking Ensemble* de Intervalos Fuzzy Ponderados. O modelo utilizado é composto pelos mesmos métodos individuais do modelo *Stacking Ensemble*. O modelo SEIFP, é um modelo de previsão que possui a sua saída ponderada de acordo com o tipo de dia a ser predito. O processo de treinamento e teste do modelo SEIFP é apresentado pelo diagrama na Figura 6.10.

A Etapa 1 é responsável pelo cálculo do GHI médio (GHI_{avg}). Após o cálculo do GHI_{avg} é aplicada a função de pertinência *fuzzy*, que realiza o cálculo do grau de pertinência, assim gerando os pesos (w_1 , w_2 e w_3) para cada um dos modelos. Utilizou-se a função de pertinência na forma trapezoidal baseada em três conjuntos *fuzzy* (Figura 6.11).

Os trapézios representam um conjunto *fuzzy* para cada tipo de dia. Os parâmetros utilizados para a construção de cada um dos trapézios, são apresentados na Tabela 6.4. Conforme abordado anteriormente, os trapézios possuem quatro parâmetros, a base 1 (k_a), ombro 1 (k_b), ombro 2 (k_c) e a base 2 (k_d).

A definição dos parâmetros foi baseada nos intervalos de GHI_{avg} estipulados anteriormente para a categorização dos tipos de dias para o modelo *Stacking Ensem-*

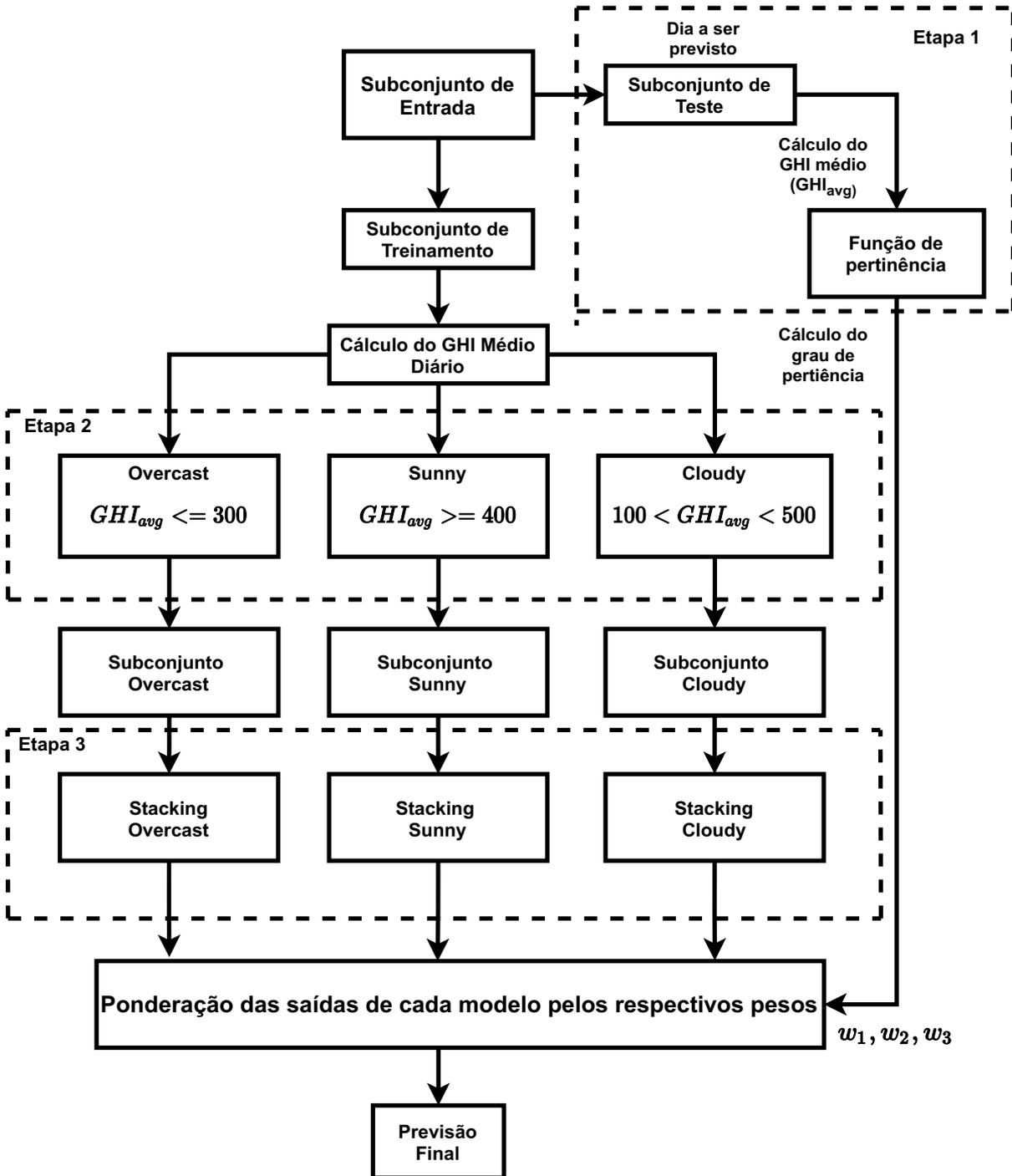


Figura 6.10 – Processo de treinamento do modelo *Stacking Ensemble* de Intervalos Fuzzy Ponderados.

Fonte: Autoria própria

ble de Intervalos Rígidos, porém foram adaptados para que fosse possível obter uma intersecção entre os intervalos considerados anteriormente. A adaptação realizada é dada pelas intersecções entre os conjuntos *fuzzy* criados. O objetivo dessa abordagem é lidar com as variações do GHI_{avg} , com o intuito de ponderar dias que podem não ser inteiramente pertencentes a um determinado tipo climático. Por exemplo, um dia que possui um GHI_{avg} igual a $120W/m^2$ pertence a intersecção entre os trapézios Nublado

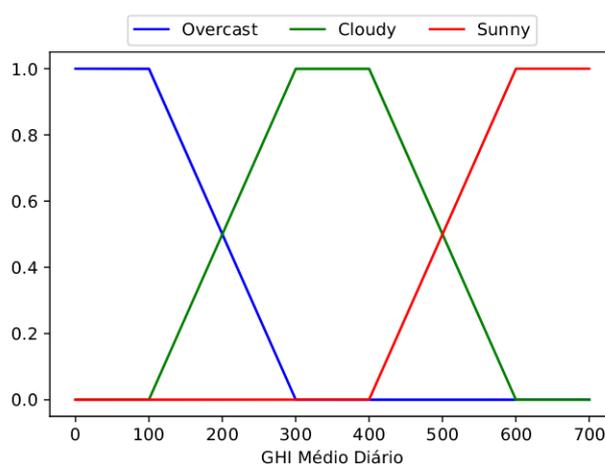


Figura 6.11 – Função de pertinência *fuzzy* na forma trapezoidal.
 Fonte: Autoria própria

Tabela 6.4 – Tabela para construção da função de pertinência fuzzy trapezoidal.

Tipo de dia	Forma	Parâmetros
Nublado	Trapezoidal	0,0,100,300
Parcialmente Nublado	Trapezoidal	100,300,400,600
Ensolarado	Trapezoidal	400,600,700,700

e Parcialmente Nublado, indicando que este dia pode não ser totalmente nublado e que existe a possibilidade de ser parcialmente nublado.

A Etapa 2 do processo de treinamento, conforme mostrado na Figura 6.10, consiste na construção dos subconjuntos de treinamento individuais para cada modelo. Esses subconjuntos individuais (*sunny*, *cloudy* e *overcast*), são construídos a partir da janela de treinamento fornecida pelos métodos de validação de avanço ancorado e não ancorado.

A Etapa 3 consiste no treinamento e teste de cada um dos modelos a partir dos subconjuntos individuais construídos. A previsão final X do modelo *Weighted Stacking Ensemble* é dada pela Equação (6.1), em que y_n representa a saída dos modelos treinados e w_n os pesos provenientes da função de pertinência *fuzzy*. Cada modelo vai gerar um previsão y , o modelo *Stacking Overcast* gerará uma previsão y_1 , o *Stacking Sunny* uma previsão y_2 e finalmente o modelo *Stacking Cloudy* gerando uma previsão y_3 . Portanto, cada saída é multiplicada pelos seus respectivos pesos e seu somatório produzirá a previsão final.

$$X = y_1w_1 + y_2w_2 + y_3w_3 \tag{6.1}$$

A próxima seção do trabalho apresentará os resultados obtidos por meio da estratégia de previsão baseada no tipo climático diário.

6.3.2 Resultados

Os resultados foram obtidos por meio de 8 métodos individuais e 2 modelos de comitês. Os modelos foram treinados separadamente para cada tipo climático diário e para a base de dados sem a discretização dos tipos de dias. Os resultados obtidos utilizando o método de validação de avanço não ancorado são apresentados na Tabela 6.5, já os resultados obtidos por meio do método de validação de avanço ancorado na Tabela 6.6. Os modelos foram avaliados por meio das métricas de desempenho RMSE, R^2 e desvio-padrão (σ).

Tabela 6.5 – Resultados obtidos pelos modelos utilizando método de validação não ancorado.

Modelo	Ensolarado		P. Nublado		Nublado		Final		N. Estratificado	
	σ	RMSE	σ	RMSE	σ	RMSE	σ	RMSE	σ	RMSE
MLP	35.67	54.39	46.70	62.46	27.92	31.63	41.60	55.60	41.52	53.67
DT	40.07	60.00	53.35	68.83	26.91	30.67	46.51	60.51	46.01	58.13
KNN	40.91	60.01	51.76	66.11	26.16	29.51	45.44	58.48	47.37	59.14
RF	72.75	135.55	80.84	130.27	48.29	69.48	73.66	119.92	76.18	126.77
SVR	34.18	50.16	45.89	59.23	28.69	31.14	40.98	52.72	41.68	52.16
LR	36.08	58.89	50.23	74.02	22.51	29.46	43.03	63.58	45.11	62.86
GB	33.93	51.85	43.80	58.12	23.11	26.06	38.53	51.30	38.76	50.00
AB	34.37	64.20	43.53	79.15	26.25	38.50	38.99	69.46	40.17	69.04
SEIR	33.27	50.56	42.92	56.58	23.78	27.05	37.96	50.26	38.33	48.70
SEIFP	-	-	-	-	-	-	36.98	46.97	-	-

Tabela 6.6 – Resultados obtidos pelos modelos utilizando método de validação ancorado.

Modelo	Ensolarado		P. Nublado		Nublado		Final		N. Estratificado	
	σ	RMSE	σ	RMSE	σ	RMSE	σ	RMSE	σ	RMSE
MLP	36.55	55.06	56.86	73.45	22.82	26.03	47.42	62.00	63.98	82.82
DT	42.77	63.98	62.88	79.83	26.91	30.67	55.27	68.44	72.56	89.87
KNN	38.99	58.52	60.29	75.45	23.51	26.93	50.35	64.02	65.64	82.31
RF	79.75	144.62	91.67	147.12	46.18	69.61	81.09	135.53	94.68	154.27
SVR	33.99	50.84	59.74	74.25	23.29	26.14	49.20	61.92	63.86	79.67
LR	36.28	60.95	66.62	91.86	23.56	30.85	54.17	76.04	70.00	99.32
GB	35.48	55.39	55.11	70.53	24.13	27.24	46.49	60.32	59.83	76.90
AB	37.63	70.36	50.09	100.90	28.78	47.07	44.31	86.45	53.82	100.69
SEIR	34.99	53.51	56.79	70.94	25.13	28.51	47.72	60.55	61.62	77.11
SEIFP	-	-	-	-	-	-	52.92	65.96	-	-

A primeira análise foi realizada a respeito dos modelos individuais, com objetivo de selecionar os três melhores modelos para construir o modelos de comitês. Sendo assim, é possível observarmos na Tabela 6.5 os modelos com os menores valores de RMSE e σ estão em negrito para cada tipo de dia. Além disso, a coluna “N. Estratificado” apresenta os resultados dos modelos para a base de dados sem a categorização

dos tipos de dias. A coluna “Final” representa a média ponderada de acordo com os tipos de dias dos valores do RMSE e σ para o modelo SEIR, sendo que para o modelo SEIFP não é calculada uma média, é apresentado apenas o seu resultado final. A média para o modelo SEIFP não é calculada devido as suas previsões não dependem da categorização dos dias, assim não havendo resultados para os tipos de dias específicos. As previsões feitas pelo modelo SEIFP são ponderadas de acordo com a teria de conjuntos *fuzzy*. Para o cálculo do resultado final apresentado na coluna “Final”, levou-se em consideração os resultados para cada tipo climático e foram considerados 90 dias ensolarados, 400 dias parcialmente nublados e 110 dias nublados, totalizando um total de 600 dias sem considerar os 30 dias presentes na primeira janela de treinamento. Os valores de R^2 não foram apresentados nas tabelas devido todos os resultados obtidos serem maiores que 0.94, dessa forma não demonstrando muito valor quando comparado o desempenho entre os modelos.

Primeiramente foram observados os resultados para os modelos utilizando método de validação avanço não ancorado (Tabela 6.5). Quando observado os valores de RMSE, as melhores performances foram alcançadas pelos modelos MLP, SVR e GB tanto as previsões para dias ensolarados quanto para os parcialmente nublados. Para os dias nublados, os modelos SVR, LR e kNN obtiveram melhores resultados. Olhando para os resultados obtidos utilizando a base de dados não categorizada em tipos de dias, os modelos MLP, SVR e GB se sobressaem em relação aos outros modelos. Os piores resultados foram obtidos pelos modelos RF e AB, no qual mantiveram a sua performance abaixo da média dos outros modelos em todas as abordagens. Olhando para os valores de desvio-padrão na Tabela 6.5, é possível observar que o modelo GB manteve o melhor desempenho ente os modelos individuais, e mesmo que a MLP não esteja entre os três melhores métodos, o valor do σ é próximo ao dos demais. Por exemplo, o modelo AB carrega um menor erro de desvio-padrão nos dias específicos, porém o seu RMSE é aproximadamente 20% maior do que para MLP, para os dias específicos e também no resultado final (o mesmo acontece na Tabela 6.6. A partir da coluna “Final” é possível realizar a comparação dos resultados de cada um dos modelos de forma geral, mostrando que a MLP, o SVR e o GB prevaleceram diante aos outros modelos individuais com os melhores resultados baseado nos valores de RMSE e também do desvio-padrão.

A Tabela 6.6 apresenta os resultados obtidos através do método de validação de avanço ancorado. Novamente os modelos MLP, SVR e GB prevaleceram com os melhores desempenhos para todos os tipos de dias, conseqüentemente obtendo um desempenho geral melhor que os demais modelos. O kNN demonstrou uma boa performance para a base de dados não discretizada, se saindo melhor que a MLP. Os piores resultados em âmbito geral, utilizando o método de validação de avanço ancorado foram obtidos pelos modelos AB e RF, mesmo o modelo AB obtendo baixos valores de

desvio-padrão comparado com o modelo RF. Os resultados individuais dos modelos para cada tipo de dia específico mostraram maiores valores de RMSE para os dias ensolarados e parcialmente nublados, e também para os resultados obtidos com a base de dados não categorizada em tipos de dias. Os melhores resultados para cada tipo de dia estão destacados em negrito nas Tabelas 6.5 e 6.6.

Após a identificação dos modelos individuais que obtiveram os melhores resultados por meio das métricas de desempenho, foram construídos os modelos de comitês utilizados no trabalho. Portanto, os modelos *Stacking Ensemble* de Intervalo Rígido (SEIR) e *Stacking Ensemble* de Intervalos *Fuzzy* Ponderados (SEIFP) são constituídos pelos modelos individuais, MLP, GB e SVR, além do meta-modelo LASSO. Os resultados obtidos pelos modelos de comitês são apresentados na parte inferior das Tabelas 6.5 e 6.6.

Os resultados apresentados na Tabela 6.5 mostram que o melhor resultado foi alcançado pelo modelo SEIFP, por meio do método de validação de avanço não ancorado. Não são demonstrados resultados utilizando a base de dados não discretizada e para os dias específicos, devido o modelo SEIFP não utilizar a categorização dos tipos de dias utilizada pelos demais modelos.

Para compararmos os resultados entre os modelos individuais e comitês, foi necessário olhar para a coluna “Final” nas Tabelas 6.5 e 6.6. Dessa forma, por meio das métricas de desempenho presentes nas tabelas foi observado que o modelo que obteve o melhor resultado foi o *Stacking Ensemble* de Intervalos *Fuzzy* Ponderados utilizando o método de validação não ancorado. Os modelos MLP, SVR e SEIR obtiveram resultados muito próximos utilizando o mesmo método de validação. Para os modelos de previsão que fizeram uso da base de dados não categorizada os melhores resultados foram obtidos pelos modelos SEIR e SVR, utilizando o método de avanço não ancorado. Os modelos que obtiveram os melhores comportamentos através do método de validação de avanço ancorado, foram SEIR e SVR.

Os métodos de validação foram utilizados com objetivo de comparação e investigação, visando entender se o tamanho dos conjuntos de dados de treinamento dos modelos poderiam afetar os resultados. Olhando para os métodos de validação, é possível observar que os resultados obtidos com o uso do método de avanço não ancorado foram nitidamente melhores do que para o método de avanço ancorado. Geralmente um conjunto de dados de treinamento com mais observações entrega a um modelo mais informações, o que deveria gerar resultados melhores, porém não foi o que aconteceu com os modelos propostos.

A validação de avanço ancorado fornece ao modelo conjuntos de treinamento maiores a cada iteração do treinamento, conseqüentemente entregando mais variações dos atributos presentes no conjunto de dados. As variações influenciam na quantidade de diferentes tipos de situações ao longo dos dias concedendo informa-

ções mais detalhadas, o que pode prejudicar o desempenho do modelo. Portanto, o maior detalhamento dos dados para janelas de treinamento com um grande número de observações carregam consigo as variações sazonais ao longo do ano, afetando as predições dos modelos a longo prazo. Os bons desempenhos alcançados pelos modelos que utilizaram o método de avanço não ancorado, são possivelmente justificados pela adaptação e reflexo sobre as mudanças sazonais ao longo do ano.

Foi realizada uma análise mais detalhada dos métodos que obtiveram os melhores desempenhos em relação aos modelos utilizados na Estratégia de Previsão Baseada em Comitê de Métodos. Entre eles estão os modelos individuais MLP e SVR, como também os modelos SEIR e SEIFP. Estes modelos foram validados por meio do método de avanço não ancorado. Essa análise abordará também o modelo SEIR validado pelo método de avanço ancorado. Será avaliado também o modelo SEIR que utilizou a base de dados não categorizada em tipos de dias e que utilizou o método de validação de avanço não ancorado. Os modelos validados por meio do método de validação de avanço não ancorado serão referenciados em suas siglas com o sufixo “N” (por exemplo, SEIRN), e os modelos validados pelo método de avanço ancorado com sufixo “A” (por exemplo, SEIRA). O modelo SEIR que utilizou a base de dados não categorizada e a validação não ancorada será referenciado como SEIRNN.

Na Figura 6.12 é apresentada as predições médias semanais em conjunto com o valor de potência medido. Pode-se observar que em alguns momentos entre as semanas 80 e 90, como também 110 e 120 o modelo SEIRA tem dificuldade em acompanhar a curva de potência. Os outros modelos possuem comportamentos similares, porém é visível que entre as semanas 10 e 20 e as semanas 30 e 40, ocorre uma pequena variação em que os modelos SEIFPN e SEIRN possuem predições mais acuradas, acompanhando a curva de potência. Nas primeiras e últimas 10 semanas e na semana 90, é possível observar que todos os modelos possuem dificuldade para acompanhar a curva de potência. Para que fosse possível obter um melhor entendimento a respeito dos erros encontrados na Figura 6.12, foi realizada uma análise dos erros individuais de cada modelo.

Na Figura 6.13 são mostrados os erros individuais obtidos para cada um dos melhores modelos. São apresentados os erros para os grupos de tipo de dia, do resultado final dos modelos e também para os modelos que utilizaram a base de dados sem a categorização dos dias. Pode-se observar que o valor final do RMSE é afetado por *outliers* em todos os casos. A grandeza de alguns erros individuais para os tipos climáticos dos dias chama atenção, como no modelo SVRN para os dias nublados e para MLPN em dias parcialmente nublados.

A análise visual dos *boxplots* para os erros individuais revelou grande número de erros nos modelos para os tipos climáticos, o que ocasionou em uma análise do comportamento dos modelos nos tipos específicos de dias. A Figura 6.14 apresenta os

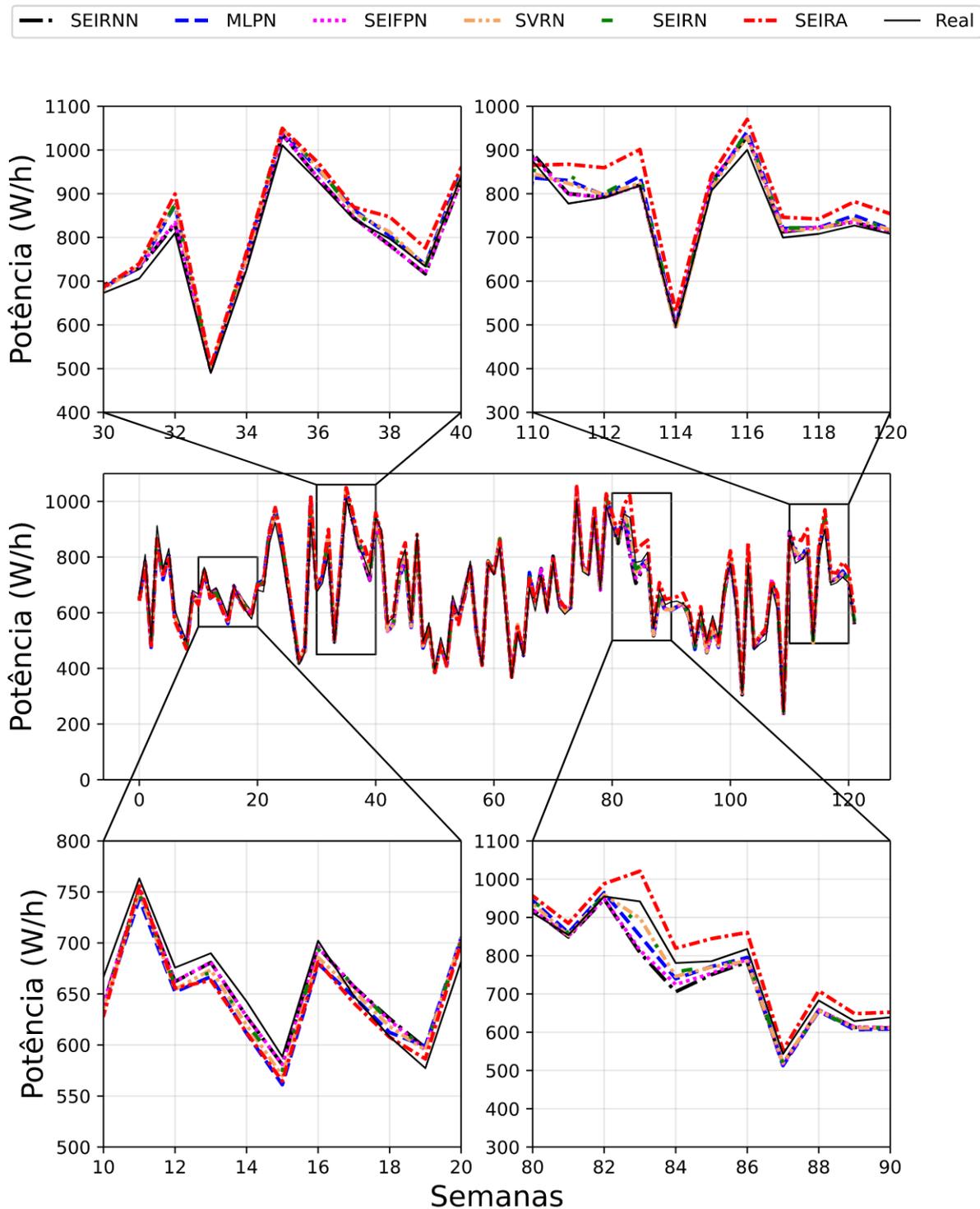


Figura 6.12 – Média das previsões semanais para os modelos que obtiveram o melhor desempenho.

Fonte: Autoria própria

resultados individuais dos modelos selecionados para cada tipo climático de dia. Essa figura tem como objetivo identificar o comportamento dos modelos em dias específicos.

Na Figura 6.14 é possível observamos que no dia 06/04/2020 todos os modelos

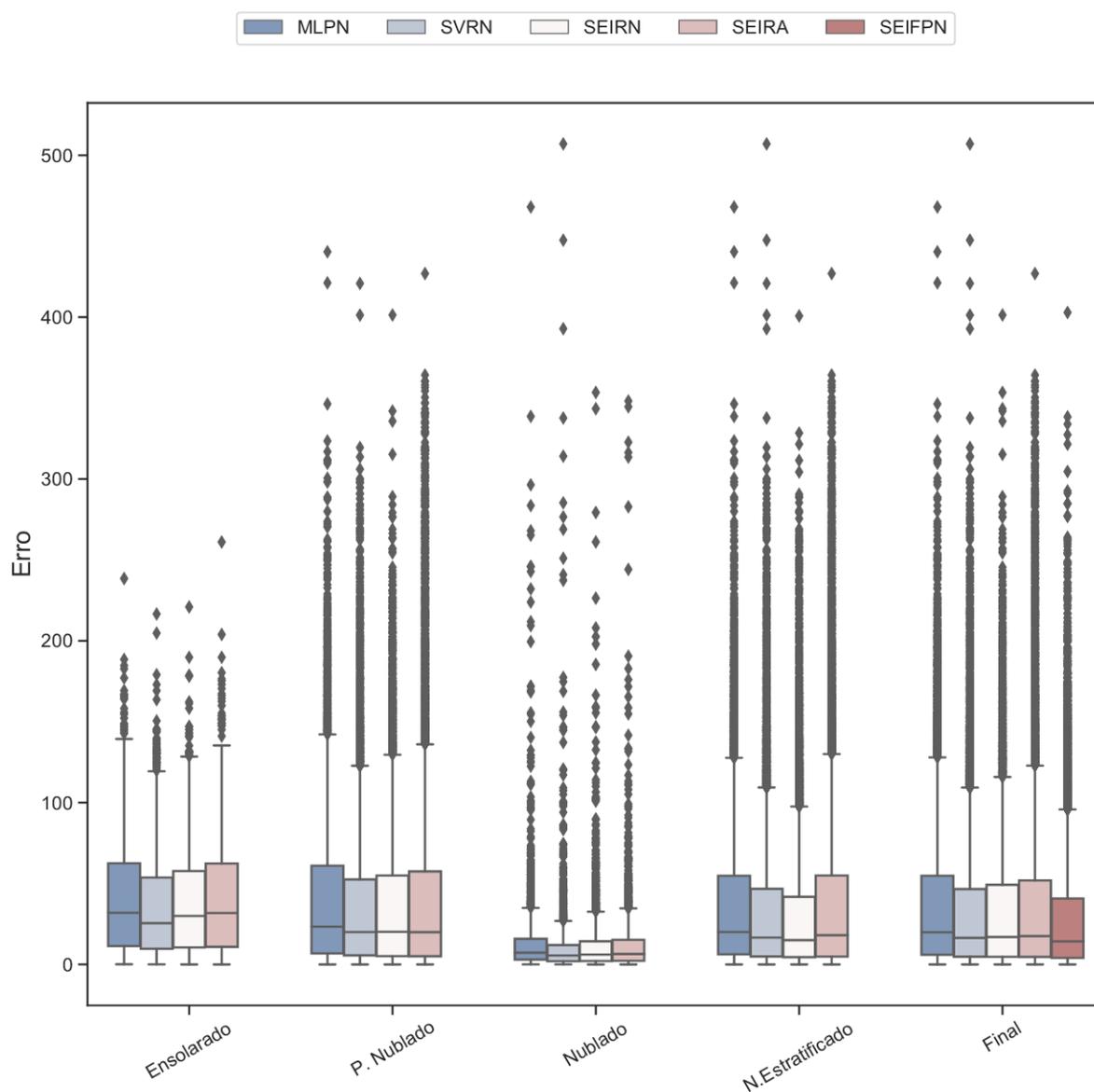


Figura 6.13 – *Boxplot* descrevendo os erros individuais de cada modelo para os tipos climáticos de dias.

Fonte: Autoria própria

tiveram dificuldades de acompanhar a curva de potência de saída entre as 6:30 e 12:00. O dia 06/04/2020 é do tipo nublado e apresenta uma alta variabilidade nos valores de irradiância, o que pode ter afetado o desempenho dos modelos. O dia 10/12/2018 é ensolarado, no qual os modelos conseguiram realizar boas previsões. No entanto, entre as 10:30 e 12:00 houve uma variação no GHI e DNI, o que afetou o desempenho dos modelos que não conseguiram realizar boas previsões. O dia 27/04/2018 possui o perfil de um dia parcialmente nublado que possui um alto índice de variação das componentes de irradiância ao longo do dia. O alto índice de variação ao longo do dia faz com que os modelos tenham dificuldades em realizar boas previsões, como é possível observar em todos os tipos climáticos de dias. Entre todos os modelos seleti-

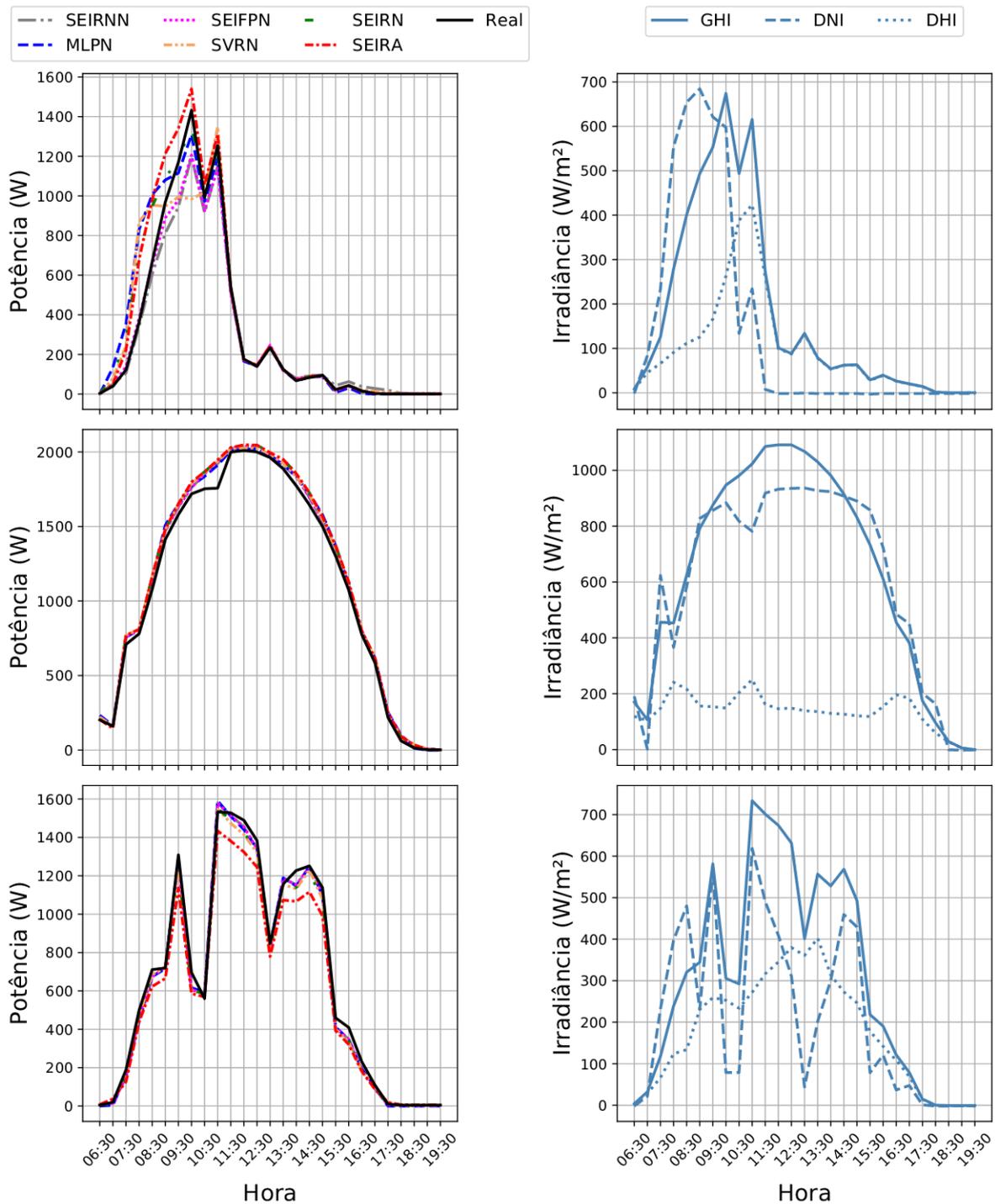


Figura 6.14 – Valores esperados e preditos para potência de saída e valores de irradiação para diferentes tipos dias. Da primeira linha para a última: 06/04/2020, 10/12/2018 e 27/04/2018.

Fonte: Autoria própria

onados para a análise detalhada desenvolvida, as piores previsões foram realizadas pelo modelo SEIRA utilizando o método de validação de avanço ancorado, mostrando um desempenho pior nos dias nublados e parcialmente nublados.

7 CONCLUSÕES

Neste trabalho foram avaliadas duas estratégias de previsão de geração de energia fotovoltaica à curto prazo. As estratégias são baseadas em métodos de aprendizado de máquina. Foi utilizada uma base de dados históricas contendo informações sobre dados meteorológico e climáticos, como também de potência gerada. Os dados foram coletados por meio de uma estação solarimétrica localizada no Laboratório Fotovoltaica da Universidade Federal de Santa Catarina, situado na cidade de Florianópolis.

Primeiramente foi realizada uma análise exploratória da base de dados histórica disponível. Através da análise e pré-processamento dos dados foram identificadas as relações entre as variáveis meteorológicas e a potência gerada pela planta fotovoltaica utilizada para a coleta dos dados. Sendo assim foram apresentadas duas estratégias de previsões, uma das estratégias baseada em diferentes subconjuntos de entrada e outra baseada no tipo climático dos dias. A análise realizada considerou os dados diários com os dados agregados em intervalos de 30 minutos. Os modelos foram construídos considerando janelas de dados passados que se estenderam por 30 dias para o método de validação de avanço não ancorado, e assim, avaliados para a previsão do dia seguinte. Foi considerado também o método de validação ancorado, no qual as janelas de dados passados de 30 dias se expandem a cada predição, adicionando o dia avaliado a janela de dados.

Com a implementação das estratégias de previsão propostas foram alcançados bons resultados, nos quais a Estratégia de Previsão Baseada em Modelos Individuais e a Estratégia de Previsão Baseada em Comitês de Métodos se mostraram capazes de realizar a previsão de geração de energia fotovoltaica. A estratégia de previsão baseada em diferentes subconjuntos de entrada passou por um pré-processamento para remover registros com medições espúrias devido a problemas com o registrador de dados. Para essa estratégia utilizou-se apenas o método de validação não ancorado para a validação dos modelos. Foram avaliados os métodos de aprendizado de máquina MLP, SVR, kNN, DT, e LR. A análise dos resultados apontou a MLP e o SVR como os métodos que obtiveram os melhores desempenhos. No entanto, o kNN e DT sofreram mais com a natureza ruidosa dos dados dos sensores. Considerando os diferentes conjuntos de características, embora o GHI esteja altamente correlacionada com a saída de potência, o uso de características adicionais se demonstrou uma boa alternativa. Especificamente, ao utilizar a MLP com todas as características foi obtida uma redução de 30% no RMSE em relação aos outros subconjuntos de atributos.

Para a Estratégia de Previsão Baseada em Comitês de Métodos, foram considerados três tipos de dias, ensolarados, parcialmente nublados e nublados. Foi avaliada a utilização dos dois métodos de validação abordados na metodologia do trabalho. Utilizou-se além dos modelos individuais utilizados na Estratégia de Previsão Base-

ada em Modelos Individuais os modelos de comitês de métodos *Stacking Ensemble* de Intervalos Rígidos e também um modelo *Stacking Ensemble* de Intervalos Fuzzy Ponderados. Os modelos MLP e SVR demonstraram bons resultados no desenvolvimento desta estratégia, no entanto os melhores resultados foram obtidos por meio dos modelos *Stacking Ensemble* de Intervalos Fuzzy Ponderados e *Stacking Ensemble* de Intervalos Rígidos. Foram avaliados os modelos com a utilização da base de dados sem a discretização dos dias e nessa ocasião o modelo *Stacking Ensemble* de Intervalos Rígidos e SVR demonstraram o melhor desempenho. O método de validação de avanço não ancorado quando utilizado demonstrou melhores resultados para todos os modelos.

As estratégias apresentadas no presente trabalho demonstraram resultados satisfatórios e a capacidade de realizar a previsão da potência de saída gerada pela planta fotovoltaica. Pode-se concluir que a aplicação das estratégias de previsão abordadas podem servir de uma solução alternativa para produtores de energia fotovoltaica e para as concessionárias de energia elétrica, tornando as operações de gerenciamento mais seguras e efetivas.

Para trabalhos futuros tem-se a intenção de realizar a previsão de potência de saída de plantas fotovoltaicas em outros horizontes de previsão. A utilização de uma estação solarimétrica de baixo custo para realizar a coleta de dados climáticos e meteorológicos. O desenvolvimento de uma estação única que realize a previsão de geração de energia e também a coleta dos dados em tempo real. Um estudo futuro pode possibilitar a implantação dos modelos de previsão em diferentes plantas fotovoltaicas em outras localidades.

REFERÊNCIAS

- AGUIAR, L. M. et al. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, Elsevier, v. 97, p. 599–610, 2016.
- ALLENDE, H.; VALLE, C. Ensemble methods for time series forecasting. In: _____. *Claudio Moraga: A Passion for Multi-Valued Logic and Soft Computing*. Cham: Springer International Publishing, 2017. p. 217–232. ISBN 978-3-319-48317-7. Disponível em: <https://doi.org/10.1007/978-3-319-48317-7_13>.
- ALMONACID, F. et al. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Conversion and Management*, Elsevier Ltd, v. 85, p. 389–398, 2014. ISSN 01968904. Disponível em: <<http://dx.doi.org/10.1016/j.enconman.2014.05.090>>.
- ALZAHIRANI, A. et al. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Computer Science*, Elsevier B.V., v. 114, p. 304–313, 2017. ISSN 18770509. Disponível em: <<https://doi.org/10.1016/j.procs.2017.09.045>>.
- ANTONANZAS, J. et al. Review of photovoltaic power forecasting. *Solar Energy*, Elsevier Ltd, v. 136, p. 78–111, 2016. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2016.06.069>>.
- ARIZE, D.; RIOS, T. N. A comparison study on time series forecasting given smart grid load uncertainties. *Proceedings - 2019 Brazilian Conference on Intelligent Systems, BRACIS 2019*, IEEE, p. 257–262, 2019.
- AWAD, M.; KHANNA, R. Support vector regression. In: _____. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA: Apress, 2015. p. 67–80. ISBN 978-1-4302-5990-9. Disponível em: <https://doi.org/10.1007/978-1-4302-5990-9_4>.
- BACHOUR, D.; PEREZ-ASTUDILLO, D. Ground measurements of Global Horizontal Irradiation in Doha, Qatar. *Renewable Energy*, Elsevier Ltd, v. 71, p. 32–36, 2014. ISSN 09601481. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2014.05.005>>.
- BAŞOLU, M. E. et al. Performance analyzes of different photovoltaic module technologies under zmit, Kocaeli climatic conditions. *Renewable and Sustainable Energy Reviews*, v. 52, p. 357–365, 2015. ISSN 18790690.
- BERGER, T. B. L. Image Comparison from two cloud cover sensor in infrared and visible spectral regions. n. 2, p. 5, 2005.
- BEZDEK, J. C. *Pattern recognition with fuzzy objective function algorithms*. [S.l.]: Springer Science & Business Media, 2013.
- BHATTACHARYA, T.; CHAKRABORTY, A. K.; PAL, K. Effects of Ambient Temperature and Wind Speed on Performance of Monocrystalline Solar Photovoltaic Module in Tripura, India. *Journal of Solar Energy*, v. 2014, p. 1–5, 2014. ISSN 2356-7635.
- BLANC, P. et al. Direct normal irradiance related definitions and applications: The circumsolar issue. *Solar Energy*, v. 110, p. 561–577, 2014. ISSN 0038092X.

BRANO, V. L.; CIULLA, G.; FALCO, M. D. Artificial neural networks to predict the power output of a PV panel. *International Journal of Photoenergy*, v. 2014, 2014. ISSN 1110662X.

BREIMAN, L. Classification Algorithms and Regression Trees. *Classification and Regression Trees*, p. 246–280, 1984.

BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

BRUTON, T. M. General trends about photovoltaics based on crystalline silicon. *Solar Energy Materials and Solar Cells*, v. 72, n. 1-4, p. 3–10, 2002. ISSN 09270248.

BUGALA, A. et al. Short-term forecast of generation of electric energy in photovoltaic systems. *Renewable and Sustainable Energy Reviews*, v. 81, n. November 2016, p. 306–312, 2018. ISSN 18790690.

CAMPBELL. *Campbell Scientific - Registrador de dados (Datalogger) para medição e controle*. 2021. <<https://www.campbellsci.com.br/cr6>>. Accessed: 2021-05-12.

CAÑETE, C.; CARRETERO, J.; CARDONA, M. Sidrach-de. Energy performance of different photovoltaic module technologies under outdoor conditions. *Energy*, Elsevier Ltd, v. 65, p. 295–302, 2014. ISSN 03605442. Disponível em: <<http://dx.doi.org/10.1016/j.energy.2013.12.013>>.

CARRASCO, M. et al. Low-Cost Solar Irradiance Sensing for PV Systems. *Energies*, v. 10, n. 7, 2017. ISSN 19961073.

CAZORLA, A.; OLMO, F.; ALADOS-ARBOLEDAS, L. Development of a sky imager for cloud cover assessment. *JOSA A*, Optical Society of America, v. 25, n. 1, p. 29–39, 2008.

CHAAR, L. E.; LAMONT, L. A.; ZEIN, N. E. Review of photovoltaic technologies. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 15, n. 5, p. 2165–2175, 2011. ISSN 13640321. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2011.01.004>>.

CHEN, C. et al. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, Elsevier Ltd, v. 85, n. 11, p. 2856–2870, 2011. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2011.08.027>>.

CHOLLET, F. et al. *Keras*. 2015. <<https://keras.io>>.

CHOW, S. K.; LEE, E. W.; LI, D. H. Short-term prediction of photovoltaic energy generation by intelligent approach. *Energy and Buildings*, Elsevier B.V., v. 55, p. 660–667, 2012. ISSN 03787788. Disponível em: <<http://dx.doi.org/10.1016/j.enbuild.2012.08.011>>.

COIMBRA, C. F.; KLEISSL, J.; MARQUEZ, R. *Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation*. Elsevier, 2013. 171–194 p. ISBN 9780123971777. Disponível em: <<http://dx.doi.org/10.1016/B978-0-12-397177-7.00008-5>>.

DAS, U. K. et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 81, n. April 2017, p. 912–928, 2018. ISSN 18790690. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2017.08.017>>.

De Barros, R. C. et al. Low-cost solar irradiance meter using LDR sensors. *2018 13th IEEE International Conference on Industry Applications, INDUSCON 2018 - Proceedings*, IEEE, p. 72–79, 2019.

De Giorgi, M. G.; CONGEDO, P. M.; MALVONI, M. Photovoltaic power forecasting using statistical methods: Impact of weather data. *IET Science, Measurement and Technology*, v. 8, n. 3, p. 90–97, 2014. ISSN 17518822.

De Giorgi, M. G. et al. Short-term power forecasting by statistical methods for photovoltaic plants in south Italy. *4th IMEKO TC19 Symposium on Environmental Instrumentation and Measurements 2013: Protection Environment, Climate Changes and Pollution Control*, p. 171–175, 2013.

DEKKER, J. et al. Investigating the effects of solar modelling using different solar irradiation data sets and sources within South Africa. *Solar Energy*, 2012. ISSN 0038092X.

DING, M.; WANG, L.; BI, R. An ANN-based approach for forecasting the power output of photovoltaic system. *Procedia Environmental Sciences*, v. 11, n. PART C, p. 1308–1315, 2011. ISSN 18780296.

DIVINA, F. et al. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies*, v. 11, n. 4, p. 1–31, 2018. ISSN 19961073.

DOBREVA, P.; DYK, E. E. V.; VORSTER, F. J. New approach to evaluating predictive models of photovoltaic systems. *Solar Energy*, v. 204, p. 134–143, 2020. ISSN 0038-092X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0038092X20304035>>.

DONDERS, A. R. T. et al. Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, v. 59, n. 10, p. 1087–1091, 2006. ISSN 08954356.

Chapter 2 - basic notions in fuzzy set theory. In: DUBOIS D., P. H.; YAGER, R. R. (Ed.). *Readings in Fuzzy Sets for Intelligent Systems*. Morgan Kaufmann, 1993. p. 21–26. ISBN 978-1-4832-1450-4. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9781483214504500067>>.

EHSAN, R. M.; SIMON, S. P.; VENKATESWARAN, P. R. Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. *Neural Computing and Applications*, Springer London, v. 28, n. 12, p. 3981–3992, 2017. ISSN 09410643.

ESCOBAR, R. A. et al. Solar energy resource assessment in Chile: Satellite estimation and ground station measurement. *Energy Procedia*, v. 57, n. 2014, p. 1257–1265, 2014. ISSN 18766102.

ESPRESSIF. *Módulo WiFi NodeMCU ESP-32 + Bluetooth*. 2019. <<https://www.huinfinito.com.br/home/1411-modulo-wifi-nodemcu-esp-32-bluetooth.html>>. [Online; Acessado em 05/05/2021].

FEISTER, U. et al. Comparison of macroscopic cloud data from ground-based measurements using VIS/NIR and IR instruments at Lindenberg, Germany. *Atmospheric Research*, Elsevier B.V., v. 96, n. 2-3, p. 395–407, 2010. ISSN 01698095. Disponível em: <<http://dx.doi.org/10.1016/j.atmosres.2010.01.012>>.

FERNANDEZ-JIMENEZ, L. A. et al. Short-term power forecasting system for photovoltaic plants. *Renewable Energy*, Elsevier Ltd, v. 44, p. 311–317, 2012. ISSN 09601481. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2012.01.108>>.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, Elsevier, v. 55, n. 1, p. 119–139, 1997.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

GANAI, M. A. et al. Ensemble deep learning: A review. 2021. Disponível em: <<http://arxiv.org/abs/2104.02395>>.

GHEDIRA, H.; EISSA, Y. Global tilt irradiation map of the united arab emirates. *Energy Procedia*, Elsevier, v. 57, p. 1199–1205, 2014.

GIELEN, D. et al. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, Elsevier, v. 24, n. January, p. 38–50, 2019. ISSN 2211467X.

GÖKMEN, N. et al. Investigation of wind speed cooling effect on PV panels in windy locations. *Renewable Energy*, v. 90, p. 283–290, 2016. ISSN 18790682.

GRADITI, G.; FERLITO, S.; ADINOLFI, G. Comparison of Photovoltaic plant power production prediction methods using a large measured dataset. *Renewable Energy*, v. 90, p. 513–519, 2016. ISSN 18790682.

GREEN, M. A. Thin-film solar cells: Review of materials, technologies and commercial status. *Journal of Materials Science: Materials in Electronics*, v. 18, n. SUPPL. 1, p. 15–19, 2007. ISSN 09574522.

GUEYMARD, C. A.; MYERS, D. R. Solar radiation measurement: progress in radiometry for improved modeling. In: *Modeling Solar Radiation at the Earth's Surface*. [S.l.]: Springer, 2008. p. 1–27.

GUEYMARD, C. A.; MYERS, D. R. Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy*, Elsevier, v. 83, n. 2, p. 171–185, 2009.

HAYKIN, S. *Neural Networks and Learning Machines*. [S.l.: s.n.], 2008. v. 3. 906 p. ISSN 14337851. ISBN 9780131471399.

HAYKIN, S. S. et al. *Neural networks and learning machines/Simon Haykin*. [S.l.]: New York: Prentice Hall,, 2009.

HICKEL, B. M. O Impacto no Desempenho de Sistemas Fotovoltaicos Causado pelo Acúmulo de Sujeria Sobre os Módulo FV - Metodologia e Avaliação Através de Curvas IxV em Campo. p. 110, 2017.

HIDALGO, F. G.; MARTINEZ, R. F.; VIDAL, E. F. Design of a Low-Cost Sensor for Solar Irradiance. *oceanoptics.com*, p. 1–8, 2013.

HINTON, G. E.; SEJNOWSKI, T. J. et al. *Unsupervised learning: foundations of neural computation*. [S.l.]: MIT press, 1999.

HOFF, T. E.; PEREZ, R. Quantifying PV power Output Variability. *Solar Energy*, Elsevier Ltd, v. 84, n. 10, p. 1782–1793, 2010. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2010.07.003>>.

HOSSAIN, M. et al. Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. *Journal of Cleaner Production*, Elsevier Ltd, v. 167, p. 395–405, 2017. ISSN 09596526. Disponível em: <<http://dx.doi.org/10.1016/j.jclepro.2017.08.081>>.

HUANG, Y. et al. Comparative study of power forecasting methods for PV stations. *2010 International Conference on Power System Technology: Technological Innovations Making Power Grid Smarter, POWERCON2010*, p. 1–6, 2010.

IEA. *International Energy Agency, Data and Statistics 2017*. 2017. <<https://www.iea.org/data-and-statistics/data-tables/?country=WORLD&energy=Electricity&year=2017>>. Accessed: 2020-03-17.

INMAN, R. H.; CHU, Y.; COIMBRA, C. F. Cloud enhancement of global horizontal irradiance in California and Hawaii. *Solar Energy*, Elsevier Ltd, v. 130, p. 128–138, 2016. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2016.02.011>>.

INMET, Instituto Nacional de Meteorologia. *Dados Históricos Anuais de Meteorologia*. 2021. <<https://portal.inmet.gov.br/dadoshistoricos>>. Acessado em 18/03/2021.

IRENA. *Renewable Energy Statistics 2020. Renewable hydropower (including mixed plants)*. [s.n.], 2020. 38 p. ISBN 978-92-9260-246-8. Disponível em: <www.irena.org>.

IRENA, I. R. E. A. *Solar Energy*. 2020. <<https://www.irena.org/solar>>. [Online; Acessado em 20/05/2021].

ISO, I. 9060: 1990: Specification and classification of instruments for measuring hemispherical solar and direct solar radiation. *Geneva, Switzerland*, 1990.

KAPLANI, E.; KAPLANIS, S. Thermal modelling and experimental assessment of the dependence of PV module temperature on wind velocity and direction, module orientation and inclination. *Solar Energy*, Elsevier Ltd, v. 107, p. 443–460, 2014. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2014.05.037>>.

KARDAKOS, E. G. et al. Application of time series and artificial neural network models in short-term forecasting of PV power generation. *Proceedings of the Universities Power Engineering Conference*, p. 2–7, 2013.

KESTREL. *Kestrel Heat Index Reference Guide*. [S.l.], 2009. Disponível em: <<https://kestrelmeters.com/products/kestrel-3000-wind-meter>>.

KHAN, I. et al. Photovoltaic power forecasting based on elman neural network software engineering method. In: *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. [S.l.: s.n.], 2017. p. 747–750.

KIM, K. Y.; KIM, B. J.; YI, G. S. Reuse of imputed data in microarray analysis increases imputation efficiency. *BMC Bioinformatics*, v. 5, p. 1–9, 2004. ISSN 14712105.

KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2014.

KRUAKAEW, R.; BANJERDPONGCHAI, D.; HOONCHAREON, N. Cloud detection on ground-based sky images with brightness reduction of circumsolar region. *ECTI-CON 2018 - 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, p. 329–332, 2019.

KUDO, M. et al. Forecasting electric power generation in a photovoltaic power system for an energy network. *Electrical Engineering in Japan (English translation of Denki Gakkai Ronbunshi)*, v. 167, n. 4, p. 16–23, 2009. ISSN 04247760.

KWAK, S. K.; KIM, J. H. Statistical data preparation: Management of missing values and outliers. *Korean Journal of Anesthesiology*, v. 70, n. 4, p. 407–411, 2017. ISSN 20057563.

LAVE, M. et al. Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States. *IEEE Journal of Photovoltaics*, IEEE, v. 5, n. 2, p. 597–606, 2015. ISSN 21563381.

LEVA, S. et al. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation*, Elsevier B.V., v. 131, p. 88–100, 2017. ISSN 03784754.

LIU, L. na et al. [A comparison of proteomic analysis of *Helicobacter pylori* in patients with gastritis and gastric cancer between areas of high and low incidence of gastric cancer]. *Beijing da xue xue bao. Yi xue ban = Journal of Peking University. Health sciences*, v. 43, n. 6, p. 827–832, 2011. ISSN 1671167X.

LORENZ, E. et al. Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 2, n. 1, p. 2–10, 2009. ISSN 21511535.

MADETI, S. R.; SINGH, S. N. Monitoring system for photovoltaic plants: A review. *Renewable and Sustainable Energy Reviews*, v. 67, p. 1180–1207, 2017. ISSN 18790690.

MAHMOUD, S.; MOHAMED, H. High accuracy modeling method for photovoltaic arrays. In: *Renewable and Sustainable Energy II*. [S.l.]: Trans Tech Publications Ltd, 2012. (Advanced Materials Research, v. 512), p. 238–241.

MAHTTA, R.; JOSHI, P. K.; JINDAL, A. K. Solar power potential mapping in India using remote sensing inputs and environmental parameters. *Renewable Energy*, Elsevier Ltd, v. 71, p. 255–262, 2014. ISSN 09601481. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2014.05.037>>.

MALVONI, M.; De Giorgi, M. G.; CONGEDO, P. M. Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia*, Elsevier B.V., v. 126, p. 651–658, 2017. ISSN 18766102. Disponível em: <<https://doi.org/10.1016/j.egypro.2017.08.293>>.

MANTELLI, S. L.; MARTINS, G. L.; RUTHER, R. Uncertainty analysis of solar monitoring station: A case study. *Proceedings of the ISES Solar World Congress 2019 and IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry 2019*, p. 2135–2140, 2020.

MARQUEZ, R.; COIMBRA, C. F. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, Elsevier Ltd, v. 85, n. 5, p. 746–756, 2011. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2011.01.007>>.

MARTÍNEZ, M. A.; ANDÚJAR, J. M.; ENRIQUE, J. M. A new and inexpensive pyranometer for the visible spectral range. *Sensors (Switzerland)*, v. 9, n. 6, p. 4615–4634, 2009. ISSN 14248220.

MARTINS, F.; SOUZA, M.; PEREIRA, E. Comparative study of satellite and ground techniques for cloud cover determination. *Advances in Space Research*, Elsevier, v. 32, n. 11, p. 2275–2280, 2003.

MARTINS, F. R.; PEREIRA, E. B.; ECHER, M. P. D. S. Levantamento dos recursos de energia solar no Brasil com o emprego de satélite geoestacionário-o Projeto Swera (Solar energy resources assesment using geostationary satellites in brazil-Swera Project). n. 2, p. 145–159, 2004. Disponível em: <www.sbfisica.org.br>.

MASTECH. *Digital Clamp Meters Datasheet*. [S.l.], 2019. 81 p. Disponível em: <<https://www.atel-electronics.eu/produto.php?hash=03948>>.

MATHIESEN, P.; KLEISSL, J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. *Solar Energy*, Elsevier, v. 85, n. 5, p. 967–977, 2011.

MCCANN, M. J. et al. A review of thin-film crystalline silicon for solar cell applications. Part 1: Native substrates. *Solar Energy Materials and Solar Cells*, v. 68, n. 2, p. 135–171, 2001. ISSN 09270248.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.

MEDUGU, D. W.; BURARI, F. W.; ABDULAZEEZ, A. A. Construction of a reliable model pyranometer for irradiance measurements. *African Journal of Biotechnology*, v. 9, n. 12, p. 1719–1725, 2010. ISSN 16845315.

MEKHILEF, S.; SAIDUR, R.; KAMALISARVESTANI, M. Effect of dust, humidity and air velocity on efficiency of photovoltaic cells. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 16, n. 5, p. 2920–2925, 2012. ISSN 13640321. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2012.02.012>>.

MELLIT, A.; Massi Pavan, A.; LUGHI, V. Short-term forecasting of power production in a large-scale photovoltaic plant. *Solar Energy*, Elsevier Ltd, v. 105, p. 401–413, 2014. ISSN 0038092X.

MENDES-MOREIRA, J. a. et al. Ensemble approaches for regression: A survey. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 45, n. 1, dez. 2012. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/2379776.2379786>>.

MITCHELL, T. *Machine Learning*. McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Disponível em: <<https://books.google.com.br/books?id=EoYBngEACAAJ>>.

MYERS, D. R. *Solar radiation: Practical modeling for renewable energy applications*. [S.l.: s.n.], 2017. 1–159 p. ISBN 9781466503274.

MYLES, A. J. et al. An introduction to decision tree modeling. *Journal of Chemometrics*, v. 18, n. 6, p. 275–285, 2004. ISSN 08869383.

NAQA, I. E.; MURPHY, M. J. Machine Learning in Radiation Oncology. *Machine Learning in Radiation Oncology*, p. 3–11, 2015.

NONNENMACHER, L.; KAUR, A.; COIMBRA, C. F. Verification of the SUNY direct normal irradiance model with ground measurements. *Solar Energy*, Elsevier Ltd, v. 99, p. 246–258, 2014. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2013.11.010>>.

OMUBO-PEPPLE, V.; ISRAEL-COOKEY, C.; ALAMINOKUMA, G. Effects of temperature, solar flux and relative humidity on the efficient conversion of solar energy to electricity. *European Journal of Scientific Research*, v. 35, n. 2, p. 173–180, 2009.

OMUBO-PEPPLE, V.; TAMUNOBERETON-ARI.; BRIGGS-KAMARA. Influence of Meteorological Parameters on the Efficiency of Photovoltaic Module in Some Cities in the Niger Delta of Nigeria. v. 3, n. 1, p. 107–113, 2013. ISSN 2226-5724.

ORTEGA, A. et al. The state of solar energy resource assessment in Chile. *Renewable Energy*, Elsevier, v. 35, n. 11, p. 2514–2524, 2010.

PADOVAN, A.; DEL COL, D. Measurement and modeling of solar irradiance components on horizontal and tilted planes. *Solar Energy*, Elsevier Ltd, v. 84, n. 12, p. 2068–2084, 2010. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2010.09.009>>.

PADOVAN, A. et al. DNI estimation procedures for the assessment of solar radiation availability in concentrating systems. *Energy Procedia*, Elsevier B.V., v. 57, p. 1140–1149, 2014. ISSN 18766102. Disponível em: <<http://dx.doi.org/10.1016/j.egypro.2014.10.100>>.

PAK, S. I.; OH, T. H. Correlation and simple linear regression. *Journal of Veterinary Clinics*, v. 27, n. 4, p. 427–434, 2010. ISSN 1598298X.

PAOLI, C. et al. Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, Elsevier Ltd, v. 84, n. 12, p. 2146–2160, 2010. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2010.08.011>>.

PARIDA, B.; INIYAN, S.; GOIC, R. A review of solar photovoltaic technologies. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 15, n. 3, p. 1625–1636, 2011. ISSN 13640321. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2010.11.032>>.

PARMAR, P. R. D. Solar radiation and pyranometer : Review. *Journal Of Information, Knowledge And Research IN Electrical Engineering*, p. 122–125, 2011.

- PDUGARIA, S. et al. Assessment of estimation methods of DNI resource in solar concentrating systems. *Solar Energy*, Elsevier Ltd, v. 121, p. 103–115, 2015. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2015.07.043>>.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011.
- PEDRO, H. T. C.; COIMBRA, C. F. M. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, Elsevier Ltd, v. 86, n. 7, p. 2017–2028, 2012. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2012.04.004>>.
- PELISSON, A. et al. Comparative study of photovoltaic power forecasting methods. In: *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*. Porto Alegre, RS, Brasil: SBC, 2020. p. 555–566. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/eniac/article/view/12159>>.
- PEREZ-ASTUDILLO, D.; BACHOUR, D. Dni, ghi and dhi ground measurements in doha, qatar. *Energy Procedia*, Elsevier, v. 49, p. 2398–2404, 2014.
- PRECISION, B. *Operating Instructions Digital Lightmeter*. [S.l.], 2000. Disponível em: <<https://www.bkprecision.com/products/environmental-testers/615-light-meter.html>>. Acesso em: 15 ago. 2020.
- RAJESH, R.; Carolin Mabel, M. A comprehensive review of photovoltaic systems. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 51, p. 231–248, 2015. ISSN 18790690.
- RANA, M.; KOPRINSKA, I.; AGELIDIS, V. G. Solar power forecasting using weather type clustering and ensembles of neural networks. *Proceedings of the International Joint Conference on Neural Networks*, IEEE, v. 2016-October, p. 4962–4969, 2016.
- RANSTAM, J.; COOK, J. A. LASSO regression. *British Journal of Surgery*, v. 105, n. 10, p. 1348, 2018. ISSN 13652168.
- RAZA, M. Q.; NADARAJAH, M.; EKANAYAKE, C. On recent advances in PV output power forecast. *Solar Energy*, Elsevier Ltd, v. 136, p. 125–144, 2016. ISSN 0038092X.
- RELVA, S. G. et al. Comprehensive analysis of solarimetry elements for primary energy forecasting methodologies related to photovoltaic power plants.
- RIBEIRO, M. H. D. M.; COELHO, L. d. S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing Journal*, Elsevier B.V., v. 86, p. 105837, 2020. ISSN 15684946. Disponível em: <<https://doi.org/10.1016/j.asoc.2019.105837>>.
- ROKACH, L.; MAIMON, O. Supervised learning. In: _____. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2010. p. 133–147. ISBN 978-0-387-09823-4. Disponível em: <https://doi.org/10.1007/978-0-387-09823-4_8>.
- RONZIO, D. A.; COLLINO, E.; BONELLI, P. A survey on different radiative and cloud schemes for the solar radiation modeling. *Solar Energy*, Elsevier Ltd, v. 98, p. 153–166, 2013. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2013.04.015>>.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

ROUHOLAMINI, A. et al. Temperature of a photovoltaic module under the influence of different environmental conditions – experimental investigation. *International Journal of Ambient Energy*, v. 37, n. 3, p. 266–272, 2016. ISSN 21628246. Disponível em: <<https://doi.org/10.1080/01430750.2014.952842>>.

ROY, G.; HAYMAN, S.; JULIAN, W. Sky analysis from CCD images: cloud cover. *Lighting Research and Technology*, v. 33, n. 4, p. 211–221, 2006. ISSN 1477-1535.

ROYSTON, P. Multiple imputation of missing values: Update of ice. *Stata Journal*, v. 5, n. 4, p. 527–536, 2005. ISSN 1536867X.

SCHAPIRE, R. E. The Strength of Weak Learnability. *Machine Learning*, v. 5, n. 2, p. 197–227, 1990. ISSN 15730565.

SCIENTIFIC, P. Wsd-100 wind speed and direction sensor. In: . [s.n.], 2011. Disponível em: <<https://www.pace-sci.com/wind-speed.htm>>.

SHI, J. et al. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, v. 48, n. 3, p. 1064–1069, 2012. ISSN 00939994.

SHINDE, P. P.; SHAH, S. A Review of Machine Learning and Deep Learning Applications. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, IEEE, p. 1–6, 2018.

Terms and definitions - a report from the committee on measurements and standards of the american society for mass spectrometry. In: SIUZDAK, G. (Ed.). *Mass Spectrometry for Biotechnology*. San Diego: Academic Press, 1996. p. 141 – 148. ISBN 978-0-12-647471-8. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780126474718500107>>.

SOUZA-ECHER, M. P. et al. A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, v. 23, n. 3, p. 437–447, 2006. ISSN 07390572.

STATISTA. *Statista - Global Business Platform, World Energy Outlook 2019*. 2019. <<https://www.statista.com/outlook/256/115/household-appliances/brazil>>. Accessed: 2020-03-17.

STOFFEL, T. et al. NREL is operated for DOE by the Alliance for Sustainable Energy, LLC CONCENTRATING SOLAR POWER Best Practices Handbook for the Collection and Use of Solar Resource Data. n. September, 2010.

STÖKLER, S.; SCHILLINGS, C.; KRAAS, B. Solar resource assessment study for pakistan. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 58, p. 1184–1188, 2016.

SU, X.; YAN, X.; TSAI, C. L. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 4, n. 3, p. 275–294, 2012. ISSN 19395108.

SYSTEMS, E.; SCIENCE, S. Development of a Simple and Cheap Equipment for Monitoring the Solar Irradiance on PV-modules. 2019.

TAN, R. H. G.; TAI, P. L. J.; MOK, V. H. Solar irradiance estimation based on photovoltaic module short circuit current measurement. *2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2013*, n. November, p. 26–27, 2013.

TENMARS. Luxmeters, Sound Level Meters, Breathalyzers Quick - Start Guide. In: . [s.n.], 1997. p. 1–2. Disponível em: <<https://www.tme.com/Document/603ea71c2508642ceccda2997c34319f/INB-POZ-MIER-EN.pdf>>.

TYAGI, V. V. et al. Progress in solar PV technology: Research and achievement. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 20, p. 443–461, 2013. ISSN 13640321. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2012.09.028>>.

VALENTINI, G.; MASULLI, F. Ensembles of learning machines. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 2486 LNCS, p. 3–20, 2002. ISSN 16113349.

VAPNIK, V. The nature of statistical learning. *Theory*, Springer, 1995.

VOYANT, C. et al. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, Elsevier Ltd, v. 105, p. 569–582, 2017. ISSN 18790682. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2016.12.095>>.

WANG, F. et al. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies*, v. 5, n. 5, p. 1355–1370, 2012. ISSN 19961073.

WANG, H.; MA, C.; ZHOU, L. A brief review of machine learning and its application. *Proceedings - 2009 International Conference on Information Engineering and Computer Science, ICIECS 2009*, 2009.

WMO, W. M. O. *Guide to Instruments and Methods of Observation*. [s.n.], 2018. I. 573 p. ISBN 978-92-63-10008-5. Disponível em: <https://library.wmo.int/index.php?lvl=notice_display&id=12407#XzhejjV7mUk>.

WOLPERT, D. H. Stacked generalization. *Neural networks*, Elsevier, v. 5, n. 2, p. 241–259, 1992.

ZADEH, L. Fuzzy sets. *Information and Control*, v. 8, n. 3, p. 338–353, 1965. ISSN 0019-9958. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S001999586590241X>>.

ZDRAVKOVIC, M. et al. Temperature Effects on Photovoltaic Components. *Scientific Publications of the State University of Novi Pazar*, v. 1, n. 1, p. 29–36, 2009.

ZELL, E. et al. Assessment of solar radiation resources in Saudi Arabia. *Solar Energy*, Elsevier Ltd, v. 119, p. 422–438, 2015. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2015.06.031>>.

ZHANG, J. et al. A suite of metrics for assessing the performance of solar power forecasting. *Solar Energy*, Elsevier Ltd, v. 111, p. 157–175, 2015. ISSN 0038092X. Disponível em: <<http://dx.doi.org/10.1016/j.solener.2014.10.016>>.

ZHANG, S. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, Elsevier Inc., v. 85, n. 11, p. 2541–2552, 2012. ISSN 01641212. Disponível em: <<http://dx.doi.org/10.1016/j.jss.2012.05.073>>.

ZHAO, J.; BOSE, B. K. Evaluation of membership functions for fuzzy logic controlled induction motor drive. *IECON Proceedings (Industrial Electronics Conference)*, v. 1, p. 229–234, 2002.

ZHONG, J. et al. Prediction of photovoltaic power generation based on general regression and back propagation neural network. *Energy Procedia*, Elsevier B.V., v. 152, p. 1224–1229, 2018. ISSN 18766102. Disponível em: <<https://doi.org/10.1016/j.egypro.2018.09.173>>.

ZHU, R.; GUO, W.; GONG, X. Short-term photovoltaic power output prediction based on k-fold cross-validation and an ensemble model. *Energies*, v. 12, n. 7, 2019. ISSN 19961073.