

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CÂMPUS GOIÂNIA
MESTRADO PROFISSIONAL EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

LEONARDO ALVES MESSIAS

Inteligência Artificial Aplicada à Previsão do Potencial de Geração de Energia Solar

Goiânia, 2025.



INSTITUTO FEDERAL
Goiás

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
SISTEMA INTEGRADO DE BIBLIOTECAS

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☒ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☐ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Leonardo Alves Messias

Matrícula: 20222011280100

Título do Trabalho: Inteligência Artificial Aplicada à Previsão do Potencial de Geração de Energia Solar

Autorização - Marque uma das opções

1. ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ____/____/____ (Embargo);
3. ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

- ☒ O documento está sujeito a registro de patente.
☐ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
☐ Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia, 30 / 07 / 2025
Local Data



Documento assinado digitalmente
LEONARDO ALVES MESSIAS
Data: 30/07/2025 10:14:08-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do Autor e/ou Detentor dos Direitos Autorais

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CÂMPUS GOIÂNIA
MESTRADO PROFISSIONAL EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

LEONARDO ALVES MESSIAS

Inteligência Artificial Aplicada à Previsão do Potencial de Geração de Energia Solar

Programa de Pós-Graduação *Stricto Sensu* em Tecnologia, Gestão e Sustentabilidade do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Dissertação de Mestrado Profissional. Área de Concentração: Tecnologia de Sistemas de Produção Limpa. Linha de Pesquisa: Energias Renováveis.

Orientador: Prof. Dr. Raphael de Aquino Gomes

Coorientador: Prof. Dr. José Luis Domingos

Goiânia, 2025.

M585i Messias, Leonardo Alves.

Inteligência artificial aplicada à previsão do potencial de geração de energia solar / Leonardo Alves Messias. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2025.
91 f.

Orientação: Prof. Dr. Raphael de Aquino Gomes.

Coorientação: Prof. Dr. José Luis Domingos.

Dissertação (mestrado) – Mestrado Profissional em Tecnologia, Gestão e Sustentabilidade, Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

1. Inteligência artificial. 2. Eficiência energética. 3. Energia fotovoltaica. I. Gomes, Raphael de Aquino (orientação). II. Domingos, José Luis (coorientação). III. Instituto Federal de Educação, Ciência e Tecnologia de Goiás. IV. Título.

CDD 006.3



INSTITUTO FEDERAL
Goiás

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS GOIÂNIA

PARECER 16/2025 - GYN-CMTGS/GYN-GPPEX/CP-GOIANIA/IFG

SERVIÇO PÚBLICO FEDERAL
INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
IFG CÂMPUS GOIÂNIA
MESTRADO EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

LEONARDO ALVES MESSIAS

INTELIGÊNCIA ARTIFICIAL APLICADA À PREVISÃO DO POTENCIAL DE GERAÇÃO DE ENERGIA SOLAR

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologia, Gestão e Sustentabilidade do Instituto Federal de Educação, Ciência e Tecnologia de Goiás como requisito parcial à obtenção do título de Mestre.

Área de Concentração: Tecnologia de Sistemas de Produção Limpa.

Linha de Pesquisa: Energias Renováveis

Aprovação em 25 de março de 2025.

Banca examinadora:

Presidente(a)/orientador(a) - Prof. Dr. Raphael de Aquino Gomes - PPGTGS IFG

Coorientador(a): Prof. Dr. Jose Luis Domingos - PPGTGS IFG

Avaliador(a): Prof.a Dr.a Patrícia Teixeira Leite Asano - UFABC

Avaliador(a): Prof. Dr. Thiago Augusto Mendes - PPGTGS IFG

Avaliador(a): Prof. Dr. Edipo Henrique Cremon - PPGTGS IFG

Documento assinado eletronicamente por:

- Patricia Teixeira Leite Asano, Patricia Teixeira Leite Asano - Docente - Fundacao Universidade Federal do Abc - Ufabac (07722779000106), em 30/07/2025 11:09:14.
- Jose Luis Domingos, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 28/04/2025 15:07:48.
- Edipo Henrique Cremon, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 23/04/2025 16:04:54.
- Thiago Augusto Mendes, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 26/03/2025 18:25:26.
- Raphael de Aquino Gomes, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 25/03/2025 16:57:24.

Este documento foi emitido pelo SUAP em 24/03/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifg.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 630973

Código de Autenticação: b5b93427cd



Dedicatória

Dedico este trabalho à memória do meu avô, Paulo Fernandes Messias, que sempre foi uma fonte de inspiração e apoio incondicional em minha vida. Sua presença continua viva em minhas lembranças e em cada conquista que realizo. Este é um pequeno tributo ao seu legado, que me motiva a seguir em frente com determinação e gratidão.

Agradecimentos

Agradeço a Deus, ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG), aos membros da banca pela valiosa contribuição. Minha gratidão ao Prof. Dr. Raphael de Aquino Gomes pela orientação e incentivo, assim como aos amigos e familiares pelo apoio incondicional. Por fim, estendo meus agradecimentos a todos que, de alguma forma, estiveram ao meu lado e colaboraram para a realização deste trabalho.

A ciência de hoje é a tecnologia de amanhã.

Edward Teller,
1908 – 2003.

Resumo

Título: Inteligência Artificial Aplicada à Previsão do Potencial de Geração de Energia Solar

Autor: Leonardo Alves Messias

Orientador: Dr. Raphael de Aquino Gomes

Coorientador: Prof. Dr. José Luis Domingos

A crescente preocupação com os impactos ambientais decorrentes da geração de energia a partir de combustíveis fósseis, associada ao agravamento do aquecimento global, tem impulsionado esforços para a transição rumo a uma matriz energética mais sustentável. Nesse cenário, a energia solar ocupa um papel estratégico, consolidando-se como uma das principais fontes renováveis. Neste contexto, o presente trabalho visa modelar e prever o potencial de geração de sistemas fotovoltaicos no estado de Goiás, utilizando dados meteorológicos combinados com registros reais de geração de sistemas fotovoltaicos existentes, envolvendo o desenvolvimento de modelos preditivos que serão comparados a abordagens consolidadas. Partindo dessa abordagem, busca-se contribuir para o avanço das estratégias de planejamento, otimização e gestão energética de sistemas fotovoltaicos em um contexto regional, fortalecendo o entendimento técnico-científico da integração entre variáveis meteorológicas e a eficiência na geração solar. Os resultados mostraram que o modelo Random Forest (RF) teve melhor desempenho na maioria das usinas, oferecendo previsões mais estáveis com janelas de 24 horas. O RF alcançou valores de R^2 acima de 0,90 e os menores erros médios (MAE e RMSE) em diversas usinas. Já o *Long Short-Term Memory* (LSTM) obteve bons resultados quando teve acesso a mais dados históricos, principalmente com janelas de 48 horas. Mesmo com estação meteorológica localizada a distâncias consideráveis das usinas variando de 151,9 km a 4,0 km o uso de variáveis climáticas ainda contribuiu positivamente para a melhoria das previsões.

Palavras-chave

Eficiência Energética, Previsão de Geração Solar, Inteligência Artificial

Abstract

Title: Artificial Intelligence Applied to Forecasting Solar Energy Generation Potential

Author: Leonardo Alves Messias

Advisor: Dr. Raphael de Aquino Gomes

Coadvisor: Prof. Dr. José Luis Domingos

The growing concern with environmental impacts from fossil fuel-based energy generation, coupled with the worsening of global warming, has driven efforts toward a more sustainable energy matrix transition. In this scenario, solar energy plays a strategic role, establishing itself as one of the main renewable sources. In this context, the present work aims to model and predict the generation potential of photovoltaic systems in the state of Goiás, Brazil, using meteorological data combined with real generation records from existing photovoltaic systems, involving the development of predictive models that will be compared to consolidated approaches. Based on this approach, we seek to contribute to the advancement of planning, optimization, and energy management strategies for photovoltaic systems in a regional context, strengthening the technical-scientific understanding of the integration between meteorological variables and solar generation efficiency. The results showed that the Random Forest (RF) model performed best in most plants, offering more stable predictions with 24-hour windows. RF achieved R^2 values above 0.90 and the lowest mean errors (MAE and RMSE) in several plants. On the other hand, Long Short-Term Memory (LSTM) obtained good results when it had access to more historical data, especially with 48-hour windows. Even with meteorological stations located at considerable distances from the plants, ranging from 151.9 km to 4.0 km, the use of climatic variables still contributed positively to the improvement of predictions.

Keywords

Energy Efficiency, Solar Generation Forecasting, Artificial Intelligence

Lista de Figuras

Figura 1 – Evolução da fonte solar fotovoltaica no Brasil	25
Figura 2 – Variação da radiação solar diária em Goiânia (GO) de 19/11/2023 a 29/11/2023	26
Figura 3 – Balanço meteorológico do mês de novembro de 2023 em Goiânia (GO)	27
Figura 4 – Influência da temperatura ambiente na tensão da placa	28
Figura 5 – Representação gráfica de uma floresta aleatória	33
Figura 6 – Representação simplificada do neurônio biológico	34
Figura 7 – Arquitetura de um bloco LSTM	35
Figura 8 – Funções de ativação em redes neurais	36
Figura 9 – Ajuste de hiperparâmetros usando validação cruzada	40
Figura 10 – Fluxograma de etapas da pesquisa	41
Figura 11 – Mapa geral de Goiás com localização dos sistemas fotovoltai- cos e estação meteorológica consideradas	46
Figura 12 – Valores de R^2 por janela de dados históricos - Cenário 1 - RF	59
Figura 13 – Valores de R^2 por janela de dados históricos - Cenário 1 - LSTM	60
Figura 14 – Valores de R^2 por janela de dados históricos - Cenário 2 - RF	61
Figura 15 – R^2 por janela de dados históricos - Cenário 2 - LSTM	62
Figura 16 – R^2 por janela de dados históricos - Cenário 3 - RF	63
Figura 17 – R^2 por janela de dados históricos - Cenário 3 - LSTM	65
Figura 18 – Comparação da dispersão dos valores de Ppv médio horário da Usina 1 para o maior R^2	66
Figura 19 – Comparação da dispersão dos valores de Ppv médio horário da Usina 2 para o maior R^2	66
Figura 20 – Comparação da dispersão dos valores de Ppv médio horário da Usina 3 para o maior R^2	66
Figura 21 – Comparação da dispersão dos valores de Ppv médio horário da Usina 4 para o maior R^2	67
Figura 22 – Comparação da dispersão dos valores de Ppv médio horário da Usina 5 para o maior R^2	67
Figura 23 – Comparação do resíduo dos valores de Ppv médio horário da Usina 1 para o maior R^2	69
Figura 24 – Comparação do resíduo dos valores de Ppv médio horário da Usina 2 para o maior R^2	69
Figura 25 – Comparação do resíduo dos valores de Ppv médio horário da Usina 3 para o maior R^2	70

Figura 26 – Comparação do resíduo dos valores de Ppv médio horário da Usina 4 para o maior R^2	70
Figura 27 – Comparação do resíduo dos valores de Ppv médio horário da Usina 5 para o maior R^2	71
Figura 28 – Comparação de valores reais e previstos da Usina 1 para o maior R^2	72
Figura 29 – Comparação de valores reais e previstos da Usina 2 para o maior R^2	73
Figura 30 – Comparação de valores reais e previstos da Usina 3 para o maior R^2	74
Figura 31 – Comparação de valores reais e previstos da Usina 4 para o maior R^2	75
Figura 32 – Comparação de valores reais e previstos da Usina 5 para o maior R^2	76

Lista de Tabelas

Tabela 1 – Análise das distâncias entre estações meteorológicas do INMET no estado de Goiás	44
Tabela 2 – Localização e capacidade dos sistemas fotovoltaicos utilizados na pesquisa	45
Tabela 3 – Dados gerais considerados no estudo	48
Tabela 4 – Configuração dos cenários de predição	49
Tabela 5 – Parâmetros definidos para o GridSearchCV, modelo RF	49
Tabela 6 – Parâmetros numéricos definidos para o GridSearchCV - Modelo LSTM	51
Tabela 7 – Parâmetros categóricos definidos para o GridSearchCV - Modelo LSTM	51
Tabela 8 – Correlação de Pearson entre dados de geração e dados climáticos	56
Tabela 9 – Melhores valores de RMSE por sistema fotovoltaico - RF	77
Tabela 10 – Melhores valores de MAE por sistema fotovoltaico - RF	78
Tabela 11 – Melhores valores de RMSE por sistema fotovoltaico - LSTM	79
Tabela 12 – Melhores valores de MAE por sistema fotovoltaico - LSTM	79

Lista de Quadros

Quadro 1 – Comparação entre ciência de dados e IA	31
---	----

Lista de Códigos de Programas

Código 1	–	<i>GridSearch</i> em <i>Python</i> para teste de hiperparâmetros, modelo RF.	50
Código 2	–	<i>Grid Search</i> em <i>Python</i> para teste de hiperparâmetros, modelo LSTM.	52
Código 3	–	Cálculo das métricas de erro.	53

Lista de Abreviaturas e Siglas

<i>tanh</i>	tangente hiperbólica
ANEEL	Agência Nacional de Energia Elétrica
CCEE	Câmara de Comercialização de Energia Elétrica
CEC	Carrossel de Erro Constante
IA	Inteligência Artificial
kWp	quilowatt pico
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MSE	Erro Quadrático Médio
MWp	megawatt pico
Ppv	Potência Fotovoltaica
RF	<i>Random Forest</i>
RMSE	<i>Root Mean Squared Error</i>
RN	Resolução Normativa
RNA	Rede Neural Artificial
RNN	<i>Recurrent Neural Network</i>

Lista de Símbolos

$b_{f,i,o,c}$	Vetores de viés das portas
c_t	Estado da célula no tempo t
$c^{(t)}$	Estado da célula do instante atual
$c^{(t-1)}$	Estado da célula do instante anterior
c_{t-1}	Estado da célula no tempo $t-1$
f_t	Ativação do portão de esquecimento
g	Função de ativação de entrada (usualmente tanh)
h	Função de ativação de saída (usualmente tanh)
h_t	Estado oculto no tempo t
h_{t-1}	Estado oculto no tempo $t-1$
i_t	Ativação do portão de entrada
j	Índice que percorre cada uma dessas entradas
m	Número total de entradas
o_t	Ativação do portão de saída
r	Coefficiente de correlação de Pearson
σ	Função de ativação de porta (sempre sigmoidal)
σ_c	Função de ativação tangente hiperbólica para atualização da célula
σ_g	Função de ativação sigmoide
σ_h	Função de ativação tangente hiperbólica para o estado oculto
u_k	Entrada ponderada do k -ésimo neurônio
$W_{f,i,o,c}$	Matrizes de pesos de entrada das portas
$U_{f,i,o,c}$	Matrizes de pesos recorrentes das portas
w_{kj}	Peso da conexão do j -ésimo neurônio para o k -ésimo neurônio
x_j	Entrada do j -ésimo neurônio
x_t	Entrada no tempo t
$x^{(t)}$	Entrada do instante atual
x_i, y_i	Valores individuais das variáveis no conjunto de dados x e y
$y^{(t)}$	Estado oculto do instante atual
$y^{(t-1)}$	Estado oculto do instante anterior
$\bar{x}, \bar{y}$	Médias dos conjuntos de dados x e y
$\odot$	Operação de produto Hadamard (elemento por elemento)
$\oplus$	Soma de todas as entradas
R^2	Coefficiente de Determinação
V	Tensão
P	Potência
I	Corrente

Sumário

1	Introdução	20
1.1	Objetivos.....	22
1.2	Estrutura do Trabalho	22
2	Fundamentação Teórica	24
2.1	Energia Elétrica Fotovoltaica	24
2.1.1	Cálculo de Geração Solar com Método Determinístico	27
2.1.2	Geração Distribuída	29
2.2	Tecnologias de Dados	30
2.2.1	Métodos de Inteligência Artificial	31
2.2.1.1	Floresta Aleatória	32
2.2.1.2	Redes Neurais Artificiais	33
2.2.2	Métricas de Avaliação de Erro de Desempenho	38
2.2.3	Validação Cruzada	39
3	Metodologia	41
3.1	Coleta de Dados	43
3.2	Pré-processamento de Dados.....	45
3.3	Definição dos Cenários de Predição	47
3.4	Construção do Modelo RF	48
3.5	Construção do Modelo LSTM	50
3.6	Geração das Métricas de Avaliação	52
3.7	Aplicação da Metodologia Proposta	53
3.7.1	Modelo RF	54
3.7.2	Modelo LSTM	54
4	Resultados e Discussão	56
4.1	Impacto dos Fatores Climáticos na Geração Solar	56
4.2	Avaliação do Desempenho dos Modelos nos Cenários Testados	57
4.2.1	Cenário 1	58
4.2.1.1	Análise do Desempenho dos Modelos RF e LSTM	58
4.2.2	Cenário 2	60
4.2.2.1	Análise do Desempenho dos Modelos RF e LSTM	60
4.2.3	Cenário 3	63
4.2.3.1	Análise do Desempenho dos Modelos RF e LSTM	63
4.2.4	Comparação entre os Modelos.....	65

5	Considerações Finais	80
5.1	Conclusão	80
5.2	Trabalhos Futuros.....	82
	Referências	83

Introdução

A transição para fontes de energia renováveis tem se tornado um imperativo global diante dos desafios ambientais e climáticos do século XXI. A dependência histórica de combustíveis fósseis para geração de energia elétrica além de contribuir para a emissão de gases de efeito estufa, como coloca em risco a sustentabilidade dos ecossistemas e a qualidade de vida das gerações futuras.

Nesse cenário, a diversificação da matriz elétrica mundial surge como uma estratégia fundamental para mitigar esses impactos, com destaque para o potencial das fontes renováveis, em especial a energia solar fotovoltaica. No Brasil, onde a matriz elétrica é predominantemente hidrelétrica, representando 48,4% do total da geração (ONS, 2024), a variabilidade climática e os desafios ambientais têm incentivado a busca por alternativas complementares, como a geração fotovoltaica, que se consolida como uma das principais fontes para ampliar a capacidade energética de forma sustentável.

O relatório apresentado pela Câmara de Comercialização de Energia Elétrica (CCEE) apresentou dados de 2022 que mostram um aumento de 72,6% na geração de energia provenientes de sistemas fotovoltaicos em relação ao período homólogo ao ano anterior (Casarin, 2022). O aumento da geração fotovoltaica também foi estimulado pela Resolução Normativa (RN) n.º 482/2012 (ANEEL, 2012), aprovada em 17 de abril de 2012, na qual a Agência Nacional de Energia Elétrica (ANEEL) estabeleceu os requisitos gerais de acesso aos sistemas de geração distribuída de micro e minigeração e ao sistema de compensação de energia elétrica, bem como deliberou requisitos gerais para acesso aos sistemas de geração distribuída, além de outras disposições necessárias.

A RN n.º 482 foi revogada pela RN n.º 1.000/2021 (ANEEL, 2021), e posteriormente pela RN n.º 1.059/2023 (ANEEL, 2023) com objetivo de reduzir o tempo de conexão e os custos e aumentar a abrangência da geração distribuída aos consumidores nas modalidades autoconsumo remoto e geração compartilhada. Somado a isto, também modificou os valores de potência que definem a microgeração até 75 quilowatt pico (kWp) e minigeração potência até 5 megawatt pico (MWp).

Dessa forma, conforme apontado por Gasparin (2022), as políticas públicas têm sido fundamentais para o crescimento da geração de energia solar no Brasil, uma vez que incentivos fiscais e programas de subsídios foram fundamentais para o desenvolvimento da indústria voltada à energia solar.

Devido ao aumento considerável de novas instalações, são necessários estudos voltados à eficiência energética de sistemas fotovoltaicos instalados, a fim de identificar possíveis problemas causados por materiais particulados, como poeira, fuligem, poluentes atmosféricos e resíduos orgânicos, ou falhas eletrônicas. A composição do ar no qual o equipamento estará em constante contato pode transportar, além de nitrogênio, oxigênio e gás carbônico, materiais sólidos como: poeira, ácaros, folhas e galhos. Além disso, o acúmulo de excremento de pássaros em módulos solares pode prejudicar severamente a geração, podendo danificar a placa e ocasionar avaria e mau funcionamento ao equipamento (Barbosa; Faria; Gontijo, 2018).

A quantidade de irradiação solar e a temperatura na qual o módulo estará exposto são fatores determinantes para a eficiência dos módulos fotovoltaicos (Ruviano *et al.*, 2018). Desta forma, esses elementos podem ser usados para estimar a eficiência, permitindo antecipar possíveis problemas e atenuar o desperdício de energia e, por sua vez, diminuir custos.

A análise da eficiência de sistemas fotovoltaicos pode ser tratada como um problema de previsão de geração, que se refere à capacidade de estimar o potencial energético com base em condições climáticas e operacionais (Chandola; Banerjee; Kumar, 2009). Para lidar com problemas deste tipo, uma das estratégias mais comumente utilizadas atualmente é o uso de Inteligência Artificial (IA) (Nassif *et al.*, 2021).

Dentre os trabalhos que utilizam métodos baseados em IA para prever a geração solar fotovoltaica, Haupt *et al.* (2020) apresentam um estudo realizado em um parque de energia renovável localizado no Kuwait, que avalia dados meteorológicos e de geração fotovoltaica por meio de aprendizado de máquina, a fim de prever a produção energética a curto prazo.

Sarp *et al.* (2021) trazem alguns recursos de análise de dados para prognosticar a geração fotovoltaica realizando cruzamento de informações entre ferramentas distintas, com objetivo de validar os resultados de previsão obtidos.

Conforme discutido por Tieghi *et al.* (2025), a utilização de modelos de IA tem sido extensivamente investigada na literatura para melhorar a capacidade de previsão de sistemas de geração de energia renovável, sobretudo usando as técnicas de Floresta Aleatória - *Random Forest* (RF) e redes de Memória Longa de Curto Prazo - *Long Short-Term Memory* (LSTM). Os estudos destacam que, apesar dos avanços, há espaço para investigar modelos que utilizem dados integrados de diferentes

sistemas fotovoltaicos e explorem sua aplicabilidade em cenários regionais.

Este trabalho se propõe a abordar limitações que ainda são pouco exploradas. Um dos principais aspectos investigados é o impacto da distância entre a usina fotovoltaica e a estação meteorológica no desempenho preditivo da geração de energia. Esse fator, frequentemente negligenciado, pode influenciar significativamente a confiabilidade das estimativas dos modelos preditivos, especialmente em regiões com pouca cobertura de estações meteorológicas. Além disso, o estudo busca analisar como a degradação do desempenho dos modelos ocorre à medida que o horizonte de previsão aumenta, investigando a relação entre a qualidade das previsões e diferentes janelas temporais. Outro ponto crítico abordado é a dependência excessiva de grandes volumes de dados meteorológicos na maioria dos modelos existentes. Para superar essa limitação, propõe-se uma abordagem inovadora que permita a previsão da geração de energia mesmo em cenários com dados meteorológicos limitados ou ausentes, ampliando a aplicabilidade dos modelos em regiões com infraestrutura de medição insuficiente.

1.1 Objetivos

O objetivo geral deste trabalho é avaliar modelos preditivos para estimar o potencial de geração de energia solar em sistemas fotovoltaicos, utilizando técnicas de IA aplicadas a dados meteorológicos e históricos de geração. Para atingir este objetivo, são propostos os seguintes objetivos específicos:

- Examinar modelos de IA para previsão do potencial de geração solar, avaliando cenários em que são trabalhados dados meteorológicos combinados a registros reais de geração de sistemas fotovoltaicos existentes;
- Comparar o desempenho de diferentes modelos preditivos com técnicas consolidadas, avaliando métricas de avaliação de erro de desempenho;
- Analisar a influência de variáveis meteorológicas e da distância entre a estação de coleta de dados e a fonte de geração sobre a qualidade das previsões;
- Avaliar a aplicabilidade dos modelos em cenários regionais.

1.2 Estrutura do Trabalho

O restante do trabalho é organizado conforme apresentado a seguir.

O Capítulo 2, dedicado à fundamentação teórica, visa fornecer uma base para os conceitos abordados ao longo do estudo. São discutidos temas como geração de energia por sistemas fotovoltaicos e ciência de dados.

O Capítulo 3 detalha as etapas metodológicas, incluindo a preparação dos dados, a seleção dos algoritmos, e o uso de validação cruzada. São descritos os procedimentos para a coleta de dados meteorológicos e a configuração das janelas de dados históricos, além das métricas de avaliação aplicadas.

O Capítulo 4 é dedicado à análise exploratória dos dados, incluindo a segmentação dos mesmos e a aplicação dos algoritmos. Também são apresentados e comparados os resultados obtidos.

Por fim, o Capítulo 5 aborda as conclusões e os resultados alcançados através do estudo. Este capítulo também inclui observações importantes identificadas durante o desenvolvimento da pesquisa.

Fundamentação Teórica

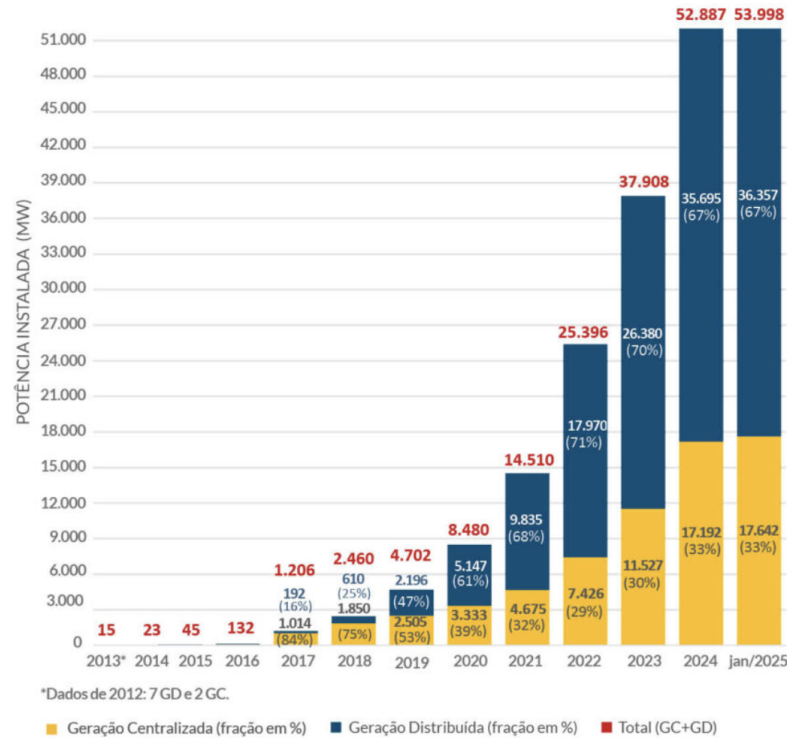
Pretende-se neste capítulo apresentar a fundamentação teórica necessária para atingir os objetivos propostos neste trabalho. São abordados conceitos vinculados a energia solar fotovoltaica, ciência de dados e tópicos complementares.

2.1 Energia Elétrica Fotovoltaica

As fontes de energia renováveis vêm crescendo a nível mundial significativamente, trazendo consigo debates sobre a produção de energia limpa (Taveira, 2023). Diante desse cenário, as dimensões territoriais e as elevadas taxas de irradiação solar no Brasil favorecem um potencial expressivo de geração fotovoltaica. Os cenários apresentados por Pereira *et al.* (2017) indicam o grande potencial energético ainda pouco explorado existente no país, devido à abundância de irradiação solar, à infraestrutura elétrica favorável e à redução dos custos das tecnologias fotovoltaicas. Esse crescimento também é evidenciado pela evolução contínua e crescente da adoção de sistemas fotovoltaicos no país, com um aumento significativo na capacidade instalada e no desenvolvimento de tecnologias solares mais eficientes, conforme destacado pela Figura 1.

A energia solar fotovoltaica é caracterizada pela conversão direta da radiação solar em eletricidade. O efeito fotovoltaico, descoberto por Becquerel em 1839, descreve a conversão de radiação solar em eletricidade, um processo realizado por células solares fotovoltaicas. Esse efeito acontece em semicondutores, que possuem bandas de energia onde os elétrons podem se mover. Ao adicionar átomos como o fósforo, que tem um elétron a mais, facilita-se a movimentação desses elétrons para a banda de condução com pouca energia térmica. O fósforo funciona como um dopante doador de elétrons, conhecido como dopante tipo *N*. Este processo ocorre por meio de um dispositivo chamado célula fotovoltaica, que opera com base no princípio do efeito fotoelétrico ou fotovoltaico. Segundo Imhoff *et al.* (2007), a célula fotovoltaica é a tecnologia central que permite essa transformação, aproveitando eficientemente a energia solar para gerar energia elétrica de maneira sustentável e renovável.

Figura 1 – Evolução da fonte solar fotovoltaica no Brasil



Fonte: ABSOLAR (2024).

As células fotovoltaicas são formadas por dois semicondutores de cargas opostas, separados por uma junção neutra. A camada tipo *N* tem um excesso de elétrons, enquanto a camada tipo *P* tem uma deficiência de elétrons. Na junção *P/N*, a exposição à luz solar libera elétrons da camada *N*, criando uma diferença de potencial que gera corrente elétrica em um circuito fechado, como descreve Nascimento (2004).

A maioria das células fotovoltaicas atuais é feita de silício, devido à sua abundância e custo acessível. Comercialmente, elas apresentam-se como lâminas circulares ou quadradas, de tonalidade azul-escuro a preta, e a parte superior contém raias cinzentas de material condutor para extrair a eletricidade produzida (Zilles, 2012).

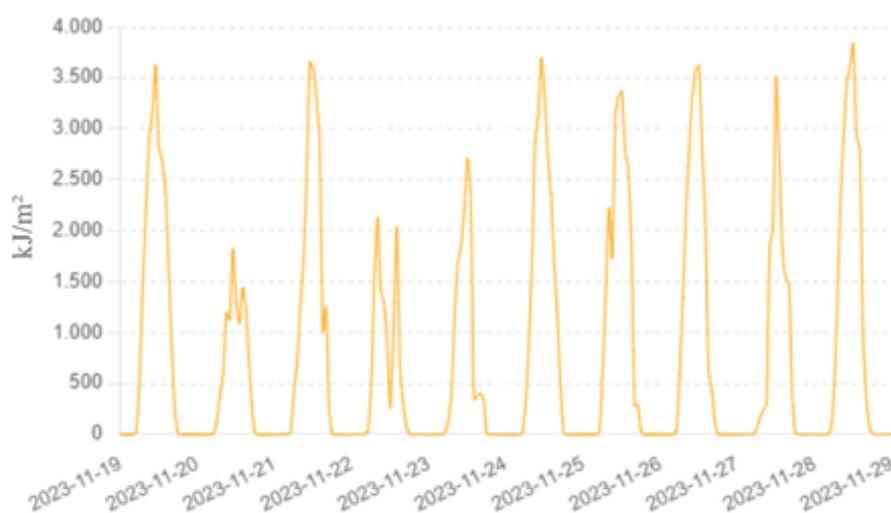
Existem três tipos principais de células de silício: monocristalino, policristalino e amorfo. As células monocristalinas são as mais eficientes, porém mais caras e demoradas para fabricar. As policristalinas são mais econômicas, mas menos eficientes. Já as células de silício amorfo, apesar de menos eficientes, são mais baratas e flexíveis (Torres *et al.*, 2019).

A geração de energia solar é amplamente afetada por variáveis meteorológicas como radiação solar, nebulosidade, umidade, chuva e temperatura, além de fatores como poluição e envelhecimento dos painéis. Estudos como os de Bonkaney

et al. (2017) e Wang, Qi e Liu (2019) ressaltam de forma significativa essa dependência, evidenciando como fatores climáticos, características do sistema e a qualidade dos dados influenciam diretamente a confiabilidade das estimativas nas previsões de energia solar. Além disso, os padrões de geração variam significativamente conforme o clima: em dias ensolarados, há uma alta e estável geração de energia, enquanto em dias nublados ou chuvosos, a produção é baixa e variável, refletindo a sensibilidade dos sistemas fotovoltaicos às condições ambientais (Cunha, 2021).

Além disso, a geração de energia em usinas solares pode ser comparada e analisada em períodos específicos para avaliar o desempenho e identificar tendências. Na Figura 2 é demonstrado o comportamento da radiação solar ao longo do tempo, em Goiânia/GO, destacando as variações diárias.

Figura 2 – *Variação da radiação solar diária em Goiânia (GO) de 19/11/2023 a 29/11/2023*

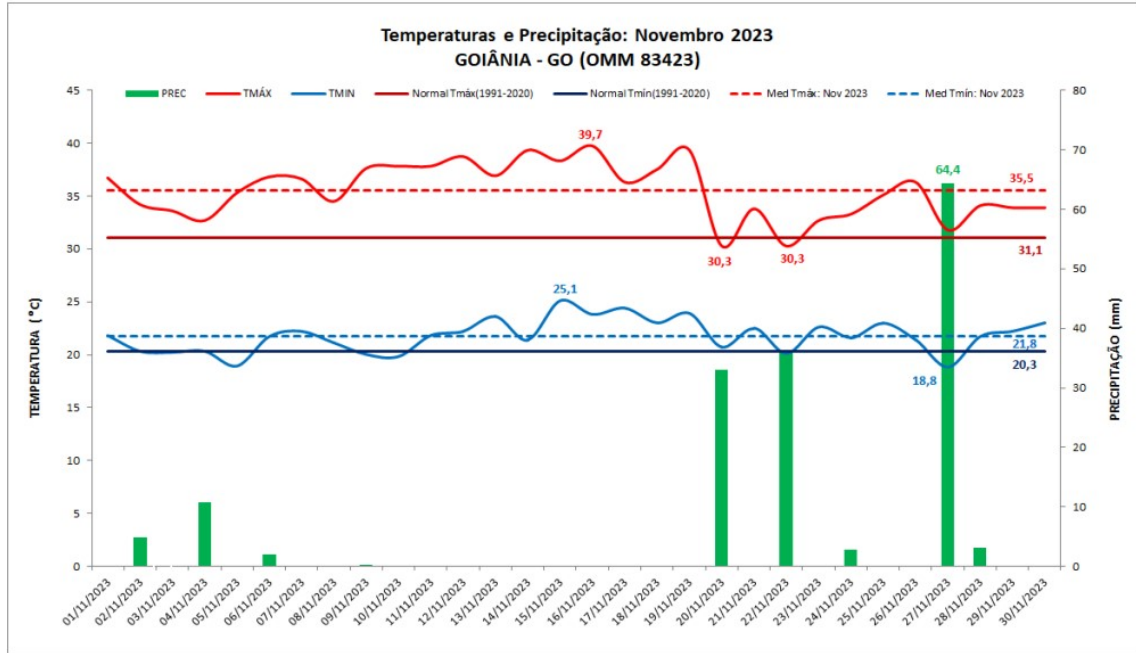


Fonte: Elaborado pelo autor (2024).

O relatório meteorológico de Goiânia para novembro de 2023, apresentado pelo INMET (2023), fornece informações detalhadas sobre precipitação e temperatura, fatores críticos para entender as variações nos dias 21, 23 e 24 de novembro (Figura 3). Nota-se que no dia 27 houve alta precipitação, com chuvas principalmente à noite.

Existe também uma relação notável entre a temperatura ambiente e a irradiação solar. Assim, à medida que a temperatura aumenta, há uma diminuição da tensão e, por consequência, uma redução na potência gerada. Essa perda de potência devido à temperatura é resultado da queda na tensão (V) (Ruviano *et al.*, 2018). Conforme a 1ª Lei de Ohm (Equação 1), essa relação resulta em uma menor produção de energia. Essa relação pode ser observada nas especificações técnicas

Figura 3 – *Balanço meteorológico do mês de novembro de 2023 em Goiânia (GO)*



Fonte: INMET (2023).

dos módulos fotovoltaicos, como o apresentado por Kyocera (2024) e ilustrado pela Figura 4.

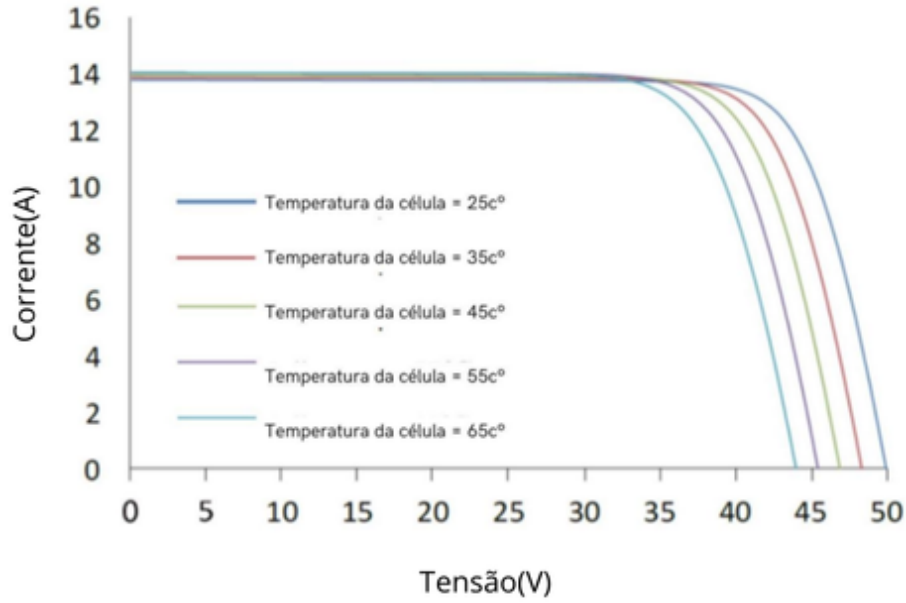
$$P = V \cdot I \quad (1)$$

Portanto, a sujidade nos módulos fotovoltaicos pode aumentar a temperatura dos mesmos, agravando ainda mais a redução da eficiência. A presença de poeira, detritos ou poluentes na superfície dos módulos reduz a absorção de luz solar e pode causar aquecimento localizado, conhecido como *hot spots*, que não só diminui a eficiência geral, como também pode danificar permanentemente os módulos (Salaazar; Macabebe, 2016). Adicionalmente, Araújo, Carvalho e Dupont (2019) relatam que a acumulação de sujeira pode ocasionar perdas significativas na produção de energia, chegando a reduzir a eficiência dos sistemas fotovoltaicos em até 16,52% em períodos prolongados sem chuva, devido ao bloqueio parcial da irradiação solar incidente sobre as células fotovoltaicas.

2.1.1 Cálculo de Geração Solar com Método Determinístico

Soares e Rech (2005) apresentam fórmulas necessárias para calcular a geração de energia solar. O modelo é desenvolvido a partir do conceito de dimensionamento do sistema fotovoltaico, considerando fatores como radiação solar, rendimento

Figura 4 – *Influência da temperatura ambiente na tensão da placa*



Fonte: (RENO Solar, 2024).

do sistema e perdas.

Inicialmente, a energia gerada por um sistema fotovoltaico depende da potência do gerador (PPV), dimensionada com base no consumo diário de energia (W) e no rendimento do sistema (η), conforme Equação 2:

$$PPV = \frac{W}{\eta \cdot Z} \quad (2)$$

onde:

- W : Energia consumida diariamente em Wh.
- η : Rendimento do sistema (ex.: 0,76 para 76%).
- Z : Fator de correção composto por três elementos:
 - Z_2 : Horas médias de radiação solar por dia (ex.: 3,5 h/dia no inverno);
 - Z_3 : Correção pelo ângulo de inclinação da placa (ex.: 1,0 se corretamente ajustado);
 - Z_4 : Correção por temperatura da célula (ex.: 0,88 para condições de operação desfavoráveis).

Exemplo: Para o inverno, $Z = Z_2 \times Z_3 \times Z_4 = 3,5 \times 1,0 \times 0,88 = 3,08$ (adimensional).

Após essa etapa, é possível determinar o número de painéis necessários (Np) com base na potência nominal de cada painel (PPF) (Equação 3):

$$Np = \frac{PPV}{PPF} \quad (3)$$

Essa relação entre a potência total (P_{pv}) e a potência individual dos módulos permite que a geração total do sistema seja associada diretamente à sua potência nominal (P_{nom}).

Por fim, ao reorganizar as variáveis e considerar que a energia gerada (E) é proporcional à radiação solar média (G) e ao desempenho do sistema (representado por PR , ou *Performance Ratio*), chega-se à Equação 4:

$$E = P_{nom} \cdot G \cdot PR \quad (4)$$

onde:

- P_{nom} : Potência nominal do sistema (soma das potências individuais dos painéis em Wp).
- G : Irradiação solar média diária (kWh/m².dia), obtida para o local e inclinação do sistema.
- PR : Fator que reflete as perdas do sistema e a eficiência operacional (ex.: 0,75 a 0,85).

Uma limitação do modelo determinístico é a sua dependência de parâmetros fixos para estimar a geração de energia solar. Essa abordagem permite capturar variações dinâmicas, como oscilações climáticas inesperadas, degradação dos painéis ao longo do tempo ou padrões complexos de sombreamento, contribuindo para previsões mais detalhadas. No entanto, essa mesma sensibilidade a múltiplos fatores pode também ampliar a propagação de incertezas e ruídos nos dados, resultando em desvios significativos nas previsões e impactando negativamente a confiabilidade e a eficiência do sistema (Lima; Gomes; Martinez, 2020; Quinelato *et al.*, 2021; Carvalho; Fantini; Siqueira, 2022).

Em contrapartida, abordagens baseadas em IA podem melhorar a previsão da geração solar ao incorporar um maior volume de dados históricos e em tempo real. Estas técnicas podem identificar padrões ocultos na radiação solar, temperatura e condições atmosféricas, ajustando dinamicamente as previsões.

2.1.2 Geração Distribuída

A implementação da RN n.º 482 facilitou ao consumidor brasileiro produzir sua própria energia elétrica usando fontes renováveis ou cogeração qualificada, podendo até mesmo distribuir o excesso para a rede local de distribuição. Isso se

refere à microgeração e minigeração distribuída de energia, inovações que combinam economia financeira, responsabilidade socioambiental e sustentabilidade.

A geração distribuída de energia elétrica é caracterizada por duas modalidades principais: a microgeração e a minigeração distribuída. A microgeração é definida como centrais geradoras com potência até 75 kWp, que empregam cogeração qualificada ou fontes renováveis de energia, conectadas à rede de distribuição por unidades consumidoras. Por outro lado, a minigeração distribuída abrange centrais com potência superior a 75 kWp, mas não excedendo 3 MWp para fontes hídricas e 5 MWp para cogeração qualificada ou outras fontes renováveis, integradas à rede da mesma forma, conforme RN n.º 1.059 (ANEEL, 2023).

No contexto de geração distribuída também é importante compreender a diferença entre sistemas de energia solar *on-grid* e *off-grid* pois estes apresentam características e aplicações distintas, cada um sendo mais adequado para diferentes cenários e necessidades.

O sistema *on-grid* é conectado à rede elétrica pública, permitindo que a energia solar não consumida seja enviada para as concessionárias de energia. Isso resulta na transformação dessa energia em créditos na conta de energia.

Por outro lado, o sistema *off-grid* é autônomo e desconectado da rede elétrica pública. Ele utiliza baterias e controladores de carga para armazenar a energia excedente gerada, proporcionando uma fonte de energia contínua e independente. Este tipo de sistema é ideal para áreas rurais ou isoladas onde a conexão com a rede elétrica é inexistente, ou pouco confiável. Além disso, é uma solução prática para locais que enfrentam interrupções frequentes de energia, garantindo um fornecimento constante e seguro de eletricidade. A robustez e a autonomia dos sistemas *off-grid* os tornam especialmente adequados para aplicações em comunidades remotas, instalações agrícolas e outras situações onde a estabilidade da rede elétrica não pode ser garantida (Bortoloto *et al.*, 2017).

2.2 Tecnologias de Dados

As tecnologias de dados englobam um amplo conjunto de métodos e ferramentas que utilizam dados para gerar indicadores, tomar decisões informadas e automatizar processos complexos. Este campo integra tanto a ciência de dados, que se concentra na coleta, análise e interpretação de grandes volumes de dados, quanto a Inteligência Artificial, que utiliza algoritmos e modelos para replicar capacidades cognitivas humanas, como aprendizado e reconhecimento de padrões.

Ciência de dados é a área que extrai valor de dados, impulsionada pela necessidade de processar, armazenar e interpretar grandes volumes de informação

gerados pela revolução tecnológica (Kotu; Deshpande, 2018). O seu objetivo é investigar e ajudar a explicar eventos e situações do dia a dia usando dados e buscando descobrir detalhes ou padrões (Hayashi, 1998). Portanto, essa ciência aplica técnicas e modelos analíticos a dados para produzir resultados probabilísticos. Em contraste, a IA emprega métodos da ciência de dados aplicada, além de outros algoritmos, para desenvolver e operar sistemas complexos que simulam a inteligência humana (AWS, 2024). Uma comparação entre as duas tecnologias é apresentada no Quadro 1, sendo que, neste trabalho, ambas serão utilizadas de forma complementar para tratar dados históricos e meteorológicos e construir modelos preditivos aplicados à estimativa de geração de energia fotovoltaica.

Quadro 1 – Comparação entre ciência de dados e IA

Item	Ciência de dados	Inteligência artificial
O que é?	O uso de modelagem estatística e algorítmica para obter <i>insights</i> a partir de dados.	Um termo de amplo espectro para aplicações baseadas em máquinas que imitam a inteligência humana.
Mais adequada para	Responder a uma pergunta a partir de um conjunto de dados.	Concluir uma tarefa humana complexa com eficiência.
Métodos	Regressão linear, regressão logística, detecção de anomalias, classificação binária, agrupamento <i>k-means</i> , análise de componentes principais e muito mais.	Reconhecimento facial, processamento de linguagem natural, aprendizado por reforço, gráficos de conhecimento, IA generativa e muito mais.
Escopo	Perguntas predefinidas que podem ser respondidas a partir dos dados.	Amplo e difícil de definir baseado em tarefas.
Implementação	Usa uma variedade de ferramentas diferentes para capturar, limpar, modelar, analisar e relatar dados.	Depende da tarefa. Normalmente, depende de componentes complexos, pré-fabricados e produzidos.

Fonte: adaptado de AWS (2024).

2.2.1 Métodos de Inteligência Artificial

Nesta seção são discutidos detalhes de dois métodos de IA utilizados neste trabalho: RF e LSTM, sendo este último um modelo derivado do conceito de Rede Neural Artificial (RNA). Cada uma dessas técnicas possui características distintas que as tornam adequadas para diferentes tipos de problemas. Enquanto a RF é co-

nhecida por sua simplicidade e robustez na criação de modelos preditivos por meio da combinação de múltiplas árvores de decisão, as RNAs são inspiradas no funcionamento do cérebro humano. Ambos métodos lidam bem com problemas complexos e não lineares, caracterizados por relações entre variáveis que não seguem proporções diretas ou constantes e que podem envolver múltiplas interações simultâneas, dificultando a modelagem por métodos estatísticos tradicionais.

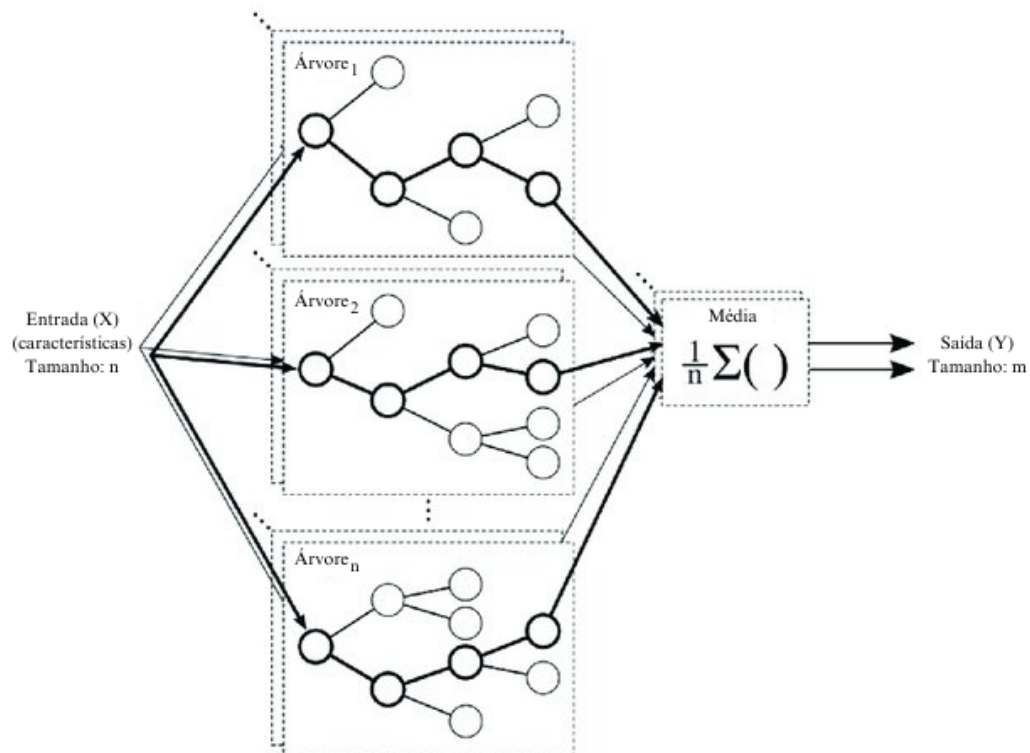
2.2.1.1 Floresta Aleatória

Comumente chamado RF, é um dos algoritmos de ML mais utilizados devido à sua simplicidade, pois opera por meio da combinação de múltiplas árvores de decisão independentes e da média de seus resultados, o que facilita a implementação e a interpretação dos modelos gerados (Osisanwo *et al.*, 2017). Desenvolvido por Breiman (2001), baseia-se na combinação de múltiplas árvores de decisão para aumentar o desempenho preditivo e reduzir o sobreajuste dos modelos preditivos (Breiman, 2001). Cada árvore segue uma lógica sequencial de “se-então” baseada nas características observadas. Essa estratégia não somente aumenta o desempenho preditivo na maioria dos cenários, como também reduz de forma significativa o risco de sobreajuste, contribuindo para a robustez e a estabilidade do modelo diante de diferentes conjuntos de dados (Breiman, 2001).

Como descrevem James *et al.* (2023), RF visa minimizar a variabilidade das previsões ao criar várias árvores de decisão independentes e combinar seus resultados. Além disso, o método destaca-se pela facilidade de treinamento e versatilidade prática.

A Figura 5 ilustra como uma árvore de decisão divide um espaço de entrada de duas dimensões em regiões menores para facilitar a classificação dos dados. Este processo é chamado de partição binária recursiva. Em cada passo, a árvore de decisão escolhe uma característica (um eixo) e um valor de corte para dividir os dados em duas partes. Cada divisão é feita paralelamente a um dos eixos, resultando em subespaços que são mais simples de analisar.

No exemplo da Figura 5, cada árvore recebe um subconjunto dos dados de entrada (X), realizando divisões sucessivas para gerar previsões individuais. A saída final (Y) é obtida por meio da média das predições de todas as árvores no caso de regressão, ou pelo voto majoritário no caso de classificação. Esse método melhora a robustez e diminui o erro médio do modelo, reduzindo o risco de sobreajuste e melhorando a generalização para novos dados. Além disso, a estrutura hierárquica da árvore de decisão permite que ela lide com dados complexos e identifique padrões de maneira estruturada.

Figura 5 – Representação gráfica de uma floresta aleatória

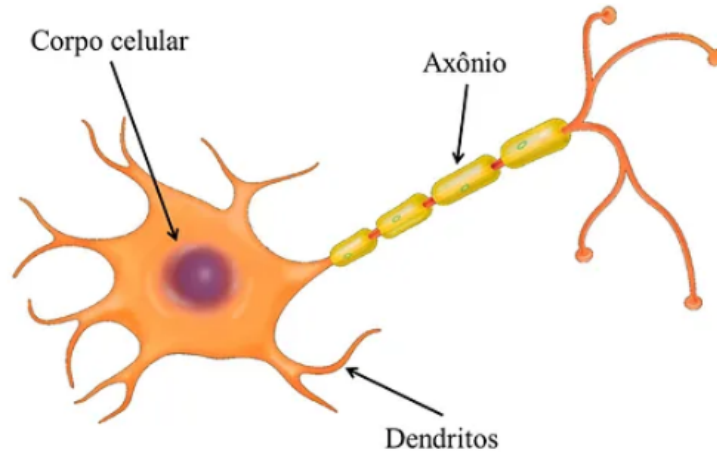
Fonte: adaptado de Parejo *et al.* (2021).

2.2.1.2 Redes Neurais Artificiais

Com o desenvolvimento do poder de processamento dos computadores e o aprimoramento de técnicas em aprendizado de máquina, diversos desafios de classificação e regressão são agora resolvidos de maneira eficiente e rápida. Neste contexto, as Redes Neurais Artificiais (RNAs) emergem como uma técnica proeminente, destacando-se pela sua capacidade de abordar problemas complexos e não lineares.

De acordo com Haykin (2001), uma RNA é um processador paralelo composto por unidades de processamento simples, capazes de adquirir e armazenar conhecimento para uso futuro. Inspiradas no funcionamento do cérebro humano, as RNA visam modelar o processo de aprendizado e resolução de problemas complexos do cérebro (Silva Filho, 2007).

O cérebro e o sistema nervoso central são, em parte, formados por neurônios. Como pode ser visto na Figura 6, estes neurônios possuem dendritos que recebem impulsos elétricos de outros neurônios e os transmitem através do corpo celular até os axônios. Através das sinapses, esses axônios podem ativar ou inibir a resposta do neurônio receptor a estímulos específicos, como apresenta Silva, Spatti e Flauzino (2010).

Figura 6 – *Representação simplificada do neurônio biológico*

Fonte: (Mattias, 2023).

Tendo como base a representação simplificada do neurônio biológico, em 1958, Frank Rosenblatt apresentou o Perceptron (Rosenblatt, 1958), considerado um dos primeiros modelos de RNA. Influenciado pelos estudos anteriores de Walter Pitts e Warren Sturgis McCulloch, o perceptron é reconhecido como o pioneiro dos modelos de neurônios computacionais, apresentando uma estrutura que emula a função de um neurônio biológico. Este modelo é uma representação simplificada do mecanismo neuronal biológico (Grübler, 2024). A Equação 5 apresenta a maneira pela qual os sinais de entrada x_j são transformados em um sinal de saída. Cada entrada x_j é multiplicada por um peso sináptico w_{kj} , único para cada conexão entre os neurônios. A soma destes produtos ponderados reflete o potencial total de ativação u_k de um neurônio específico. Aqui, m representa o número total de entradas, e j é o índice que percorre cada uma dessas entradas.

$$u_k = \sum_{j=1}^m w_{kj} \cdot x_j \quad (5)$$

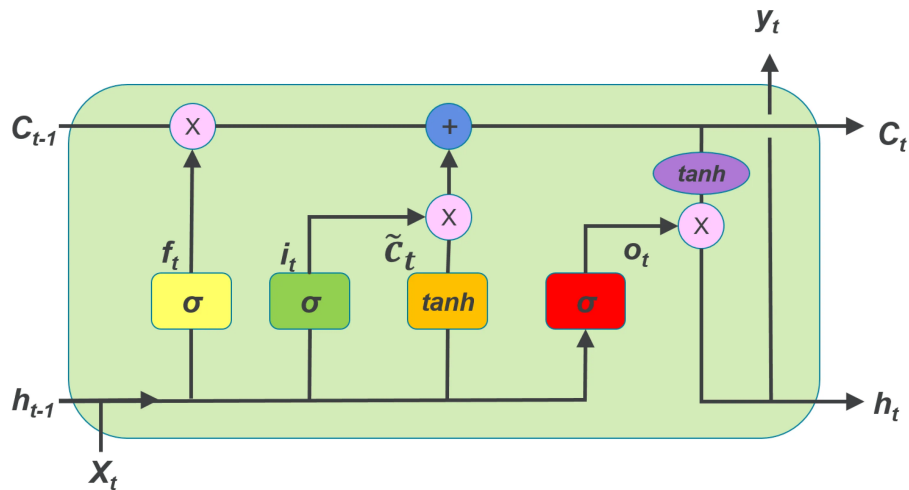
Um dos modelos de RNA para tarefas de classificação de padrões onde as entradas e saídas são sequências é a Rede Neural Recorrente, ou *Recurrent Neural Network* (RNN). Este modelo utiliza um vetor de alta dimensão (estado oculto) de dimensionalidade fixa que incorpora novas observações por meio de uma função não-linear complexa. Essas redes são extremamente expressivas e capazes de implementar cálculos limitados pela memória, o que permite alcançar desempenhos notáveis em tarefas sequenciais. No entanto, as RNNs são difíceis de treinar, especialmente em problemas com estruturas temporais de longo alcance complexas (Medsker; Jain, 1999). Esta dificuldade é acentuada pelo problema de gradientes que desaparecem ou explodem, um desafio técnico que compromete a capacidade das RNN de aprender

eficazmente em contextos de dependências temporais longas.

Para abordar esses desafios, foi desenvolvido o modelo Memória Longa de Curto Prazo ou LSTM, uma forma avançada de RNN projetada especificamente para superar os problemas de explosão e desaparecimento de gradientes. Utilizando um Carrossel de Erro Constante (CEC), o LSTM mantém o sinal de erro estável dentro de cada célula da unidade, permitindo que estas lembrem valores por períodos extensos. A inclusão de uma porta de esquecimento, uma inovação subsequente, permite que a rede redefina seu estado quando necessário.

Essencialmente, a arquitetura do LSTM é composta por blocos de memória conectados recorrentemente que regulam o fluxo de informações por meio de unidades de portão não lineares (Divyanshu, 2018), conforme ilustrado na Figura 7. Uma conexão básica nos modelos de LSTM consiste na transmissão direta de informação entre dois pontos, enquanto uma conexão de *peepholes* é um mecanismo específico das LSTM que permite ao estado da célula influenciar diretamente os Portões de Ativação (*Activation Gate*). O estado da célula, denotado por C_{t-1} para o instante anterior e C_t para o instante atual, serve como um núcleo de memória que carrega informações ao longo do tempo, sendo atualizado a cada passo para incorporar novos dados e esquecer informações irrelevantes. Paralelamente, o estado oculto, representado por h_{t-1} e h_t para os instantes anterior e atual, respectivamente, funciona como a saída do modelo em cada passo de tempo, influenciado tanto pelo estado atual da célula quanto pela nova entrada X_t .

Figura 7 – Arquitetura de um bloco LSTM

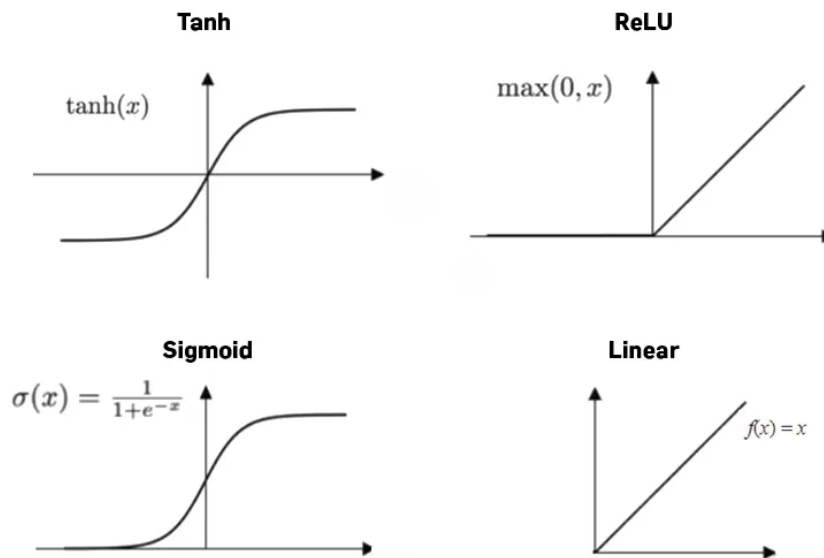


Fonte: (Divyanshu, 2018).

As operações fundamentais em uma célula LSTM incluem a multiplicação, denotada por $\times$, a qual é uma operação elemento a elemento usada para combinar vetores seletivamente, e a soma $+$, que agrega todas as entradas de maneira vetorial.

Essenciais para a modulação dos fluxos de informação são as funções de ativação: a função sigmoideal σ , utilizada nos portões da célula para decidir a quantidade de informação a ser passada, produzindo valores entre 0 e 1; a função de ativação de entrada $\tilde{C}_t$, geralmente uma tangente hiperbólica ($\tanh$), que ajusta os valores de entrada para um intervalo de -1 a 1, facilitando a adição de novas informações ao estado da célula; e a função de ativação de saída o_t , que determina quais partes do estado da célula serão transmitidas como o estado oculto h_t . A Figura 8 apresenta estas funções.

Figura 8 – Funções de ativação em redes neurais



Fonte: (Udit, 2023).

As funções de ativação σ e $\tanh$ são fundamentais para regular o fluxo de informações, mantendo os valores dentro de limites que ajudam a prevenir problemas numéricos, como o desaparecimento ou a explosão dos gradientes durante o treinamento da rede (Minami, 2018). A função sigmoideal σ , apesar de amplamente utilizada nos portões das LSTM, pode contribuir para o problema do *vanish gradient* devido à saturação nos extremos, onde sua derivada se aproxima de zero. No entanto, a arquitetura LSTM mitiga esse efeito ao combinar $\tanh$ no estado da célula e conexões diretas que permitem a retenção de informações ao longo do tempo.

Esses elementos são importantes para a capacidade das células LSTM de capturar dependências de longo prazo e mitigar os problemas supracitados, comuns em RNN tradicionais. Por meio dessas operações e estruturas, as LSTM oferecem um mecanismo para aprender e memorizar informações ao longo de sequências extensas, possibilitando sua aplicação em uma variedade de tarefas de modelagem de

sequência, como previsão de séries temporais, processamento de linguagem natural e mais.

Outro elemento importante do LSTM é o Portão de Esquecimento (*Forget Gate*), definido pela Equação 6. Ele decide quais informações o estado da célula deve descartar. Usa a função de ativação *sigmoide* (σ) na entrada f_t da Figura 7 para gerar valores entre 0 e 1 para cada número no estado da célula C_{t-1} . Se f_t é próximo de 0, a informação correspondente é “esquecida”; se é próximo de 1, é mantida. Esse portão é calculado considerando a entrada atual x_t e o estado oculto do passo de tempo anterior h_{t-1} , os quais são multiplicados por suas respectivas matrizes de pesos U_f e W_f , que são definidos durante o processo de treinamento por meio da otimização da função de perda da rede, e utilizados em cada iteração para ponderar a contribuição da entrada atual e do estado anterior na ativação do portão de esquecimento.

$$f_t = \sigma(x_t U_f + h_{t-1} W_f) \quad (6)$$

Já o Portão de Entrada (*Input Gate*) é definido pela Equação 7. Ele decide quais novas informações serão adicionadas ao estado da célula. Ele também usa a σ , na entrada i_t da Figura 7 para gerar valores entre 0 e 1, determinando a importância de novas informações contidas em x_t e h_{t-1} , sendo essas ponderadas pelas matrizes de pesos U_i e W_i , aprendidos durante o treinamento da rede, sendo responsáveis por avaliar quão relevantes são as novas informações recebidas em cada etapa antes de atualizarem o estado da célula.

$$i_t = \sigma(x_t U_i + h_{t-1} W_i) \quad (7)$$

O Portão de Saída (*Output Gate*) determina quais informações do estado da célula serão usadas na saída h_t . Utilizando a σ , decide a parte do estado da célula a ser revelada como saída, como apresenta a Equação 8. Assim como nos demais portões, essa ativação é influenciada pela entrada x_t e pelo estado oculto anterior h_{t-1} , cujas contribuições são calculadas com o auxílio das matrizes de pesos U_o e W_o , que são aprendidas durante o treinamento e definem quais partes da informação processada serão efetivamente repassadas ao próximo passo da rede.

$$o_t = \sigma(x_t U_o + h_{t-1} W_o) \quad (8)$$

A Atualização do Estado da Célula (*Cell State Update*) é realizada combinando as informações a serem esquecidas (moduladas por f_t) com as novas informações candidatas (moduladas por i_t e transformadas pela função de ativação tangente hiperbólica $\tanh$). A nova informação candidata ao estado da célula, representada

por $\tilde{C}_t$, é obtida por meio de uma transformação das entradas ponderadas por suas respectivas matrizes de pesos. O resultado dessa operação define a atualização do estado da célula conforme apresentado na Equação 9.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(\tilde{C}_t) \quad (9)$$

A Equação 10 apresenta a Atualização do Estado Oculto (*Hidden State Update*) (ou saída da célula LSTM para um passo de tempo). Ela é calculada aplicando a função de ativação tangente hiperbólica $\tanh$ ao estado da célula e depois multiplicando pelo resultado do portão de saída o_t , permitindo assim que apenas as informações relevantes sejam passadas como saída.

$$h_t = \tanh(C_t) \cdot o_t \quad (10)$$

2.2.2 Métricas de Avaliação de Erro de Desempenho

O Erro Quadrático Médio (MSE), descrito na Equação 11, e o *Root Mean Squared Error* (RMSE) são métricas amplamente utilizadas na avaliação de modelos de regressão. O RMSE corresponde à raiz quadrada da média dos quadrados das diferenças entre os valores preditos e os observados, conforme a Equação 12, em que y_i representa os valores observados da variável resposta, $\hat{y}_i$ são os valores estimados pelo modelo, n é o número total de amostras ou observações no conjunto de dados e i indica o índice de cada observação. Essa medida expressa os erros na mesma unidade dos dados originais e apresenta sensibilidade a grandes desvios, penalizando de forma mais intensa os erros de maior magnitude (Hodson, 2022).

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

O *Mean Absolute Error* (MAE) é uma medida da diferença entre duas variáveis contínuas. Esta métrica corresponde à média das distâncias absolutas entre as previsões e as observações reais, proporcionando uma medida direta e compreensível dos erros médios no mesmo sistema de unidades das variáveis sendo medidas. Diferentemente do RMSE, o MAE não eleva os erros ao quadrado e, portanto, não penaliza tanto os grandes erros, tornando-o mais robusto a *outliers* (Hodson, 2022). A Equação 13 apresenta como calcular esta métrica.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

No modelo de regressão linear simples, onde x é a variável preditora e y a variável resposta, o Coeficiente de Determinação (R^2) é uma métrica fundamental. Esta métrica indica a proporção da variabilidade de y que é explicada pela variabilidade de x . Um R^2 próximo de 1 sugere uma forte relação linear entre as variáveis, onde os pontos no diagrama de dispersão estão próximos à linha de regressão (Martins, 2018). A Equação 14 apresenta como calcular esta métrica, sendo que $\bar{y}$ é a média dos valores observados de y .

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

A correlação de Pearson, dada pela Equação 15, é representada por r e constitui uma medida estatística que quantifica a relação linear entre duas variáveis quantitativas. O cálculo desse coeficiente considera a variância compartilhada entre as variáveis, refletindo como elas se associam na distribuição de seus valores.

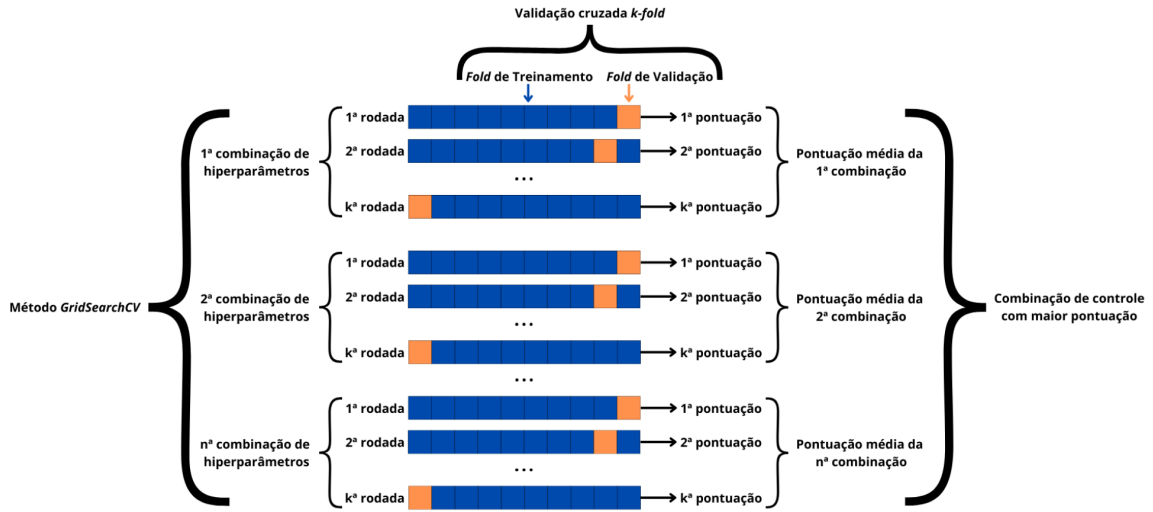
$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} \quad (15)$$

Na Equação 15, x_i e y_i são os valores das variáveis em análise, enquanto $\bar{x}$ e $\bar{y}$ representam as médias dos valores de x e y , respectivamente. O coeficiente r varia de -1 a 1, com $r = 1$ indicando uma correlação positiva perfeita, $r = -1$ indicando uma correlação negativa perfeita e $r = 0$ mostrando a ausência de relação linear. Valores próximos a 1 ou -1 sugerem uma forte dependência linear, positiva ou negativa, enquanto valores próximos a 0 indicam uma relação linear fraca. A interpretação desses valores pode variar, mas geralmente, valores entre 0,10 e 0,29 são considerados fracos, entre 0,30 e 0,49 como moderados, e de 0,50 em diante como fortes (Filho; Júnior, 2009).

2.2.3 Validação Cruzada

A validação cruzada é uma técnica utilizada para avaliar a capacidade de generalização de modelos preditivos e prevenir o sobreajuste. O processo envolve dividir o conjunto de dados em várias partes (ou subconjuntos) e, em cada iteração, utilizar um subconjunto como conjunto de validação, enquanto os demais são usados para treinar o modelo (Berrar *et al.*, 2019). Além disso, esse procedimento pode ser repetido para diferentes combinações de hiperparâmetros, como ilustrado na Figura 9.

Figura 9 – *Ajuste de hiperparâmetros usando validação cruzada*



Fonte: adaptado de (Shatnawi *et al.*, 2022).

A Figura 9 também ilustra o processo de validação cruzada em três partes usado neste trabalho para o treinamento e seleção de hiperparâmetros do modelo. Nesse contexto, para cada combinação de hiperparâmetros, a validação cruzada é executada k vezes, garantindo que todas as partes dos dados sejam utilizadas tanto para treino quanto para teste ao longo das iterações. O processo se repete para n diferentes combinações de hiperparâmetros, permitindo a seleção da configuração que apresenta o melhor desempenho médio.

No caso de validação cruzada o erro de validação cruzada é calculado pela Equação 16.

$$\text{Erro de Validação Cruzada} = \frac{1}{k} \sum_{i=1}^k E_i \quad (16)$$

onde k é o número de subconjuntos (*folds*) nos quais o conjunto de dados foi dividido, e E_i é o erro do modelo no i -ésimo subconjunto de validação.

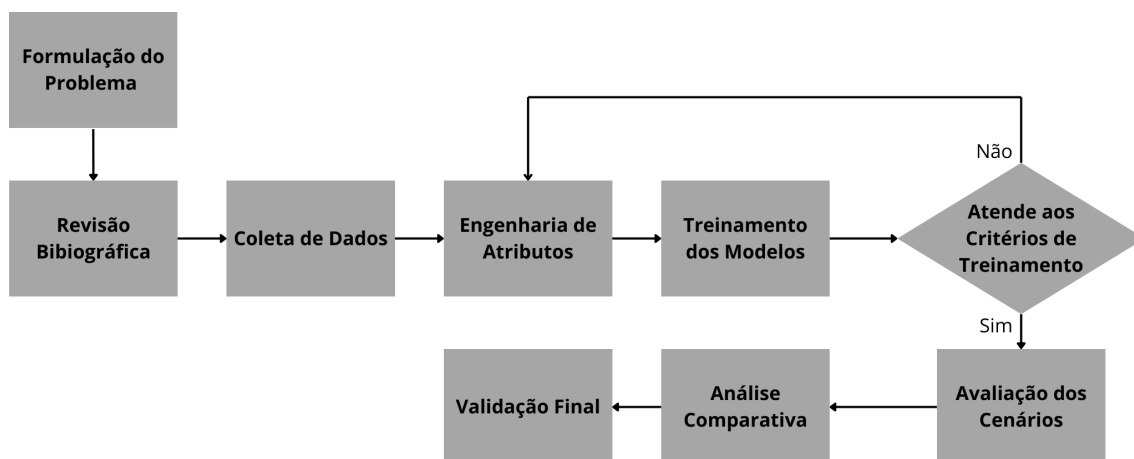
Essa abordagem ajuda a garantir que o modelo possa generalizar bem para novos dados, evitando tanto o sobreajuste quanto o subajuste.

Metodologia

Este capítulo apresenta os procedimentos metodológicos adotados para o desenvolvimento e avaliação do modelo de IA destinado ao controle da manutenção de sistemas fotovoltaicos propostos neste trabalho.

Adotando as etapas típicas no desenvolvimento de modelos de IA, estes procedimentos incluem a coleta e o tratamento de dados relevantes, a aplicação de técnicas de pré-processamento e a implementação de métodos de treinamento e teste específicos para otimizar o desempenho do modelo. No desenvolvimento do modelo são utilizados dados operacionais de usinas fotovoltaicas reais com dados meteorológicos. As etapas para desenvolvimento do trabalho podem ser observados na Figura 10.

Figura 10 – *Fluxograma de etapas da pesquisa*



Fonte: Elaborado pelo autor (2025).

Inicialmente, é realizado a identificação do problema, definido como a melhora da eficiência energética de sistemas fotovoltaicos de geração solar instaladas, baseada na análise de dados aplicados a técnicas de IA para sugerir manutenção nos sistemas. Em seguida, foi realizada uma revisão sistemática da literatura que envolve o uso de ferramentas e técnicas de IA aplicadas à matriz de energia solar.

De acordo com Severino (2017), a revisão bibliográfica é um passo fundamental para a produção de conhecimento em pesquisa científica, subsidiando a estrutura do projeto e justificando a necessidade de avaliar as contribuições existentes sobre o tema em discussão. A partir dessa análise, pode-se determinar como a pesquisa proposta pode contribuir para o entendimento sobre o objeto a ser investigado.

A primeira fase da pesquisa refere-se ao levantamento bibliográfico focado na detecção de falhas em usinas fotovoltaicas, com ênfase nos métodos que utilizam técnicas de IA para análise de dados. A segunda etapa tratou da predição de geração solar. Para tal, foi realizada a coleta dos dados meteorológicos e de produção de energia solar, visando prepará-los para o treinamento dos modelos de ML propostos. Os métodos de obtenção dos dados e escopo adotado na coleta serão discutidos na próxima seção.

Após a coleta, realizou-se um passo fundamental no processo de ML, conhecido como engenharia de atributos. Esse passo envolve a transformação dos dados brutos em *features* adequadas para os modelos, abordando aspectos como filtragem, agrupamento, escalonamento, transformações logarítmicas e de potência, além de *features* de interação. A escolha correta dessas *features* facilita a modelagem e melhora a qualidade dos resultados (Zheng; Casari, 2018). A sua importância reside no fato de que as *features* adequadas podem simplificar a complexidade do modelo, resultando em previsões mais precisas e eficientes.

Posteriormente, inicia-se o treinamento que permite que o modelo se adapte aos dados e produza como resultado previsões com base em novos dados. Ao longo do treinamento, o modelo é ajustado iterativamente até atingir um menor erro médio.

Após o treinamento dos modelos, a avaliação ocorre com base nas métricas de desempenho para modelos de regressão. Os resultados obtidos foram comparados com os resultados de artigos revisados anteriormente, para avaliar a eficiência das predições dos modelos desenvolvidos. Se os modelos não atingirem os valores esperados das métricas, a etapa de engenharia de atributos é reavaliada, permitindo ajustes e refinamentos nas *features* de entrada antes de um novo ciclo de treinamento.

Consecutivamente, uma vez treinado, testado e avaliado de forma satisfatória, o modelo poderá ser usado para prever como o sistema deve se comportar em diferentes períodos do ano e fornecer uma projeção quantitativa da energia que será gerada, visando identificar a necessidade de manutenção para o sistema, caso seja detectado um desvio entre os valores obtidos através da análise dos dados e os valores reais do desempenho do sistema.

Os experimentos e treinamentos dos modelos foram realizados em um computador com as seguintes especificações técnicas: Processador Intel® Core™

i7-11800H de 11^a Geração com clock base de 2.30 GHz, e RAM instalada de 32,0 GB. O sistema operacional é de 64 bits, baseado em arquitetura x64. O computador também possui uma placa de vídeo dedicada NVIDIA GeForce RTX 3050.

3.1 Coleta de Dados

Nesta seção são apresentados os dados de entrada necessários para o desenvolvimento e avaliação dos modelos de IA. Conforme observado na literatura, diferentes abordagens utilizam variáveis como temperatura, irradiação solar, velocidade do vento, altitude, pressão atmosférica, dados elétricos (tensão, corrente e potência) e informações geográficas, entre outras. (Wang *et al.*, 2023) Propõem um modelo híbrido de ML em duas camadas, combinando XGBoost, RF e LSTM com elevação dimensional de características e regressão quantílica para previsão determinística e probabilística, considerando dados meteorológicos e elétricos em alta frequência. Já Pan *et al.* (2024) desenvolveram uma estrutura multimodal que integra dados históricos de geração e imagens do céu por meio de atenção cruzada, visando melhorar a capacidade de previsão em condições climáticas variáveis. No entanto, neste trabalho, além dos dados meteorológicos fixos e da geração de energia das usinas em diferentes localidades, inclui-se como entrada a distância entre as usinas e o local de medição meteorológica. Essa abordagem visa investigar se a distância influencia na confiabilidade das estimativas da previsão, aspecto que não foi amplamente explorado nos estudos anteriores.

O primeiro conjunto de dados de entrada considerado consiste em dados climáticos obtidos do INMET (2024), órgão oficial do governo que disponibiliza dados meteorológicos abertos. Foi identificada apenas uma estação meteorológica em Goiânia, localizada nas coordenadas 16°38'24" S e 49°13'12" O, que corresponde ao setor Jaó. Esta estação fornece dados em intervalos de 1 hora. Os parâmetros selecionados incluem: temperatura instantânea em graus Celsius (°C), umidade relativa instantânea (em porcentagem), velocidade do vento em metros por segundo (m/s), radiação em quilojoule por metro quadrado (kJ/m²) e precipitação de chuva em milímetros por hora (mm/h).

Outro conjunto de dados de entrada considerado, consiste na geração fotovoltaica coletada de 5 sistemas fotovoltaicos reais, situados em um raio de 150 quilômetros do local da coleta de dados meteorológicos. A seleção dessas usinas foi baseada na análise das distâncias entre as estações meteorológicas do INMET no estado de Goiás, apresentada na Tabela 1. A maior distância registrada entre duas estações foi de aproximadamente 142,78 km. Assim, a escolha desse raio

assegura uma amostragem representativa da variabilidade geográfica sem extrapolar significativamente os limites observados na distribuição espacial das estações.

Equipados com inversores de frequência da marca Growatt e diversas marcas de módulos fotovoltaicos, esses sistemas foram selecionados com base na integridade dos dados, assegurando uma ausência mínima de valores faltantes durante o período de coleta, que se estendeu de 10 de agosto de 2024 a 10 de novembro de 2024. Ressalta-se, contudo, que não foi contemplado um período sazonal completo, sendo importante considerar essa dimensão temporal em estudos futuros, uma vez que a variabilidade sazonal pode influenciar de maneira relevante os resultados e a interpretação das análises. Este intervalo foi estabelecido conforme a política de retenção de dados da plataforma Growatt, que exclui registros com mais de três meses.

Tabela 1 – *Análise das distâncias entre estações meteorológicas do INMET no estado de Goiás*

Considerações	Valor (km)
Menor distância entre estações	39,56
Maior distância entre estações	142,78
Média das distâncias entre estações	88,20
Intervalo de confiança (95%)	5,95

Fonte: Elaborado pelo autor (2025).

As informações das instalações fotovoltaicas selecionadas incluem a localização e a capacidade dos inversores e módulos para cada usina, conforme descrito na Tabela 2. A localização geográfica é significativa para este estudo devido às variações microclimáticas e seu impacto na geração de energia. A Tabela 2 também apresenta a distância em linha reta de cada sistema fotovoltaico para a estação meteorológica utilizada.

Para ilustrar a dispersão das usinas fotovoltaicas, a Figura 11 fornece uma representação visual detalhada da distribuição geográfica em Goiás. A visualização permite identificar possíveis áreas para expansão futura da coleta dos dados fotovoltaicos, que podem contribuir para o planejamento estratégico de energias renováveis nas cidades.

As medições incluem a geração de energia em Potência Fotovoltaica (Ppv), o qual é uma medida de tensão no ponto de operação onde o produto da corrente e da tensão é máximo, dado em kWp. As medidas são registrados pelo fabricante dos

Tabela 2 – *Localização e capacidade dos sistemas fotovoltaicos utilizados na pesquisa*

Usina	Município	Distância (km)	Potência inversor (kW)	Potência módulos (kW)	Latitude (°)	Longitude (°)
1	Carmo do Rio Verde	151,9	5	4,92	-15,35	-49,71
2	Santa Cruz de Goiás	108,5	10	12,22	-17,32	-48,48
3	Gemeleira de Goiás	62,7	10	14,84	-16,49	-48,65
4	Trindade	30,1	3	3,68	-16,65	-49,49
5	Goiânia	4,0	3	3,88	-16,65	-49,20

Fonte: Elaborado pelo autor (2024).

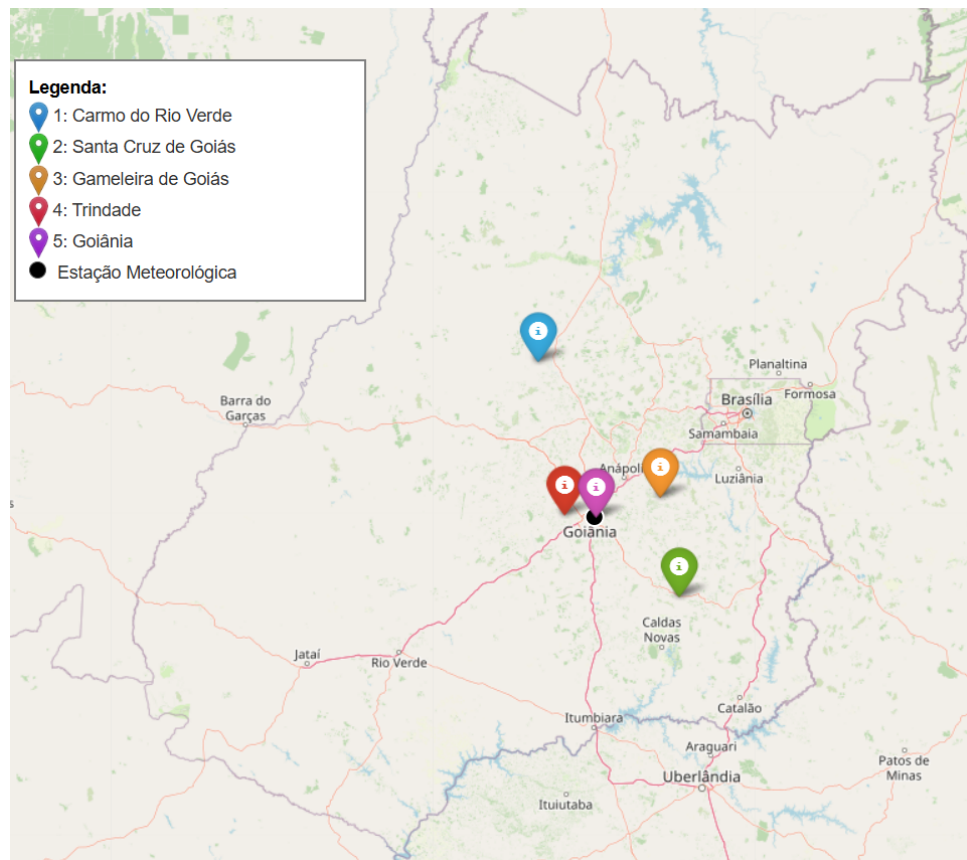
inversores em intervalos de 5 minutos a partir do momento em que a tensão mínima no inversor é atingida, impedindo o início simultâneo de todas as usinas. O estudo concentrou-se nessa grandeza, a qual é a variável de saída a ser prevista.

3.2 Pré-processamento de Dados

Na etapa de pré-processamento dos dados de geração de energia e dos dados climáticos, foram empregadas diversas bibliotecas Python. A biblioteca Pandas (McKinney *et al.*, 2011) foi utilizada para manipulação e organização dos dados permitindo a representação e o tratamento de dados tabulares de forma bidimensional. Além disso, a NumPy (Harris *et al.*, 2020) realiza a manipulação numérica, sendo utilizada em computação científica. De forma complementar, a Scikit-learn (Hackerling, 2017) foi aplicada no processo de normalização dos dados e na divisão dos conjuntos de treino, evidenciam e teste, garantindo a padronização adequada para o treinamento dos modelos.

Os dados de geração de energia de cada sistema fotovoltaico são coletados em múltiplos arquivos Excel. Utilizando a biblioteca Pandas, cada arquivo é lido separadamente, removendo as três primeiras linhas para eliminar cabeçalhos inapropriados e assegurando a leitura apenas dos dados relevantes. Para os dados climáticos, um procedimento similar é seguido, onde dados mais relevantes para o treinamento são filtrados. Foram selecionadas variáveis diretamente associadas à ge-

Figura 11 – *Mapa geral de Goiás com localização dos sistemas fotovoltaicos e estação meteorológica consideradas*



Fonte: Elaborado pelo autor (2025).

ração fotovoltaica, como temperatura e umidade instantâneas, velocidade do vento, radiação e chuva. Demais variáveis, como pressão atmosférica e direção do vento, foram excluídas por apresentarem pouca relação com a geração fotovoltaica. Assim, a seleção visa reduzir redundâncias e manter um conjunto de entrada eficiente para o modelo.

Após a limpeza inicial, os dados de Ppv são agrupados por hora utilizando a média como medida. Os dados climáticos coletados, por sua vez, já são disponibilizados em resolução horária, facilitando a integração direta e a análise comparativa com os dados de geração de energia. Essa uniformidade nos dados climáticos permite uma análise mais eficiente e precisa das correlações entre condições climáticas e a geração de energia, sem a necessidade de transformações adicionais para alinhar os períodos.

Finalmente, ambos os conjuntos de dados são integrados em um único *DataFrame*. Isso ocorre alinhando os carimbos de data e hora dos dados de Ppv e dos dados climáticos para que cada registro horário contenha tanto os dados de

energia quanto os dados climáticos correspondentes. Além disso, os dados foram normalizados para garantir que todas as variáveis possuam escalas comparáveis, evitando que atributos com magnitudes distintas impactem desproporcionalmente os modelos de ML. Para isso, utilizou-se o método de normalização *Min-Max Scaling*, que transforma os valores para o intervalo $[0,1]$ com base nos mínimos e máximos observados no conjunto de treino.

A Tabela 3 apresenta um resumo dos dados considerados no estudo em que os dados foram divididos em 60% para treinamento, 20% para validação e 20% para teste, uma abordagem comumente adotada para garantir a generalização do modelo e evitar problemas como sobreajuste e subajuste (Madhuri, 2024). As colunas de *features*, que totalizam 10 variáveis, possuem individualmente 2.224 registros, resultando em um total de 22.240 entradas analisadas no estudo. Além dessas variáveis, foi considerada como entrada a distância das usinas em relação ao local de medição meteorológica, incorporando um fator espacial à modelagem. Essa abordagem se assemelha a trabalhos de predição como propostos por Guermoui *et al.* (2022) e Carrera e Kim (2023).

3.3 Definição dos Cenários de Predição

Para a análise de predição, foram definidos três cenários distintos, conforme descrito na Tabela 4, com o objetivo de avaliar diferentes combinações de variáveis de entrada e sua influência na predição do desempenho dos sistemas fotovoltaicos.

No Cenário 1, adota-se uma abordagem híbrida, combinando dados de geração fotovoltaica de 4 sistemas e variáveis meteorológicas como entrada, enquanto o sistema fotovoltaico foco da predição permanece como saída do modelo.

No Cenário 2, utilizam-se apenas os valores de geração de 4 sistemas fotovoltaicos, colocando o sistema foco da predição como variável de saída.

Por fim, no Cenário 3, o modelo é alimentado exclusivamente com dados meteorológicos, sem incluir informações sobre os sistemas, mantendo a usina foco da predição como saída.

Na construção dos modelos, os dados pré-processados são divididos em 4 usinas para treinamento e 1 usina para teste em cada etapa de treinamento, garantindo que a usina utilizada como saída não seja incluída entre as entradas do modelo.

Tabela 3 – Dados gerais considerados no estudo

Descrição	Detalhes	Quantidade
Número de registros para cada <i>feature</i> e saída	Total de registros individuais analisados	2.224
Dados para treinamento (60%)	Registros individuais utilizados para treinamento dos modelos	1.334
Dados para validação (20%)	Registros individuais utilizados para validação dos modelos	444
Dados para teste (20%)	Registros individuais utilizados para teste dos modelos	444
Colunas de <i>features</i>	Temperatura Instantânea (°C), Umidade Instantânea (%), Velocidade do Vento (m/s), Radiação Solar (kJ/m ²), Precipitação de Chuva (mm), Distância e Ppv das Usinas 1 a 5 (considerando que, em cada previsão, as entradas incluem somente as usinas que não correspondem à usina de saída)	10
Colunas de saída	Ppv das Usinas 1 a 5 (cada modelo prevê exclusivamente uma usina por vez)	1

Fonte: Elaborado pelo autor (2025).

3.4 Construção do Modelo RF

Na construção do modelo RF, para determinar os melhores hiperparâmetros do modelo, utilizou-se a abordagem **GridSearchCV**, que realiza uma busca exaustiva por meio de uma grade especificada de parâmetros, testando todas as combinações possíveis para identificar a configuração que proporciona o melhor desempenho. O processo de **GridSearchCV** envolve a divisão dos dados em subconjuntos de treinamento e validação, utilizando um número de *folds* definido pelo usuário para realizar a validação cruzada, assegurando que os resultados não sejam influenciados por um conjunto específico de dados.

Os parâmetros ajustados incluíram: o número de estimadores (**n_estimators**), que define o número de árvores na floresta; a profundidade máxima das árvores (**max_depth**), que define a profundidade máxima de cada árvore na floresta; o número mínimo de amostras por divisão (**min_samples_split**), que especifica o número mínimo de amostras necessárias para dividir um nó; e o

Tabela 4 – *Configuração dos cenários de predição*

Cenário	Dados de entrada	Dados de saída
1	Usinas, meteorológicos e distância	Usina objetivo de predição
2	Usinas	Usina objetivo de predição
3	Meteorológicos e distância	Usina objetivo de predição

Fonte: Elaborado pelo autor (2025).

número mínimo de amostras por folha (`min_samples_leaf`), que define o número mínimo de amostras que devem estar presentes em um nó folha. escolha desses parâmetros foi baseada na literatura existente, em particular na abordagem apresentada em Cunha e Sobel (2021), que realizaram a calibração de hiperparâmetros em problemas de predição similares. No estudo, os autores testaram diferentes funções de ativação (como `relu`, `tanh`, `sigmoid` e `linear`), taxas de dropout (0.0, 0.4, 0.7), quantidades de neurônios (1 a 30), tamanhos de batch (16 e 32) e números de épocas (50 a 150), utilizando a técnica de `GridSearchCV` para encontrar a melhor configuração, conforme apresentado na Tabela 5.

Tabela 5 – *Parâmetros definidos para o GridSearchCV, modelo RF*

Parâmetro	Valor mínimo	Valor médio	Valor máximo
<code>n_estimators</code>	50	100	200
<code>max_depth</code>	10	20	∞
<code>min_samples_split</code>	2	5	10
<code>min_samples_leaf</code>	1	2	4

Fonte: Elaborado pelo autor (2024).

O Código 1 apresenta o *script* implementado em *Python*, de definição da grade de parâmetros, seguido pela inicialização do modelo RF e a execução da busca pelos hiperparâmetros ideais utilizando validação cruzada com 5 *folds*. Esse procedimento permite identificar a configuração que minimiza o MSE, resultando em um modelo com maior capacidade preditiva (Cunha; Sobel, 2021). Em comparação com o método convencional de divisão entre treinamento e teste (*Hold-Out*), a validação cruzada reduz a variância da estimativa do erro e aprimora a capacidade

de generalização do modelo, ao avaliar seu desempenho em múltiplos subconjuntos dos dados (Cunha, 2019).

Código 1 *GridSearch* em *Python* para teste de hiperparâmetros, modelo RF.

```
1 # Definir os parâmetros para o GridSearchCV
2 param_grid = {
3     'n_estimators': [50, 100, 200],
4     'max_depth': [10, 20, None],
5     'min_samples_split': [2, 5, 10],
6     'min_samples_leaf': [1, 2, 4]
7 }
8
9 # Inicializar o modelo de Random Forest e GridSearchCV
10 rf = RandomForestRegressor(random_state=42)
11 grid_search = GridSearchCV(estimator=rf, param_grid=param_grid, cv=3, n_jobs=-1, verbose=2
12                             , scoring='neg_root_mean_squared_error')
13
14 # Realizar a busca em grade
15 grid_search.fit(X_train, y_train)
16
17 # Melhores parâmetros encontrados
18 print(f"Melhores parâmetros encontrados: {grid_search.best_params_}")
```

3.5 Construção do Modelo LSTM

Na construção do modelo LSTM, os dados pré-processados utilizados foram os mesmos do modelo RF, visando permitir a comparação entre eles.

Para a análise e treinamento do modelo LSTM, utilizou-se a biblioteca Keras (Gulli; Pal, 2017), uma biblioteca de redes neurais de alto nível, escrita em Python, que executa sobre TensorFlow (Pang; Nijkamp; Wu, 2020), plataforma de código aberto para ML que usa grafos de fluxo de dados nos quais os nós representam operações matemáticas, enquanto as arestas representam os arrays de dados multidimensionais (tensores) que fluem entre eles.

A avaliação de desempenho é realizada utilizando a mesma métrica empregada no modelo RF, sendo esta o MSE. Para determinar os melhores parâmetros do modelo LSTM, utilizou-se uma abordagem de **GridSearchCV** semelhante àquela aplicada ao RF. Nesta abordagem, foram ajustados parâmetros como a função de ativação (**activation**), a taxa de dropout (**dropout_rate**), o número de neurônios (**neurons**), o otimizador (**optimizer**), o número de épocas (**epochs**) e o tamanho do lote (**batch_size**). As Tabelas 6 e 7 apresentam os parâmetros testados e a busca pelos melhores hiperparâmetros. A seleção desses parâmetros baseia-se em boas práticas descritas na literatura, como discutido por (Bischl *et al.*, 2023), que destacam que ajustar corretamente hiperparâmetros como função de ativação, dropout, número de neurônios, otimizador, épocas e batch size é indispensável para melhorar o desempenho e evitar *overfitting*.

Tabela 6 – *Parâmetros numéricos definidos para o GridSearchCV - Modelo LSTM*

Parâmetro	Valor mínimo	Valor médio	Valor máximo
model_neurons	8	16	32
model_dropout_rate	0.0	0.4	0.7
batch_size	16	32	-
epochs	64	128	244

Fonte: Elaborado pelo autor (2024).

Tabela 7 – *Parâmetros categóricos definidos para o GridSearchCV - Modelo LSTM*

Parâmetro	Opção 1	Opção 2	Opção 3	Opção 4
model_activation	relu	tanh	sigmoid	linear
model_optimizer	adam			

Fonte: Elaborado pelo autor (2024).

O Código 2 apresenta a implementação prática da busca por hiperparâmetros utilizando o `GridSearchCV` em Python. Este código automatiza o processo de experimentação, permitindo testar exaustivamente todas as combinações de parâmetros descritas anteriormente. A validação cruzada com 5 *folds* foi realizada de forma semelhante ao modelo citado anteriormente, garantindo que os resultados sejam robustos e generalizáveis (Cunha, 2019).

Código 2 *Grid Search* em *Python* para teste de hiperparâmetros, modelo LSTM.

```
1 # Definindo os hiperparâmetros a serem testados
2 param_grid = {
3     'model_neurons': [8, 16, 32],
4     'model_activation': ['relu', 'tanh', 'sigmoid', 'linear'],
5     'model_dropout_rate': [0.0, 0.4, 0.7],
6     'model_optimizer': ['adam'],
7     'batch_size': [16, 32],
8     'epochs': [64, 128, 244]
9 }
10
11 # Inicializar o modelo de LSTM e GridSearchCV
12 grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1,
13                     scoring='neg_root_mean_squared_error', cv=3)
14 grid_result = grid.fit(X_train, y_train)
15
16 # Melhor combinação de hiperparâmetros
17 print(f"Melhores parâmetros encontrados: {grid.best_params_}")
```

Os resultados obtidos com o modelo LSTM, incluindo a comparação com o modelo RF, são apresentados e discutidos no próximo capítulo, analisando o desempenho de ambos os modelos e suas implicações para a previsão do potencial de geração dos sistemas fotovoltaicos.

3.6 Geração das Métricas de Avaliação

Para a avaliação dos resultados dos algoritmos implementados, utilizaram-se como métricas de desempenho o RMSE e o MAE, considerando que são os indicadores mais utilizados na literatura (Botchkarev, 2019). Além destas duas métricas, utilizou-se também o R^2 , o qual é amplamente reconhecido como uma métrica essencial para medir a qualidade do ajuste de um modelo (Klemme *et al.*, 2020).

Para implementar essas métricas no contexto dos modelos de previsão, utilizou-se a biblioteca `sklearn`, importando-se as funções `mean_absolute_error`, `mean_squared_error` e `r2_score`, conforme apresentado no Código 3. Este trecho de código calcula as três métricas.

Código 3 Cálculo das métricas de erro.

```
1 # Importação das bibliotecas necessárias
2 from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
3
4 # Cálculo de erro dos modelos
5 mae = mean_absolute_error(y_invertidas, previsoes_invertidas)
6 rmse = np.sqrt(mean_squared_error(y_invertidas, previsoes_invertidas))
7 r2 = r2_score(y_invertidas, previsoes_invertidas)
8
9 print(f"Mean Absolute Error (MAE): {mae}")
10 print(f"Root Mean Squared Error (RMSE): {rmse}")
11 print(f"Coeficiente de Determinação (R^2): {r2}")
```

3.7 Aplicação da Metodologia Proposta

Conforme a metodologia proposta, validação e teste dos modelos, foram utilizados dados da média horária da potência de pico do sistema fotovoltaico (Ppv) e das condições meteorológicas, sendo temperatura, umidade, velocidade do vento, precipitação de chuva e radiação coletados pelo INMET em Goiânia, Goiás.

Para assegurar que o sistema de ML opere conforme o esperado, mesmo em situações não previstas durante o desenvolvimento ou no conjunto de dados de treinamento, foi inserido 20% dos dados do conjunto para validação, monitorando o desempenho do modelo ao longo do processo. Essa prática, recomendada por abordagens comuns de ML (Myllyaho *et al.*, 2021), que garante a funcionalidade correta do sistema após sua implantação. Planejou-se, em etapas subsequentes, realizar uma avaliação completa do modelo com dados de teste independentes para validar sua generalização..

Para a previsão da geração de energia solar, adotou-se janelas de dados históricos com horizontes de 1, 6, 12, 24, 48 e 72 horas, ou seja, o modelo considera as informações registradas nessas quantidades de horas anteriores ao momento da previsão para estimar os valores futuros, em uma abordagem semelhante a abordada por Yang, Lee e Lee (2020) que exploram a aplicação de redes neurais híbridas para previsão de poluentes atmosféricos e demonstram que um *lookback* adequado proporciona um desempenho superior na previsão de séries temporais ambientais. A partir dessa abordagem, a utilização de múltiplos horizontes para aprendizado de padrões atmosféricos se torna um fator determinante na modelagem de previsão da geração solar, uma vez que as algumas grandezas meteorológicas locais possuem uma forte correlação com a geração de energia fotovoltaica, conforme será apresentado a seguir.

Para o horizonte de previsão, foi utilizado um valor fixo de 24 horas à frente, considerando que previsões de curto prazo são amplamente empregadas

em aplicações operacionais. Qin *et al.* (2023) relatam que modelos de previsão de radiação solar foram avaliados para horizontes de 1 dia, observando uma redução da variabilidade dos erros ao longo de intervalos mais curtos.

Além disso, Perez *et al.* (2013) destacam que previsões com horizontes de até algumas horas são utilizadas no planejamento e na operação de sistemas de energia, por fornecerem informações alinhadas com a necessidade de resposta imediata dos operadores.

Os parâmetros definidos foram utilizados somente durante o teste do modelo, possibilitando avaliar como o modelo se comporta em diferentes horizontes de tempo, o que permite entender sua capacidade de generalização e grau de ajuste em previsões de curto, médio e longo prazo.

3.7.1 Modelo RF

Por meio dos dados de entrada, o modelo RF desenvolve, constrói várias árvores de decisão, cada uma gerada a partir de uma amostra aleatória dos dados. Cada árvore faz uma previsão com base nas variáveis fornecidas. As previsões de todas as árvores são então combinadas (geralmente por votação majoritária) para determinar a previsão final do modelo. Esse processo permite ao RF capturar padrões complexos e melhorar a capacidade preditiva das previsões.

Mediante o uso da função `GridSearchCV` da biblioteca `Scikit-Learn` e validação cruzada e considerando as métricas de avaliação, foram obtidos os melhores parâmetros para o modelo: `max_depth` de 20, `min_samples_leaf` de 1, `min_samples_split` de 2 e `n_estimators` de 100. Esses parâmetros otimizados garantiram um ajuste mais preciso do modelo aos dados de treinamento, melhorando assim a capacidade preditiva da RF.

3.7.2 Modelo LSTM

A análise do modelo LSTM envolveu a aplicação de métodos de avaliação de parâmetros e validação semelhantes aos usados no treinamento do modelo RF. O modelo LSTM foi escolhido por sua capacidade de capturar dependências de longo prazo em dados temporais (Yu *et al.*, 2019), tornando-o especialmente eficaz para prever a geração de energia solar em horizontes de médio e longo prazo. A habilidade do LSTM de processar sequências de dados e reter informações importantes ao longo do tempo permite uma análise mais precisa das variáveis meteorológicas que impactam a produção de energia.

Usando a estratégia apresentada, os melhores parâmetros encontrados foram: `batch_size` de 16, `epochs` de 64, `activation tanh`, `dropout_rate` de 0.0, `neurons` 15, e `optimizer adam`.

Resultados e Discussão

Neste capítulo, são apresentados os resultados da aplicação dos dois modelos de aprendizado de máquina (RF e LSTM), utilizados para comparar a capacidade de prever a geração de energia solar com base em dados reais de usinas fotovoltaicas. Embora o estudo de caso tenha sido realizado no estado de Goiás, a abordagem proposta pode ser adaptada e aplicada em qualquer região que disponha de dados equivalentes, permitindo avaliar a capacidade de generalização dos modelos em diferentes contextos geográficos.

4.1 Impacto dos Fatores Climáticos na Geração Solar

Para entender a relação entre a produção energética dos sistemas solares e as condições climáticas, aplicou-se a correlação de Pearson às variáveis selecionadas, conforme mostra a Tabela 8.

Tabela 8 – *Correlação de Pearson entre dados de geração e dados climáticos*

Usina	Temperatura instantânea (°C)	Umidade instantânea (%)	Velocidade do vento (m/s)	Precipitação de chuva (mm)	Radiação (kJ/m²)
1	-0,176676	0,136261	0,004202	-0,036161	0,899704
2	-0,149012	0,152510	-0,003245	-0,051249	0,881538
3	-0,149921	0,139474	0,043786	-0,042970	0,904016
4	-0,275037	0,196493	-0,073434	-0,068595	0,941835
5	-0,131536	0,150764	0,040747	-0,056457	0,921448

Fonte: Elaborado pelo autor (2025).

A análise desses coeficientes apresenta importantes percepções sobre os fatores que mais influenciam a produção de energia nas diferentes localizações. Observa-se uma correlação fortemente positiva com a radiação solar, acima de 0,88 para todas as usinas, indicando que a radiação solar é o principal determinante da produção de energia. A correlação da temperatura com a geração de energia foi negativa para todas as usinas, indicando uma redução na eficiência dos módulos fotovoltaicos com o aumento da temperatura (Lemes, 2023). Por outro lado, a umidade relativa, velocidade do vento e a precipitação de chuva apresentam correlações próximas de zero, indicando uma influência pouco significativa na geração de energia. Esse comportamento pode ser explicado pelo fato de que, na análise de Pearson, a correlação mede a relação linear entre as variáveis. Como estas grandezas podem afetar a produção fotovoltaica de forma não linear e indireta, por meio da formação de nuvens ou acúmulo de sujeira nos painéis esses efeitos podem não ser capturados adequadamente pela correlação linear, resultando em valores próximos de zero.

Ao considerar as distâncias dos sistemas fotovoltaicos em relação ao ponto de coleta dos dados climáticos, pode-se notar que usinas mais próximas, como a Usina 5 (4,0 km) e a Usina 4 (30,1 km), apresentam altas correlações com a radiação solar (0,921448 e 0,941835, respectivamente), corroborando a hipótese de que a proximidade com a fonte dos dados climáticos pode contribuir para uma maior capacidade explicativa (R^2) na previsão de geração de energia. No entanto, existem casos em que sistemas fotovoltaicos mais distantes apresentam correlação superior às mais próximas, como a Usina 1, que, apesar de estar mais distante, exibe uma correlação ligeiramente maior com a radiação do que a Usina 2 (0,899704 contra 0,881538). Isso indica que fatores como a inclinação dos módulos, condições atmosféricas locais e calibração dos sensores também influenciam a capacidade de previsão.

Esses resultados destacam a importância de considerar múltiplos fatores, incluindo a localização e as características técnicas das usinas, ao se desenvolver modelos preditivos de desempenho energético.

4.2 Avaliação do Desempenho dos Modelos nos Cenários Testados

Nesta seção, apresenta-se a análise comparativa do desempenho dos modelos RF e LSTM nos três cenários definidos para a predição da geração de energia em sistemas fotovoltaicos. Cada cenário representa um conjunto de variáveis distintas, permitindo avaliar como diferentes fontes de informação impactam a capacidade de

previsão.

Os resultados são apresentados em termos do coeficiente de determinação (R^2) para diferentes janelas históricas de dados. As Figuras 12 a 17 ilustram o desempenho dos modelos em cada um dos cenários testados.

4.2.1 Cenário 1

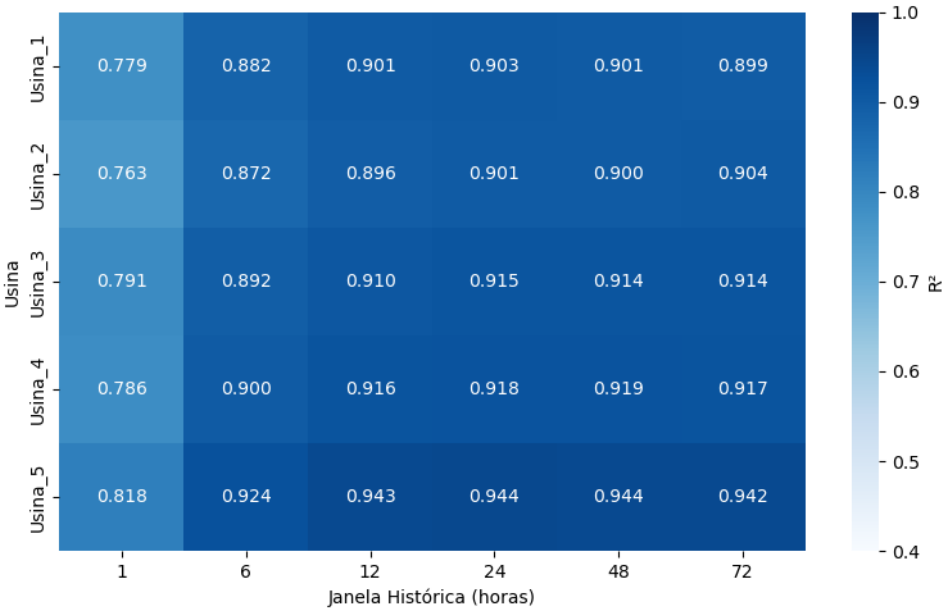
O Cenário 1 foi projetado para avaliar o impacto da inclusão de variáveis meteorológicas, histórico de geração (Ppv Médio Horário) e da distância dos sistemas fotovoltaicos em relação ao ponto de coleta de dados meteorológicos. Neste cenário, foram utilizados dados de cinco usinas, cujas distâncias variam entre 4,0 km e 151,9 km, conforme a apresentado anteriormente. O objetivo foi identificar padrões de desempenho entre os modelos RF e LSTM, bem como verificar a influência da proximidade da usina aos dados climáticos na adequação do modelo de previsões.

4.2.1.1 Análise do Desempenho dos Modelos RF e LSTM

O modelo RF, ilustrado na Figura 12, apresentou um desempenho superior ao LSTM, Figura 13, com valores de R^2 variando entre 0,763 e 0,944, dependendo da janela histórica utilizada. O melhor resultado foi observado na Usina 5, onde o valor de R^2 igual a 0,944 foi alcançado com uma janela de 24 e 48 horas. Esse desempenho pode ser atribuído à proximidade da Usina 5 em relação ao ponto de coleta de dados meteorológicos, que está a apenas 4,0 km de distância.

O modelo LSTM, apresentado na Figura 13, obteve valores de R^2 mais baixos em comparação ao RF, oscilando entre 0,641 e 0,949 (Figura 13). O melhor desempenho também foi registrado na Usina 5, com R^2 igual a 0,949 para uma janela de 72 horas. Diferentemente do RF, a LSTM apresentou maior variação nos valores de R^2 entre as diferentes janelas históricas, indicando que o modelo é mais sensível ao volume de dados utilizados para treinamento. Por exemplo, na Usina 2, o desempenho foi inferior para janelas curtas (como 1 hora), mas melhorou significativamente à medida que a quantidade de dados históricos aumentou.

Figura 12 – *Valores de R^2 por janela de dados históricos - Cenário 1 - RF*



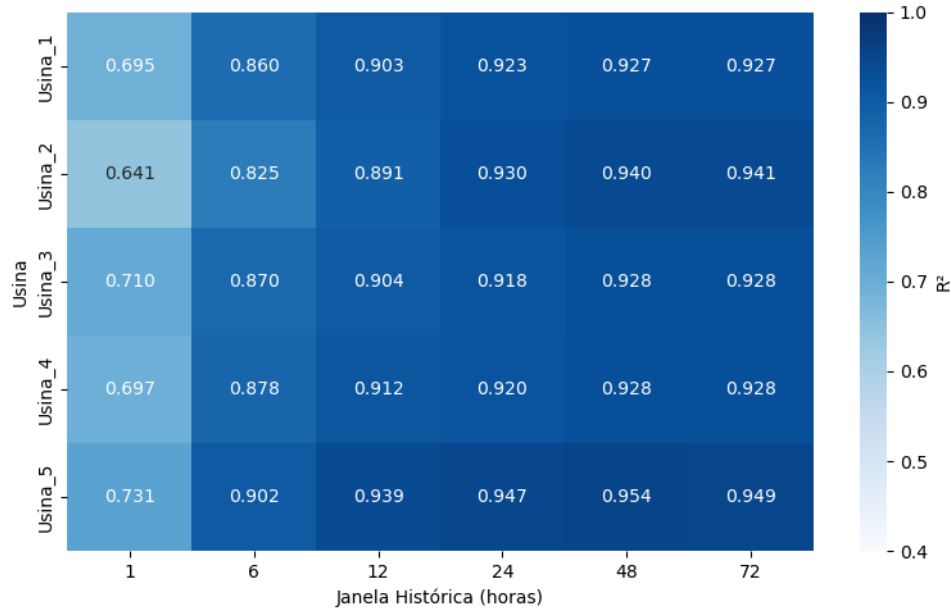
Fonte: Elaborado pelo autor (2025).

No caso da Usina 5, a maior aderência dos dados climáticos ao local de geração contribuiu para a obtenção do maior R^2 , validando a eficácia desse modelo para previsões em usinas próximas ao ponto de coleta.

Em contraste, usinas localizadas a distâncias superiores, como a Usina 1 (151,9 km) e a Usina 2 (108,5 km), apresentaram desempenhos inferiores, evidenciando o impacto da variabilidade espacial dos dados climáticos na avaliação do erro médio das previsões. Embora o modelo RF tenha demonstrado uma tendência de melhora do R^2 com o aumento da janela histórica para a maioria das usinas, essa melhora foi menos expressiva em usinas mais distantes, possivelmente devido a diferenças climáticas locais não captadas pelos dados meteorológicos utilizados.

A Figura 13apresenta o resultado da aplicação do modelo LSTM, considerando diferentes janelas históricas e cenários de entrada, com destaque para a análise comparativa entre as usinas.

Figura 13 – *Valores de R^2 por janela de dados históricos - Cenário 1 - LSTM*



Fonte: Elaborado pelo autor (2025).

Portanto, os resultados sugerem que, a escolha da usina de referência e a inclusão de dados meteorológicos são fatores críticos. O melhor geral para diferentes janelas históricas foi o Cenário 1, sendo essa abordagem especialmente significativa para sistemas fotovoltaicos situados próximos às estações meteorológicas de referência.

4.2.2 Cenário 2

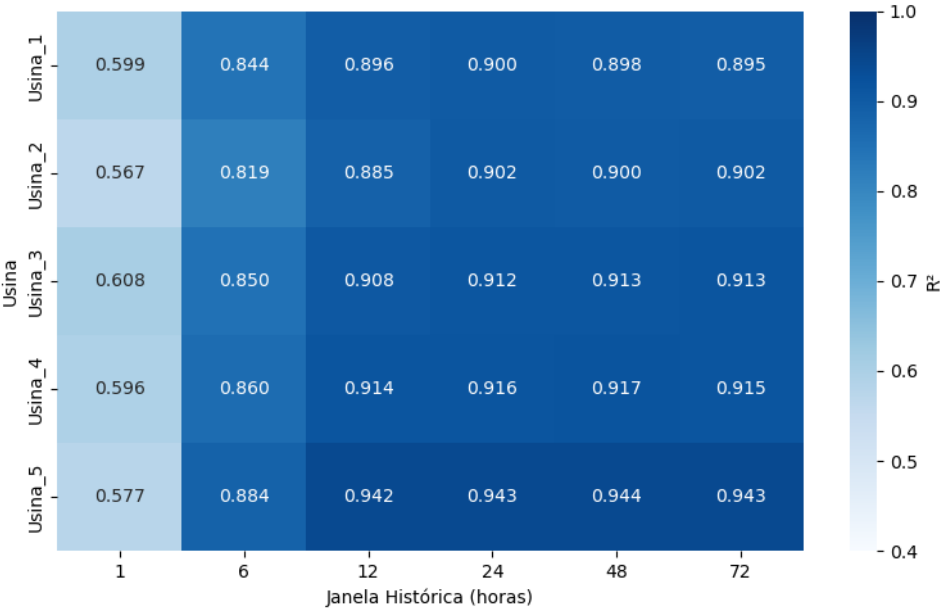
O Cenário 2 foi estruturado para avaliar o impacto da utilização dos dados reais de sistemas fotovoltaicos (Ppv Médio Horário) e da distância dos sistemas fotovoltaicos em relação ao ponto de coleta de dados meteorológicos como entrada do modelo. Diferente do Cenário 1, neste cenário não foram consideradas variáveis meteorológicas diretamente, tornando a análise focada na influência da posição geográfica das usinas sobre a confiabilidade das estimativas nas previsões.

4.2.2.1 Análise do Desempenho dos Modelos RF e LSTM

O modelo RF, ilustrado na Figura 14, apresentou valores de R^2 variando entre 0,567 e 0,944, dependendo da janela histórica utilizada. Assim como no Cenário 1, o melhor desempenho foi observado na Usina 5, onde o valor de R^2 igual à 0,944 foi alcançado com uma janela de 48 horas. No entanto, no Cenário 2, que não utilizou dados climáticos diretamente, o modelo também obteve resultados

satisfatórios para a Usina 3, com R^2 igual à 0,913 para uma janela de 48 horas. Esse comportamento pode estar relacionado à posição intermediária da Usina 3 (62,7 km) em relação ao ponto de coleta meteorológico. Sendo uma usina situada entre os extremos de distância, o modelo pode ter encontrado maior facilidade em estimar padrões de geração ao capturar tendências tanto de usinas próximas quanto de usinas mais distantes. Em contraste, usinas muito afastadas, como a Usina 1 (151,9 km), tiveram desempenho inferior, com valor de R^2 de 0,895, possivelmente devido à maior variabilidade climática local, que não pôde ser compensada durante o treinamento.

Figura 14 – *Valores de R^2 por janela de dados históricos - Cenário 2 - RF*

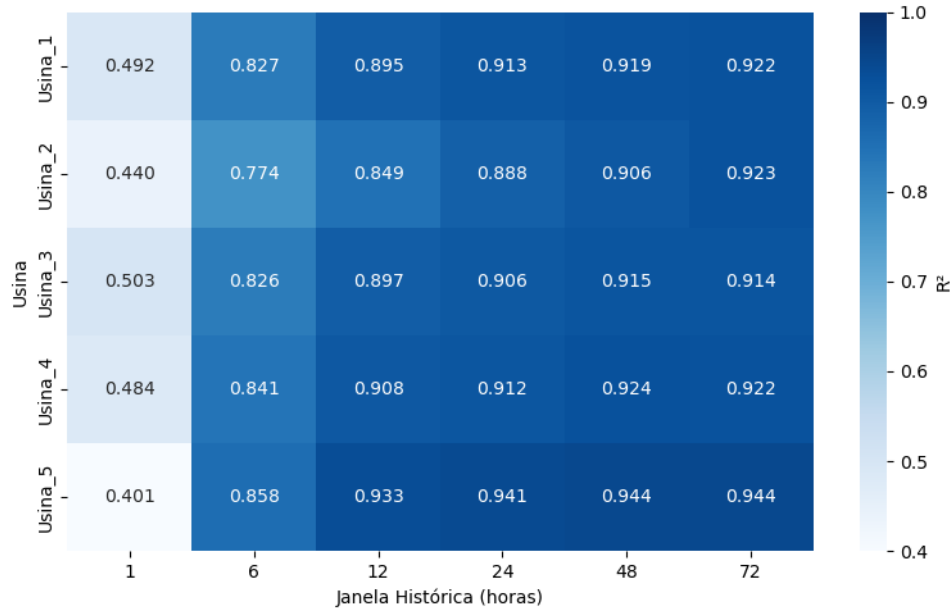


Fonte: Elaborado pelo autor (2025).

A Figura 15 apresenta os valores de R^2 obtidos pelo modelo LSTM no Cenário 2, variando entre 0,593 e 0,937, dependendo da usina e da janela histórica analisada. A melhor previsão foi registrada na Usina 5, com um coeficiente de determinação de R^2 igual a 0,937 para a janela de 48 horas. Esse desempenho indica que, sem a utilização de variáveis meteorológicas, o modelo depende da regularidade dos padrões históricos de geração para realizar suas estimativas.

Em comparação ao modelo RF, o LSTM demonstrou resultados bastante competitivos, com desempenho superior em diversas combinações de usinas e janelas, especialmente nas usinas mais próximas da estação meteorológica. Por exemplo, na Usina 1, os valores de R^2 foram satisfatórios apenas com maiores quantidades de dados históricos, evidenciando a dificuldade do modelo em capturar a variabilidade da geração em janelas curtas sem informações externas complementares.

Figura 15 – R^2 por janela de dados históricos - Cenário 2 - LSTM



Fonte: Elaborado pelo autor (2025).

Outro aspecto observado foi a relativa estabilidade dos valores de R^2 para as usinas 3, 4 e 5 em janelas superiores a 24 horas, embora os resultados tenham sido ligeiramente inferiores aos do modelo RF. Isso sugere que o modelo LSTM conseguiu detectar padrões de longo prazo nessas usinas, mesmo sem acesso a variáveis climáticas.

Observou-se também uma menor variação no desempenho entre as usinas 1 a 5 conforme o aumento da janela histórica, sobretudo nas janelas de 48h e 72h. Essa uniformidade pode ser atribuída à ausência de dados meteorológicos no treinamento, o que evita interferências externas e força o modelo a basear suas previsões exclusivamente em padrões históricos de geração. Contudo, essa limitação também reduz a capacidade do modelo de capturar variações abruptas ligadas ao clima, prejudicando o desempenho em ambientes com maior variabilidade atmosférica.

Os resultados do Cenário 2 mostram que o modelo LSTM apresentou um desempenho consistente, especialmente nas usinas 4 e 5, com valores de R^2 acima de 0,90 em janelas de 24 a 72 horas. A Usina 5 obteve os melhores resultados, com destaque para a janela de 48 horas ($R^2 = 0,937$), possivelmente devido à regularidade de seus padrões de geração. Dessa forma, a análise reforça que, mesmo sem dados meteorológicos, o modelo pode apresentar bom desempenho em locais onde a geração de energia segue comportamentos mais estáveis ao longo do tempo.

4.2.3 Cenário 3

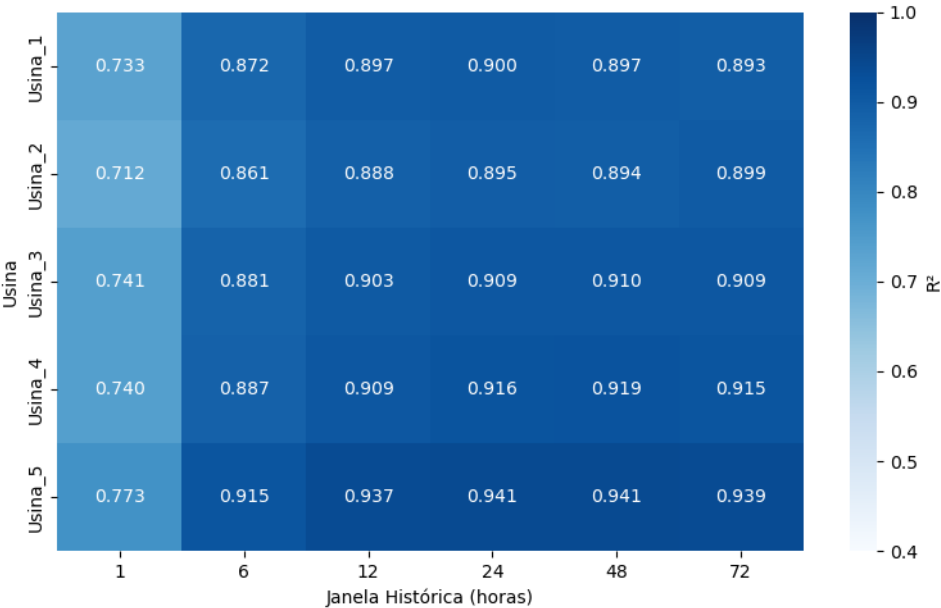
O Cenário 3 foi projetado para avaliar a capacidade dos modelos de previsão da geração fotovoltaica utilizando exclusivamente variáveis meteorológicas e a distância das usinas em relação ao ponto de coleta de dados climáticos. Diferente dos cenários anteriores, não foram considerados os dados históricos de geração, tornando este cenário uma avaliação direta da influência dos fatores ambientais e da posição geográfica das usinas na redução de erro das previsões.

4.2.3.1 Análise do Desempenho dos Modelos RF e LSTM

Na sequência são descritos os resultados referentes ao desempenho do RF na previsão da geração fotovoltaica, considerando diferentes configurações de janelas históricas.

A Figura 16 apresenta os valores de R^2 obtidos pelo modelo RF no Cenário 3, variando entre 0,712 e 0,941, dependendo do sistema fotovoltaico e da janela histórica analisada. O melhor desempenho foi observado na Usina 5, onde R^2 igual à 0,941 foi alcançado para janelas de 24 e 48 horas. Este resultado confirma que a proximidade ao ponto de coleta de dados meteorológicos ainda desempenha um papel significativo na previsão da geração, mesmo sem a inclusão de informações sobre o histórico de produção.

Figura 16 – R^2 por janela de dados históricos - Cenário 3 - RF



Fonte: Elaborado pelo autor (2025).

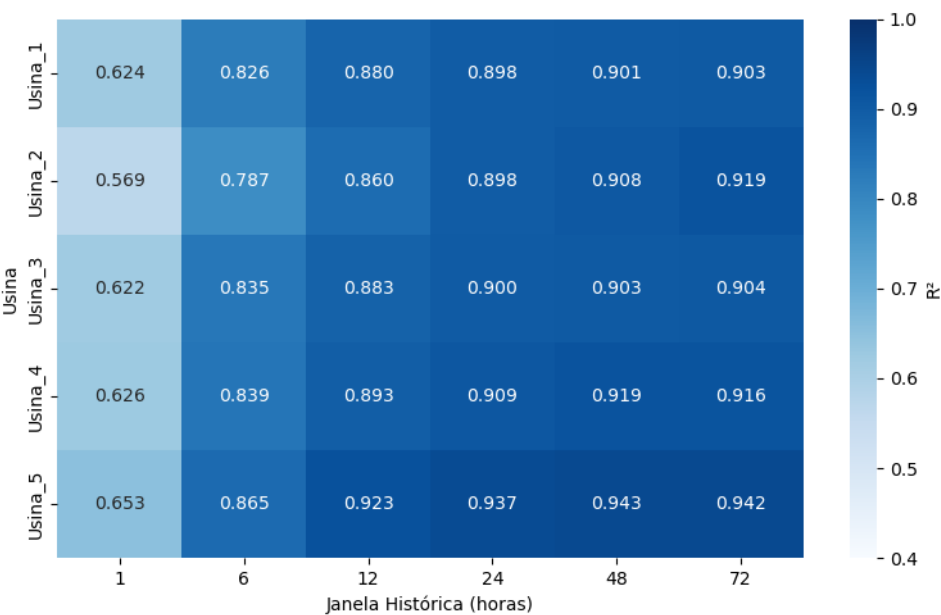
O modelo RF mostrou um comportamento consistente ao longo das diferentes janelas históricas, com uma tendência clara de melhora à medida que a quantidade de dados aumentava. As usinas 4 e 5 apresentaram os melhores resultados para janelas entre 12 e 48 horas, com destaque para a Usina 5, que alcançou um R^2 de 0,941. Isso sugere que o modelo conseguiu capturar bem as variações climáticas nessas regiões. Em contraste, a Usina 2 obteve os valores mais baixos de R^2 , especialmente para janelas curtas, indicando que a distância em relação ao ponto de coleta meteorológica pode ter impactado negativamente a qualidade das previsões.

A Figura 17 ilustra os resultados do modelo LSTM no Cenário 3, com valores de R^2 variando entre 0,569 e 0,942. Comparado ao RF, o desempenho da LSTM foi mais instável, apresentando maior sensibilidade à variação das janelas históricas. O melhor desempenho foi registrado na Usina 5, com R^2 igual a 0,942 para uma janela de 72 horas, sugerindo que a rede neural conseguiu extrair informações relevantes a partir dos dados meteorológicos, embora com comportamento mais oscilante em relação ao modelo RF.

Um aspecto notável foi a melhora progressiva nos valores de R^2 à medida que se aumentavam as janelas históricas, especialmente em usinas como a Usina 2, onde o coeficiente saltou de 0,569 para 0,919 entre as janelas de 1 e 72 horas. Esse comportamento sugere que o modelo LSTM foi capaz de extrair padrões temporais relevantes a partir das variáveis climáticas e da distância, mesmo na ausência do histórico de geração. Tal resultado evidencia que, com maior volume de dados atmosféricos disponíveis, o desempenho da rede neural pode se aproximar ao observado nos cenários que incluíam dados de produção, ainda que com uma sensibilidade mais acentuada às características específicas de cada usina.

Os resultados do Cenário 3 demonstram que, embora as variáveis meteorológicas e a distância das usinas sejam informações relevantes, sua utilização isolada impacta de forma diferenciada o grau de ajuste das previsões, dependendo da localização do sistema fotovoltaico. Um dos achados mais significativos deste cenário foi o bom desempenho registrado na Usina 1, com R^2 igual a 0,903, e na Usina 4, com R^2 de 0,919, mesmo sem a inclusão de dados históricos de geração. Esse comportamento sugere que, para usinas mais distantes, a modelagem baseada exclusivamente em variáveis meteorológicas e distância pode ser eficaz, desde que os dados climáticos utilizados sejam representativos das condições locais. Esses resultados contrastam com os cenários anteriores, nos quais usinas mais afastadas apresentaram desempenho inferior, indicando que, em determinadas situações, informações meteorológicas bem caracterizadas podem ser suficientes para gerar previsões consistentes.

Figura 17 – R^2 por janela de dados históricos - *Cenário 3 - LSTM*



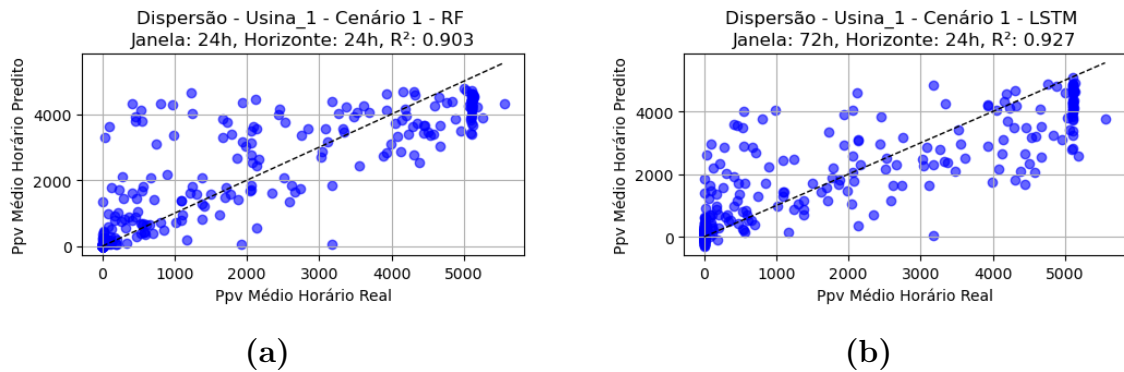
Fonte: Elaborado pelo autor (2025).

No entanto, é importante considerar que a quantidade de dados de treinamento pode ter influenciado esses resultados. Modelos baseados apenas em variáveis ambientais podem apresentar desempenhos distintos à medida que novos dados são incorporados, especialmente se houver maior variabilidade nas condições climáticas ao longo do tempo. Além disso, a sazonalidade da radiação solar e das condições meteorológicas pode impactar na qualidade dos ajustes das previsões, uma vez que períodos de alta incidência solar tendem a apresentar padrões mais regulares de geração, enquanto períodos chuvosos ou nublados podem introduzir maior variabilidade e incerteza nos modelos. Assim, futuras avaliações com conjuntos de dados ampliados e que contemplem diferentes estações do ano podem fornecer uma visão mais precisa sobre a real eficácia do uso exclusivo de dados meteorológicos e distância na previsão da geração fotovoltaica.

4.2.4 Comparação entre os Modelos

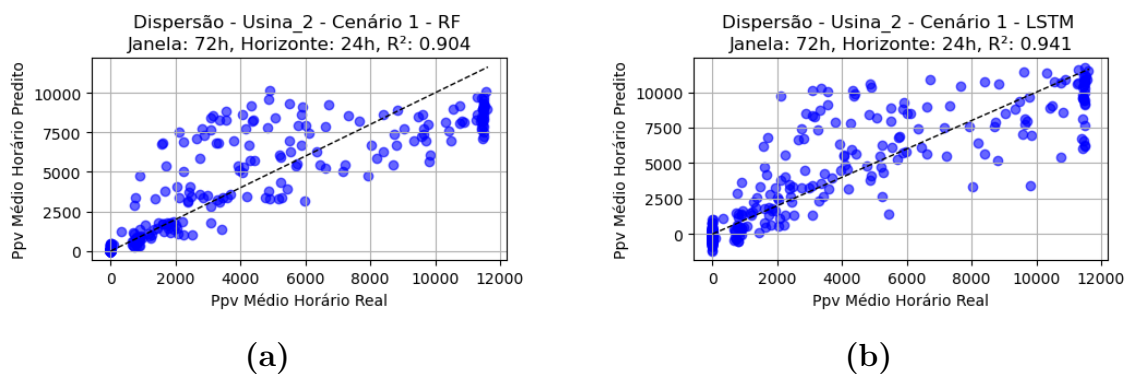
Os resultados têm como objetivo determinar qual abordagem apresenta o menor erro quadrado médio e absoluto e coeficiente de determinação em diferentes janelas de previsão, levando em consideração as particularidades de cada usina e o contexto específico de cada cenário analisado.

Figura 18 – *Comparação da dispersão dos valores de Ppv médio horário da Usina 1 para o maior R^2*



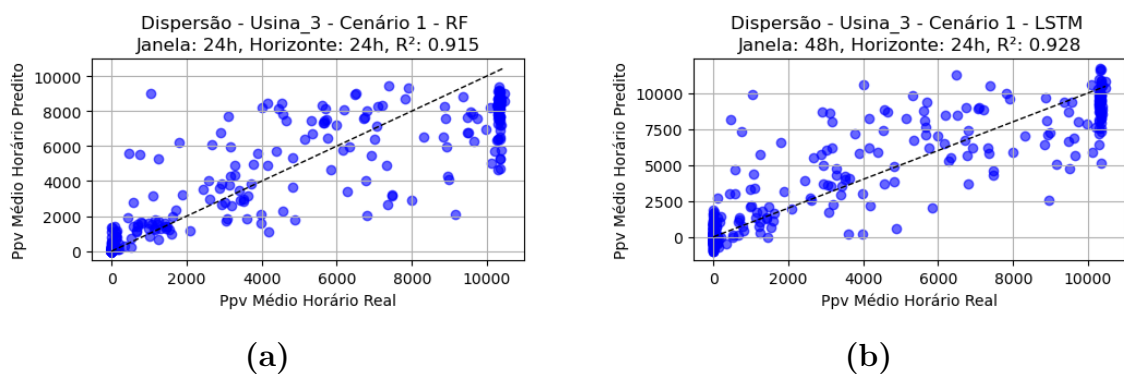
Fonte: Elaborado pelo autor (2025).

Figura 19 – *Comparação da dispersão dos valores de Ppv médio horário da Usina 2 para o maior R^2*



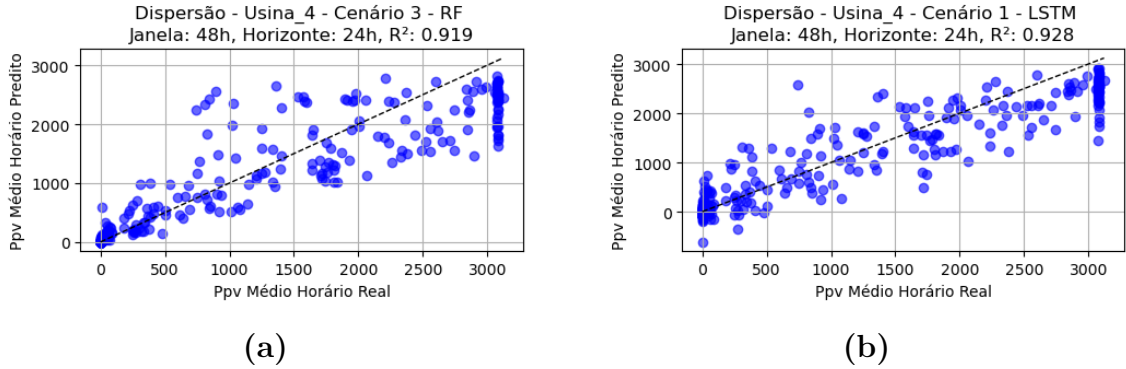
Fonte: Elaborado pelo autor (2025).

Figura 20 – *Comparação da dispersão dos valores de Ppv médio horário da Usina 3 para o maior R^2*



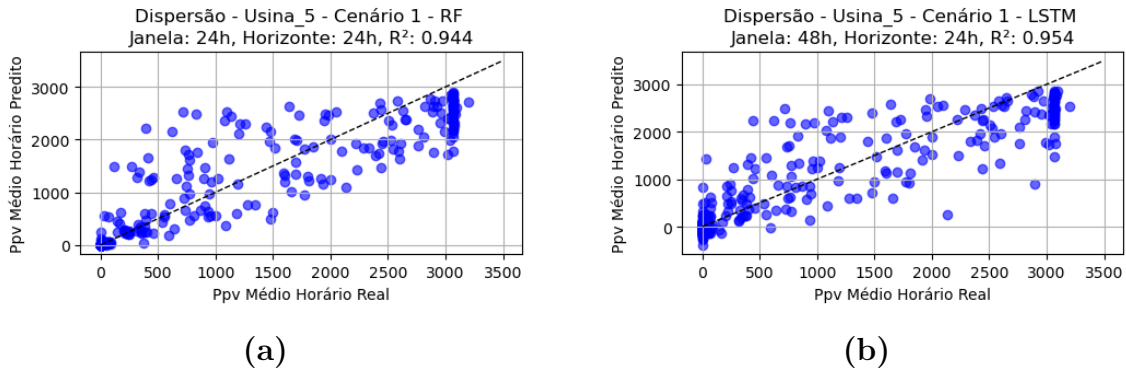
Fonte: Elaborado pelo autor (2025).

Figura 21 – *Comparação da dispersão dos valores de P_{pv} médio horário da Usina 4 para o maior R^2*



Fonte: Elaborado pelo autor (2025).

Figura 22 – *Comparação da dispersão dos valores de P_{pv} médio horário da Usina 5 para o maior R^2*



Fonte: Elaborado pelo autor (2025).

O coeficiente de determinação (R^2) variou entre 0,750 e 0,954, indicando que ambos os modelos apresentaram um bom desempenho em diferentes cenários. No entanto, observou-se que, na maioria dos casos, o modelo RF tende a superar o LSTM. Por exemplo, na Usina 5, o RF alcançou um R^2 de 0,943 com uma janela de previsão de 24 horas, enquanto o LSTM obteve um R^2 ligeiramente superior de 0,954 com uma janela de 48 horas. Esse resultado sugere que, embora o RF seja eficaz para previsões de P_{pv} médio horário especialmente em usinas onde a relação entre as variáveis envolvidas é mais bem capturada por métodos baseados em árvores de decisão o LSTM pode alcançar desempenhos ainda melhores quando ajustado corretamente, considerando janelas históricas mais amplas.

O modelo LSTM demonstrou desempenho superior particularmente na Usina 5, com valores de R^2 acima de 0,94 em diferentes cenários e janelas (por exemplo, $R^2 = 0,944$ com 72 horas no Cenário 2 e $R^2 = 0,943$ com 48 horas no Cenário 3). Esses resultados reforçam a capacidade das redes LSTM de capturar padrões

temporais complexos e sazonais, especialmente em locais com comportamento mais regular de geração solar.

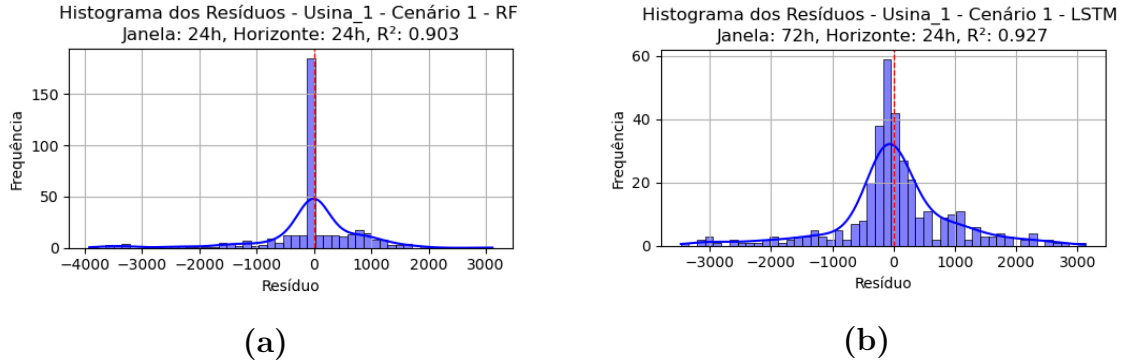
A janela de tempo utilizada para as previsões variou de 1 a 72 horas, e observou-se uma tendência de melhora no desempenho dos modelos à medida que o número de horas consideradas na janela histórica aumentava. Os piores resultados, tanto para o modelo RF quanto para o LSTM, ocorreram quando apenas uma hora anterior foi utilizada como entrada, indicando que uma quantidade muito limitada de dados históricos compromete a capacidade de generalização dos modelos. O desempenho dos modelos mostrou-se mais sensível às características específicas de cada usina do que à janela em si. Na Usina 2, por exemplo, uma janela de 72 horas com RF resultou em um R^2 de 0,903, enquanto o LSTM, com a mesma janela e usina, alcançou um R^2 de 0,941. Isso evidencia que a escolha da janela deve ser feita com base tanto no modelo quanto nas particularidades do local de aplicação.

Portanto, a eficácia dos modelos está intimamente ligada às características específicas de cada usina, como localização geográfica, condições climáticas, capacidade instalada e padrões de geração de energia. Ambos os modelos são promissores, mas sua aplicação ideal depende de uma análise criteriosa do contexto e dos dados históricos disponíveis.

A Usina 1, conforme apresenta a Figura 23, com potência máxima em torno de 5 kWp, apresentou resíduos mais controlados no modelo RF, com variações entre aproximadamente -3919 W e 3117 W, o que representa cerca de 78% da capacidade nominal. Já o modelo LSTM apresentou uma dispersão ainda mais ampla, com resíduos variando entre -4803 W e 3193 W, chegando a 96% da potência máxima. Esses resultados evidenciam uma maior instabilidade no LSTM e maior dificuldade em capturar os padrões de geração com boa qualidade de ajuste. Essa diferença sugere que, para essa usina, o modelo RF oferece previsões mais confiáveis e menos propensas a erros extremos.

Curiosamente, embora o LSTM tenha apresentado um coeficiente de determinação (R^2) ligeiramente superior ao do RF, essa métrica reflete a capacidade do modelo em explicar a variabilidade geral dos dados, não necessariamente sua confiabilidade preditiva ponto a ponto. Dessa forma, o LSTM consegue capturar melhor as tendências globais da geração, mas ainda enfrenta dificuldades em estabilizar seus resíduos em torno de zero.

Figura 23 – *Comparação do resíduo dos valores de Ppv médio horário da Usina 1 para o maior R^2*

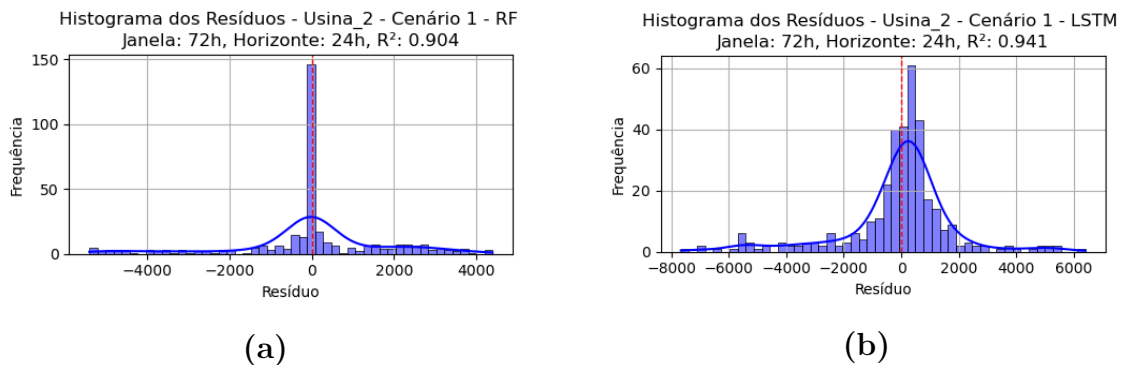


Fonte: Elaborado pelo autor (2025).

A Usina 2, conforme Figura 24, possui uma potência máxima de aproximadamente 10 kWp. O modelo RF apresentou resíduos variando entre -5413 W e 4399 W, o que corresponde a cerca de 54% e 44% da capacidade nominal, respectivamente. Em contrapartida, o modelo LSTM demonstrou uma dispersão significativamente maior, com resíduos que variam de -8838 W a 7818 W, ultrapassando 88% da potência total da usina.

Essa instabilidade pode estar associada à natureza da arquitetura do LSTM, que depende fortemente da qualidade da série temporal e de seu pré-processamento. Além disso, problemas de convergência, sensibilidade à normalização dos dados ou ruídos no sinal de entrada podem levar à geração de previsões com desvios acentuados.

Figura 24 – *Comparação do resíduo dos valores de Ppv médio horário da Usina 2 para o maior R^2*

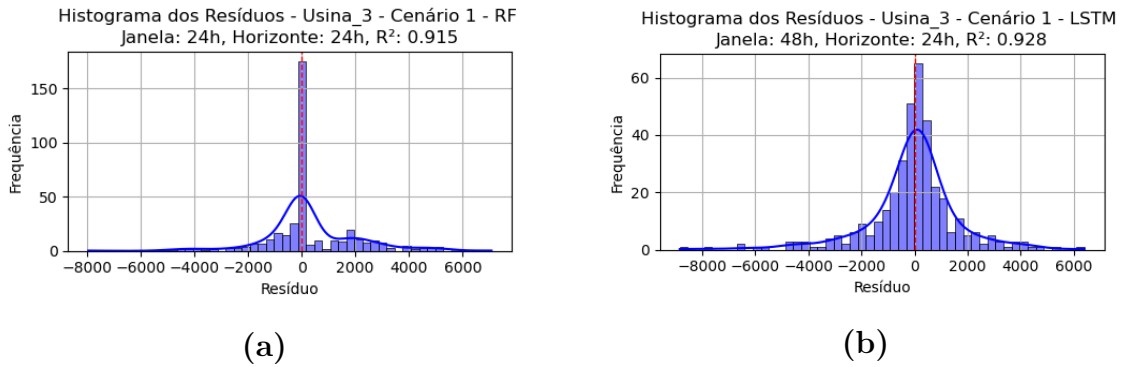


Fonte: Elaborado pelo autor (2025).

A Figura 25 representa a Usina 3, com capacidade instalada de 10 kWp. O modelo RF apresentou uma faixa de resíduos entre -7939 W e 7070 W, o que equivale a até 79% da potência nominal. Já o modelo LSTM exibiu resíduos ainda mais extremos, variando entre -9137 W e 6983 W, atingindo até 91% da capacidade

total da usina. Apesar disso, a diferença de comportamento entre os modelos é relevante: o RF apresenta uma distribuição de resíduos com menor assimetria e mais estabilidade em torno de zero, enquanto o LSTM tende a apresentar oscilações mais acentuadas, especialmente para horizontes temporais mais longos.

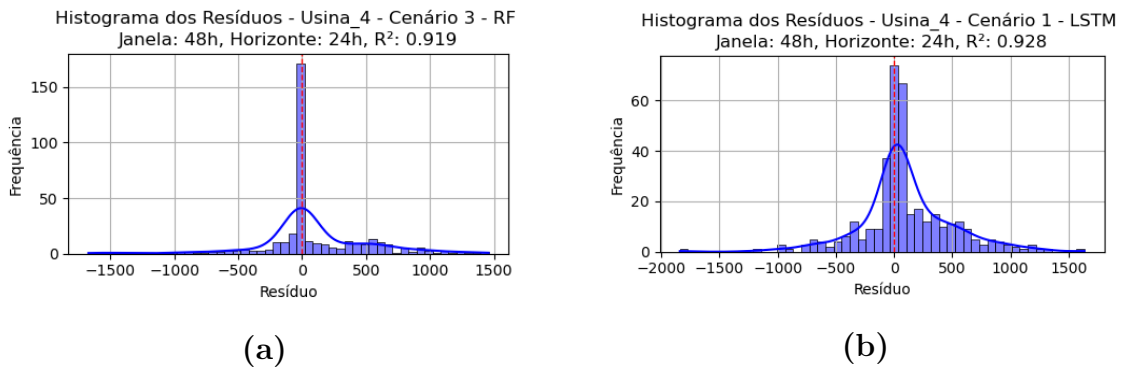
Figura 25 – *Comparação do resíduo dos valores de Ppv médio horário da Usina 3 para o maior R^2*



Fonte: Elaborado pelo autor (2025).

A Usina 4, menor em capacidade, com cerca de 3,5 kWp, também apresenta um contraste interessante entre os modelos. O RF mantém resíduos em torno de 1671 W, o que corresponde a aproximadamente 47,7% da potência total da usina. No entanto, o LSTM apresenta resíduos que atingem 2046 W, cerca de 58,4% da potência nominal, indicando que suas previsões possuem uma variabilidade maior, como apresentado na Figura 26.

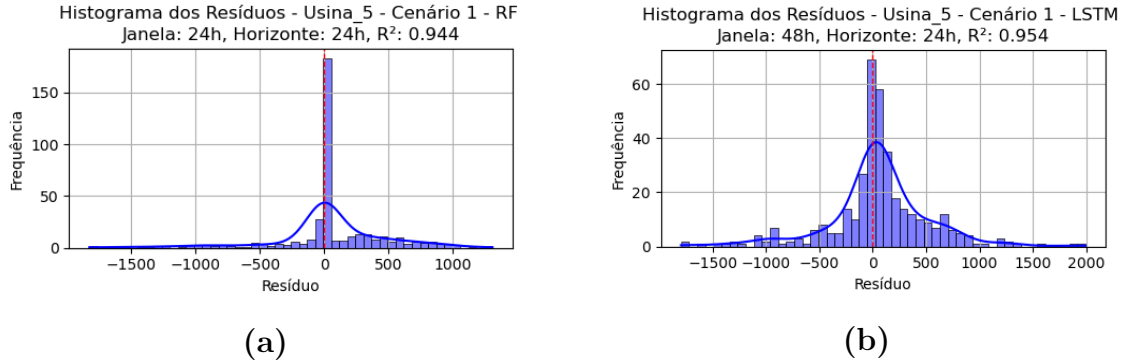
Figura 26 – *Comparação do resíduo dos valores de Ppv médio horário da Usina 4 para o maior R^2*



Fonte: Elaborado pelo autor (2025).

Por fim, a Usina 5, representada na Figura 27, com capacidade de 3 kWp, apresentou resíduos de até 1825 W no modelo RF, o que representa um erro relativo de aproximadamente 60,8%. O modelo LSTM, por outro lado, exibiu resíduos variando até 2508 W, correspondendo a cerca de 83,6% da potência nominal.

Figura 27 – *Comparação do resíduo dos valores de Ppv médio horário da Usina 5 para o maior R^2*



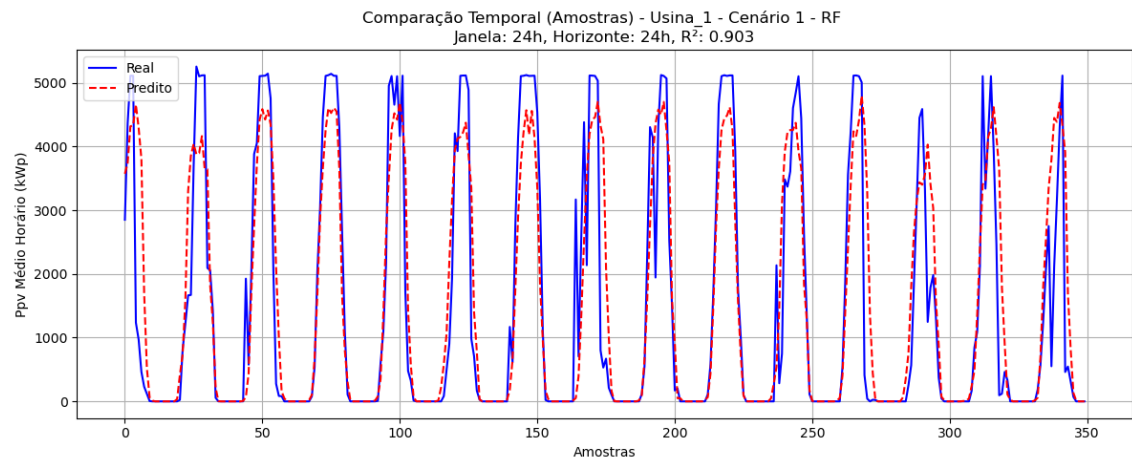
Fonte: Elaborado pelo autor (2025).

Portanto, ao analisar esses resultados, pode-se observar que o modelo RF mantém erros menores e mais estáveis, enquanto o LSTM apresenta resíduos significativamente mais amplos e demonstra dificuldades em estabilizar as previsões próximas de zero. O impacto desses erros se torna mais crítico em usinas de menor porte, onde desvios superiores a 30% da potência nominal podem comprometer o planejamento da geração e o gerenciamento da distribuição de energia.

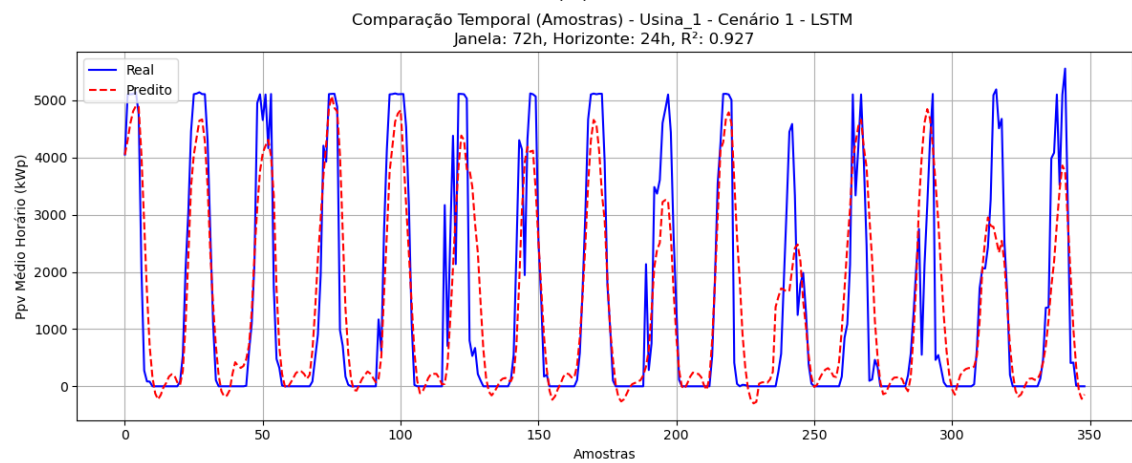
A superioridade do RF se destaca especialmente na sua capacidade de capturar padrões não-lineares da geração solar com maior robustez e menor sensibilidade a ruídos. Por outro lado, o LSTM, embora capaz de alcançar bons valores de R^2 , mostrou-se mais dependente do volume e da diversidade dos dados de entrada, o que pode ter sido um fator limitante, considerando a janela temporal de apenas três meses utilizada no treinamento. Essa limitação pode ter afetado sua capacidade de generalização e contribuído para os altos resíduos observados.

A subestimação observada em grande parte das previsões realizadas pelo modelo LSTM, conforme mostrado nas Figuras 28 a 32, sugere que o modelo pode ter sido treinado predominantemente com dados de períodos de menor geração. Isso resultou em previsões sistematicamente abaixo dos valores reais em momentos de maior irradiação, especialmente na Usina 2, onde mais de 59% dos resíduos foram negativos. Tal comportamento é evidenciado pela maior concentração de resíduos negativos, indicando uma limitação na capacidade do modelo de capturar picos de geração solar.

Figura 28 – *Comparação de valores reais e previstos da Usina 1 para o maior R^2*



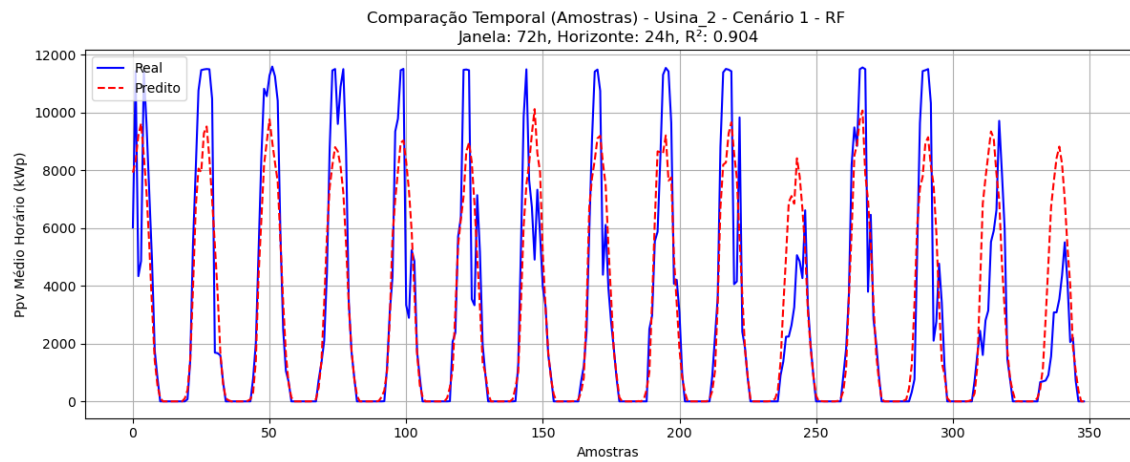
(a)



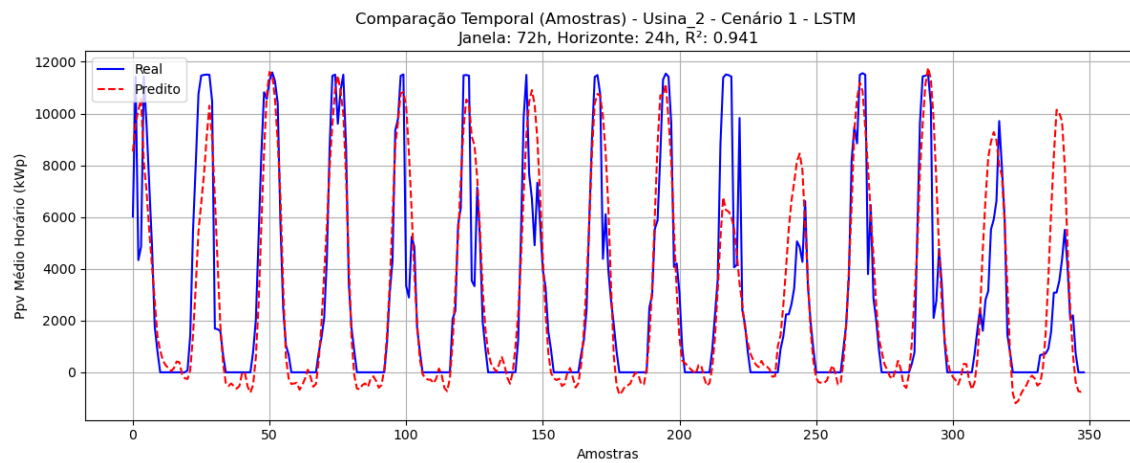
(b)

Fonte: Elaborado pelo autor (2025).

Figura 29 – *Comparação de valores reais e previstos da Usina 2 para o maior R^2*



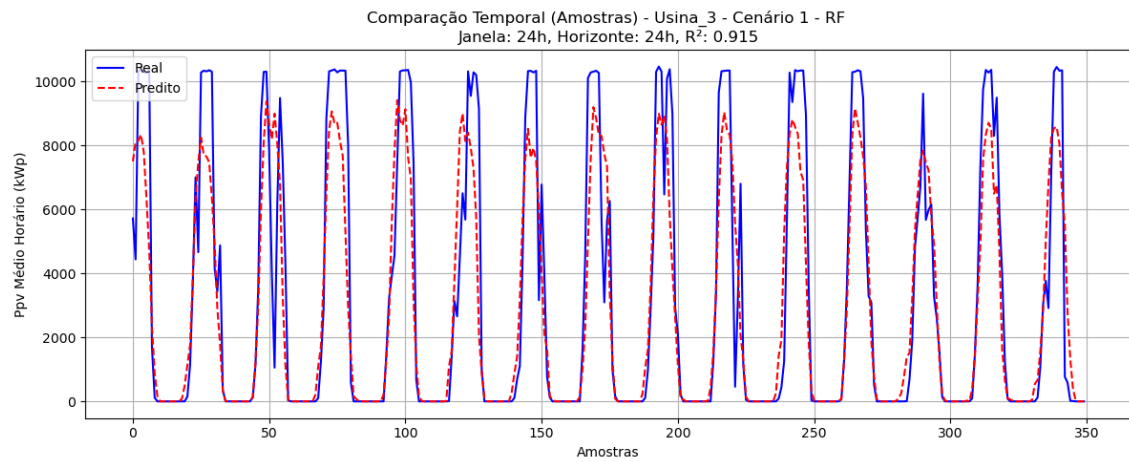
(a)



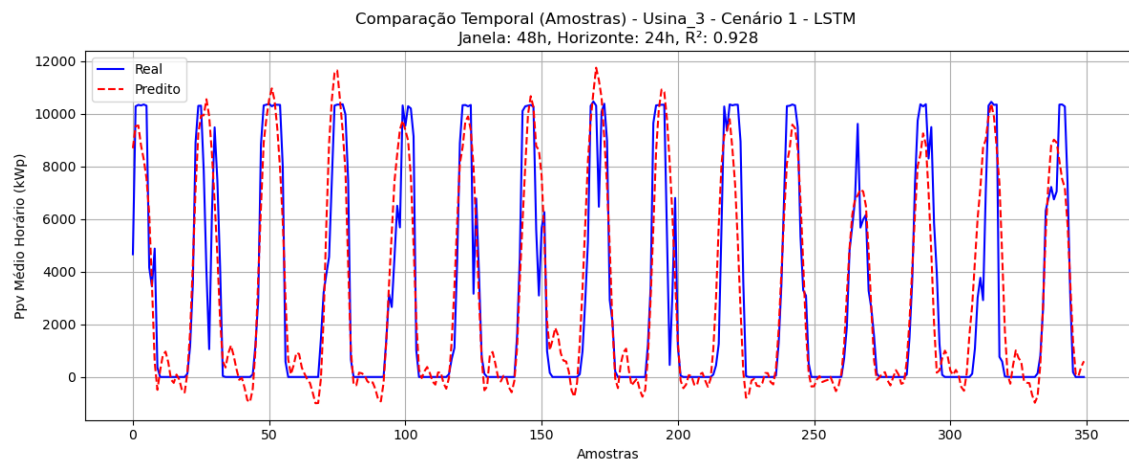
(b)

Fonte: Elaborado pelo autor (2025).

Figura 30 – *Comparação de valores reais e previstos da Usina 3 para o maior R^2*



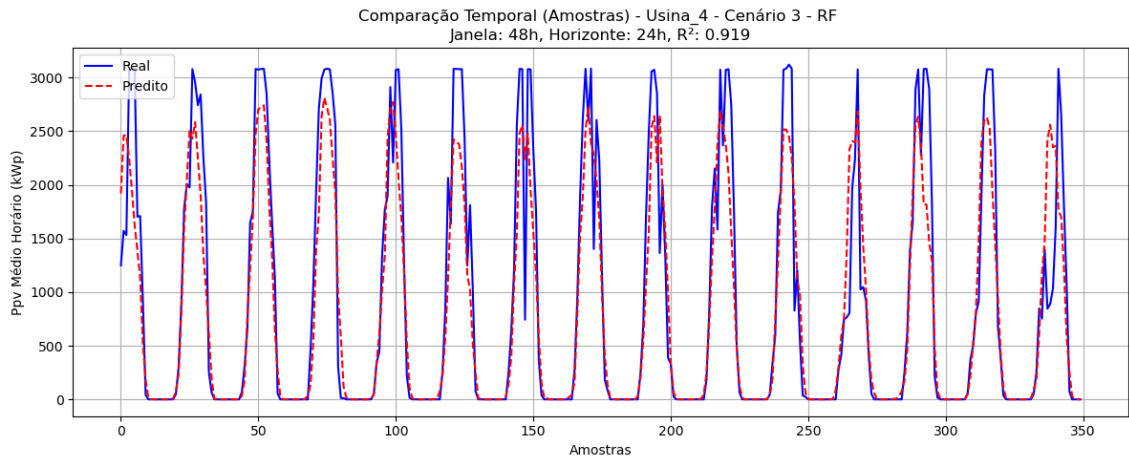
(a)



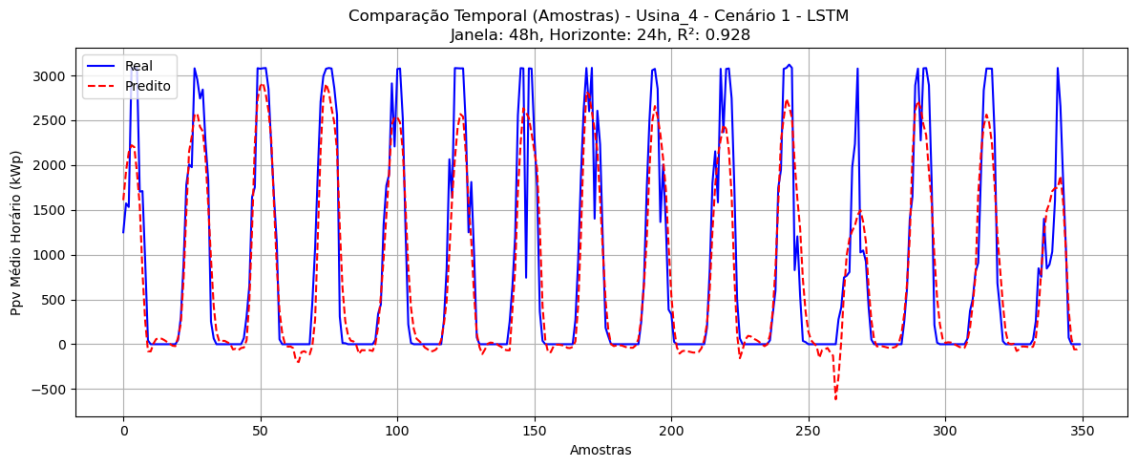
(b)

Fonte: Elaborado pelo autor (2025).

Figura 31 – *Comparação de valores reais e previstos da Usina 4 para o maior R^2*



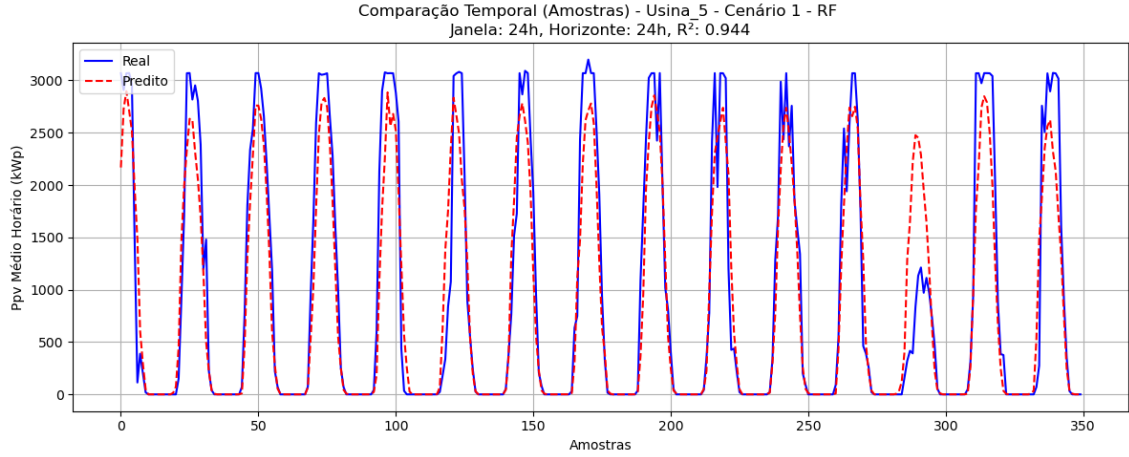
(a)



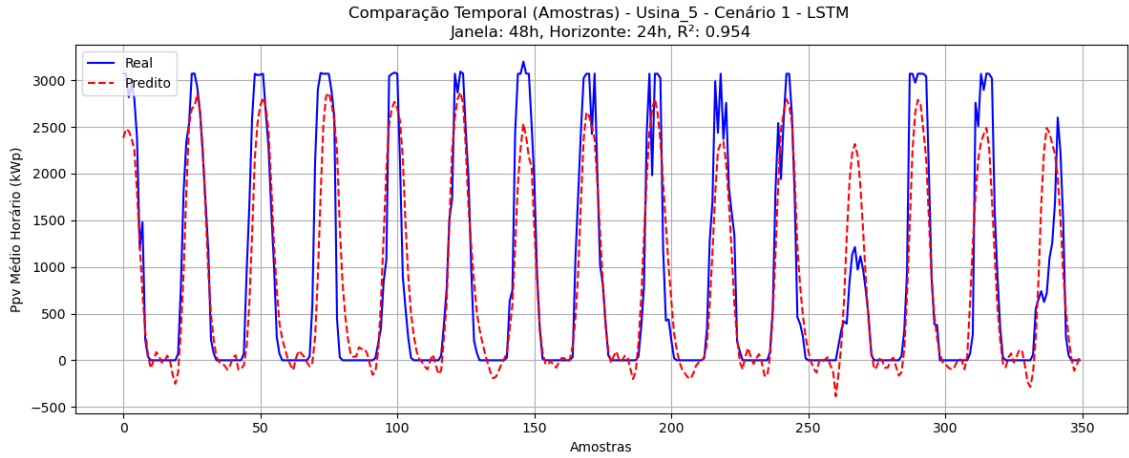
(b)

Fonte: Elaborado pelo autor (2025).

Figura 32 – *Comparação de valores reais e previstos da Usina 5 para o maior R^2*



(a)



(b)

Fonte: Elaborado pelo autor (2025).

A dispersão dos erros aumenta com o crescimento do horizonte de previsão, como evidenciado nas Figuras 28 a 32, indicando que a incerteza climática impacta significativamente os modelos, resultando no acúmulo de erros ao longo do tempo. Além disso, observa-se que o modelo LSTM apresenta maiores dificuldades em estabilizar as previsões em torno de zero, especialmente durante os períodos de transição entre baixa e alta geração, onde ocorrem oscilações mais acentuadas nos resíduos. Esse comportamento pode ser atribuído à natureza sequencial e dependente do histórico do LSTM, que tende a propagar erros ao longo das etapas temporais. Quando o conjunto de treinamento não inclui exemplos suficientes desses momentos de transição, o modelo apresenta dificuldades para convergir corretamente, comprometendo a precisão das estimativas nesses intervalos críticos.

A avaliação das Tabelas 9 a 12, que apresentam o desempenho dos modelos, revela que não existe um cenário único ideal para todas as usinas e métricas,

evidenciando a necessidade de ajustes específicos conforme o contexto analisado. O RF destacou-se por sua maior eficiência em janelas curtas e intermediárias, especialmente na minimização do MAE, com destaque para a janela de 24 horas, que foi a melhor escolha para quatro das cinco usinas.

Esse resultado indica que a abordagem estatística do RF é capaz de capturar padrões recorrentes de geração com boa precisão local, favorecendo previsões mais estáveis. No entanto, ao considerar o RMSE, observou-se que janelas mais longas, como 72 horas, apresentaram melhores resultados em algumas usinas, como a Usina 2, sugerindo que horizontes históricos mais amplos podem ser benéficos para mitigar grandes erros pontuais em contextos de maior variabilidade.

Tabela 9 – *Melhores valores de RMSE por sistema fotovoltaico - RF*

Usina	Cenário (RMSE)	Melhor RMSE	Janela histórica (RMSE)
1	1	567,57	24
2	1	1146,58	72
3	1	1163,87	24
4	3	275,37	48
5	1	256,86	24

Fonte: Elaborado pelo autor (2025).

O Cenário 1, que incorpora o maior número de variáveis, demonstrou um desempenho consistente em todas as usinas no modelo LSTM. Por exemplo, na Usina 2, esse cenário obteve o menor MAE (569,50) e RMSE (914,70), ambos com uma janela histórica de 48 horas, indicando que a inclusão de variáveis meteorológicas e de Ppv médio horário contribuiu para previsões mais estáveis. Situação semelhante foi observada na Usina 5, com os menores erros também registrados no Cenário 1. Isso reforça a hipótese de que a complexidade do modelo LSTM beneficia-se da riqueza de variáveis, especialmente em contextos com padrões climáticos mais dinâmicos.

O Cenário 1, que incorpora múltiplas variáveis, incluindo dados meteorológicos e o P_{pv} médio horário, demonstrou melhor desempenho na maioria das usinas, tanto para o RMSE quanto para o MAE, especialmente no modelo LSTM. Por exemplo, na Usina 3, esse cenário apresentou o menor RMSE (1046,53) e o menor MAE (626,90), ambos com uma janela histórica de 48 horas. Esse resultado sugere que a complexidade do modelo, ao considerar informações climáticas e operacionais, favo-

Tabela 10 – *Melhores valores de MAE por sistema fotovoltaico - RF*

Usina	Cenário (MAE)	Melhor MAE	Janela histórica (MAE)
1	2	234,24	24
2	2	546,94	24
3	1	554,60	24
4	1	110,38	24
5	1	106,73	24

Fonte: Elaborado pelo autor (2025).

rece a capacidade de capturar variações na geração de energia com maior precisão, principalmente em horizontes temporais intermediários, como janelas de 48 horas.

Mesmo com a estação meteorológica localizada a uma distância considerável de algumas usinas, o Cenário 1 ainda apresentou os melhores resultados, especialmente no modelo LSTM, o que reforça a relevância das variáveis climáticas no desempenho dos modelos. Esse achado indica que, embora a proximidade da fonte de dados possa influenciar a qualidade das medições, a inclusão dessas informações contribui de forma significativa para a robustez e estabilidade das previsões.

A análise do tamanho da janela histórica revela que essa variável exerce influência significativa sobre o desempenho dos modelos. Para o RMSE, que é sensível a grandes desvios, janelas de 48 a 72 horas proporcionaram os melhores resultados, especialmente no modelo LSTM. De forma semelhante, o MAE, que mede o erro absoluto médio, também apresentou seus menores valores com janelas de 48 horas em grande parte das usinas. Esse comportamento indica maior estabilidade nas previsões e menor suscetibilidade a oscilações pontuais. Esses achados reforçam a importância de ajustar tanto a configuração do cenário quanto o horizonte histórico conforme o perfil e a dinâmica de cada sistema fotovoltaico.

Tabela 11 – *Melhores valores de RMSE por sistema fotovoltaico - LSTM*

Usina	Cenário (RMSE)	Melhor RMSE	Janela Histórica (RMSE)
1	1	491,85	48
2	1	914,70	48
3	1	1046,53	48
4	1	242,34	72
5	1	238,71	48

Fonte: Elaborado pelo autor (2025).

Tabela 12 – *Melhores valores de MAE por sistema fotovoltaico - LSTM*

Usina	Cenário (MAE)	Melhor MAE	Janela Histórica (MAE)
1	1	281,56	48
2	1	569,50	48
3	1	626,90	48
4	1	144,29	72
5	1	140,42	24

Fonte: Elaborado pelo autor (2025).

Considerações Finais

Este trabalho resultará em diversas contribuições significativas para o estudo e avaliação de métodos aplicados à predição de geração de energia solar, analisando o impacto de diferentes cenários de entrada nessa predição. Evidencia-se que não existe uma janela temporal ou conjunto de *features* universalmente ideal para prever o comportamento de sistemas fotovoltaicos. Portanto, é fundamental avaliar a quantidade e a qualidade dos dados fornecidos, bem como as características específicas das usinas, para garantir previsões precisas e confiáveis. Além disso, este estudo contribui para a manutenção preditiva de usinas fotovoltaicas, com foco na otimização da eficiência operacional e na redução de custos.

Uma das principais contribuições deste trabalho é o desenvolvimento de abordagens inovadoras que integram dados meteorológicos e o Ppv Médio Horário com técnicas de IA para prever a geração de energia solar. Esse monitoramento preditivo permite além de antecipar falhas e implementar ações corretivas antes que ocorram perdas significativas na geração de energia, também aumenta a confiabilidade e a vida útil dos sistemas fotovoltaicos. Somado a isso, o estudo permite a criação de sistemas capazes de estimar a capacidade de geração em regiões que não possuem equipamentos meteorológicos, superando a limitação dos métodos determinísticos tradicionais. Essa abordagem abre novas possibilidades para a expansão e otimização de usinas fotovoltaicas em áreas carentes de infraestrutura de medição, contribuindo para a democratização do acesso à energia solar.

5.1 Conclusão

Este estudo analisou a previsão da geração de energia fotovoltaica utilizando modelos de aprendizado de máquina, comparando o desempenho do RF e da LSTM em diferentes usinas e janelas temporais de previsão. Os resultados indicaram que o RF demonstrou maior estabilidade e menor erro médio em várias usinas, com valores de R^2 variando entre aproximadamente 0,903 e 0,944, enquanto o LSTM apresentou desempenhos ligeiramente superiores em alguns casos específicos, alcançando R^2 de

até 0,954, mas com maior dispersão de erros. Essa diferença pode ser atribuída ao fato de que o RF é mais eficiente na captura de padrões não lineares com menor sensibilidade ao volume de dados, enquanto o LSTM depende de um histórico mais longo e estruturado para otimizar sua capacidade de modelagem sequencial.

A análise dos resíduos revelou que o LSTM teve maior dificuldade em estabilizar as previsões próximas de zero, especialmente em usinas de menor porte e durante transições entre períodos de baixa e alta geração. Além disso, observou-se uma tendência à subestimação sistemática, com predominância de resíduos negativos, o que sugere que o modelo pode ter sido treinado com dados concentrados em períodos de menor geração. Isso resultou em previsões abaixo dos valores reais em momentos de maior irradiância solar. Essa limitação reforça a necessidade de expandir a base de treinamento para incluir um ciclo sazonal completo, garantindo que o modelo aprenda padrões mais representativos ao longo do ano.

A dispersão dos erros aumentou com o horizonte de previsão, refletindo a crescente incerteza climática e a dificuldade dos modelos em capturar padrões de longo prazo a partir de 24 horas. No entanto, janelas intermediárias de 48 horas apresentaram o melhor equilíbrio entre estabilidade e capacidade preditiva, especialmente no modelo LSTM.

Os resultados das métricas utilizadas indicam que não há um único modelo ideal para todas as usinas e métricas. O RF mostrou-se mais eficiente para o MAE em janelas de 24 horas, enquanto o LSTM apresentou melhor desempenho em RMSE e MAE com janelas de 48 horas, demonstrando sensibilidade ao volume e à diversidade dos dados. Para o RMSE, observou-se que janelas mais longas tendem a minimizar os grandes desvios, evidenciando a necessidade de ajustes conforme a métrica e o tipo de variação que se deseja capturar.

A análise comparativa entre os modelos reforça a necessidade de ajustes específicos conforme as características de cada usina. Estratégias como o uso de técnicas que combinem previsões de diferentes modelos podem contribuir para melhorar a confiabilidade das estimativas. Além disso, a incorporação de dados exógenos, como índices de nebulosidade e padrões atmosféricos, pode ajudar a capturar variações abruptas na geração de energia solar. Ajustes nos hiperparâmetros do LSTM, incluindo otimização da taxa de aprendizado e do número de camadas recorrentes, também podem aprimorar sua capacidade de generalização e reduzir os erros observados.

Por fim, a implementação de metodologias que integrem modelos de aprendizado com dados meteorológicos mais detalhados e a ampliação do período de coleta de dados são medidas fundamentais para aprimorar a previsibilidade da geração fotovoltaica. O estudo demonstrou que o RF é, no cenário atual, uma alternativa

robusta para previsões de curto e médio prazo, enquanto o LSTM apresenta potencial competitivo, desde que acompanhado de um maior refinamento no treinamento. Com a adoção de técnicas mais robustas e a inclusão de um volume maior de dados, espera-se que as previsões futuras sejam mais precisas e estáveis, favorecendo a gestão eficiente da geração solar e sua integração à rede elétrica.

5.2 Trabalhos Futuros

Com base nos resultados obtidos e nas limitações identificadas neste estudo, algumas direções podem ser exploradas em pesquisas futuras para aprimorar a previsão da geração solar e otimizar a operação de sistemas fotovoltaicos. A seguir, são apresentadas sugestões de trabalhos que podem contribuir para avanços nessa área:

- Quantificação da sujeira nos módulos fotovoltaicos: Desenvolvimento de métodos para medir o acúmulo de poeira e resíduos nos painéis solares, utilizando sensores ou visão computacional, a fim de otimizar estratégias de manutenção e minimizar perdas de eficiência.
- Análise econômica da limpeza dos painéis: Avaliação da relação custo-benefício da limpeza dos módulos fotovoltaicos, determinando a frequência ideal para maximizar a eficiência energética e reduzir despesas operacionais.
- Estimativa da capacidade de geração regional: Desenvolvimento de modelos baseados em dados de usinas próximas e informações meteorológicas para prever o potencial de geração solar em regiões sem monitoramento direto, ampliando o desempenho preditivo das análises energéticas.
- Desenvolvimento de uma plataforma web para avaliação de usinas fotovoltaicas: Criação de um sistema online que forneça alertas de desempenho para sistemas fotovoltaicos, permitindo que os usuários comparem o desempenho de suas usinas com outras próximas, utilizando dados meteorológicos e históricos de geração.

Referências

AMAZON WEB SERVICES. **The Difference Between Data Science and AI**. 2024. <https://aws.amazon.com/pt/compare/the-difference-between-data-science-and-ai/>. Acessado em: 27/06/2024. Disponível em: <https://aws.amazon.com/pt/compare/the-difference-between-data-science-and-ai/>. 31

ANEEL. **Resolução Normativa nº 482/2012**. 2012. www.aneel.gov.br. Acessado em: 18/12/2024. 20

_____. **RESOLUÇÃO NORMATIVA ANEEL Nº 1.000, DE 7 DE DEZEMBRO DE 2021**. 2021. www.aneel.gov.br. Estabelece as Regras de Prestação do Serviço Público de Distribuição de Energia Elétrica; revoga as Resoluções Normativas ANEEL nº 414, de 9 de setembro de 2010; nº 470, de 13 de dezembro de 2011; nº 901, de 8 de dezembro de 2020 e dá outras providências. Disponível em: <https://www.aneel.gov.br/>. 20

_____. **Resolução Normativa ANEEL nº 1.059, de 7 de Fevereiro de 2023**. 2023. www.aneel.gov.br. Acessado em: 18/12/2024. 20, 30

ARAÚJO, Danielly Norberto; CARVALHO, Paulo César Marques de; DUPONT, Ivonne Montero. Efeitos da acumulação de sujeira sobre o desempenho de módulos fotovoltaicos. **Revista Tecnologia**, v. 40, n. 2, 2019. 27

ASSOCIAÇÃO BRASILEIRA DE ENERGIA SOLAR FOTOVOLTAICA. **Energia Solar Fotovoltaica no Brasil: Infográfico ABSOLAR**. jul 2024. Infográfico. Atualizado em 03/07/2024, n.º 69. Disponível em: <https://www.absolar.org.br/mercado/infografico/>. 25

BARBOSA, Elismar Ramos; FARIA, Merlim dos Santos Ferreira; GONTIJO, Fabio de Brito. Influência da sujeira na geração fotovoltaica. *In: VII Congresso Brasileiro de Energia Solar-CBENS 2018*. [S.l.: s.n.], 2018. 21

BERRAR, Daniel *et al.* **Cross-validation**. 2019. 39

BISCHL, Bernd; BINDER, Martin; LANG, Michel; PIELOK, Tobias; RICHTER, Jakob; COORS, Stefan; THOMAS, Janek; ULLMANN, Theresa; BECKER, Marc; BOULESTEIX, Anne-Laure *et al.* Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 13, n. 2, p. e1484, 2023. 50

BONKANEY, Abdou Latif; MADOUGOU, Saidou; ADAMOU, Rabani *et al.* Impact of climatic parameters on the performance of solar photovoltaic (pv) module in niamey. **Smart Grid and Renewable Energy**, Scientific Research Publishing, v. 8, n. 12, p. 379, 2017. 26

BORTOLOTO, Valter A; SOUZA, Andre J; GOIS, Guilherme; MARTINS, Marcio A; BERGHE, Murilo J; MONTANHA, Gustavo Kimura. Geração de energia solar on grid e off grid. *In: VI JORNACITEC-Jornada Científica e Tecnológica*. [S.l.: s.n.], 2017. 30

BOTCHKAREV, Alexei. A new typology design of performance metrics to measure errors in machine learning regression algorithms. **Interdisciplinary Journal of Information, Knowledge, and Management**, v. 14, p. 045–076, 2019. 52

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

CARRERA, Berny; KIM, Kwanho. Deep learning-based approach for short-term solar power forecasting. *In: SPRINGER. International Conference on Green Energy and Environment Engineering*. [S.l.], 2023. p. 119–127. 47

CARVALHO, M. J. d.; FANTINI, D. G.; SIQUEIRA, M. B. B. d. Métodos de previsão solar intra-hora. *In: Anais do Congresso Brasileiro de Energia Solar - CBENS*. [s.n.], 2022. Disponível em: <https://doi.org/10.59627/cbens.2022.1179>. 29

CASARIN, Ricardo. **Geração de energia solar cresce 72,6% em novembro, aponta InfoMercado da CCEE**. 2022. Disponível em: <https://www.portalsolar.com.br/noticias/operacao-e-expansao/oem/geracao-de-energia-solar-no-brasil-tem-crescimento-de-65-9-em-abril>. Acesso em: 08 dez. 2022. 20

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: A survey. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 41, n. 3, p. 1–58, 2009. 21

CUNHA, Bruno Aguilar da. Previsão intra-diária de geração fotovoltaica usando redes neurais recorrentes do tipo lstm e dados históricos de energia. Universidade Estadual Paulista (Unesp), 2021. 26

CUNHA, Henrique Queiroz; SOBEL, Leonardo Farias. Aplicação de redes neurais artificiais na predição de geração de energia fotovoltaica no nordeste do brasil. **Revista de Engenharia e Pesquisa Aplicada**, v. 6, n. 5, p. 73–80, 2021. 49

CUNHA, João Paulo Zanola. **Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos**. Tese (Doutorado) — Universidade de São Paulo, 2019. 50, 51

DIVYANSHU. **LSTM and its Equations**. 2018. <https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af>. Acessado em: 29/05/2024. 35

FILHO, Dalson Britto Figueiredo; JÚNIOR, José Alexandre Silva. Desvendando os mistérios do coeficiente de correlação de pearson (r). **Revista Política Hoje**, v. 18, n. 1, p. 115–146, 2009. 39

GASPARIN, Fernanda Bach. A influência de políticas públicas para o progresso da geração solar fotovoltaica e diversificação da matriz energética brasileira. **Revista Virtual de Química**, v. 14, n. 1, 2022. 21

GRÜBLER, Murillo. Entendendo o funcionamento de uma rede neural artificial. **Brasil AI on Medium**, 2024. Acessado em: 28/01/2024. Disponível em: <https://medium.com/brasil-ai/entendendo-o-funcionamento-de-uma-rede-neural-artificial-4463fcf44dd0>. 34

GUERMOUI, Mawloud; BENKACIALI, Said; GAIRAA, Kacem; BOUCHOUICHA, Kada; BOULMAIZ, Tayeb; BOLAND, John W. A novel ensemble learning approach for hourly global solar radiation forecasting. **Neural Computing and Applications**, Springer, p. 1–23, 2022. 47

GULLI, Antonio; PAL, Sujit. **Deep learning with Keras**. [S.l.]: Packt Publishing Ltd, 2017. 50

HACKELING, Gavin. **Mastering Machine Learning with scikit-learn**. [S.l.]: Packt Publishing Ltd, 2017. 45

HARRIS, Charles R; MILLMAN, K Jarrod; WALT, Stéfan J Van Der; GOMMERS, Ralf; VIRTANEN, Pauli; COURNAPEAU, David; WIESER, Eric; TAYLOR, Julian; BERG, Sebastian; SMITH, Nathaniel J *et al.* Array programming with NumPy.

Nature, Nature Publishing Group UK London, v. 585, n. 7825, p. 357–362, 2020. 45

HAUPT, Sue Ellen; MCCANDLESS, Tyler C; LEE, Jared C; KOSOVIC, Branko; ALESSANDRINI, Stefano; DETTLING, Susan; HUSSAIN, Tahani; AL-RASHEEDI, Majed. Combining physical modeling with artificial intelligence for solar power forecasting. *In: IEEE. 2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. [S.l.], 2020. p. 2051–2053. 21

HAYASHI, Chikio. What is data science? fundamental concepts and a heuristic example. *In: SPRINGER. Data Science, Classification, and Related Methods: Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96), Kobe, Japan, March 27–30, 1996*. [S.l.], 1998. p. 40–51. 31

HAYKIN, Simon. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2001. 33

HODSON, Timothy O. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. **Geoscientific Model Development Discussions**, Göttingen, Germany, v. 2022, p. 1–10, 2022. 38

IMHOFF, Johninon *et al.* Desenvolvimento de conversores estáticos para sistemas fotovoltaicos autônomos. Universidade Federal de Santa Maria, 2007. 24

INMET, Instituto Nacional de Meteorologia – **Balanço do Mês de Novembro de 2023 em Goiânia (GO)**. Brasília: [s.n.], dezembro 2023. Relatório meteorológico. 26, 27

INSTITUTO NACIONAL DE METEOROLOGIA. **Mapas Interativos - Instituto Nacional de Meteorologia**. 2024. Acessado em: 25 ago. 2024. Disponível em: <https://mapas.inmet.gov.br/>. 43

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert; TAYLOR, Jonathan. Tree-based methods. *In: An introduction to statistical learning: with applications in python*. [S.l.]: Springer, 2023. p. 331–366. 32

KLEMME, Florian; PRINZ, Jannik; SANTEN, Victor M van; HENKEL, Jörg; AM-ROUCH, Hussam. Modeling emerging technologies using machine learning: Challenges and opportunities. *In: Proceedings of the 39th International Conference on Computer-Aided Design*. [S.l.: s.n.], 2020. p. 1–9. 52

KOTU, Vijay; DESHPANDE, Bala. **Data science: concepts and practice**. [S.l.]: Morgan Kaufmann, 2018. 31

Kyocera. **KK275P-3CD3CG High Efficiency Multicrystalline Photovoltaic Module**. 6 Takeda Tobadono-cho Fushimi-ku, Kyoto 612-8501, Japan: [s.n.], 2024. <http://www.kyocera.com/>. Technical Specification Sheet. 27

LEMES, Vinicius Gustavo de Jesus Ferreira. **Análise da eficiência da geração de energia elétrica em placas solares, a partir da interferência da temperatura sobre a placa**. Dissertação de Mestrado. Dissertação (Mestrado) — Universidade Estadual de Campinas, Faculdade de Tecnologia, Limeira, SP, Novembro 2023. Disponível em: <http://lattes.cnpq.br/7639509960443216>. 57

LIMA, G. A.; GOMES, M. P.; MARTINEZ, L. Previsão de irradiação solar usando modelos de séries temporais. In: **Anais do Simpósio Brasileiro de Sistemas Elétricos 2020**. [s.n.], 2020. Disponível em: <https://doi.org/10.48011/sbse.v1i1.2224>. 29

MADHURI. **The Roles of Data Sets in Machine Learning Projects: A Guide to Data Splitting**. 2024. Accessed: 2025-02-16. Disponível em: <https://medium.com/@madhuri15/the-roles-of-data-sets-in-machine-learning-projects-a-guide-to-data-splitting-454beb468a49>. 47

MARTINS, EGM. Coeficiente de determinação. **Revista Ciência Elementar**, v. 6, n. 1, p. 24, 2018. 39

MATTIAS, Alef. **Criando um Perceptron do Zero com Python**. 2023. Acessado em: 16 fev. 2025. Disponível em: <https://medium.com/@alef.mattias/criando-um-perceptron-do-zero-com-python-f23bba2375bf>. 34

MCKINNEY, Wes *et al.* pandas: a foundational python library for data analysis and statistics. **Python for high performance and scientific computing**, Seattle, v. 14, n. 9, p. 1–9, 2011. 45

MEDSKER, Larry; JAIN, Lakhmi C. **Recurrent neural networks: design and applications**. [S.l.]: CRC press, 1999. 34

MINAMI, Shotaro. Predicting equity price with corporate action events using lstm-rnn. **Journal of Mathematical Finance**, Scientific Research Publishing, v. 8, n. 1, p. 58–63, 2018. 36

MYLLYAHÖ, Lalli; RAATIKAINEN, Mikko; MÄNNISTÖ, Tomi; MIKKONEN, Tommi; NURMINEN, Jukka K. Systematic literature review of validation methods

for ai systems. **Journal of Systems and Software**, Elsevier, v. 181, p. 111050, 2021. 53

NASCIMENTO, Cássio Araújo do. Princípio de funcionamento da célula fotovoltaica. **Monografia. 21p. Disponível em http://www.solenerg.com.br/files/monografia_cassio.pdf. Acesso em**, v. 10, 2004. 25

NASSIF, Ali Bou; TALIB, Manar Abu; NASIR, Qassim; DAKALBAB, Fatima Mohamad. Machine learning for anomaly detection: A systematic review. **Ieee Access**, IEEE, v. 9, p. 78658–78700, 2021. 21

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **O Sistema em Números**. 2024. <https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>. Acessado em: 2 de setembro de 2024. 20

OSISANWO, FY; AKINSOLA, JET; AWODELE, O; HINMIKAIYE, JO; OLAKANMI, O; AKINJOBI, J *et al.* Supervised machine learning algorithms: classification and comparison. **International Journal of Computer Trends and Technology (IJCTT)**, v. 48, n. 3, p. 128–138, 2017. 32

PAN, Chen; LIU, Yuqiao; OH, Yeonjae; LIM, Changgyoon. Short-term photovoltaic power forecasting using pv data and sky images in an auto cross modal correlation attention multimodal framework. **Energies**, MDPI, v. 17, n. 24, p. 6378, 2024. 43

PANG, Bo; NIJKAMP, Erik; WU, Ying Nian. Deep learning with tensorflow: A review. **Journal of Educational and Behavioral Statistics**, SAGE Publications Sage CA: Los Angeles, CA, v. 45, n. 2, p. 227–248, 2020. 50

PAREJO, Antonio; BRACCO, Stefano; PERSONAL, Enrique; LARIOS, Diego Francisco; DELFINO, Federico; LEÓN, Carlos. Short-term power forecasting framework for microgrids using combined baseline and regression models. **Applied Sciences**, MDPI, v. 11, n. 14, p. 6420, 2021. 33

PEREIRA, Enio Bueno; MARTINS, Fernando Ramos; GONÇALVES, André Rodrigues; COSTA, Rodrigo Santos; LIMA, Francisco J. Lopes de; RÜTHER, Ricardo; ABREU, Samuel Luna de; TIEPOLO, Gerson Máximo; PEREIRA, Silvia Vitorino; SOUZA, Jefferson Gonçalves de. **Atlas Brasileiro de Energia Solar**. 2. ed. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. Disponível em: <http://doi.org/10.34024/978851700089>. ISBN 978-85-17-00089-8. Disponível em: https://labren.ccst.inpe.br/atlas_2017.html. 24

PEREZ, Richard; LORENZ, Elke; PELLAND, Sophie; BEAUHARNOIS, Mark; KNOWE, Glenn Van; JR, Karl Hemker; HEINEMANN, Detlev; REMUND, Jan;

MÜLLER, Stefan C; TRAUNMÜLLER, Wolfgang *et al.* Comparison of numerical weather prediction solar irradiance forecasts in the us, canada and europe. **Solar Energy**, Elsevier, v. 94, p. 305–326, 2013. 54

QIN, Shujing; LIU, Zhihe; QIU, Rangjian; LUO, Yufeng; WU, Jingwei; ZHANG, Baozhong; WU, Lifeng; AGATHOKLEOUS, Evgenios. Short-term global solar radiation forecasting based on an improved method for sunshine duration prediction and public weather forecasts. **Applied Energy**, Elsevier, v. 343, p. 121205, 2023. 54

QUINELATO, D. D. D.; BATISTA, E. A.; SANTOS, R. R. d.; ALLATTA, E.; BRITO, M. A. G. d. Projeto de um estimador de perdas por sujidade em módulos fotovoltaicos utilizando modelo preditivo. **Eletrônica de Potência**, v. 26, n. 1, p. 84–93, 2021. Disponível em: <https://doi.org/10.18618/rep.2021.1.0046>. 29

RENO Solar. **Curva I-V do módulo fotovoltaico RENO C 555**. jun 2024. Figura extraída do datasheet técnico: Módulo Fotovoltaico Monocristalino RENO C 555. Disponível em: RENO C 555 - 06.2024, p. [inserir número da página se houver]. 28

ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958. 34

RUVIARO, Raiana Spat; DALTROZO, Jenifer Godoy; GARLET, Liège; LOURENÇO, Willian Magalhães de; SANTOS, Ísis Portolan dos. Análise da variação da eficiência do módulo fotovoltaico em função da temperatura. *In: VII Congresso Brasileiro de Energia Solar-CBENS 2018*. [S.l.: s.n.], 2018. 21, 26

SALAZAR abril M.; MACABEBE, Erees Queen B. Detecção de hotspots em módulos fotovoltaicos utilizando termografia infravermelha. *In: Web de conferências MATEC*. [S.l.]: EDP Ciências, 2016. p. 10015. 27

SARP, Salih; KUZLU, Murat; CALI, Umit; ELMA, Onur; GULER, Ozgur. An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool. *In: IEEE. 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. [S.l.], 2021. p. 1–5. 21

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. [S.l.]: Cortez editora, 2017. 42

SHATNAWI, Amjed; ALKASSAR, Hana Mahmood; AL-ABDALY, Nadia Moneem; AL-HAMDANY, Emadaldeen A; BERNARDO, Luís Filipe Almeida; IMRAN, Hamza. Shear strength prediction of slender steel fiber reinforced concrete beams using a gradient boosting regression tree method. **Buildings**, MDPI, v. 12, n. 5, p. 550, 2022. 40

SILVA FILHO, João Inácio da. Introdução às células neurais artificiais paraconsistentes. **Seleção Documental: Inteligência Artificial e novas Tecnologias**, Universidade Santa Cecília, n. 8, p. 5–14, 2007. 33

SILVA, Ivan Nunes Da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. Redes neurais artificiais para engenharia e ciências aplicadas-curso prático. **São Paulo: Artliber**, 2010. 33

SOARES, Giovane Gai; RECH, Cassiano. Metodologia de dimensionamento de um gerador fotovoltaico para sistemas autônomos. **UNIJUI, Departamento de Tecnologia–DETEC, Rio Grande do Sul**, 2005. 27

TAVEIRA, Gabriela de Lima. Energia renovável e tecnologia na américa latina: desafios para a implementação da ods 7-energia limpa e acessível. Universidade Federal de Uberlândia, 2023. 24

TIEGHI, Camila Piacitelli; CANEPPELE, Fernando de Lima; PAI, Alexandre Dal; GODINHO, Emmanuel Zullo; ALMEIDA, Carlos Frederico Meschini; MALAGUETA, Diego Cunha; FRIGO, Murilo Miceno. Aplicações e desafios da inteligência artificial na previsão da radiação solar: uma revisão sistemática. **Revista Brasileira de Climatologia**, v. 36, n. 21, p. 170–201, fev. 2025. Disponível em: <https://ojs.ufgd.edu.br/rbclima/article/view/18872>. 21

TORRES, Douglas Guedes Batista; LUCA, Bruno Hermes de; MALACARNE, João André; SOUZA, Joel; CASAROLLI, Deivid; SILVA, Eduardo Dalla Costa; GUERRA, Thiago. Células fotovoltaicas: desenvolvimento e as três gerações. **Revista Técnico-Científica**, n. 17, 2019. 25

UDIT. Activation functions in deep learning: Understanding the role of activation functions in neural networks. **Medium**, 2023. Disponível em: <https://itsudit.medium.com/activation-functions-in-deep-learning-understanding-the-role-of-activation-functions-in-neural-423694f7f54e>. 36

WANG, Hui; YAN, Su; JU, Danyang; MA, Nan; FANG, Jun; WANG, Song; LI, Haijun; ZHANG, Tianyu; XIE, Yipeng; WANG, Jun. Short-term photovoltaic power

forecasting based on a feature rise-dimensional two-layer ensemble learning model. **Sustainability**, MDPI, v. 15, n. 21, p. 15594, 2023. 43

WANG, Kejun; QI, Xiaoxia; LIU, Hongda. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. **Applied Energy**, Elsevier, v. 251, p. 113315, 2019. 26

YANG, Guang; LEE, HwaMin; LEE, Giyeol. A hybrid deep learning model to forecast particulate matter concentration levels in seoul, south korea. **Atmosphere**, MDPI, v. 11, n. 4, p. 348, 2020. 53

YU, Yong; SI, Xiaosheng; HU, Changhua; ZHANG, Jianxun. A review of recurrent neural networks: LSTM cells and network architectures. **Neural computation**, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 31, n. 7, p. 1235–1270, 2019. 54

ZHENG, Alice; CASARI, Amanda. **Feature engineering for machine learning: principles and techniques for data scientists**. [S.l.]: "O'Reilly Media, Inc.", 2018. 42

ZILLES, Roberto. Energia solar fotovoltaica. **USP, São Paulo**, 2012. 25