

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO POLITÉCNICO  
CURSO DE BACHARELADO EM ENGENHARIA MECÂNICA

HUGO KERSBAUMER KNUPP

PREVISÃO DE GERAÇÃO ENERGIA FOTOVOLTAICA NO BRASIL POR MEIO  
DE MODELOS DE APRENDIZADO DE MÁQUINA

MACAÉ  
2023

HUGO KERSBAUMER KNUPP

PREVISÃO DE GERAÇÃO ENERGIA FOTOVOLTAICA NO BRASIL POR MEIO  
DE MODELOS DE APRENDIZADO DE MÁQUINA

Trabalho de conclusão de curso de graduação apresentado ao Instituto Politécnico da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Engenharia Mecânica.

Orientador: Profa. Janaína Santanna Gomide

MACAÉ

2023

## CIP - Catalogação na Publicação

K74

Knupp, Hugo Kersbaumer

Previsão de geração energia fotovoltaica no Brasil por meio de modelos de  
aprendizado de máquinas / Hugo Kersbaumer Knupp - Macaé, 2023.  
63 f.

Orientador(a): Janáina Sant'anna Gomide.

Trabalho de conclusão de curso (graduação) - Universidade Federal do Rio de  
Janeiro, Instituto Politécnico, Bacharel em Engenharia Mecânica, 2023.

1. Aprendizado de máquina. 2. Energia solar. 3. Geração de energia  
fotovoltaica . I. Gomide, Janaina Santa'anna, orient. II. Título.

CDD 624

HUGO KERSBAUMER KNUPP

PREVISÃO DE GERAÇÃO ENERGIA FOTOVOLTAICA NO BRASIL POR MEIO  
DE MODELOS DE APRENDIZADO DE MÁQUINA

Trabalho de conclusão de curso de graduação apresentado ao Instituto Politécnico da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Engenharia Mecânica.

Aprovado em 19 de Julho de 2023

BANCA EXAMINADORA:

---

Janaína Sant'Anna Gomide  
Doutora (UFRJ)

---

Diego Cunha Malgueta  
Doutor (UFRJ)

---

Raquel Jahara Lobosco  
Doutora (UFRJ)

---

Leila Weitzel Coelho da Silva  
Doutora (UFF)

Dedico este trabalho a todas as pessoas que estiveram ao meu lado ao longo desta jornada, me apoiando e incentivando a perseguir meus sonhos.

Aos meus pais, que me deram o suporte necessário, e que sempre acreditaram em mim. Seu amor incondicional e apoio constante foram essenciais.

À minha família, que sempre esteve presente, me motivando e encorajando a enfrentar os desafios com determinação e perseverança. Seus valores e exemplos de vida foram fundamentais para o meu crescimento pessoal e acadêmico.

Aos meus amigos, companheiros de jornada, que compartilharam risadas, desafios e momentos inesquecíveis ao longo da graduação. Sua amizade e apoio foram essenciais para superar os obstáculos e celebrar as conquistas.

E, finalmente, agradeço a mim mesmo pela dedicação, esforço e determinação em concluir este trabalho. Cada desafio superado e cada conquista alcançada foram frutos do meu comprometimento e perseverança.

Esta dedicação é um reconhecimento a todos aqueles que contribuíram direta ou indiretamente para o meu crescimento e sucesso acadêmico. Meu profundo agradecimento a todos vocês.

## AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos à Universidade Federal do Rio de Janeiro (UFRJ) por me proporcionar uma formação acadêmica sólida e por oferecer recursos e suporte ao longo da minha jornada acadêmica. A UFRJ é uma instituição de renome que me permitiu explorar e aprofundar meus conhecimentos na área de mecânica e computação.

Em especial, gostaria de agradecer à minha orientadora, Professora Janaína Gomide, pela dedicação, orientação e apoio durante todo o desenvolvimento deste trabalho. Sua expertise, orientação precisa e valiosos insights foram fundamentais para o sucesso deste projeto. Agradeço por sua paciência, motivação e pela oportunidade de aprender com uma profissional tão experiente e dedicada.

Também gostaria de expressar minha gratidão a todos os meus professores ao longo da graduação. Seus ensinamentos, orientações e feedbacks contribuíram significativamente para o meu crescimento acadêmico e profissional. Agradeço por compartilharem seu conhecimento e experiência, e por terem me inspirado a buscar a excelência em meus estudos.

Por fim, gostaria de agradecer aos meus familiares e amigos por seu apoio constante e encorajamento ao longo desta jornada. Suas palavras de incentivo e compreensão foram essenciais para superar desafios e alcançar meus objetivos acadêmicos.

A todos que contribuíram de alguma forma para o desenvolvimento deste trabalho, meu mais sincero agradecimento. Sem o apoio e suporte de cada um de vocês, este trabalho não seria possível. Sou imensamente grato por fazer parte de uma comunidade acadêmica tão enriquecedora.

Obrigado!

*“Ninguém é tão grande que não possa aprender,  
nem tão pequeno que não possa ensinar.”*

**Esopo**

## RESUMO

A crescente preocupação com a sustentabilidade e a busca por fontes de energia limpas e renováveis impulsionam o desenvolvimento de tecnologias voltadas para o aproveitamento dos recursos naturais. A energia solar fornece uma fonte de energia alternativa limpa e abundante para atender às necessidades energéticas. O objetivo desse trabalho é fornecer previsões da geração de energia fotovoltaica no Brasil aplicando modelos de aprendizado de máquina. A metodologia adotada envolveu a coleta e pré-processamento de dados climáticos e de geração solar, a seleção e implementação dos modelos de aprendizado de máquina, bem como a avaliação do desempenho dos modelos por meio de métricas como acurácia, erro médio absoluto (MAE) e erro médio quadrático (RMSE). Os resultados obtidos demonstraram que os modelos de árvores de decisão e florestas aleatórias apresentam um desempenho promissor na previsão de geração solar. Esses modelos mostraram-se capazes de fornecer previsões precisas, com um coeficiente de determinação superior a 60%, e uma boa concordância entre as previsões e as leituras reais de geração solar. A utilização desses modelos de aprendizado de máquina na previsão de geração solar traz vantagens significativas em relação às abordagens tradicionais, permitindo a captura de relações complexas entre os dados climáticos e a geração solar, bem como a adaptação a mudanças e a incorporação de novos dados ao longo do tempo. Este estudo contribui para a área de previsão de geração solar, fornecendo insights valiosos sobre o uso de modelos de aprendizado de máquina nesse contexto. As previsões precisas obtidas por meio desses modelos são essenciais para o planejamento estratégico, operação de usinas solares e otimização do sistema elétrico, contribuindo para a transição energética e a sustentabilidade. Como sugestões para pesquisas futuras, destaca-se a investigação de outros algoritmos de aprendizado de máquina, a incorporação de dados adicionais e a consideração de fatores socioeconômicos e políticas energéticas na modelagem. Essas abordagens podem aprimorar ainda mais as previsões de geração solar e contribuir para um sistema energético mais eficiente e sustentável.

**Palavras-chave:** aprendizado de máquina, previsão de geração solar, árvores de decisão, florestas aleatórias, energia renovável.



## ABSTRACT

This work addresses the topic "Machine learning models for solar generation forecasting using weather data" and investigates the performance of decision tree models and random forests in this task. The aim of the study is to provide accurate and reliable forecasts of solar generation, aiming at the efficient integration of this renewable energy source into the power grid. The methodology adopted involved the collection and pre-processing of climate and solar generation data, the selection and implementation of the machine learning models, as well as the evaluation of the models' performance through metrics such as accuracy, mean absolute error (MAE) and root mean square error (RMSE). The results obtained showed that the decision tree and random forest models have a promising performance in solar generation forecasting. These models were shown to provide accurate predictions, with a coefficient of determination higher than 60%, and a good agreement between the predictions and the actual solar generation readings. The use of these machine learning models in solar generation forecasting brings significant advantages over traditional approaches, allowing the capture of complex relationships between climate data and solar generation, as well as adaptation to changes and the incorporation of new data over time. This study contributes to the field of solar generation forecasting by providing valuable insights into the use of machine learning models in this context. The accurate forecasts obtained through these models are essential for strategic planning, solar plant operation and power system optimization, contributing to the energy transition and sustainability. Suggestions for future research include the investigation of other machine learning algorithms, the incorporation of additional data, and the consideration of socioeconomic factors and energy policies in the modeling. These approaches can further improve solar generation forecasts and contribute to a more efficient and sustainable energy system.

**Keywords:** machine learning, solar generation forecasting, decision trees, random forests, renewable energy.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Usina solar fotovoltaica . . . . .	20
Figura 2 – Demonstração do efeito fotovoltaico . . . . .	21
Figura 3 – Painel solar de perovskita . . . . .	24
Figura 4 – Exemplo de árvore de decisão . . . . .	31
Figura 5 – Exemplo de floresta aleatória . . . . .	32
Figura 6 – Processo seguido pelo trabalho . . . . .	35
Figura 7 – Mapa de distribuição das usinas de geração fotovoltaica pelo país . . . . .	39
Figura 8 – Mapa de distribuição das estações meteorológicas do Brasil . . . . .	41
Figura 9 – Ilustração da conexão entre os bancos de dados . . . . .	42
Figura 10 – Gráfico de seno e cosseno para horas do dia . . . . .	43
Figura 11 – Gráfico da divisão de treino e teste das usinas . . . . .	45
Figura 12 – Gráfico da divisão de treino e teste das usinas . . . . .	46
Figura 13 – Gráfico CCDF do $R^2$ dos modelos de árvores de decisão e florestas aleatórias . . . . .	47
Figura 14 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro I . . . . .	48
Figura 15 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro I . . . . .	49
Figura 16 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro II . . . . .	50
Figura 17 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro II . . . . .	51
Figura 18 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro III . . . . .	51
Figura 19 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro III . . . . .	52

## LISTA DE TABELAS

Tabela 1 – Dados referentes a localização da estação climática . . . . .	37
Tabela 2 – Dados meteorológicos referentes a estação climática . . . . .	37
Tabela 3 – Dados de geração, código CEG e tipo de fonte energética . . . . .	37
Tabela 4 – Código CEG e coordenadas das usinas . . . . .	38
Tabela 5 – Nome das usinas e código CEG . . . . .	38
Tabela 6 – Data de início, fim e grupo das usinas . . . . .	40
Tabela 7 – Exemplos de linhas do banco de dados unificado das localizações das estações climáticas . . . . .	40
Tabela 8 – Banco de dados antes do processamento . . . . .	41
Tabela 9 – Banco de dados após da normalização . . . . .	43
Tabela 10 – Resultados dos modelos de Árvores de decisão . . . . .	60
Tabela 11 – Resultados dos modelos de florestas aleatórias . . . . .	61

## LISTA DE ABREVIATURAS E SIGLAS

CEG	Código de Identificação do Empreendimento de Geração de Energia Elétrica
ANEEL	Agência Nacional de Energia Elétrica
ABNT	Associação Brasileira de Normas Técnicas
INMET	Instituto Nacional de Meteorologia
CCEE	Câmara de Comercialização de Energia Elétrica
ONS	Operador Nacional do Sistema Elétrico
IEEE	Instituto de engenheiros eletricitas e eletrônicos

## LISTA DE SÍMBOLOS

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence
$\$$	subcampo

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>14</b>
1.1	OBJETIVOS . . . . .	15
1.1.1	Objetivo geral . . . . .	15
1.1.2	Objetivos específicos . . . . .	15
1.2	ORGANIZAÇÃO DA MONOGRAFIA . . . . .	15
<b>2</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>16</b>
2.1	ESTRATÉGIAS DE APRENDIZADO POR TRANSFERÊNCIA PARA PREVISÃO DE GERAÇÃO SOLAR NA ESCASSEZ DE DADOS . .	16
2.2	MODELO DE REDES NEURAIS ENCADEADAS PARA PREVISÃO DE GERAÇÃO SOLAR . . . . .	16
2.3	ESTIMATIVA DE POTENCIAL ENERGÉTICO UTILIZANDO AL- GORITMOS DE PREVISÃO DE DADOS CLIMÁTICOS . . . . .	17
2.4	MODELOS DE VETOR DE SUPORTE DE REGRESSÃO, UTILI- ZADOS EM CONJUNTO POR UM MODELO DE ÁRVORES ALE- ATÓRIAS PARA PREVISÃO DE GERAÇÃO SOLAR . . . . .	17
2.5	PREVISÃO DE GERAÇÃO FOTOVOLTAICA A PARTIR DE DA- DOS METEOROLÓGICOS UTILIZANDO REDE LSTM . . . . .	18
<b>3</b>	<b>REFERENCIAL TEÓRICO . . . . .</b>	<b>19</b>
3.1	ENERGIA SOLAR E GERAÇÃO FOTOVOLTAICA . . . . .	19
3.1.1	Introdução à Energia Solar . . . . .	19
3.1.2	Princípios da Geração Fotovoltaica . . . . .	19
3.1.3	Vantagens e Benefícios da Energia Solar . . . . .	21
3.1.4	Desafios e Limitações da Energia Solar . . . . .	22
3.1.5	Tendências e Avanços na Energia Solar . . . . .	23
3.2	PREVISÃO DE GERAÇÃO SOLAR . . . . .	24
3.2.1	Introdução à Previsão de Geração Solar . . . . .	24
3.2.2	Desafios e Complexidades da Previsão de Geração Solar . . .	25
3.2.3	Métricas e Métodos de Avaliação de Previsão . . . . .	26
3.2.4	Técnicas de Previsão de Geração Solar . . . . .	27
3.2.5	Aplicações e Casos de Uso da Previsão de Geração Solar . . .	28
3.3	APRENDIZADO DE MÁQUINA . . . . .	29
3.3.1	Introdução ao Aprendizado de Máquina . . . . .	29
3.3.2	Modelos de Árvores de Decisão . . . . .	30
3.3.3	Florestas Aleatórias ( <i>Random Forests</i> ) . . . . .	31

3.3.4	Pré-processamento de Dados para Modelos de Árvores de Decisão e Florestas Aleatórias . . . . .	33
3.3.5	Avaliação e Interpretação dos Modelos . . . . .	34
4	<b>MATERIAIS E MÉTODOS</b> . . . . .	<b>35</b>
4.1	DEFINIÇÃO DO PROBLEMA E SELEÇÃO DOS DADOS . . . . .	35
4.2	PRÉ-PROCESSAMENTO DOS DADOS E UNIÃO DOS DATASETS	38
4.2.1	<b>Conectando os bancos de dados</b> . . . . .	<b>38</b>
4.2.2	<b>Pré-Processamento dos dados</b> . . . . .	<b>41</b>
4.3	ESCOLHA DO ALGORITMO/MODELO DE APRENDIZADO DE MÁQUINA . . . . .	44
4.4	TREINAMENTO DO MODELO . . . . .	45
4.5	AVALIAÇÃO DO MODELO . . . . .	46
4.6	ANÁLISE DOS RESULTADOS . . . . .	46
4.6.1	<b>Resultados encontrados</b> . . . . .	<b>47</b>
4.6.2	<b>Comparação com trabalhos relacionados</b> . . . . .	<b>48</b>
4.7	DISCUSSÃO DOS RESULTADOS E LIMITAÇÕES . . . . .	48
5	<b>CONCLUSÃO</b> . . . . .	<b>53</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>54</b>
	<b>APÊNDICE A – TABELAS DOS VALORES DE RESULTADOS DAS USINAS</b> . . . . .	<b>59</b>

## 1 INTRODUÇÃO

A crescente preocupação com a sustentabilidade e a busca por fontes de energia limpas e renováveis impulsionam o desenvolvimento de tecnologias voltadas para o aproveitamento dos recursos naturais. Sob tais circunstâncias, a energia solar surgiu como uma solução promissora que fornece uma fonte de energia alternativa limpa e abundante para atender às necessidades energéticas (IRENA, Renewable energy statistics 2023). No entanto, a geração de energia solar varia com o tempo e as estações, portanto métodos eficientes devem ser usados para prever seu rendimento. É nesse contexto que o aprendizado de máquina ganha destaque, permitindo a criação de modelos capazes de analisar dados climáticos e deles extrair informações valiosas, ajudando a prever com precisão a geração de energia solar.

O aprendizado de máquina é um subcampo da inteligência artificial que busca desenvolver algoritmos e técnicas que permitem que os computadores aprendam e tomem decisões com base em dados sem serem explicitamente programados (MURPHY, 2012). Os avanços na tecnologia de aprendizado de máquina possibilitaram a análise de grandes volumes de dados climáticos e a extração de padrões complexos, ajudando a prever a geração de energia solar com mais precisão e confiabilidade. Algoritmos de aprendizado de máquina, como regressão linear, redes neurais artificiais e árvores de decisão, têm sido amplamente usados para prever a geração de energia solar com base em dados climáticos, como radiação solar, temperatura, velocidade do vento e umidade (FOROOTAN et al., 2022).

Além disso, é importante contextualizar o histórico de geração solar no Brasil em relação a outras fontes energéticas. Nos últimos anos, tem havido um crescimento significativo na capacidade instalada de usinas solares no país, impulsionado por políticas governamentais de incentivo à geração de energia renovável (ANEEL, 2023b). Apesar desse progresso, a geração solar ainda representa uma parcela menor em comparação a outras fontes energéticas, como hidrelétrica e termelétrica. No entanto, o potencial de expansão da energia solar é imenso, considerando o vasto território brasileiro e sua alta incidência de radiação solar ao longo do ano (PEREIRA et al., 2017).

A integração entre modelos de energia solar e aprendizado de máquina oferece oportunidades significativas para desenvolver sistemas de previsão de geração solar mais eficientes e precisos. Combinando dados climáticos e algoritmos de aprendizado de máquina, é possível obter modelos capazes de prever a quantidade de energia solar produzida em condições atmosféricas específicas (IRENA, Renewable energy statistics 2023). Essas previsões desempenham um papel crucial no planejamento e gerenciamento de sistemas de energia solar, permitindo a otimização da distribuição e armazenamento de energia, além de ajudar a melhorar a confiabilidade e a estabilidade da geração de energia solar. Portanto,



este esforço de pesquisa visa explorar diferentes modelos de aprendizado de máquina para prever a geração de energia solar usando dados climáticos, com o objetivo de promover a eficiência e a expansão das energias renováveis (MURPHY, 2012).

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

O objetivo geral deste trabalho é avaliar a utilização de modelos de aprendizado de máquina para previsão de geração fotovoltaica no Brasil.

### 1.1.2 Objetivos específicos

1. Realizar a extração dos dados necessários para a construção dos modelos
2. Realizar o tratamento adequado dos dados
3. Analisar os modelos e escolher os modelos mais adequados para a situação
4. Treinar e avaliar os modelos gerados
5. Comparar os resultados obtidos com a literatura para comprovar o atingimento do objetivo geral

## 1.2 ORGANIZAÇÃO DA MONOGRAFIA

Além do primeiro capítulo que se trata da introdução ao tema, a monografia está organizada da seguinte maneira: O 2º Capítulo aborda os trabalhos relacionados e os principais pontos que influenciaram esse trabalho. No 3º Capítulo será abordado todo o referencial teórico necessário para a construção desse trabalho. O 4º Capítulo retrata todo o processo de elaboração dos modelos e análise de dados e resultados. O 5º Capítulo conclui e traz reflexões a respeito das dificuldades e oportunidades encontradas durante o trabalho para trabalhos futuros.

## 2 TRABALHOS RELACIONADOS

### 2.1 ESTRATÉGIAS DE APRENDIZADO POR TRANSFERÊNCIA PARA PREVISÃO DE GERAÇÃO SOLAR NA ESCASSEZ DE DADOS

O primeiro artigo que será citado é um artigo sobre estratégias de *transfer learning* para previsão de geração solar sob escassez de dados. O artigo escrito por (SARMAS et al., 2022), foi publicado na renomada revista Nature e é bem pertinente a construção desse trabalho.

Esse estudo faz uso do modelo de aprendizado LSTM (*Long Short-Term Memory*), com três estratégias de *transfer learning* para prever com precisão a produção de usinas geração solar fotovoltaica. Esse estudo foi realizado com dados referentes a usinas de diferentes cidades de Portugal(Lisboa, Faro, Setubal e Braga).

O estudo fez uma abordagem diferente no pré-processamento dos dados, aplicando uma transformação *One-Hot* para o processamento dos meses e uma transformação seno/cosseno para as horas adicionando esses dois como features na entrada dos dados para o modelo.

O estudo concluiu que a aplicação de *transfer learning* para as estações onde havia poucos dados históricos trazia resultados significativos quando comparados a modelos treinados apenas com os dados históricos da estação.

### 2.2 MODELO DE REDES NEURAI ENCADEADAS PARA PREVISÃO DE GERAÇÃO SOLAR

O segundo artigo que foi analisado para o trabalho foi um artigo que analisa um modelo de rede neural encadeada para previsão de geração solar. Escrito por (GAJEK; SCHIENDORFER; REIF, 2019), foi publicado em uma conferência internacional, que ocorreu na Itália em Siena.

Esse artigo utiliza modelos de rede neural encadeada e leva em consideração a adição frequente de novas plantas. Essa modelagem em 2 passos permite uma melhor integração de novas plantas e atinge previsões de alta qualidade comparáveis a modelos de dados climáticos tradicionais.

O estudo encontrou grandes dificuldades na disponibilidade e qualidade dos dados apresentados e portanto, trabalhou com um conjunto limitado de dados, para prever a irradiação solar e a partir de um modelo físico, calcular a geração para cada usina.

Os resultados encontrados mostraram ganhos consideráveis em relação ao modelo de base, e até ganhos em relação a uma modelagem mais tradicional com dados climáticos, e principalmente o tratamento de dados de baixa qualidade.

### 2.3 ESTIMATIVA DE POTENCIAL ENERGÉTICO UTILIZANDO ALGORITMOS DE PREVISÃO DE DADOS CLIMÁTICOS

O próximo artigo analisado foi sobre a estimativa do potencial de geração solar da Nigéria, utilizando modelos de aprendizado de máquina baseados no clima. Foi escrito por (MADUABUCHI et al., 2023), e foi publicado pela editora MDPI em fevereiro de 2023.

O trabalho teve como objetivo a análise de viabilidade de instalação de plantas solares na Nigéria, uma vez que não há investimento significativo nessa área no país. E isso foi feito através da previsão de todas as variáveis analisadas, baseados nos dados coletados pela NASA.

O artigo gerou os modelos com base nas características: Temperatura, umidade relativa, pressão atmosférica, velocidade do vento, precipitação e incidência solar direta e indireta. O tipo de modelo utilizado no estudo foi rede neural e foi usado MSE para avaliação da performance do modelo.

As conclusões encontraram a quantidade ideal de neurônios para o funcionamento adequado da rede neural e além disso demonstrou a viabilidade na implementação de modelos similares na análise de viabilidade de construção de usinas solares na Nigéria.

### 2.4 MODELOS DE VETOR DE SUPORTE DE REGRESSÃO, UTILIZADOS EM CONJUNTO POR UM MODELO DE ÁRVORES ALEATÓRIAS PARA PREVISÃO DE GERAÇÃO SOLAR

O quarto artigo relacionado a ser analisado foi sobre modelos de florestas aleatórias para *ensemble learning* de modelos de regressão de vetor de suporte aplicado a previsão solar. O artigo escrito por (ABUELLA; CHOWDHURY, 2017) foi publicado em uma conferência da IEEE em 2017.

O objetivo deste estudo foi comparar o método de *ensemble learning* utilizando um modelo de florestas aleatórias, comparado a outras formas de combinação de modelos e combinação de dados climáticos, para avaliação da efetividade do método.

O trabalho foi realizado com 2 anos de dados (Abril de 2012 até Maio de 2014), de uma usina localizada na Austrália. Foram consideradas 14 características climáticas diferentes, com método de normalização *Min-Max*.

As conclusões mostraram que de fato há um ganho significativo no emprego de modelos de florestas aleatórias para *ensemble learning* de modelos de geração solar, e que o cálculo de médias dos valores calculados pelos modelos, não deve ser empregado pois não traz valor algum comparativamente.

## 2.5 PREVISÃO DE GERAÇÃO FOTOVOLTAICA A PARTIR DE DADOS METEOROLÓGICOS UTILIZANDO REDE LSTM

O próximo estudo relacionado que foi analisado é sobre a utilização de uma rede neural LSTM (*Long Short-Term Memory*) para a previsão da tensão e corrente gerada pela usina de geração solar. Foi escrito por (BASTOS et al., 2020) e publicado no Congresso Brasileiro de Energia Solar.

O artigo se propôs a avaliar a performance de uma rede neural LSTM, com aprendizado baseado em dados climáticos obtidos por uma estação solarimétrica de baixo custo. O estudo em questão usou a radiação solar, temperatura dos painéis, temperatura ambiente, umidade, velocidade do vento e quantidade de chuva (precipitação).

A base de dados do artigo foi criada com base em 30 dias de dados gerados a cada segundo, as linhas com dados faltantes foram eliminadas e com isso o total de dados caiu para 24 dias de dados. A métrica utilizada para avaliação da qualidade do modelo de predição foi o erro absoluto.

Por fim as conclusões do artigo mostram que o modelo convergiu, e que os modelos de LSTM demonstram serem promissores na previsão de geração solar e que mesmo com uma estação solarimétrica de baixo custo foi possível obter os dados necessários para uma boa modelagem.

### 3 REFERENCIAL TEÓRICO

#### 3.1 ENERGIA SOLAR E GERAÇÃO FOTOVOLTAICA

##### 3.1.1 Introdução à Energia Solar

A geração solar é conhecida por ser uma fonte de energia limpa e sustentável. Com o aproveitamento da radiação solar, é possível gerar eletricidade e calor de maneira ecológica, o que reduz de maneira considerável a emissão de gases do efeito estufa. De acordo com a Agência Nacional de Energia Elétrica (ANEEL), a energia solar tem um papel fundamental na diversificação da matriz energética brasileira e vem ganhando destaque nas últimas décadas. A (ANEEL, 2023b) também destaca que a geração de energia solar fotovoltaica vem sendo uma boa opção para o modelo de geração distribuída, que traz benefícios tanto para os usuários quanto para o sistema de transmissão como um todo. Assim, a importância da energia solar tem crescido ao longo dos anos, impulsionada pela preocupação com as mudanças climáticas e pela busca de alternativas sustentáveis e renováveis. Em seu Balanço Energético Nacional de 2022, a Empresa de Pesquisa Energética (EPE) destacou o grande potencial da energia solar no Brasil devido à alta incidência solar do país. A (EPE, 2023) destacou ainda que a energia solar pode ajudar a diversificar a matriz energética do país e reduzir a dependência de fontes não renováveis de energia. Além disso, a energia solar desempenha um papel fundamental na geração distribuída, permitindo que residências e empresas gerem sua própria energia por meio de sistemas fotovoltaicos montados em telhados.

A energia solar representa uma solução promissora para a transição para um sistema energético mais sustentável. De acordo com a Agência Internacional de Energia (IEA), a energia solar experimentou um crescimento significativo globalmente, impulsionado por avanços tecnológicos, custos de instalação mais baixos e incentivos governamentais. A energia solar fotovoltaica tem sido amplamente adotada em diversos países, contribuindo significativamente para a geração de energia renovável. De acordo com a (IEA, 2021), espera-se que a energia solar desempenhe um papel cada vez mais importante no futuro, à medida que os países buscam reduzir sua dependência de combustíveis fósseis e mitigar os efeitos das mudanças climáticas.

##### 3.1.2 Princípios da Geração Fotovoltaica

Segundo (ANEEL, 2023b) a geração de energia fotovoltaica é baseada em princípios fundamentais que envolvem a conversão direta de energia solar em eletricidade por meio de painéis solares. Esses painéis solares consistem em células fotovoltaicas que usam o efeito fotovoltaico para converter a luz solar em eletricidade, como ilustrado na figura 2. O

Figura 1 – Usina solar fotovoltaica



Figura retirada de (MÓDULOENERGIA, 2022)

efeito fotovoltaico ocorre devido às propriedades semicondutoras dos materiais presentes nas células solares. Quando a luz solar atinge os semicondutores em um painel solar, os elétrons são excitados e criam uma corrente contínua. Este processo, conhecido como o princípio fundamental da geração de energia fotovoltaica, tem se mostrado uma solução de geração de energia eficiente e sustentável. Um painel solar fotovoltaico consiste em várias células solares interligadas para formar um módulo fotovoltaico. De acordo com (BÜHLER, 2011), cada célula solar é feita de camadas de material semicondutor, como o silício, que possuem diferentes propriedades elétricas. As células solares de silício cristalino são as mais comuns e amplamente utilizadas devido à sua eficiência e disponibilidade comercial. Essas células empregam uma estrutura em camadas de junção p-n, onde a camada n é rica em elétrons e a camada p é deficiente em elétrons. Essa estrutura pode gerar um campo elétrico que guia os elétrons excitados pela luz solar para gerar uma corrente elétrica utilizável. (MOCELIN, 2014) adiciona que os semicondutores desempenham um papel crucial na conversão da luz solar em eletricidade nos painéis fotovoltaicos. O silício, amplamente utilizado na fabricação de células solares, é um semicondutor intrínseco, o que significa que ele possui elétrons fracamente ligados em sua estrutura cristalina. Quando a luz solar atinge a célula solar de silício, fótons são absorvidos pelos elétrons do material semicondutor, fornecendo energia suficiente para que eles escapem de suas órbitas atômicas e se tornem elétrons livres. Esses elétrons livres são então capturados pelo campo elétrico criado na junção p-n, resultando em uma corrente elétrica que pode

Figura 2 – Demonstração do efeito fotovoltaico

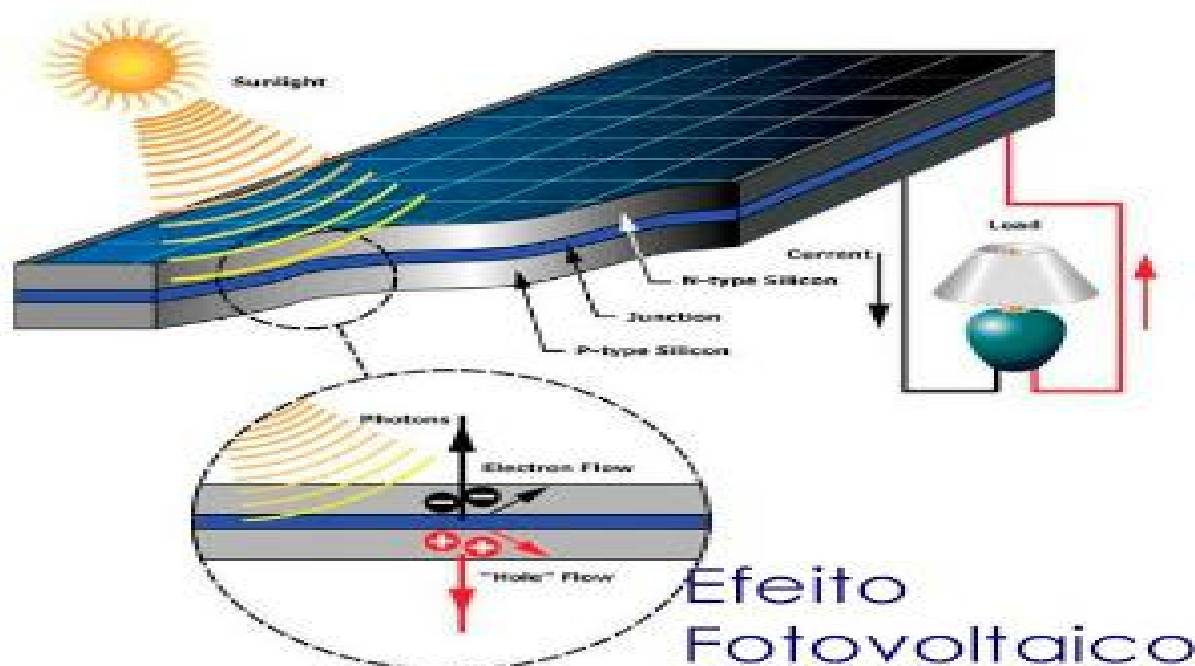


Figura retirada de (COMINTECH, 2021)

ser aproveitada para alimentar dispositivos elétricos.

A geração fotovoltaica desempenha um papel importante na transição para fontes de energia mais sustentáveis. As células solares utilizadas nos painéis fotovoltaicos são projetadas para otimizar a conversão da luz solar em eletricidade, permitindo a geração de energia limpa e renovável. Essa tecnologia como citado pela (ONS, 2023a) está em constante evolução, com pesquisas e desenvolvimentos em andamento para melhorar a eficiência das células solares, reduzir os custos de produção e aumentar a vida útil dos painéis solares. A crescente adoção da geração fotovoltaica tem contribuído significativamente para a diversificação da matriz energética global e para a redução das emissões de gases de efeito estufa.

### 3.1.3 Vantagens e Benefícios da Energia Solar

A energia solar tem muitas vantagens e benefícios significativos em relação às fontes de energia tradicionais. Segundo a (ANEEL, 2023b) uma das principais vantagens da energia solar é que ela reduz o impacto no meio ambiente. Ao contrário dos combustíveis fósseis, a energia solar não emite gases de efeito estufa durante a operação, contribuindo diretamente para a redução da emissão de poluentes e combate às mudanças climáticas. Além disso, a energia solar é uma fonte de energia renovável, inesgotável, inesgotável e amplamente disponível, o que a torna uma opção atraente para a transição para um

sistema energético mais sustentável. Conforme a (IEA, 2021) mostrou, a adoção da energia solar também pode trazer vantagens significativas em termos de autonomia energética e independência de combustíveis fósseis. Ao investir em sistemas de energia solar, indivíduos e comunidades podem gerar sua própria eletricidade, reduzindo sua dependência de redes centralizadas e a volatilidade dos preços dos combustíveis fósseis. Isso proporciona maior segurança energética e a capacidade de atender a demanda de eletricidade mesmo em locais remotos ou quando o fornecimento de energia convencional é interrompido.

Além dos benefícios ambientais e de autonomia, a adoção da energia solar traz vantagens econômicas significativas. O custo dos sistemas solares fotovoltaicos diminuiu significativamente nas últimas décadas, tornando-os cada vez mais acessíveis e economicamente viáveis. Além disso, segundo (DANTAS; POMPERMAYER, 2018), a energia solar pode reduzir as contas de eletricidade a longo prazo, porque os sistemas solares podem gerar eletricidade gratuitamente após o investimento inicial. Além disso, a energia solar contribui para a geração de empregos no setor de energia renovável, crescimento econômico e geração de renda.

A energia solar é uma opção sustentável e rentável para gerar eletricidade. As suas propriedades limpas e renováveis, aliadas aos benefícios ambientais, à autonomia energética e aos ganhos econômicos associados, tornam-no numa solução atrativa para a transição para um sistema energético mais sustentável e resiliente. Segundo (REN23, 2023), a crescente adoção da energia solar em todo o mundo reflete uma crescente conscientização sobre a importância da sustentabilidade e a necessidade de reduzir a dependência de combustíveis fósseis.

#### **3.1.4 Desafios e Limitações da Energia Solar**

Embora a energia solar seja uma fonte de energia promissora e sustentável, seu uso enfrenta importantes desafios e limitações. Segundo a (IEA, 2021), um dos desafios mais significativos é a variação diária e sazonal na disponibilidade solar. A geração de energia solar depende da quantidade de luz solar disponível em um determinado momento, que varia com o dia e as estações. Isso significa que a geração de energia solar não é constante e mecanismos de compensação, como armazenamento de energia, são necessários para garantir o fornecimento contínuo de energia.

Outro desafio é a dependência das condições climáticas. A energia solar é capturada de forma mais eficiente em áreas de alta radiação solar, como os trópicos e desertos. Em áreas com clima nublado ou baixa radiação solar, a produção de energia solar pode ser limitada. Por exemplo, segundo (REN23, 2023), em altas latitudes ou em locais frequentemente cobertos por nuvens, a eficiência dos sistemas de energia solar será reduzida. Isso requer uma consideração cuidadosa da localização e do potencial solar ao planejar a implementação de sistemas fotovoltaicos.



A necessidade de armazenamento de energia é outro desafio enfrentado pela energia solar. A geração de energia solar é intermitente porque a disponibilidade de luz solar muda com o tempo. Para fornecer eletricidade na ausência de luz solar, é necessário armazenar o excesso de energia gerada durante os períodos de maior irradiância. Apesar dos avanços nas tecnologias de armazenamento de energia, como as baterias, ainda existem desafios a serem superados em termos de eficiência, custo e escala desses sistemas (CASTRO, 2020). Além dos desafios técnicos, a viabilidade econômica da energia solar também depende de políticas e incentivos governamentais. Segundo (ESMAP, 2022), o apoio político, como tarifas de alimentação e esquemas de incentivo, desempenha um papel crucial na promoção do desenvolvimento solar. A falta de uma política consistente de longo prazo pode afetar negativamente o investimento em sistemas solares e a ampla adoção da tecnologia. No entanto, deve-se notar que, à medida que o custo da energia solar continua caindo, a viabilidade econômica dessa fonte de energia aumenta.

### **3.1.5 Tendências e Avanços na Energia Solar**

Nos últimos anos, avanços significativos em novos materiais e tecnologias fotovoltaicas impulsionaram o desenvolvimento da energia solar. Os pesquisadores estão explorando materiais como células solares de perovskita (Figura 3) com alta eficiência de conversão e baixos custos de produção (SOLARMAGAZINE, 2022). Além disso, o desenvolvimento de tecnologias de silício cristalino, como células solares de alta eficiência e módulos solares de dupla face, melhorou as capacidades de geração de energia solar. Esses avanços tecnológicos estão impulsionando o crescimento da capacidade solar instalada globalmente, que segundo a (IEA, 2021) aumentará quase 45% até 2027. Outra tendência importante em energia solar é a integração de sistemas de armazenamento de energia. O armazenamento de energia permite que a eletricidade gerada pelo sol seja armazenada para uso posterior, tornando a geração de energia solar mais confiável e flexível. A tecnologia das baterias, especialmente as baterias de íon-lítio, tem se mostrado crucial neste processo. Além disso, segundo (BLOOMBERGNEF, 2022), o desenvolvimento de sistemas de armazenamento de energia em larga escala, como usinas de armazenamento de energia, está estabilizando a rede e maximizando o uso da energia solar. Estima-se que até 2030, a capacidade global de armazenamento de energia crescerá a uma taxa anual de cerca de 23%.

A implementação de projetos de grande escala tornou-se uma tendência significativa no desenvolvimento da energia solar. Em todo o mundo, estão surgindo usinas e parques solares de grande escala, capazes de gerar energia em larga escala. Esses projetos aproveitam a escala para reduzir o custo de instalação e produção de eletricidade solar. Por exemplo, segundo a (NASA, 2022), a usina de energia solar de Bahdla, na Índia, é uma das maiores usinas de energia solar do mundo, com capacidade de 2,2 GW. A implementação desses projetos ajuda a aumentar a capacidade instalada de energia solar e fortalecer sua posição como fonte de energia limpa e sustentável.

Figura 3 – Painel solar de perovskita

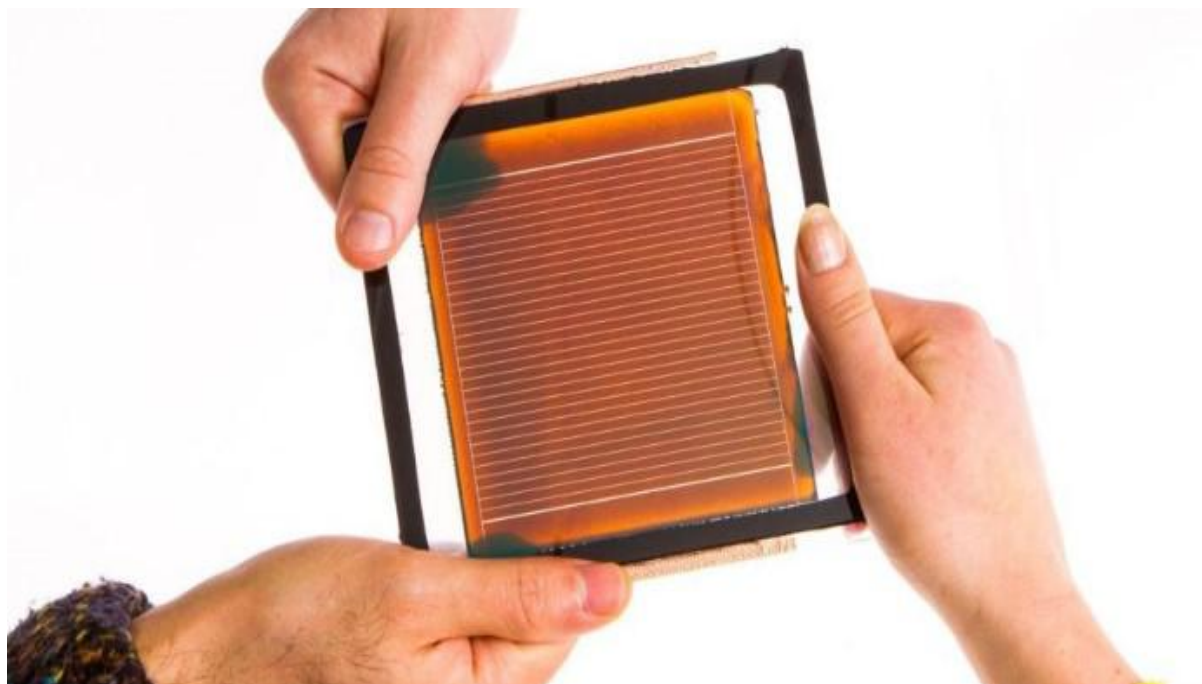


Figura retirada de (PORTALENERGIA, 2021)

## 3.2 PREVISÃO DE GERAÇÃO SOLAR

### 3.2.1 Introdução à Previsão de Geração Solar

A previsão da geração solar desempenha um papel vital na integração eficiente da energia solar na rede. Devido à variabilidade e intermitência da energia solar, é necessário prever a quantidade de energia produzida em um determinado período de tempo. A previsão da geração solar permite o planejamento adequado da demanda, a coordenação energética e a otimização de todo o sistema energético. Segundo a (IEA, 2021) estima-se que os serviços de previsão de energia solar tenham uma taxa de crescimento anual de 29% até 2027 devido à necessidade de lidar com a variabilidade e incerteza da geração de energia solar.

A previsão precisa da geração de energia solar pode trazer benefícios significativos para a operação de parques solares. Com previsões confiáveis, é possível planejar as operações da planta, ajustar cronogramas de manutenção e otimizar o uso de recursos. Além disso, previsões precisas podem permitir a integração eficiente da energia solar com outras fontes de energia, como eólica e hidrelétrica, contribuindo para a estabilidade e confiabilidade do sistema de energia. O estudo de (LIMA, 2016) demonstrou que a previsão de geração energética de fontes instáveis incorreta aumentou o custo de geração em 2,20 euros por MWh gerado, desse modo a previsão correta pode reduzir os custos operacionais de usinas de energia solar.

Além dos benefícios operacionais, a previsão precisa da geração de energia solar é fundamental para decisões estratégicas na expansão da capacidade solar e no planejamento do sistema de energia. Com base em previsões confiáveis, os planejadores podem fazer ajustes apropriados na capacidade solar instalada, levando em consideração a demanda futura e a disponibilidade de recursos. Além disso, conforme (LIMA, 2016), previsões precisas podem levar a uma melhor integração da energia solar nos mercados de energia, facilitando a negociação de contratos de longo prazo e o desenvolvimento de políticas de incentivo.

Avanços na tecnologia de previsão e acesso a dados meteorológicos mais precisos levaram a melhorias na precisão das previsões de energia solar. Segundo, (BASTOS et al., 2020), modelos matemáticos avançados combinados com técnicas de aprendizado de máquina são capazes de analisar fatores como radiação solar, cobertura de nuvens e padrões climáticos para fornecer previsões mais confiáveis. Essas tecnologias estão cada vez mais disponíveis e amplamente utilizadas na indústria solar para melhorar a eficiência e a confiabilidade da geração de energia solar.

### 3.2.2 Desafios e Complexidades da Previsão de Geração Solar

O aprendizado de máquina é o ramo da inteligência artificial que visa desenvolver algoritmos que podem aprender com os dados para tomar decisões e fazer previsões. De acordo com (JAMES et al., 2017), existem duas abordagens principais para aprendizado de máquina: aprendizado supervisionado e não supervisionado. No aprendizado supervisionado, um algoritmo é treinado usando um conjunto de dados rotulado onde cada exemplo tem a resposta esperada. No aprendizado não supervisionado, os algoritmos exploram padrões nos dados sem rotulagem prévia. Esses princípios são a base para o desenvolvimento de modelos de aprendizado de máquina que podem ser aplicados a uma ampla gama de problemas.

No aprendizado de máquina, há vários métodos altamente populares. Por exemplo, o aprendizado por regressão é um procedimento de análise de dados que busca encontrar uma relação de causa e efeito entre as variáveis de entrada e de saída. Segundo o trabalho escrito por (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), as redes neurais artificiais são modelos que se originam do comportamento do cérebro humano e que são capazes de aprender e se tornarem mais generalizadas a partir de uma grande quantidade de informação. Uma árvore de decisão é um dispositivo de navegação que representa uma sucessão de decisões possíveis e suas consequências. Os modelos de série temporal são usados em conjunto com dados temporais para fazer projeções sobre o futuro baseadas em características da série anterior.

O processo de aprendizado de máquina envolve várias etapas. De acordo com (MÜLLER; GUIDO, 2017), o pré-processamento de dados é uma etapa importante na preparação de dados brutos e sua conversão em um formato adequado para análise. Isso inclui

lidar com valores omissos, padronizar e codificar variáveis categóricas e outros procedimentos. A seleção de recursos é outra etapa importante, selecionando as variáveis mais relevantes para o modelo para melhorar sua capacidade de generalização. Em seguida, a avaliação do modelo para medir o desempenho do modelo em dados não vistos usando métricas como exatidão, precisão, *recall* e *F1-score*.

O aprendizado de máquina está se tornando cada vez mais importante e aplicado em muitos campos, como medicina, finanças, marketing e indústria. Por exemplo, segundo (KUMAR et al., 2023), estima-se que a aplicação de algoritmos de aprendizado de máquina na previsão e diagnóstico de tuberculose possa reduzir os custos hospitalares em até 95% dos casos. Além disso, segundo (RESEARCH, 2023), o mercado global de Machine Learning está crescendo e deve registrar uma taxa de crescimento anual composto (*CAGR*) de aproximadamente 34.8% durante o período 2023-2030. Esses dados demonstram a relevância e o impacto do aprendizado de máquina na sociedade atual.

### 3.2.3 Métricas e Métodos de Avaliação de Previsão

Na avaliação de previsões de geração solar, várias métricas são comumente utilizadas para medir a precisão das previsões. O erro médio absoluto (MAE) é uma métrica que calcula a média das diferenças absolutas entre as previsões e os valores reais. Ele fornece uma medida do tamanho médio do erro cometido pelas previsões. Conforme (JAMES et al., 2017), o erro médio quadrático (RMSE) é outra métrica amplamente empregada, que calcula a raiz quadrada da média dos erros quadrados. Essa métrica é útil para medir o desvio padrão dos erros, penalizando erros maiores de forma mais significativa. O coeficiente de correlação também é usado para avaliar a relação linear entre as previsões e os valores reais, fornecendo uma medida da qualidade da previsão em relação ao comportamento geral dos dados.

Além das métricas, os métodos de avaliação também desempenham um papel importante na análise da precisão das previsões de geração solar. Segundo (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), a validação cruzada é um método amplamente utilizado para avaliar o desempenho dos modelos de previsão. Nesse método, o conjunto de dados é dividido em subconjuntos, sendo que cada um é utilizado como conjunto de teste em diferentes iterações. Isso permite uma avaliação mais robusta do modelo, reduzindo o impacto da aleatoriedade nos resultados. Outro método comum é a divisão dos dados em conjuntos de treinamento e teste. Nesse caso, uma parte dos dados é usada para treinar o modelo e a outra parte é reservada para testar o desempenho em dados não vistos. Esse método permite avaliar como o modelo se comporta em situações reais de previsão.

É importante ressaltar que a escolha das métricas e métodos de avaliação deve levar em consideração as características específicas do problema de previsão de geração solar e as metas do projeto. Por exemplo, ao avaliar previsões em diferentes períodos do ano, é importante considerar a sazonalidade da geração solar e ajustar as métricas e métodos de

acordo. Além disso, é necessário considerar os requisitos e restrições específicos do projeto, como a disponibilidade de dados e a frequência de atualização das previsões. Segundo (MÜLLER; GUIDO, 2017), a escolha correta das métricas e métodos garantirá uma avaliação precisa e confiável das previsões de geração solar.

Estudos destacam a importância da escolha correta de métricas e métodos de avaliação para melhorar as previsões de geração solar. Por exemplo, a pesquisa de (AYKO; KESER, 2021) sobre sistemas de previsão de energia solar concluiu que o uso de métricas como MAE e RMSE pode avaliar com precisão a precisão da previsão e ajudar a identificar melhorias no modelo. Além disso, o uso da validação cruzada permite uma avaliação robusta do desempenho do modelo em diferentes cenários.

Dados do mundo real também demonstram a importância do uso correto de métricas e métodos de avaliação para prever a geração de energia solar. Por exemplo, um estudo de (ABUELLA; CHOWDHURY, 2017) de um sistema de previsão de geração solar em um parque solar mostrou que a relação entre os valores previstos e reais pode ser avaliada com precisão usando a raiz do erro quadrático médio. Além disso, a aplicação da validação cruzada confirmou a consistência e confiabilidade do modelo de previsão em diferentes períodos e condições climáticas.

### 3.2.4 Técnicas de Previsão de Geração Solar

Na previsão de energia solar, várias técnicas são aplicadas para melhorar a precisão da previsão. Métodos tradicionais, como modelos estatísticos e séries temporais, têm sido amplamente utilizados. Segundo (JAMES et al., 2017), modelos estatísticos, como regressão linear, capturam a relação linear entre as variáveis de entrada e a saída. Esses modelos são fáceis de implementar e interpretar, mas podem falhar em capturar as complexas relações não lineares presentes nos dados de geração solar. Por outro lado, técnicas de séries temporais como os modelos ARIMA (Autoregressive Integrated Moving Average) são especialmente projetadas para lidar com dados seriais e podem considerar tendências sazonais e temporais de geração de energia solar.

Além dos métodos tradicionais, as técnicas avançadas de aprendizado de máquina também desempenham um papel importante na previsão da geração de energia solar. Por exemplo, como exposto no trabalho de (BASTOS et al., 2020), redes neurais são capazes de aprender padrões complexos e não lineares em dados, tornando-os adequados para prever a geração de energia solar, que está sujeita a muitos fatores inter-relacionadas. Conforme demonstrado no estudo de (ABUELLA; CHOWDHURY, 2017), uma máquina de vetores de suporte (SVM) é um algoritmo que pode mapear dados em um espaço de alta dimensão, permitindo distinguir entre diferentes classes de geração de energia solar. Além disso, conforme descrito por (SRIVASTAVA; TIWARI; GIRI, 2019), os algoritmos de floresta aleatória são capazes de lidar com grandes quantidades de dados e capturar relacionamentos não lineares complexos.

Cada técnica de previsão de geração solar tem vantagens e limitações. Os modelos estatísticos, como descrito por (JAMES et al., 2017), são fáceis de interpretar e implementar, mas podem não ser adequados para capturar relacionamentos não lineares complexos. As técnicas de séries temporais levam em consideração a natureza sequencial dos dados, mas podem não ser capazes de lidar com dados de energia solar com comportamento imprevisível. Por outro lado, como explicitado por (MÜLLER; GUIDO, 2017), técnicas avançadas de aprendizado de máquina são capazes de aprender padrões complexos e se adaptar a diferentes cenários, mas podem exigir mais recursos computacionais e serem mais difíceis de interpretar.

Ao escolher o método de previsão de geração solar mais adequado, como explicitado por (GAJEK; SCHIENDORFER; REIF, 2019), é importante considerar as características dos dados disponíveis, como a disponibilidade de geração solar histórica, a frequência das medições e a presença de sazonalidade e tendências. Além disso, os recursos de computação disponíveis e as capacidades reais de execução do modelo também precisam ser considerados. A seleção adequada da tecnologia de previsão garantirá resultados de previsão de energia solar precisos e confiáveis.

Estudos têm demonstrado a relevância e eficácia das diferentes técnicas de previsão de geração solar. Por exemplo, uma pesquisa comparativa de técnicas de previsão de geração solar feita por (SRIVASTAVA; TIWARI; GIRI, 2019), em um parque solar mostrou que as florestas aleatórias e os modelos arvores de decisão obtiveram desempenho superior em relação a outros modelos, como MARS e CART. Esses resultados foram baseados em métricas como o MAE e o RMSE, indicando a capacidade dessas técnicas em fornecer previsões precisas de geração solar. Além disso, estudos como o de (BASTOS et al., 2020) têm mostrado que o uso de técnicas avançadas de aprendizado de máquina, como redes neurais, tem levado a melhorias significativas na precisão das previsões de geração solar em comparação com abordagens tradicionais.

### **3.2.5 Aplicações e Casos de Uso da Previsão de Geração Solar**

A previsão de energia solar tem muitas aplicações práticas e desempenha um papel fundamental na operação e gerenciamento de usinas de energia solar. Por exemplo, a previsão é usada para otimizar o despacho de energia, determinando a quantidade de energia injetada na rede em diferentes intervalos de tempo. Com previsões precisas, os desvios entre a geração de energia solar real e a geração de energia planejada podem ser minimizados, aumentando a eficiência das usinas de energia solar e reduzindo os custos operacionais. Assim, (KRAAS et al., 2011) mostra que o uso de previsões de geração solar pode reduzir os custos operacionais de usinas solares na Espanha em até 50

A previsão de geração solar também desempenha um papel importante na integração de energia renovável na rede elétrica. Com as previsões, segundo (SARMAS et al., 2022), é possível prever os níveis de geração solar com antecedência e planejar a operação

do sistema elétrico de forma a garantir a estabilidade e a segurança da rede. Além disso, as previsões permitem uma melhor coordenação entre as diferentes fontes de energia renovável, como a energia eólica e a energia hidrelétrica, facilitando a integração eficiente dessas fontes no sistema elétrico.

Casos de sucesso na aplicação de previsões de geração solar têm sido relatados em diferentes partes do mundo. Por exemplo, um estudo realizado por (SRIVASTAVA; TIWARI; GIRI, 2019) na Índia mostrou que a utilização de previsões precisas permite o aumento da eficiência do despacho de energia, resultando em uma redução significativa nos custos operacionais e um aumento na receita anual da usina. Além disso, a utilização de previsões permitiu uma melhor coordenação com a rede elétrica, facilitando a integração de energia solar no sistema elétrico.

### 3.3 APRENDIZADO DE MÁQUINA

#### 3.3.1 Introdução ao Aprendizado de Máquina

O aprendizado de máquina é uma área da inteligência artificial que se dedica ao desenvolvimento de algoritmos e modelos capazes de aprender a partir dos dados. Conforme descrito por (MURPHY, 2012), seu objetivo é extrair padrões e informações úteis dos dados, permitindo que os modelos sejam treinados para fazer previsões e tomar decisões com base nesses padrões identificados.

O processo de construção de um modelo de aprendizado de máquina geralmente segue algumas etapas essenciais. Segundo (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), primeiro, é necessário definir o problema e coletar os dados relevantes para a tarefa em questão. Os dados podem ser estruturados, como tabelas ou bancos de dados, ou não estruturados, como textos ou imagens.

Em seguida, os dados são pré-processados para garantir sua qualidade e prepará-los para o treinamento do modelo. Assim como escrito no livro de (GÉRON, 2022), isso pode incluir a remoção de valores ausentes, a normalização das escalas, a codificação de variáveis categóricas, entre outras técnicas de pré-processamento.

Após o pré-processamento dos dados, é hora de selecionar um algoritmo de aprendizado de máquina adequado para o problema em questão. Ainda segundo (GÉRON, 2022), existem diversos algoritmos disponíveis, cada um com suas próprias características e aplicações. Alguns exemplos comuns são regressão linear, árvores de decisão, redes neurais, entre outros.

O próximo passo é treinar o modelo selecionado usando os dados de treinamento disponíveis. Segundo (JAMES et al., 2017), durante o treinamento, o modelo aprende a partir dos exemplos fornecidos, ajustando seus parâmetros e otimizando sua capacidade de fazer previsões. O desempenho do modelo é avaliado usando métricas apropriadas, como acurácia, precisão ou erro médio.

Por fim, o modelo treinado pode ser utilizado para fazer previsões ou tomar decisões com base nos dados de entrada. Como descrito por (MÜLLER; GUIDO, 2017), essas previsões ou decisões são obtidas aplicando-se o modelo aos dados de teste ou dados futuros. É importante monitorar o desempenho do modelo em tempo real e fazer ajustes conforme necessário para garantir a precisão contínua das previsões.

### 3.3.2 Modelos de Árvores de Decisão

Os modelos de árvores de decisão são algoritmos de aprendizado de máquina amplamente utilizados devido à sua simplicidade e interpretabilidade. Alinhado com o que foi descrito por (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), esses modelos são baseados na construção de uma estrutura hierárquica de decisões, em que cada nó interno representa uma escolha entre diferentes atributos e cada folha representa uma decisão final. A árvore é construída através do processo de divisão de nós, onde cada nó é dividido com base em um atributo que melhor separa os dados de acordo com algum critério.

Existem vários algoritmos utilizados para construir árvores de decisão, como o ID3 (Iterative Dichotomiser 3), o C4.5 e o CART (Classification and Regression Trees). Conforme descrito por (QUINLAN, 2014), esses algoritmos diferem em suas estratégias de seleção de atributos e critérios de divisão de nós. O ID3 e o C4.5 utilizam a entropia ou o ganho de informação como critério de seleção, enquanto o CART usa o índice de impureza de Gini. Esses algoritmos são capazes de lidar tanto com problemas de classificação quanto com problemas de regressão.

Uma das principais vantagens dos modelos de árvores de decisão é sua interpretabilidade. Conforme pode-se ver no livro de (GÉRON, 2022) A estrutura em forma de árvore permite visualizar todas as tomadas de decisão, tornando o modelo facilmente compreensível, inclusive para pessoas não familiarizadas com o campo do aprendizado de máquina. Além disso, as árvores de decisão são capazes de lidar com dados não lineares e são robustas em relação a outliers e dados faltantes.

É possível mostrar a árvore de decisão completa, incluindo todas as tomadas de decisão, o que facilita ainda mais a interpretação e análise do modelo. Ainda segundo (GÉRON, 2022), a visualização da árvore permite entender como cada atributo contribui para a tomada de decisão e identificar os caminhos percorridos para cada instância, podemos ver um exemplo na figura 4. Essa capacidade de visualização torna as árvores de decisão uma ferramenta poderosa para análise exploratória de dados e para a tomada de decisões fundamentadas.

Como escrito por (JAMES et al., 2017) ao utilizar árvores de decisão para problemas de regressão, algumas métricas comuns são utilizadas para avaliar o desempenho do modelo, como o erro médio absoluto (MAE), o erro médio quadrático (RMSE) e o coeficiente de determinação ( $R^2$ ). O MAE mede a média dos erros absolutos entre as previsões e os



Figura 4 – Exemplo de árvore de decisão

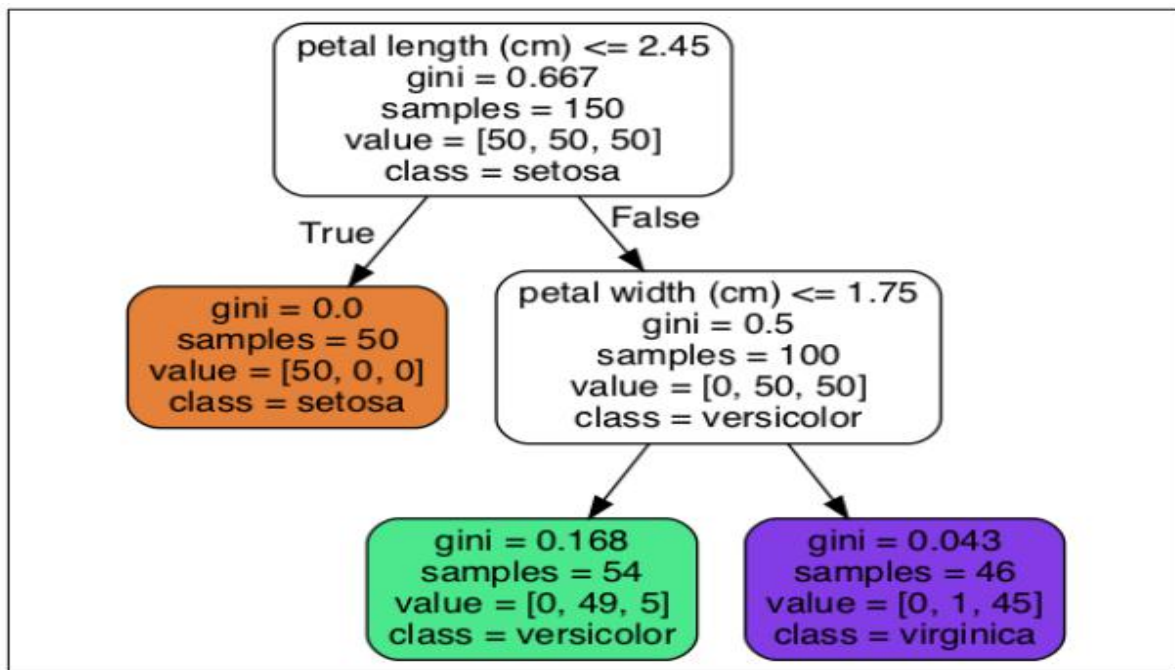


Figura retirada do livro de (Géron, 2022)

valores reais, o RMSE mede a raiz quadrada da média dos erros ao quadrado, e o  $R^2$  mede a proporção da variabilidade dos dados que é explicada pelo modelo.

Embora as árvores de decisão tenham muitas vantagens, elas também apresentam algumas limitações. Assim como citado no livro de (Müller; Guido, 2017), a medida que a árvore cresce, pode ocorrer superajuste (overfitting) aos dados de treinamento, levando a um desempenho inferior em dados não vistos. Além disso, a construção de uma árvore muito profunda pode levar a uma estrutura complexa e difícil de interpretar. Para mitigar esses problemas, é possível utilizar técnicas de poda (pruning) ou limitar a profundidade máxima da árvore.

### 3.3.3 Florestas Aleatórias (*Random Forests*)

Conforme construído por (Breiman, 2001), as florestas aleatórias são modelos de aprendizado de máquina compostos por um conjunto de árvores de decisão independentes (como se fossem uma "floresta" mesmo), chamadas de "árvores de decisão de base". Cada árvore é treinada em um subconjunto aleatório dos dados de treinamento, utilizando um processo conhecido como bootstrapping, que podemos ver na figura 5. Essa abordagem cria diversidade entre as árvores, permitindo que elas capturem diferentes aspectos dos dados. Uma das principais vantagens das florestas aleatórias em relação às árvores de decisão individuais é a redução do sobreajuste (overfitting) e a melhoria da generalização.

Figura 5 – Exemplo de floresta aleatória

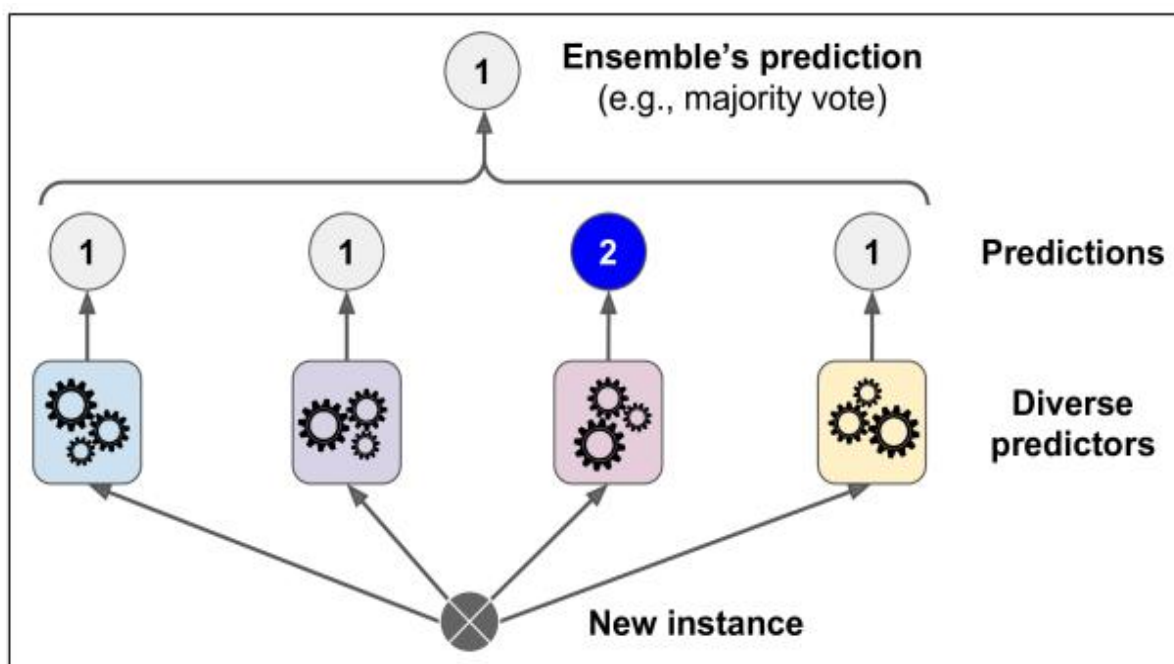


Figura retirada do livro de (GÉRON, 2022)

Segundo (HASTIE; TIBSHIRANI; FRIEDMAN, 2017) Enquanto as árvores de decisão tendem a se ajustar excessivamente aos dados de treinamento, as florestas aleatórias conseguem mitigar esse problema, ao combinar as previsões de várias árvores. Isso resulta em um modelo mais robusto e capaz de se adaptar melhor a dados novos e não vistos anteriormente.

A aleatoriedade desempenha um papel fundamental na construção das florestas aleatórias. Conforme escrito por (GÉRON, 2022) Além de utilizar subconjuntos aleatórios dos dados de treinamento, as árvores de decisão individuais também realizam divisões aleatórias durante o processo de construção. Essa aleatoriedade introduz variabilidade nos modelos, permitindo que eles capturem relações mais complexas e evitem depender excessivamente de características específicas.

Ainda segundo (GÉRON, 2022), durante a fase de predição, as florestas aleatórias agregam as previsões de todas as árvores de decisão para obter uma previsão final. Para problemas de regressão, a previsão final pode ser a média das previsões de todas as árvores. Já para problemas de classificação, a previsão final é obtida por votação, onde cada árvore contribui com um voto. Essa estratégia de agregação melhora a estabilidade e a confiabilidade das previsões.

As métricas mais comuns para avaliar o desempenho das florestas aleatórias em problemas de regressão são similares as das árvores de decisão. Conforme o livro de (JAMES et al., 2017). O erro médio absoluto (MAE) que mede a diferença média entre as previ-

sões e os valores reais. O erro médio quadrático (RMSE) que é a raiz quadrada da média dos erros ao quadrado, sendo uma métrica mais sensível a grandes erros. E por fim, o coeficiente de determinação ( $R^2$ ) mede a proporção da variância dos valores dependentes explicada pelas previsões do modelo.

Além dos parâmetros comuns de árvores de decisão, como a profundidade máxima e o critério de divisão, as florestas aleatórias também possuem parâmetros específicos. Como foi escrito por (GÉRON, 2022), o número de árvores na floresta é um parâmetro importante, onde um número maior de árvores pode melhorar a precisão, mas aumenta o tempo de treinamento e a complexidade do modelo. Outro parâmetro relevante é o número de características consideradas em cada divisão, conhecido como número máximo de características.

### 3.3.4 Pré-processamento de Dados para Modelos de Árvores de Decisão e Florestas Aleatórias

O pré-processamento de dados desempenha um papel crucial no sucesso da aplicação de modelos de árvores de decisão e florestas aleatórias. Essas etapas visam preparar os dados para o treinamento dos modelos, garantindo a qualidade e a consistência das informações utilizadas. Segundo (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), uma consideração importante é a normalização ou padronização dos atributos. Isso envolve a escala dos valores dos atributos para um intervalo específico. A normalização é particularmente importante quando os atributos possuem escalas diferentes, pois garante que todos os atributos sejam considerados igualmente pelo modelo. Métodos comuns de normalização incluem a normalização min-max e a padronização com base na média e desvio padrão.

O tratamento de dados ausentes também é uma etapa crítica. Segundo (GÉRON, 2022), é comum que os conjuntos de dados reais possuam valores ausentes em determinadas instâncias ou atributos. Nesses casos, é necessário decidir como lidar com esses valores. Isso pode envolver técnicas como remoção das instâncias com dados ausentes, substituição por valores médios ou mais frequentes, ou o uso de técnicas de imputação mais avançadas, como regressão ou K-NN (K-vizinhos mais próximos).

Ainda segundo (GÉRON, 2022), outra consideração é a codificação de variáveis categóricas, que são atributos que assumem valores qualitativos em vez de quantitativos. As árvores de decisão e florestas aleatórias geralmente requerem que os atributos sejam numéricos. Portanto, é necessário converter as variáveis categóricas em representações numéricas adequadas. Isso pode ser feito usando técnicas como codificação one-hot, onde cada valor categórico é convertido em uma nova coluna binária, ou codificação ordinal, onde os valores categóricos são mapeados para números inteiros.

Essas etapas de pré-processamento de dados podem afetar o desempenho dos modelos de árvores de decisão e florestas aleatórias. Como escrito por (HASTIE; TIBSHIRANI; FRIEDMAN, 2017), a normalização adequada dos atributos pode evitar a dominância

de certos atributos devido a suas escalas e garantir que todos os atributos contribuam igualmente para a construção das árvores. O tratamento adequado de dados ausentes evita a perda de informações valiosas e minimiza o impacto negativo nos modelos. A codificação correta de variáveis categóricas permite que o modelo capture as informações corretas e faça previsões precisas.

É importante ressaltar que as etapas de pré-processamento de dados podem variar de acordo com o conjunto de dados e o problema em questão. De acordo com o livro escrito por (MÜLLER; GUIDO, 2017), é recomendado realizar uma análise cuidadosa dos dados e considerar abordagens específicas para cada caso. Além disso, o pré-processamento deve ser aplicado tanto aos dados de treinamento quanto aos dados de teste para garantir a consistência dos resultados.

### **3.3.5 Avaliação e Interpretação dos Modelos**

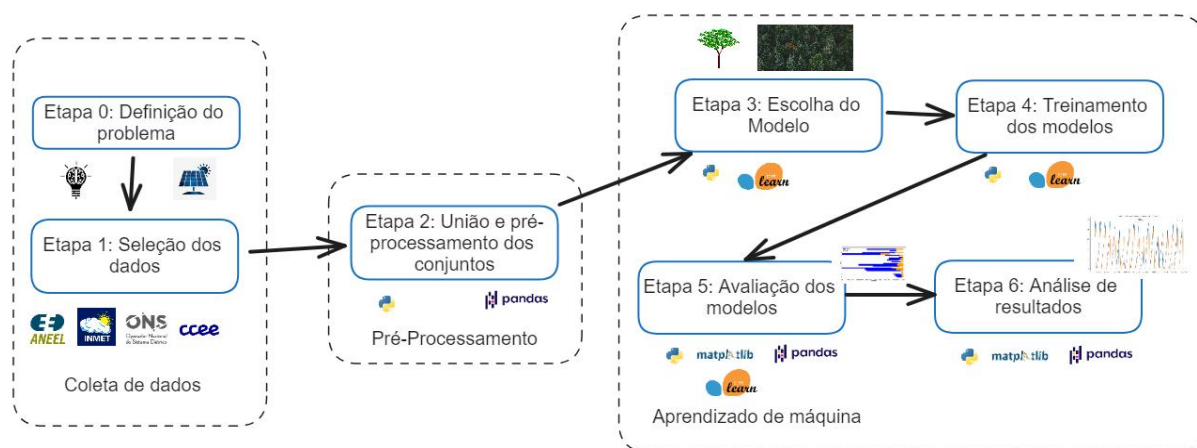
A avaliação de modelos de árvore de decisão e floresta aleatória envolve o uso de métricas comuns para avaliar sua precisão e desempenho. Essas métricas, entre outras coisas, fornecem uma visão abrangente do desempenho do modelo. Além de avaliar as métricas de desempenho, também é importante avaliar a robustez e a estabilidade da árvore de decisão e dos modelos de floresta aleatória. Alternativamente, de acordo com (BROWNLEE, 2016), isso pode ser feito analisando a variação nos resultados ao executar o modelo várias vezes em diferentes conjuntos de dados ou por técnicas de validação cruzada. Essas análises fornecem informações sobre a consistência dos resultados e a confiabilidade das previsões do modelo.

Além da avaliação do modelo e métricas de interpretação, é importante enfatizar a importância do monitoramento contínuo e atualização da árvore de decisão e modelos de floresta aleatória na previsão de geração de energia solar. Conforme citado por (GAJEK; SCHIENDORFER; REIF, 2019), as condições meteorológicas e a disponibilidade de dados podem mudar ao longo do tempo, portanto, o modelo deve ser ajustado e recalibrado para garantir previsões precisas. A reavaliação periódica dos modelos e a incorporação de novos dados podem ajudar a melhorar a qualidade das previsões.

## 4 MATERIAIS E MÉTODOS

Neste capítulo será exposto em detalhes como foi conduzido todo o processo até a obtenção dos resultados finais na criação dos modelos. O processo de criação de um modelo de aprendizado de máquina envolve várias etapas básicas, desde a seleção de um banco de dados até a análise dos resultados. A seguir, o processo será dividido em seis etapas principais:

Figura 6 – Processo seguido pelo trabalho



### 4.1 DEFINIÇÃO DO PROBLEMA E SELEÇÃO DOS DADOS

Prever a geração de energia solar é um desafio fundamental na busca por sistemas energéticos mais sustentáveis e eficientes. Para desenvolver modelos de aprendizado de máquina capazes de fazer tais previsões, é fundamental ter uma compreensão clara do problema em questão. Nesse contexto, o objetivo é usar dados climáticos relevantes para estimar a quantidade de energia solar produzida em determinadas condições atmosféricas. Esses dados incluem informações como radiação solar, temperatura, velocidade do vento e umidade, que desempenham um papel fundamental na captura dos padrões e impactos que afetam a geração de energia solar. A seleção cuidadosa de dados representativos e abrangentes é fundamental para treinar e testar modelos com precisão, permitir previsões mais confiáveis e promover o uso da energia solar como fonte primária de energia.

Os seguintes conjuntos foram considerados na elaboração do banco de dados principal que irá alimentar os modelos:

1. Os dados climáticos utilizados neste estudo são provenientes do banco de dados do INMET (Instituto Nacional de Meteorologia). Esses dados incluem informações como velocidade do vento, incidência solar e umidade do ar. São fornecidos como

um conjunto de arquivos no formato .csv, onde os dados são separados por vírgulas. Cada arquivo corresponde a uma estação meteorológica e contém dados de um ano inteiro. Antes dos dados, há uma pequena tabela com informações sobre o local, como coordenadas geográficas, nome, cidade, estado, altitude e código identificador. Os dados foram coletados de 2001 a março de 2023, fornecendo uma extensa série histórica para análise e modelagem. É possível ver nas tabelas 1 e 2 uma pequena amostra de um dos bancos de dados que está disponível online<sup>1</sup>.

2. Os dados horários de geração de energia das usinas solares fotovoltaicas nacionais são provenientes do banco de dados disponibilizado no site da CCEE (Câmara de Comercialização de Energia Elétrica). Os dados são organizados em um arquivo tipo Excel onde todas as informações geradas são reunidas em uma única planilha. Para facilitar a organização, os dados são separados por mês, sendo cada mês representado por um arquivo separado. O banco de dados contém informações sobre a produção horária de cada usina solar, juntamente com seus respectivos nomes. A coleta de dados abrange o período de setembro de 2020 a março de 2023, fornecendo uma série histórica abrangente para análise e modelagem. É possível ver na tabela 3 uma pequena amostra de um dos bancos de dados
3. Os dados referentes às usinas solares, como o CEG (Cadastro dos Empreendimentos Geradores) e as coordenadas geográficas, são obtidos nos bancos de dados disponibilizados pela ANEEL (Agência Nacional de Energia Elétrica). Esses dados estão contidos em um arquivo que lista todas as usinas solares cadastradas no Brasil com seus respectivos dados. O banco de dados é composto por informações detalhadas sobre as plantas, permitindo uma análise abrangente de suas características e localização geográfica.

É possível ver na tabela 4 uma pequena amostra de um dos bancos de dados.

4. Além disso, outra base de dados foi obtida no site do ONS (Operador Nacional dos Sistemas Elétricos) para estabelecer a relação entre os nomes das usinas (presentes na segunda base de dados) e os códigos CEG (Cadastro de Empreendimentos de Geração, presentes na terceira base de dados). Essa ligação é crítica para unificar os dados dos dois bancos em um conjunto único e coeso. Dessa forma, o banco de dados do ONS integra e sincroniza as informações, proporcionando uma visão consolidada das usinas solares e estações meteorológicas.

É possível ver na tabela 5 uma pequena amostra de um dos bancos de dados

---

<sup>1</sup> Dados climáticos do site do INMET: (INMET, 2023)

Tabela 1 – Dados referentes a localização da estação climática

Tipo de informação	Dado referente a estação
REGIAO:	SE
UF:	RJ
ESTACAO:	MACAE
CODIGO (WMO):	A608
LATITUDE:	-22,3761111
LONGITUDE:	-41,81194444
ALTITUDE:	28
DATA DE FUNDACAO:	22/09/2006

Fonte: (INMET, 2023)

Tabela 2 – Dados meteorológicos referentes a estação climática

Data	Hora UTC	Radiação	...	Umidade
01/01/2023	1000 UTC	946,3	...	88
01/01/2023	1100 UTC	1634,7	...	88
01/01/2023	1200 UTC	2014,5	...	90
01/01/2023	1300 UTC	3014	...	90
01/01/2023	1400 UTC	3659,4	...	90

Fonte: (INMET, 2023)

Tabela 3 – Dados de geração, código CEG e tipo de fonte energética

...	Sigla da Usina	Fonte	...	Hora	Dia	Geração
...	SANTA CRUZ-FURNAS	Térmica bi-Combustível - gás/óleo	...	1	01/02/2023	0,000
...	SANTA CRUZ-FURNAS	Térmica bi-Combustível - gás/óleo	...	2	01/02/2023	0,000
...	SANTA CRUZ-FURNAS	Térmica bi-Combustível - gás/óleo	...	3	01/02/2023	0,000
...	SANTA CRUZ-FURNAS	Térmica bi-Combustível - gás/óleo	...	4	01/02/2023	0,000
...	SANTA CRUZ-FURNAS	Térmica bi-Combustível - gás/óleo	...	5	01/02/2023	0,000

Fonte: (INMET, 2023)

Tabela 4 – Código CEG e coordenadas das usinas

...	Código CEG	...	Num. coord. Norte	Num. coord. Leste
...	PCH.PH.MG.000008-6.1	...	-20,12479858	-43,87020250
...	PCH.PH.MG.000009-4.1	...	-20,13187300	-43,87693500
...	PCH.PH.MG.000010-8.1	...	-20,13754468	-43,89192620
...	UHE.PH.RS.000012-4.1	...	-29,06475278	-51,67494167
...	CGH.PH.RS.000015-9.1	...	-27,82377500	-52,06999722

Fonte: (ANEEL, 2023a)

Tabela 5 – Nome das usinas e código CEG

...	Nome Usina	...	Código CEG
...	UFV B JL 11	...	UFV.RS.BA.034153-3.01
...	UFV B JL 4	...	UFV.RS.BA.034158-4.01
...	UFV Bom Jesus da Lapa I	...	UFV.RS.BA.032892-8.01
...	UFV Bom Jesus da Lapa II	...	UFV.RS.BA.032893-6.01
...	UFV São Pedro II	...	UFV.RS.BA.033782-0.01

Fonte: (ONS, 2023b)

## 4.2 PRÉ-PROCESSAMENTO DOS DADOS E UNIÃO DOS DATASETS

Para começar se fez necessário definir o período que será tratado. Para isso é possível observar que o banco de dados de geração tem seu início em setembro de 2020, o que faz com que ele seja o limite inferior, uma vez que o banco de dados de clima possui dados até 2001. O limite superior é comum aos dois bancos de dados (Março de 2023) e portanto não é necessário nenhuma alteração.

Após baixar todas as bases de dados relevantes, os bancos de dados foram organizados em uma estrutura de pastas, para que pudessem ser facilmente acessadas pelo *python*. Em seguida foi necessário definir a melhor maneira de construir o banco de dados.

### 4.2.1 Conectando os bancos de dados

Como mencionado anteriormente para realizar a conexão entre o segundo e o terceiro banco de dados, foi necessário um banco de dados de apoio, pois o segundo banco de dados possuía apenas a informação do nome da usina enquanto que o terceiro banco de dados possuía apenas o CEG. Assim o quarto banco de dados foi retirado do site da (ONS, 2023b) que possuía uma correlação entre o nome da usina (existente no segundo banco de dados) e o código CEG (existente no terceiro banco de dados). Assim foi possível estabelecer uma conexão.



Após conectar o segundo, quarto e terceiro banco de dados, foi possível criar o banco de dados chamado de "NomeDaUsina\_geracao.csv", que foi criado para cada usina de geração que foi possível obter uma correlação entre os bancos de dados. Após uma análise, os conjuntos de dados foram agrupados de acordo com o início e o fim dos bancos de dados, obtendo assim 111 usinas e 19 grupos, conforme a tabela 6.

Após a geração desse banco de dados foi possível também gerar um mapa com a distribuição de todas as usinas disponíveis, conforme figura 7.

Figura 7 – Mapa de distribuição das usinas de geração fotovoltaica pelo país



Fonte: Dados extraídos da (ANEEL, 2023a)

O próximo passo foi gerar um banco de dados unificado de todas as estações meteorológicas com as informações necessárias para decidir as 3 mais relevantes para cada usina. Isso foi feito capturando as informações geográficas de todas as estações disponíveis em 2020 (começo do banco de dados de geração). É possível ver um exemplo do banco construído na tabela 7, e para melhor visualização, o mapa da distribuição das usinas na figura 8

O próximo passo foi a escolha de qual das 3 estações mais próximas geograficamente da usina em questão seria a melhor. Para isso foi feita uma contagem de itens nulos e faltantes, caso a primeira estação tenha  $>75\%$  de itens nulos/faltantes, a segunda estação é automaticamente escolhida, caso a segunda também tenha  $>75\%$  de itens nulos/faltantes, é escolhida a terceira estação e por fim a usina é descartada do conjunto. Nenhuma usina chegou a ser descartada, e para 81% das usinas a estação mais próxima foi a escolhida, no restante a segunda estação foi a escolhida.

Tabela 6 – Data de início, fim e grupo das usinas

Grupo	Nome da usina	Data início	Data fim
0	AC III; AC IX; CASTILHO 1 a 5	2022-12-01	2023-02-28
1	AC IV; AC VI; AC V	2022-11-01	2023-02-28
2	AC VIII; AC VII; JANAÚBA 11	2022-10-01	2023-02-28
3	AC XV; AC X	2023-02-01	2023-02-28
4	ALEX I e III a VIII	2021-09-04	2023-02-28
5	ALEX IX; ALEX X	2021-08-01	2023-02-28
6	BJL 11, BOA HORA 1 a 3; COREMAS II e III; FONTES SOLAR II; HORIZONTE MP 1, 2 e 11; ITUVERAVA 1 a 4 e 6; NOVA OLINDA 8 a 14; PIRAPORA 2 a 7, 9 e 10; SOL DO FUTURO I a III; SÃO PEDRO II e IV; VAZANTE 1 a 3	2020-09-01	2023-02-28
7	BOM NOME 1-5 e 1-6	2021-12-01	2023-02-28
8	BRIGIDA 2	2021-05-01	2023-02-28
9	COREMAS IV; LUZIA 3	2022-05-01	2023-02-28
10	JANAÚBA 10; JANAÚBA 5 a 7	2022-09-01	2023-02-28
11	JANAÚBA 2, 6 a 9 e 12; LUZIA 2	2022-07-01	2023-02-28
12	JANAÚBA 8,13 e 14	2022-08-01	2023-02-28
13	JANAÚBA 15 a 17	2023-01-01	2023-02-28
14	JANAÚBA 1,3 e 4; LAVRAS 1 a 5	2022-06-01	2023-02-28
15	JUAZEIRO SOLAR I a IV	2020-09-01	2022-08-10
16	Pereira Barreto I a V; Sol do Sertão VIII, XIII, XXXIII e XXXV; Terra do Sol XI	2021-07-01	2023-02-28
17	SERRA DO MEL I e II	2022-08-15	2022-12-03
18	Sol do Sertão XII e XIV; Terra do Sol VII	2021-09-01	2023-02-28

Fonte: (CCEE, 2023)

Tabela 7 – Exemplos de linhas do banco de dados unificado das localizações das estações climáticas

Estação	Latitude	Longitude
BRASILIA	-15,78944444	-47,92583332
BRAZLANDIA	-15,59972221	-48,13111111
AGUAS EMENDADAS	-15,596491	-47,625801
GAMA (PONTE ALTA)	-15,93527777	-48,13749999
PARANOIA (COOPA-DF)	-16,012222	-47,557417

Fonte: (INMET, 2023)

Figura 8 – Mapa de distribuição das estações meteorológicas do Brasil



Fonte: Dados extraídos da (INMET, 2023)

Tabela 8 – Banco de dados antes do processamento

<i>datetime</i>	Precipitação	Pressão Atmosférica	radiação	...
2022-12-01 09:00:00	0.0	957,7	168,0	...
2022-12-01 10:00:00	0.0	958,2	722,1	...
2022-12-01 11:00:00	0.0	958,9	1751,7.0	...
2022-12-01 12:00:00	0.0	959,5	2699,6	...
2022-12-01 13:00:00	0.0	959,4	3463,0	...

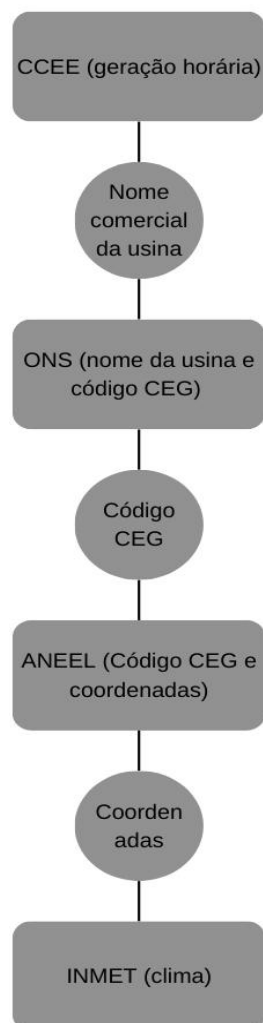
Após a junção dos bancos de dados de geração e de clima, foi gerado um banco de dados chamado "NomeDaUsina\_*full.csv*", esse banco de dados possui todos os dados necessários e está indexado por *datetime*. Na figura 9, há um esquema representando a junção dos bancos de dados. Nesse ponto, existe apenas um banco de dados por usina com todas as informações necessárias para a criação dos modelos.

#### 4.2.2 Pré-Processamento dos dados

Antes de começar a trabalhar com os dados, o banco de dados está como na tabela 8.

Para que os modelos aprendam de maneira adequada é necessário que os dados estejam normalizados, para que isso ocorra foram necessárias algumas transformações nos dados. Para a escolha da metodologia citada abaixo foi tomado como fonte o artigo de (SARMAS et al., 2022).

Figura 9 – Ilustração da conexão entre os bancos de dados



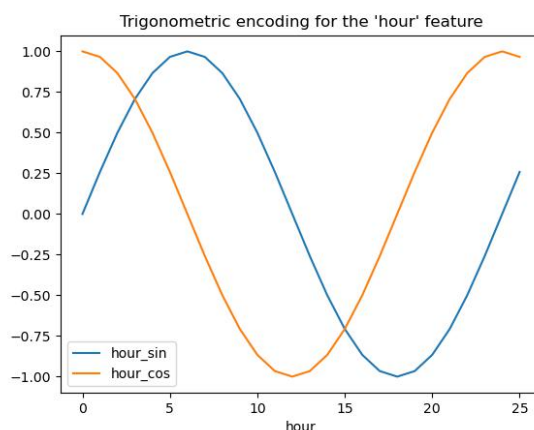
Para transformar o datetime, por não se tratar de uma grandeza linear, foi utilizado uma técnica diferente. Os meses foram transformados utilizando o método *One-Hot*, que basicamente converte um dado categórico em  $n$  tabelas (uma pra cada categoria possível). Assim foram criadas 12 colunas (de 1 a 12) para cada mês do ano, e caso aquele mês seja o atual, ele será 1, caso contrário 0. Já para transformar as horas, foi utilizado uma transformação de cosseno, uma vez que ele aproxima o comportamento (inversamente, como é possível observar na figura 10<sup>2</sup>) da geração solar, e isso facilita o aprendizado do modelo.

O restante dos dados foi transformado utilizando uma transformação *MinMax*, que consiste na aplicação da fórmula

$$X' = (X - X_{min}) / (X_{max} - X_{min})$$

<sup>2</sup> (SCIKIT-LEARN, 2007-2023)

Figura 10 – Gráfico de seno e cosseno para horas do dia



Fonte: Documentação da biblioteca Scikit-Learn

Tabela 9 – Banco de dados após da normalização

Geração	Radiação	Temperatura	Umidade	OneHot	Cos(Hora)
175.9	0.037581	0.470588	0.544444	...	0.078481
1046.8	0.223647	0.523220	0.477778	...	0.019960
1905.5	0.407106	0.557276	0.388889	...	0.000000
2614.9	0.558668	0.594427	0.344444	...	0.019960
3058.3	0.653399	0.643963	0.255556	...	0.078481

onde  $X'$  é o valor normalizado,  $X$  é o valor de entrada, e por fim  $X_{max}$  e  $X_{min}$  são respectivamente o valor máximo e mínimo do conjunto.

Além disso, após consulta ao (SARMAS et al., 2022), decidiu-se por remover algumas colunas, como a precipitação, temperatura do ponto de orvalho, e coordenadas. Ao fim desses ajustes, os dados possuíam agora os seguintes valores já ajustados:

1. Geração
2. Radiação
3. Temperatura
4. Umidade
5. One-Hot dos meses (12 colunas)
6. Cosseno da Hora

E para demonstrar os dados citados, tem-se a tabela 9.

### 4.3 ESCOLHA DO ALGORITMO/MODELO DE APRENDIZADO DE MÁQUINA

Existem vários algoritmos e modelos de aprendizado de máquina disponíveis, cada um com suas próprias características e adequados para diferentes tipos de problemas. Nesta etapa, foi escolhido o algoritmo/modelo mais adequado para o problema de previsão de geração solar. Algoritmos comuns incluem regressão linear, redes neurais artificiais, árvores de decisão e algoritmos de séries temporais.

A previsão de energia solar com base em dados climáticos é um importante desafio para maximizar o uso da energia solar e otimizar os sistemas de geração de energia fotovoltaica. Nesse contexto, o uso de modelos de aprendizado de máquina, como árvores de decisão para regressão e florestas aleatórias de regressão, tem se mostrado promissor. Neste artigo vamos considerar os aspectos técnicos e sua adequação aos dados disponíveis, justificando o uso desses dois tipos de modelos para criar modelos de previsão de geração solar.

Primeiro, as árvores de decisão são modelos de aprendizado de máquina com uma estrutura intuitiva e facilmente interpretável. Essas estruturas de árvore são construídas de acordo com uma série de regras de decisão baseadas nas propriedades dos dados de entrada. Para previsão de energia solar, árvores de decisão para regressão podem ser usadas para identificar padrões e relações entre variáveis climáticas e geração de energia solar. Além disso, a interpretabilidade da árvore de decisão permite analisar como cada variável afeta a previsão, ajudando a identificar fatores relevantes e entender o processo de geração de energia solar.

As florestas aleatórias de regressão, por outro lado, consistem em conjuntos de árvores de decisão e são capazes de lidar com problemas mais complexos. O método envolve a criação de várias árvores de decisão, cada uma usando uma amostra aleatória de dados disponíveis e um subconjunto de atributos. As previsões são feitas pela média das previsões individuais de cada árvore. Os benefícios do uso de florestas aleatórias para prever a geração de energia solar incluem overfitting reduzido, que é a capacidade de generalizar melhor os resultados para dados inéditos. Além disso, as florestas aleatórias são capazes de lidar com dados ruidosos e têm uma boa capacidade de lidar com variáveis não lineares.

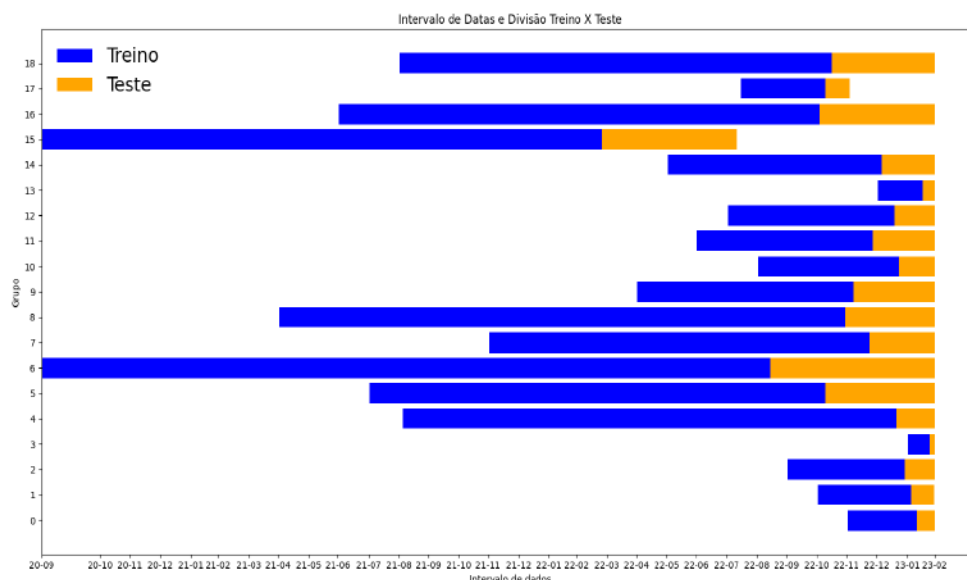
Ambos os modelos têm vantagens significativas na previsão da geração de energia solar a partir de dados climáticos. As árvores de decisão para regressão são adequadas quando se busca entender o processo de previsão de forma mais interpretável e visualizar o impacto de cada variável no resultado final. Por outro lado, florestas aleatórias para regressão são mais adequadas quando você deseja lidar com problemas mais complexos, dados ruidosos e obter uma melhor generalização de resultados. A escolha entre esses dois modelos dependerá da natureza dos dados disponíveis, do objetivo da previsão e da importância da interpretabilidade e da capacidade de lidar com a complexidade, desse modo iremos realizar uma comparação entre os dois.

#### 4.4 TREINAMENTO DO MODELO

Nesta etapa, os dados são divididos em conjuntos de treinamento, validação e teste. O conjunto de treinamento é usado para alimentar o modelo de aprendizado de máquina, permitindo que ele aprenda os padrões presentes nos dados. O conjunto de validação é usado para ajustar os hiperparâmetros do modelo e avaliar seu desempenho durante o treinamento. O conjunto de teste é utilizado para avaliar o desempenho após o treino e validação, e é a partir dele que são calculadas as métricas (Essa parte será descrita na próxima seção). O objetivo é encontrar a configuração de modelo que produza os melhores resultados.

Cada conjunto de dados de cada usina individual foi dividida em teste e treino de acordo com o tamanho do conjunto. A divisão foi realizada de maneira sequencial para que se mantenha a temporalidade dos dados, mantendo assim uma característica muito relevante para o aprendizado que é a sequência em que os dados aparecem. O tamanho que foi escolhido para a divisão entre treino e teste foi de 80% dos dados para treino e 20% separado para teste, essa divisão é comum na literatura como em (SARMAS et al., 2022) e (SRIVASTAVA; TIWARI; GIRI, 2019). Dessa maneira é possível testar os dados em uma quantidade satisfatória de situações, porém sem perder muito na fase de aprendizado do modelo, garantindo assim um ajuste satisfatório. Como é possível ver, realizando as divisões para todos os grupos de usinas, tem-se a figura 11.

Figura 11 – Gráfico da divisão de treino e teste das usinas

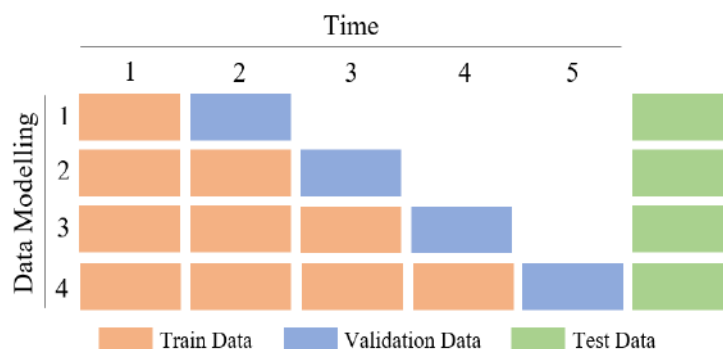


Após a primeira divisão entre treino e teste se faz necessário dividir o conjunto de treino em treino e validação. Essa subdivisão foi realizada utilizando o *TimeSeriesSplit*

que realiza a validação cruzada para o banco de dados em series temporais e calcula a melhor divisão entre teste e validação simultaneamente.

Podemos entender melhor como funciona essa divisão de acordo com o gráfico da figura 12, que foi retirado da página de (BANACHEWICZ, 2022), no Kaggle.

Figura 12 – Gráfico da divisão de treino e teste das usinas



Fonte: Kaggle

Após a divisão para validação, ocorreu o treinamento dos modelos, que foi feito utilizando *GridSearchCV*, método que otimiza a escolha de hiperparâmetros, calcula todas as métricas do aprendizado e escolhe o melhor estimador. Após rodar o método para todas as usinas possíveis obtivemos todos os melhores modelos para todas as usinas. Com os modelos criados, podemos dar seguimento para a avaliação dos modelos.

#### 4.5 AVALIAÇÃO DO MODELO

Após o treinamento, o modelo precisa ser avaliado para verificar sua capacidade de generalização. Isso envolve a utilização de um conjunto de testes separado, que contém dados não vistos durante o treinamento, para avaliar o desempenho do modelo. Para um modelo de regressão, como é o caso, foram utilizadas métricas como erro médio quadrático (MSE), raiz do erro médio quadrático (RMSE), erro médio absoluto (MAE) e coeficiente de determinação ( $R^2$ ) para medir o desempenho do modelo em relação aos resultados esperados.

Foram eliminados modelos com menos de 10000 linhas consideradas pelo modelo para garantir uma maior quantidade de dados para treino, validação e teste.

#### 4.6 ANÁLISE DOS RESULTADOS

Por fim, é importante realizar uma análise aprofundada dos resultados obtidos pelo modelo. Isso pode incluir a identificação de padrões ou insights úteis, a interpretação dos pesos ou importâncias atribuídos às variáveis de entrada, e a comparação com outros



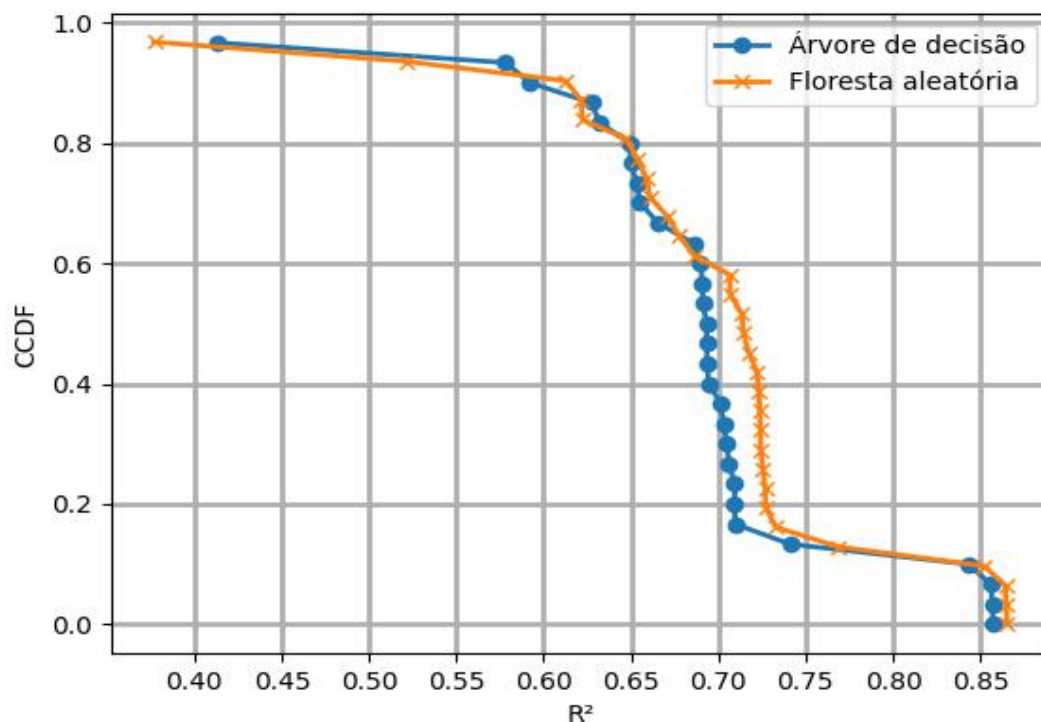
modelos ou métodos existentes. A análise dos resultados pode fornecer insights valiosos para melhorar o modelo ou direcionar futuras pesquisas.

#### 4.6.1 Resultados encontrados

O gráfico 13, mostra a distribuição acumulada dos  $R^2$  de todos os modelos de árvore de decisão e florestas aleatórias treinados. As tabelas 10 e 11 se encontram no anexo 5, e exibem as métricas de avaliação dos melhores modelos de arvores de decisão e florestas aleatórias de cada usina selecionada. Os modelos foram ordenados pela métrica do coeficiente de determinação ( $R^2$ ). O coeficiente de determinação foram consideravelmente satisfatórios para quase todas as usinas avaliadas, segundo (BROWNLEE, 2016), ( $R^2$ ) menores que 0,5 e próximos de 0 são considerados um ajuste ruim.

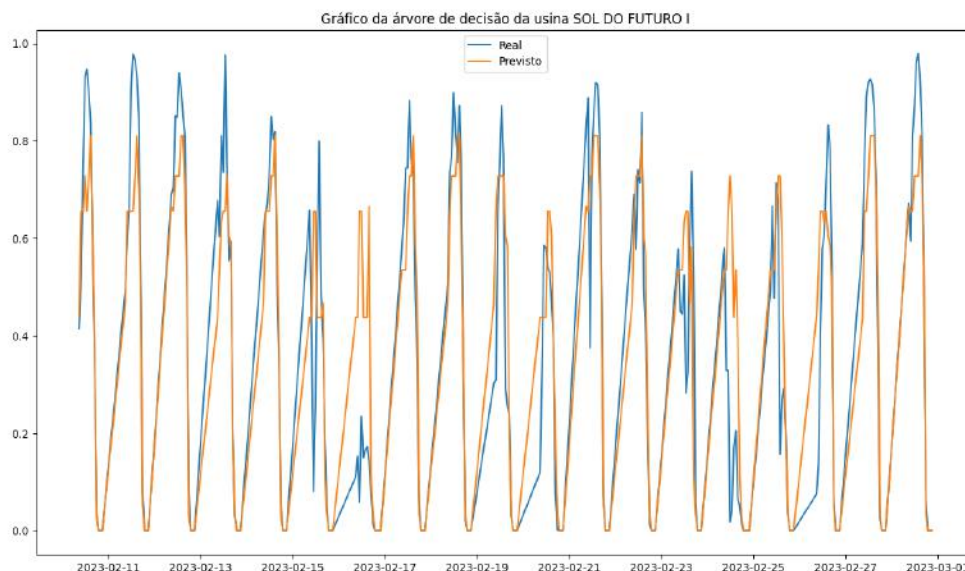
Além disso, é possível ver que 80% dos modelos para a geração de energia fotovoltaica tiveram ( $R^2$ ) maior que 65% para os algoritmos de árvores de decisão e florestas aleatorias. As curvas de CCDF para os modelos de florestas aleatorias e árvores de decisão estão bem próximas, dessa forma podemos observar que apresentaram resultados similares aos modelos de árvores de decisão. Mais de 20% dos modelos de florestas aleatórias apresentam  $R^2$  maiores que 72% para previsão da geração de energia.

Figura 13 – Gráfico CCDF do  $R^2$  dos modelos de árvores de decisão e florestas aleatórias



Foram gerados também gráficos demonstrando a capacidade de previsão dos 3 melhores modelos de cada tipo (Árvores de decisão e Florestas aleatórias).

Figura 14 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro I



Como é possível ver nos gráficos, e é reforçado pela tabela, as previsões estão bem ajustadas ao gráfico, porém sem incorrer em sobre-ajuste.

#### 4.6.2 Comparação com trabalhos relacionados

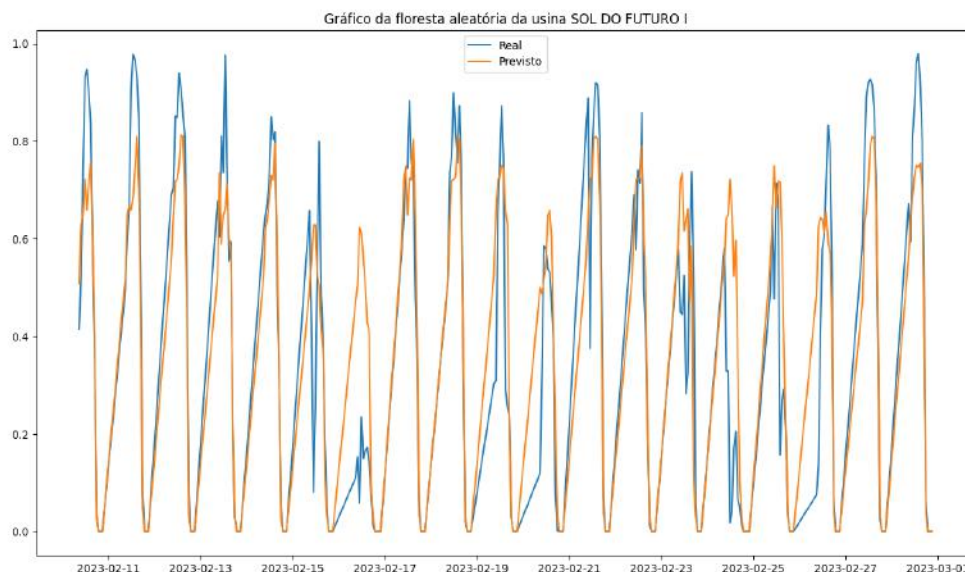
Comparando com o trabalho de (SARMAS et al., 2022), o RMSE encontrado por eles em todos os modelos (chamado por eles de nRMSE) é maior ao encontrado nas tabelas 10 e 11 (26.97% no melhor caso deles, contra 26,81% no pior caso desse trabalho), entretanto os coeficientes de determinação dos modelos deles foram superiores aos resultados encontrados neste trabalho(), o que pode indicar um ajuste inferior deste trabalho, ou um sobreajuste dos modelos construídos por eles.

Se compararmos os resultados obtidos anteriormente, nas tabelas 11 e 10 com o trabalho de (SRIVASTAVA; TIWARI; GIRI, 2019), podemos ver que o RMSE encontrado foi bem melhor que da literatura, até nos piores casos. Na literatura tivemos RMSE de 83.39% para um dia a frente, nesse trabalho o pior caso foi 26.81%

### 4.7 DISCUSSÃO DOS RESULTADOS E LIMITAÇÕES

A análise dos resultados obtidos neste estudo revelou um desempenho promissor dos modelos de árvores de decisão e florestas aleatórias na previsão de geração solar com

Figura 15 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro I



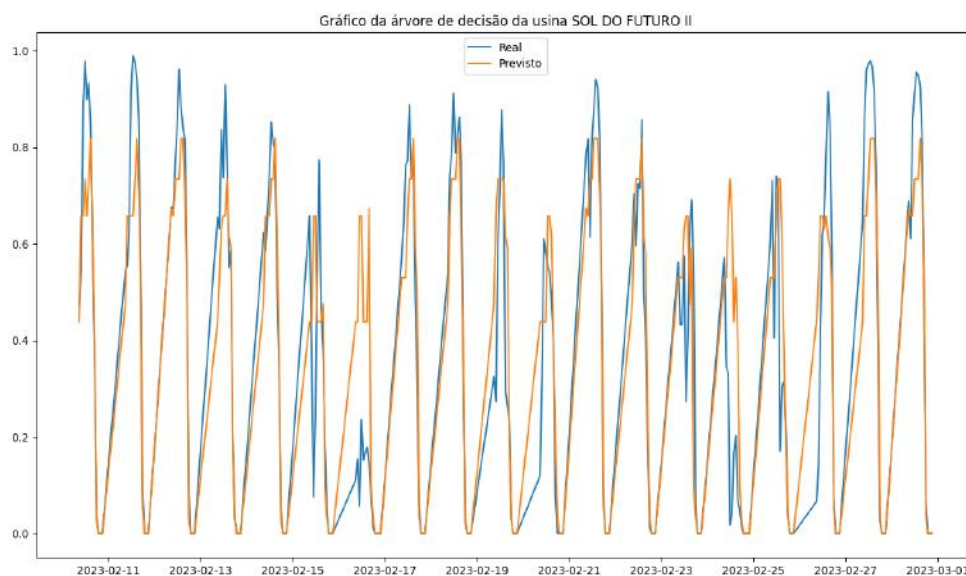
base em dados climáticos. As métricas de desempenho utilizadas, como acurácia, erro médio absoluto (MAE) e erro médio quadrático (RMSE), demonstraram a eficácia desses modelos em fornecer previsões precisas.

Conforme visto na seção 4.6, ao comparar os resultados obtidos com as referências da literatura, constatou-se que os modelos propostos alcançaram resultados competitivos. Em termos de RMSE, nossos modelos superaram as métricas reportadas em estudos semelhantes, mostrando um coeficiente de determinação superior a 60% em grande parte dos modelos. Além disso, o MAE e o RMSE, indicadores de erro, foram consistentemente baixos, sugerindo que as previsões se aproximaram das leituras reais de geração solar.

É importante ressaltar que, apesar dos bons resultados, este estudo também enfrentou algumas limitações e desafios. Um dos principais desafios foi a disponibilidade de dados climáticos e de geração solar de alta qualidade e com resolução temporal adequada. Embora tenhamos selecionado fontes confiáveis, alguns dados estavam indisponíveis ou apresentavam lacunas que exigiram estratégias adicionais para seu tratamento. A falta de dados históricos de geração solar também pode ter impactado a precisão das previsões em períodos de alta variabilidade climática.

Em relação aos dados, é importante buscar fontes de dados mais completas e atualizadas, com maior resolução temporal e espacial, para aprimorar a precisão das previsões. Além disso, a inclusão de dados contextuais, como informações sobre o consumo de energia, demanda elétrica e capacidade instalada de usinas solares, pode contribuir para um modelo mais abrangente e preciso.

Figura 16 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro II



Para reprodutibilidade dos resultados obtidos acima os códigos utilizados no trabalho foram anexados no GitHub

Figura 17 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro II

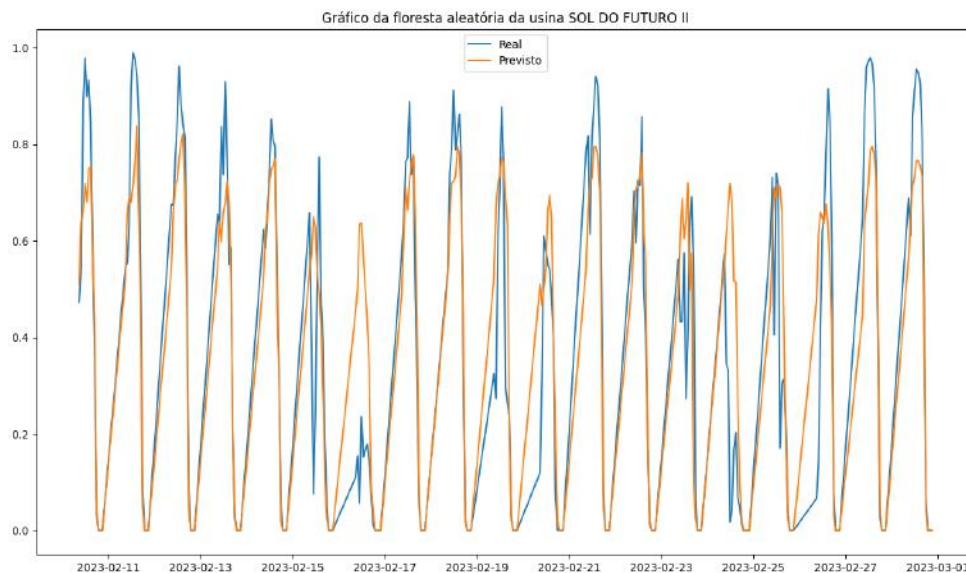


Figura 18 – Gráfico de comparação entre previsto e real - Árvore de decisão - Sol do Futuro III

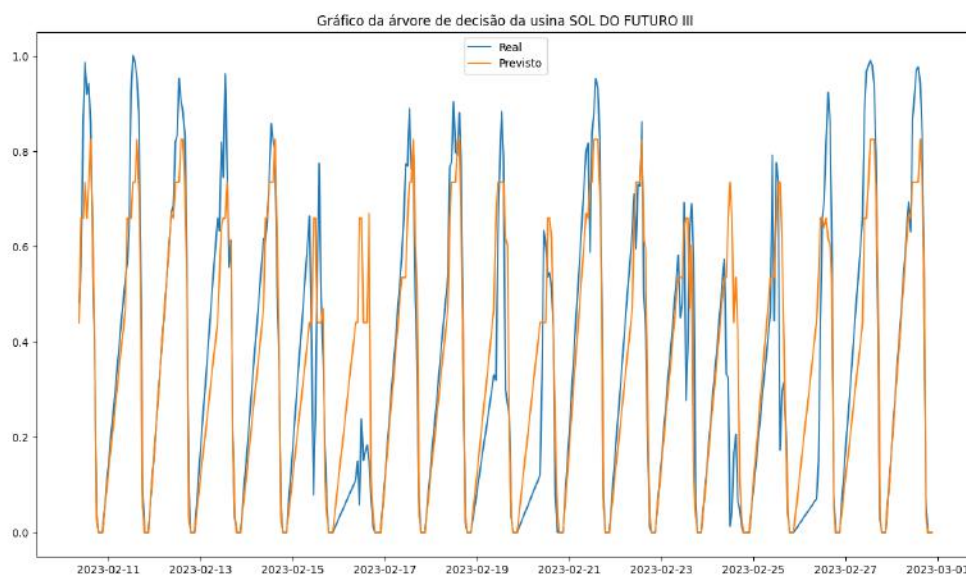
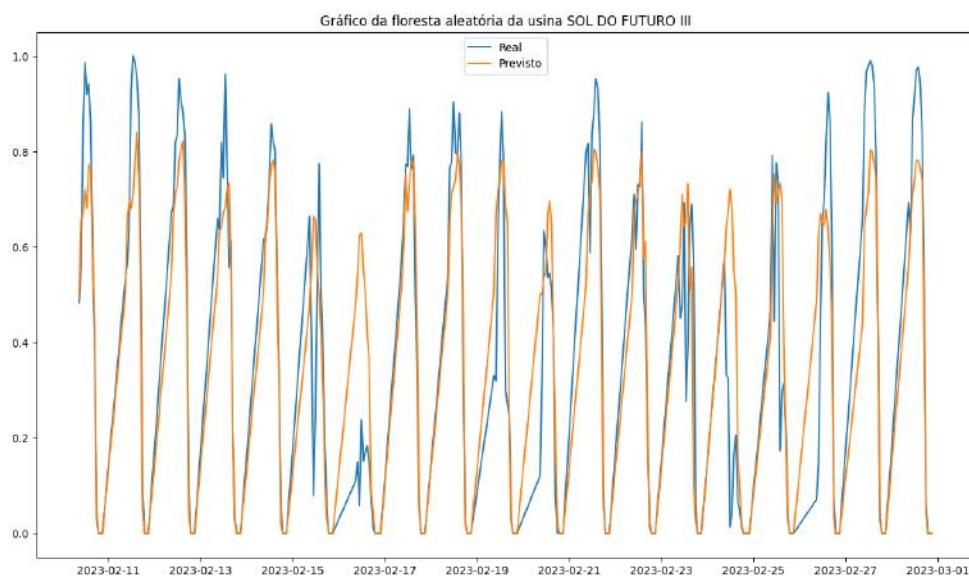


Figura 19 – Gráfico de comparação entre previsto e real - Floresta aleatória - Sol do Futuro III



## 5 CONCLUSÃO

Em conclusão, este estudo ressaltou a importância do uso de modelos de aprendizado de máquina na previsão de geração solar com base em dados climáticos. Ao longo deste trabalho, foi demonstrado que os modelos de árvores de decisão e florestas aleatórias apresentam resultados promissores, fornecendo previsões precisas e confiáveis.

A utilização de modelos de aprendizado de máquina traz vantagens significativas em relação às abordagens tradicionais de previsão de geração solar. Esses modelos têm a capacidade de capturar relações complexas entre os dados climáticos e a geração solar, permitindo uma análise mais detalhada e uma melhor compreensão dos fatores que influenciam a produção de energia solar. Além disso, sua flexibilidade e adaptabilidade permitem a incorporação de novos dados e o ajuste do modelo ao longo do tempo, contribuindo para previsões mais precisas e atualizadas.

No contexto de transição energética e sustentabilidade, a previsão de geração solar desempenha um papel crucial. Essas previsões são essenciais para o planejamento estratégico de políticas energéticas, permitindo uma alocação eficiente de recursos e facilitando a integração harmoniosa da energia solar na rede elétrica. Com base nos resultados obtidos neste estudo, fica claro que os modelos de aprendizado de máquina são ferramentas poderosas para auxiliar nesse processo.

No entanto, há ainda áreas a serem exploradas e aprimoradas. Como possíveis direções para pesquisas futuras, sugere-se a aplicação de outros algoritmos de aprendizado de máquina, como redes neurais ou métodos baseados em séries temporais, para comparar e avaliar seu desempenho em relação aos modelos utilizados neste estudo. Além disso, a incorporação de dados adicionais, como informações sobre o consumo de energia, perfil de demanda elétrica e capacidade instalada de usinas solares, pode contribuir para uma modelagem mais completa e precisa.

## REFERÊNCIAS

- ABUELLA, M.; CHOWDHURY, B. Random forest ensemble of support vector regression models for solar power forecasting. 2017. Disponível em: <https://ieeexplore.ieee.org/document/8086027>. Acesso em: 5 jun.2023.
- ANEEL. Dados abertos de geração solar. 2023. Disponível em: <https://dadosabertos.aneel.gov.br/dataset/siga-sistema-de-informacoes-de-geracao-da-aneel/resource/11ec447d-698d-4ab8-977f-b424d5deee6a>. Acesso em: 2 jun.2023.
- ANEEL. Energia solar. 2023. Disponível em: <http://www.aneel.gov.br/energias-renovaveis/energia-solar>. Acesso em: 2 jun.2023.
- AYKO, O.; KESER, S. B. A comparison of machine learning algorithms for forecasting solar irradiance in eskisehir, turkey. 2021. Disponível em: <https://dergipark.org.tr/tr/download/article-file/1974603>. Acesso em: 7 jul.2023.
- BANACHEWICZ, K. Validation methods for time series. 2022. Disponível em: <https://www.kaggle.com/code/konradb/ts-10-validation-methods-for-time-series>. Acesso em: 10 jul.2023.
- BASTOS, I. G. P. et al. Previsão de geração fotovoltaica a partir de dados meteorológicos utilizando rede lstm. 2020. Disponível em: <https://anaiscbens.emnuvens.com.br/cbens/article/view/761>. Acesso em: 5 jun.2023.
- BLOOMBERGNEF. The future of energy storage: 2022 market outlook. 2022. Disponível em: <https://about.bnef.com/blog/1h-2023-energy-storage-market-outlook>. Acesso em: 19 jun.2023.
- BREIMAN, L. Random forests. 2001. Disponível em: <https://link.springer.com/article/10.1023/a:1010933404324>. Acesso em: 9 jul.2023.
- BROWNLEE, J. **Machine Learning Mastery with Python**. [S.l.: s.n.], 2016. 170 p.
- BÜHLER, A. J. Estudo de técnicas de determinação experimental e pósprocessamento de curvas características de módulos fotovoltaicos. p. 195, 2011. Tese (Doutorado) - Curso de Engenharia Mecânica, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. Disponível em: <https://www.lume.ufrgs.br/handle/10183/30213>. Acesso em: 2 jun.2023.
- CASTRO, P. B. de. Análise econômica de sistema fotovoltaico híbrido com baterias. 2020. PUC Campinas. Disponível em: [https://repositorio.sis.puc-campinas.edu.br/bitstream/handle/123456789/14647/ceatec\\_engenhariaeletrica\\_tcc\\_castro\\_pb.pdf?sequence=1&isAllowed=y](https://repositorio.sis.puc-campinas.edu.br/bitstream/handle/123456789/14647/ceatec_engenhariaeletrica_tcc_castro_pb.pdf?sequence=1&isAllowed=y). Acesso em: 26 jun.2023.
- CCEE. \*checar\*dados históricos anuais de meteorologia. 2023. Disponível em: <https://www.ccee.org.br/web/guest/dados-e-analises/dados-geracao>. Acesso em: 2 jun.2023.
- COMINTECH. Como funciona a usina solar? como irei utilizar? 2021. Disponível em: <https://comintech.com.br/como-funciona-energia-solar>. Acesso em: 12 jul.2023.



- DANTAS, S. G.; POMPERMAYER, F. M. Viabilidade econômica de sistemas fotovoltaicos no Brasil e possíveis efeitos no setor elétrico. 2018. Disponível em: [https://repositorio.ipea.gov.br/bitstream/11058/8400/1/TD\\_2388.pdf](https://repositorio.ipea.gov.br/bitstream/11058/8400/1/TD_2388.pdf). Acesso em: 19 jun.2023.
- EPE. **Balço Energético Nacional 2022**. [s.n.], 2023. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-675/topico-638/BEN2022.pdf>. Acesso em: 2 jun.2023.
- ESMAP, E. S. M. A. P. Rise 2022, regulatory indicators for sustainable energy. 2022. Washington, DC. Disponível em: <https://rise.esmap.org/data/files/reports/2022/RISE\%202022\%20Report\%20Building\%20Resilience.pdf>. Acesso em: 27 jun.2023.
- FOROOTAN, M. M. et al. Machine learning and deep learning in energy systems: A review. **Sustainability**, 2022. Disponível em: <https://www.mdpi.com/2071-1050/14/8/4832>.
- GAJEK, C.; SCHIENDORFER, A.; REIF, W. A chained neural network model for photovoltaic power forecast. 2019.
- GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S.l.]: "O'Reilly Media, Inc.", 2022.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning Data Mining, Inference, and Prediction**. New York: Springer, 2017. 426 p.
- IEA. Renewables 2022: Analysis and forecast to 2027. 2021. Disponível em: <https://iea.blob.core.windows.net/assets/ada7af90-e280-46c4-a577-df2e4fb44254/Renewables2022.pdf>. Acesso em: 3 jun.2023.
- INMET. Dados históricos anuais de meteorologia. 2023. Disponível em: <https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 2 jun.2023.
- IRENA. Métodos de imputação de dados aplicados na Área da Saúde. Renewable energy statistics 2023. International Renewable Energy Agency, Abu Dhabi. Disponível em: [https://mc-cd8320d4-36a1-40ac-83cc-3389-cdn-endpoint.azureedge.net/-/media/Files/IRENA/Agency/Publication/2023/Jul/IRENA\\_Renewable\\_energy\\_statistics\\_2023.pdf?rev=7b2f44c294b84cad9a27fc24949d2134](https://mc-cd8320d4-36a1-40ac-83cc-3389-cdn-endpoint.azureedge.net/-/media/Files/IRENA/Agency/Publication/2023/Jul/IRENA_Renewable_energy_statistics_2023.pdf?rev=7b2f44c294b84cad9a27fc24949d2134). Acesso em: 11 jul.2023.
- JAMES, G. et al. **An Introduction to Statistical Learning with Applications in R**. New York: Springer, 2017. 426 p.
- KRAAS, B. et al. Economic assessment of a concentrating solar power forecasting system for participation in the Spanish electricity market. 2011. Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1874063](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1874063). Acesso em: 7 jul.2023.
- KUMAR, K. et al. Artificial intelligence and machine learning based intervention in medical infrastructure: A review and future trends. 2023. Disponível em: <https://www.mdpi.com/2227-9032/11/2/207>. Acesso em: 7 jul.2023.
- LIMA, M. A. F. B. Teoria do portfólio aplicada à previsão de recursos solar e eólico. 2016. Disponível em: [https://repositorio.ufc.br/bitstream/riufc/22068/1/2016\\_dis\\_mafblima.pdf](https://repositorio.ufc.br/bitstream/riufc/22068/1/2016_dis_mafblima.pdf). Acesso em: 26 jun.2023.

MADUABUCHI, C. et al. Renewable energy potential estimation using climatic-weather-forecasting machine learning algorithms. 2023.

MOCELIN, A. R. Qualificação profissional e capacitação laboratorial em sistemas fotovoltaicos. p. 300, 2014. Tese (Doutorado) - Programa Pós-graduação em Energia, Universidade de São Paulo, São Paulo, 2014. Disponível em: <https://www.teses.usp.br/teses/disponiveis/86/86131/tde-22072014-154625/pt-br.php>. Acesso em: 2 jun.2023.

MURPHY, K. P. **Machine Learning, A probabilistic perspective**. [S.l.]: Massachusetts Institute of Technology, 2012. 1067 p.

MÓDULOENERGIA. Painel solar ou telha solar: qual é a melhor opção? 2022. Disponível em: <https://moduloenergia.com/painel-solar-e-telha-solar/>. Acesso em: 12 jul.2023.

MÜLLER, A. C.; GUIDO, S. **Introduction to Machine Learning with Python**. Sebastopol, CA: O'Reilly Media, 2017. 378 p.

NASA. Soaking up sun in the thar desert. 2022. Disponível em: <https://earthobservatory.nasa.gov/images/149442/soaking-up-sun-in-the-thar-desert>. Acesso em: 20 jun.2023.

ONS. Fontes renováveis tem participação superior a 90% nos quatro primeiros meses de 2023. 2023. Disponível em: <https://www.ons.org.br/Paginas/Noticias/20230529-Fontes-renov%C3%A1veis-tem-participa%C3%A7%C3%A3o-superior-a-90-nos-quatro-primeiros-meses-de-2023-.aspx>. Acesso em: 31 mai.2023.

ONS. Histórico da geração eólica e fotovoltaica. 2023. Disponível em: <https://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao/dados-de-gera%C3%A7%C3%A3o-e-%C3%B3lica-e-solar>. Acesso em: 2 jun.2023.

PEREIRA, E. B. et al. Atlas brasileiro de energia solar. **São José dos Campos: Inpe**, v. 1, 2017.

PORTALENERGIA. Importantes avanços nos painéis solares com camada fina de perovskita. 2021. Disponível em: <https://www.portal-energia.com/importantes-avancos-paineis-solares-camada-fina-perovskita/>. Acesso em: 12 jul.2023.

QUINLAN, J. R. **C4. 5: programs for machine learning**. [S.l.]: Elsevier, 2014.

REN23. Renewables 2023, global status report. 2023. Disponível em: [https://www.ren21.net/wp-content/uploads/2019/05/GSR-2023\\_Energy-Supply-Module.pdf](https://www.ren21.net/wp-content/uploads/2019/05/GSR-2023_Energy-Supply-Module.pdf). Acesso em: 19 jun.2023.

RESEARCH, G. G. V. Machine learning market size, share i& trends analysis report by component (hardware, software, services), by enterprise size, by end-use (advertising i& media, healthcare, retail), by region, and segment forecasts, 2023 - 2030. 2023. Disponível em: <https://www.grandviewresearch.com/industry-analysis/machine-learning-market>. Acesso em: 7 jul.2023.

SARMAS, E. et al. Transfer learning strategies for solar power forecasting under data scarcity. 2022. Disponível em: <https://www.nature.com/articles/s41598-022-18516-x>. Acesso em: 5 jun.2023.

SCIKIT-LEARN. Time-related feature engineering. 2007–2023. Disponível em: [https://scikit-learn.org/stable/auto\\_examples/applications/plot\\_cyclical\\_feature\\_engineering.html](https://scikit-learn.org/stable/auto_examples/applications/plot_cyclical_feature_engineering.html). Acesso em: 9 jul.2023.

SOLARMAGAZINE. Perovskite solar cells: An in-depth guide + comparisons with other techs. 2022. Disponível em: <https://solarmagazine.com/solar-panels/perovskite-solar-cells/>. Acesso em: 22 jun.2023.

SRIVASTAVA, R.; TIWARI, A. N.; GIRI, V. K. Solar radiation forecasting using mars, cart, m5, and random forest model: A case study for india. 2019. Disponível em: [https://www.researchgate.net/publication/337002951\\_Solar\\_radiation\\_forecasting\\_using\\_MARS\\_CART\\_M5\\_and\\_random\\_forest\\_model\\_A\\_case\\_study\\_for\\_India](https://www.researchgate.net/publication/337002951_Solar_radiation_forecasting_using_MARS_CART_M5_and_random_forest_model_A_case_study_for_India). Acesso em: 7 jul.2023.

## APÊNDICES

**APÊNDICE A – TABELAS DOS VALORES DE RESULTADOS DAS USINAS**

Tabela 10 – Resultados dos modelos de Árvores de decisão

Nome da usina	Nº linhas	MAE	MSE	RMSE	$R^2$	$R^2$ validação
SOL DO FUTURO I	11060	0.0874	0.0193	0.1389	0.8574	0.8040
SOL DO FUTURO II	11060	0.0879	0.0196	0.1402	0.8569	0.8015
SOL DO FUTURO III	11060	0.0895	0.0202	0.1422	0.8561	0.8001
COREMAS II	11294	0.0996	0.0213	0.1462	0.8431	0.8059
NOVA OLINDA 14	10929	0.1054	0.0257	0.1603	0.7411	0.7393
PIRAPORA 6	10854	0.1296	0.0356	0.1887	0.7096	0.8091
VAZANTE 3	10854	0.1305	0.0357	0.1889	0.7094	0.8211
NOVA OLINDA 13	10929	0.1420	0.0385	0.1962	0.7094	0.7219
PIRAPORA 7	10854	0.1310	0.0364	0.1909	0.7061	0.8063
PIRAPORA 10	10854	0.1291	0.0353	0.1880	0.7049	0.8070
PIRAPORA 9	10854	0.1316	0.0365	0.1910	0.7030	0.8081
PIRAPORA 2	10854	0.1307	0.0374	0.1935	0.7013	0.8171
COREMAS III	11294	0.1353	0.0411	0.2028	0.6945	0.4016
PIRAPORA 5	10854	0.1315	0.0373	0.1932	0.6936	0.8067
NOVA OLINDA 12	10929	0.1443	0.0394	0.1987	0.6933	0.6957
VAZANTE 2	10854	0.1356	0.0377	0.1942	0.6932	0.8033
PIRAPORA 4	10854	0.1330	0.0386	0.1966	0.6911	0.8080
VAZANTE 1	10854	0.1321	0.0372	0.1929	0.6908	0.8075
NOVA OLINDA 11	10929	0.1362	0.0350	0.1872	0.6899	0.4939
PIRAPORA 3	10854	0.1347	0.0394	0.1986	0.6856	0.8003
ITUVERAVA 3	10800	0.1476	0.0431	0.2077	0.6656	0.6180
ITUVERAVA 2	10800	0.1495	0.0451	0.2123	0.6553	0.6334
ITUVERAVA 6	10800	0.1514	0.0454	0.2132	0.6532	0.6248
ITUVERAVA 4	10800	0.1576	0.0480	0.2192	0.6499	0.5913
NOVA OLINDA 9	10929	0.1382	0.0407	0.2018	0.6491	0.6909
NOVA OLINDA 8	10929	0.1336	0.0414	0.2036	0.6322	0.6933
ITUVERAVA 1	10800	0.1553	0.0481	0.2194	0.6272	0.6154
NOVA OLINDA 10	10929	0.1691	0.0525	0.2292	0.5921	0.3947
HORIZONTE MP 1	10800	0.1609	0.0497	0.2229	0.5781	0.5385
HORIZONTE MP 11	10800	0.1361	0.0420	0.2051	0.4127	0.1908
HORIZONTE MP 2	10800	0.2072	0.0726	0.2695	0.391	0.3163

Tabela 11 – Resultados dos modelos de florestas aleatórias

Nome da usina	Nº linhas	MAE	MSE	RMSE	$R^2$	$R^2$ validação
SOL DO FUTURO I	11060	0.0835	0.0182	0.1350	0.8654	0.8174
SOL DO FUTURO II	11060	0.0843	0.0185	0.1362	0.8649	0.8175
SOL DO FUTURO III	11060	0.0857	0.0190	0.1380	0.8644	0.8144
COREMAS II	11294	0.0961	0.0201	0.1419	0.8523	0.8177
NOVA OLINDA 14	10929	0.1024	0.0229	0.1515	0.7685	0.7601
VAZANTE 3	10854	0.1245	0.0327	0.1810	0.7331	0.8438
PIRAPORA 4	10854	0.1257	0.0341	0.1848	0.7271	0.8329
PIRAPORA 6	10854	0.1247	0.0335	0.1830	0.7268	0.8303
PIRAPORA 9	10854	0.1254	0.0337	0.1837	0.7255	0.8307
PIRAPORA 2	10854	0.1254	0.0345	0.1859	0.7242	0.8364
PIRAPORA 5	10854	0.1265	0.0336	0.1833	0.7241	0.8285
VAZANTE 2	10854	0.1275	0.0339	0.1842	0.7238	0.8274
PIRAPORA 7	10854	0.1252	0.0343	0.1853	0.7232	0.8291
NOVA OLINDA 13	10929	0.1397	0.0368	0.1919	0.7221	0.7488
PIRAPORA 3	10854	0.1283	0.0354	0.1883	0.7174	0.8254
VAZANTE 1	10854	0.1272	0.0343	0.1854	0.7143	0.8291
PIRAPORA 10	10854	0.1258	0.0343	0.1853	0.7135	0.8296
COREMAS III	11294	0.1316	0.0394	0.1987	0.7068	0.4077
NOVA OLINDA 12	10929	0.1414	0.0377	0.1943	0.7065	0.7196
NOVA OLINDA 11	10929	0.1358	0.0354	0.1883	0.6862	0.5000
NOVA OLINDA 8	10929	0.1248	0.0363	0.1906	0.6776	0.7379
ITUVERAVA 3	10800	0.1472	0.0424	0.2061	0.6710	0.6267
ITUVERAVA 2	10800	0.1485	0.0443	0.2106	0.6609	0.6462
ITUVERAVA 6	10800	0.1500	0.0447	0.2115	0.6588	0.6513
ITUVERAVA 1	10800	0.1508	0.0447	0.2116	0.6534	0.6294
ITUVERAVA 4	10800	0.1600	0.0484	0.2200	0.6473	0.6019
NOVA OLINDA 9	10929	0.1492	0.0438	0.2094	0.6222	0.7154
HORIZONTE MP 1	10800	0.1521	0.0445	0.2111	0.6214	0.5783
NOVA OLINDA 10	10929	0.1655	0.0499	0.2234	0.6124	0.4164
HORIZONTE MP 11	10800	0.1232	0.0343	0.1852	0.5211	0.2846
HORIZONTE MP 2	10800	0.2061	0.0719	0.2681	0.376	0.3896