



University of
Texas Libraries



e-revist@s



Centro Universitário Santo Agostinho

revistafsa

www4.fsnet.com.br/revista

Rev. FSA, Teresina, v. 22, n. 3, art. 9, p. 170-182, mar. 2025

ISSN Impresso: 1806-6356 ISSN Eletrônico: 2317-2983

<http://dx.doi.org/10.12819/2025.22.3.9>

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

WZB
Wissenschaftszentrum Berlin
für Sozialforschung



Previsão da Potência Elétrica de um Sistema Fotovoltaico: Uma Análise Comparativa entre Modelos de Aprendizado de Máquina

Prediction of Electric Power of a Photovoltaic System: A Comparative Analysis Between Machine Learning Models

Aldino Normelio Brun Polo

Engenheiro Químico pela Universidade Estadual do Oeste do Paraná
Mestrado no Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA)
E-mail: aldinopolo@alunos.utfpr.edu.br

José Airton Azevedo dos Santos

Doutor em Engenharia Elétrica, Universidade Federal de Santa Catarina
Professor da Universidade Tecnológica Federal do Paraná
E-mail: airton@utfpr.edu.br

Endereço: Aldino Normelio Brun Polo

Universidade Tecnológica Federal do Paraná – Campus
Medianeira - Avenida Brasil, 4232 CEP 85884-000 -
Caixa Postal 271 - Medianeira - PR – Brasil

Endereço: José Airton Azevedo dos Santos

Universidade Tecnológica Federal do Paraná – Campus
Medianeira - Avenida Brasil, 4232 CEP 85884-000 -
Caixa Postal 271 - Medianeira - PR – Brasil

Editor-Chefe: Dr. Tonny Kerley de Alencar Rodrigues

Artigo recebido em 29/01/2025. Última versão recebida em 13/02/2025. Aprovado em 14/02/2025.

Avaliado pelo sistema Triple Review: a) Desk Review pelo Editor-Chefe; e b) Double Blind Review (avaliação cega por dois avaliadores da área).

Revisão: Gramatical, Normativa e de Formatação



RESUMO

Este trabalho tem como objetivo comparar, na previsão da potência elétrica gerada por um sistema fotovoltaico, os desempenhos de previsão dos algoritmos XGBoost (*Xtreme Gradient Boosting Regressor*) e LighGBR (*Light Gradient Boosting Regressor*). Uma base de dados obtida de um sistema fotovoltaico localizado na região oeste paranaense foi utilizada para realizar esta comparação. A base de dados, com 21604 observações, apresenta as seguintes variáveis: potência elétrica, temperatura do sistema fotovoltaico e irradiação solar. Modelos de previsão XGBoost e LighGBR, otimizados por meio da biblioteca de ajuste de hiperparâmetros Optuna, foram implementados na linguagem de programação Python. Resultados de previsão foram comparados por meio das métricas: R^2 (*Determination coefficient*), MAE (*Mean Absolute Percent Error*) e RMSE (*Root Mean Squared Error*). Verificou-se, para resultados de previsão de um, dois e quatro dias, que o modelo XGBoost apresentou, com relação ao modelo LighGBR, um melhor desempenho.

Palavras-chave: Aprendizado de máquina. Modelos. Optuna. Python.

ABSTRACT

This work aims to compare the performance of the XGBoost (*Xtreme Gradient Boosting Regressor*) and LighGBR (*Light Gradient Boosting Regressor*) algorithms in predicting the electrical power generated by a photovoltaic system. A database obtained from a photovoltaic system located in the western region of Paraná was used to perform this comparison. The database presents 21,604 observations and the following variables: electrical power, photovoltaic system temperature, and solar irradiation. XGBoost and LighGBR forecasting models, optimized through the Optuna hyperparameter tuning library, were implemented in Python. Forecasting results were compared using the following metrics: R^2 (*Determination coefficient*), MAE (*Mean Absolute Percent Error*), and RMSE (*Root Mean Squared Error*). It was found that, for one-, two- and four-day forecasting results, the XGBoost model presented better performance than the LighGBR model.

Keywords: Machine learning. Models. Optuna. Python.

1 INTRODUÇÃO

A energia solar fotovoltaica, energia limpa, renovável e sustentável, tem tido, nos últimos anos, um grande crescimento no Brasil devido à instalação de painéis fotovoltaicos em residências, empresas e fazendas solares de grande porte. Sua geração alcançou 50,6TWh em 2023, apresentando um crescimento de 68,1%. Sua capacidade instalada obteve 37843MW, expandindo 54,8% em relação ao ano anterior (EPE, 2024). Contudo, as condições meteorológicas incertas, intermitentes e variáveis afetam diretamente a integração da produção de energia fotovoltaica à rede de energia elétrica.

A capacidade de prever com precisão a energia produzida por sistemas fotovoltaicos é um dos principais desafios para integração fotovoltaica em massa (AL-ALI *et al.*, 2023). Nesse contexto, vários trabalhos apresentados na literatura têm utilizado técnicas de *machine learning* e *deep learning* para previsão da energia gerada por sistemas fotovoltaicos. Dentre esses trabalhos, podem-se citar: Zhou *et al.* (2023) que propuseram, para previsão de potência de saída de um sistema fotovoltaico, a utilização de um modelo híbrido RFR (*Random Forest Regressor*) – XGBoost (*Xtreme Gradient Boosting Regressor*). Resultados experimentais demonstraram que o modelo RFR-XGBoost obteve um desempenho de previsão superior aos modelos RFR e XGBoost. Al-Ali *et al.* (2023) utilizaram, para previsão de produção de energia solar, uma combinação das redes CNN-LSTM (*Convolutional Neural network - Long Short-Term Memory*). Seus resultados mostraram que o modelo proposto pode ser utilizado como uma técnica de previsão confiável, que facilita a integração de energia solar a redes elétricas. Garip *et al.* (2023) utilizaram redes neurais LSTM, BiLSTM e *stacked LSTM* para previsão de energia fotovoltaica. Concluíram que a rede *stacked LSTM* apresentou melhor desempenho de previsão. No entanto, relataram que mais pesquisas devem ser realizadas para aprimorar o modelo de previsão. Xu *et al.* (2024) propuseram a utilização de modelos GRU (*Gated Recurrent Unit*) e XGBoost para melhorar o desempenho da previsão horária de energia solar de longo prazo. Relataram que o método proposto demonstra desempenho superior quando comparado a modelos de previsão de séries de longo prazo de última geração. Já Agga *et al.* (2022) propuseram a utilização da rede CNN-LSTM para previsão da energia solar fotovoltaica de curto prazo. Concluíram que o modelo CNN-LSTM, em comparação com os modelos CNN, LSTM e MLP (*Multilayer Perception*), obteve o melhor desempenho de previsão.

Nesse contexto, este trabalho tem como objetivo comparar, na previsão da potência elétrica gerada por um sistema fotovoltaico, o desempenho de previsão dos algoritmos

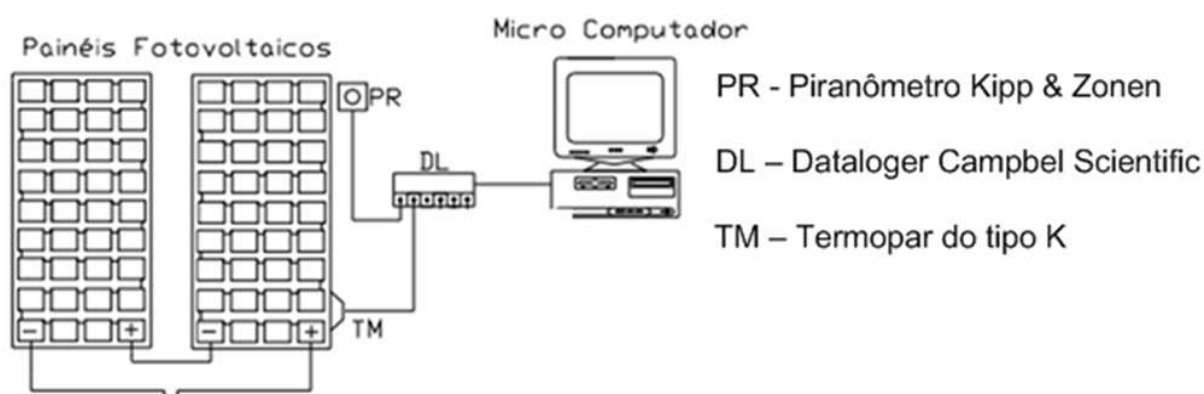
XGBoost (*Xtreme Gradient Boosting Regressor*) e LighGBR (*Light Gradient Boosting Regressor*).

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Sistema Fotovoltaico

Os componentes básicos do sistema fotovoltaico, localizado no Campus Medianeira da Universidade Tecnológica Federal do Paraná (UTFPR), são apresentados na Figura 1, indicando o conjunto fotovoltaico e o sistema de controle.

Figura 1 – Sistema fotovoltaico



Fonte: Oliveira *et al.* (2023).

2.2 Algoritmos

Xtreme Gradient Boosting Regressor (XGBoost): O XGBoost, proposto por Chen & Guestrin em 2016, é um algoritmo de conjunto baseado em árvore de decisão com aumento do gradiente (*gradient boosting*). As árvores, no *boosting*, são construídas na sequência, de modo que a árvore subsequente aprenda e reduza os erros da anterior. Os erros são minimizados através de um algoritmo de gradiente descendente. Nesse algoritmo, o tempo de treinamento é melhorado por meio de processamento paralelo (SOLANO *et al.*, 2022).

Light Gradient Boosting Machine (LighGBM): O algoritmo LighGBM, proposto pela Microsoft MSRA em 2016, é uma técnica de *machine learning* utilizada tanto para regressão (*Light Gradient Boosting Regressor* (LighGBR)) quanto para classificação. O LighGBR, como o algoritmo XGBoost, também é um algoritmo baseado em árvores de decisão. Esse algoritmo aplica técnicas de *gradient boosting* para construir um conjunto de árvores de decisão, com foco específico em eficiência e velocidade. Utiliza, para otimizar o processo de

treinamento e fornecer modelos de previsão de alto desempenho, a estratégia de crescimento de árvore (*leafwise*). Esse algoritmo tem ganhado, devido à sua velocidade e capacidade de lidar com grandes conjuntos de dados, muito destaque na área de *machine learning* (SRIDHAR, 2024; TANG *et al.*, 2020).

2.3 Optuna

Hiperparâmetros são variáveis que controlam o processo de aprendizagem de um modelo de *machine learning*. Essas variáveis podem ser otimizadas por meio de algoritmos de busca. O Optuna é uma biblioteca de otimização de hiperparâmetros do Python. Ele acelera, por meio da poda de testes pouco promissores, o tempo de otimização dos hiperparâmetros. O Optuna pesquisa os melhores hiperparâmetros dos modelos de previsão por meio de algoritmos de otimização bayesiana (DENG *et al.*, 2024; LIM, 2022; AGGARWAL, 2020).

2.4 Validação Cruzada

Para divisão do conjunto de dados em subconjuntos de treinamento e validação, utilizou-se a técnica de validação cruzada. Essa técnica divide os dados em k partições mutuamente excludentes, de modo que cada partição é utilizada somente uma vez para validação, enquanto as demais são utilizadas para treinamento. Esse processo é realizado k vezes, alternando de forma circular a partição de validação (JUNIOR *et al.*, 2019).

Após vários experimentos, $k=10$ foi escolhido para este estudo, o que significa que em cada rodada uma partição é selecionada como conjunto de validação e as nove restantes como conjunto de treinamento. Esse processo se repete dez vezes, cada vez escolhendo um subconjunto diferente como conjunto de validação.

3 MATERIAIS E MÉTODOS

Nesta seção serão abordados os procedimentos adotados para realização do trabalho.

3.1 Base de Dados

A base de dados, obtida por meio do sistema de aquisição de dados (micrologger Campbell CD23X), é formada pelas variáveis: Tempo (Time), Irradiação Solar Global na

mesma inclinação do painel (Irrad), Temperatura de Operação do painel (Temp) e Potência Elétrica (Pot). Cada uma das quatro variáveis apresenta 21604 observações. Os seis primeiros registros do conjunto de dados são apresentados na Figura 2.

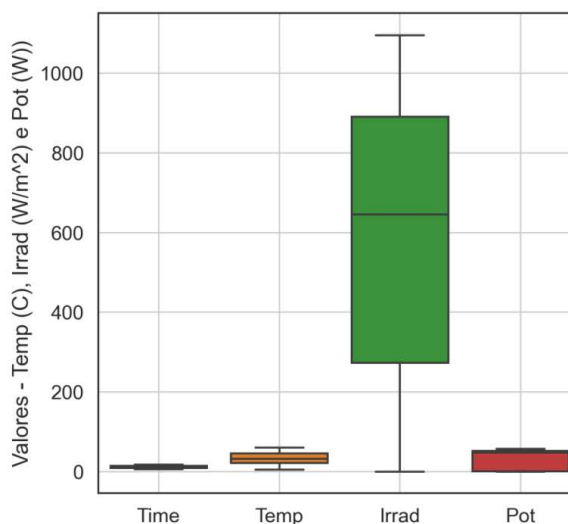
Figura 2 – Seis primeiros registros da base de dados

	Time	Temp	Irrad	Pot
0	6.0	5.767	0.000	0.0
1	6.0	5.999	0.000	0.0
2	6.0	7.080	1.266	0.0
3	6.0	7.530	0.000	0.0
4	6.0	8.170	0.000	0.0
5	6.0	10.260	0.000	0.0

Fonte: Os autores.

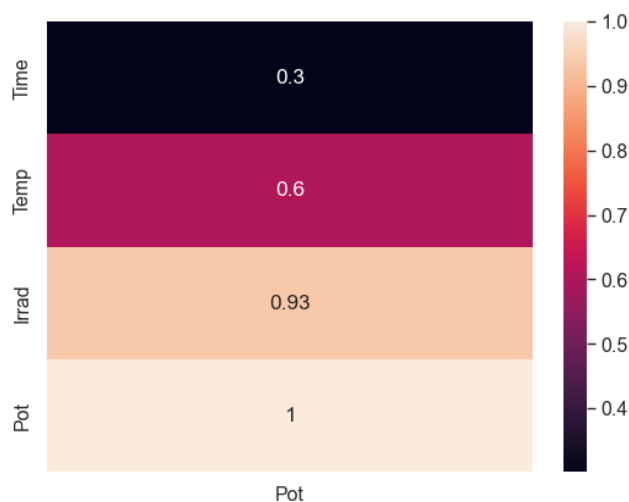
Os *Boxplots* dos dados das quatro variáveis são apresentados na Figura 3.

Figura 3 – Boxplots das variáveis: Time, Temp, Irrad e Pot



Fonte: Os autores.

Na Figura 4 apresentam-se a correlação das variáveis Time, Temp, Irrad com a variável Pot. Observa-se, dos coeficientes de correlação apresentados na figura, que a variável Irradiação Solar (Irrad) tem maior impacto sobre a potência elétrica gerada pelo sistema fotovoltaico (0,93).

Figura 4 – Correlações das variáveis Irrad, Temp e Time com a variável Pot

Fonte: Os autores.

3.2 Métricas

Neste estudo, para avaliação do desempenho dos modelos XGBoost e LighGBR, utilizaram-se as métricas R^2 , MAE e RMSE (Tabela 1). A raiz do erro quadrático médio (RMSE) é uma forma padrão de medir o erro de um modelo na previsão de dados quantitativos. O erro absoluto médio (MAE) é uma métrica que fornece a média da diferença absoluta entre a previsão do modelo e o valor observado. O R^2 é uma métrica que representa a eficácia do ajuste de um modelo de regressão. Essas métricas são muito utilizadas em avaliação de desempenho de modelos de previsão (MAGHRAOUI *et al.*, 2023; LU *et al.*, 2020).

Tabela 1 – Métricas R^2 , MAE e RMSE

Métrica	Equação	Melhor Resultado
R^2	$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$	1
MAE	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	0
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	0

Fonte: Os autores.

Onde: y_i é o valor real do período i , \hat{y}_i é a previsão para o período i e n é o número de observações.

3.3 Etapas do Trabalho

A estrutura proposta para fazer a previsão da potência elétrica do sistema fotovoltaico foi dividida em três etapas. Na primeira etapa, foi realizada a coleta e a limpeza dos dados obtidos do sistema fotovoltaico. Aplicou-se também, para evitar viés em direção a valores extremos, normalização aos dados. Nessa etapa os dados dos conjuntos de teste foram separados da base de dados. Na segunda etapa, treinamento e validação, os dados foram divididos por meio da técnica de validação cruzada nos conjuntos de dados de treinamento e validação. Nessa etapa, utilizou-se o algoritmo de busca Optuna para encontrar os melhores hiperparâmetros dos modelos XGBoost e LighGBR. Na etapa de teste, terceira etapa, os resultados de previsão para os conjuntos de teste dos modelos XGBoost e LighGBR são avaliados por meio das métricas R^2 , MAE e RMSE.

4 RESULTADOS E DISCUSSÕES

Inicialmente, neste estudo, realizou-se uma análise descritiva dos dados obtidos do sistema fotovoltaico (Figura 5). Para os dados de potência elétrica (Pot), gerada pelo sistema fotovoltaico, foram observados uma potência média de 32,926W e desvio padrão de 22,920W. Observou-se, por meio desses valores, um alto coeficiente de variação (69,6%).

Figura 5 – Análise descritiva dos dados treinamento e validação

	Time	Temp	Irrad	Pot
count	18720.000000	18720.000000	18720.000000	18720.000000
mean	12.005921	33.207788	576.512329	32.926504
std	3.465139	13.269075	342.951542	22.920676
min	6.000000	5.416000	0.000000	0.000000
25%	9.000000	22.310000	273.275000	1.048537
50%	12.000000	32.535000	645.150000	47.823130
75%	15.000000	45.890000	891.000000	51.547905
max	18.000000	60.060000	1095.000000	57.680700

Fonte: Os autores.

4.1 Seleção de Hiperparâmetros

Escolheram-se, para realizar os ajustes, quatro hiperparâmetros do XGBoost e cinco do LighGBR, deixando os outros hiperparâmetros dos modelos com valores padrão. A descrição dos hiperparâmetros é apresentada na Tabela 2.

Tabela 2 – Descrição dos Parâmetros

Parâmetro	Descrição
<i>max_depth</i>	Profundidade máxima da árvore.
<i>min_child_weight</i>	Soma mínima do peso da instância.
<i>learning_rate</i>	Taxa de aprendizagem.
<i>num_leaves</i>	Máximo de folhas por árvore.
<i>n_estimators</i>	Número de árvores.
<i>colsample_bytree</i>	Proporção de subamostragem de recursos.
<i>min_child_weight</i>	Soma mínima do peso da instância.

Fonte: Os autores.

Os intervalos de buscas e os melhores valores dos hiperparâmetros encontrados por meio do algoritmo de busca Optuna são apresentados na Tabela 3.

Tabela 3 – Parâmetros dos modelos

Algoritmo	Hiperparâmetro	Intervalo de Busca	Valores Ótimos
XGBoost	<i>learning_rate</i>	[0.0001,1]	0,826
	<i>max_depth</i>	[1,10]	5
	<i>min_child_weight</i>	[1,8]	4
	<i>colsample_bytree</i>	[0.001,0.1]	0,805
LighGBR	<i>learning_rate</i>	[0.0001,0.1]	0,089
	<i>num_leaves</i>	[25,50]	37
	<i>n_estimators</i>	[50,300]	229
	<i>min_child_weight</i>	[0.001,0.1]	0,48
	<i>colsample_bytree</i>	[0.001,0.1]	0,935

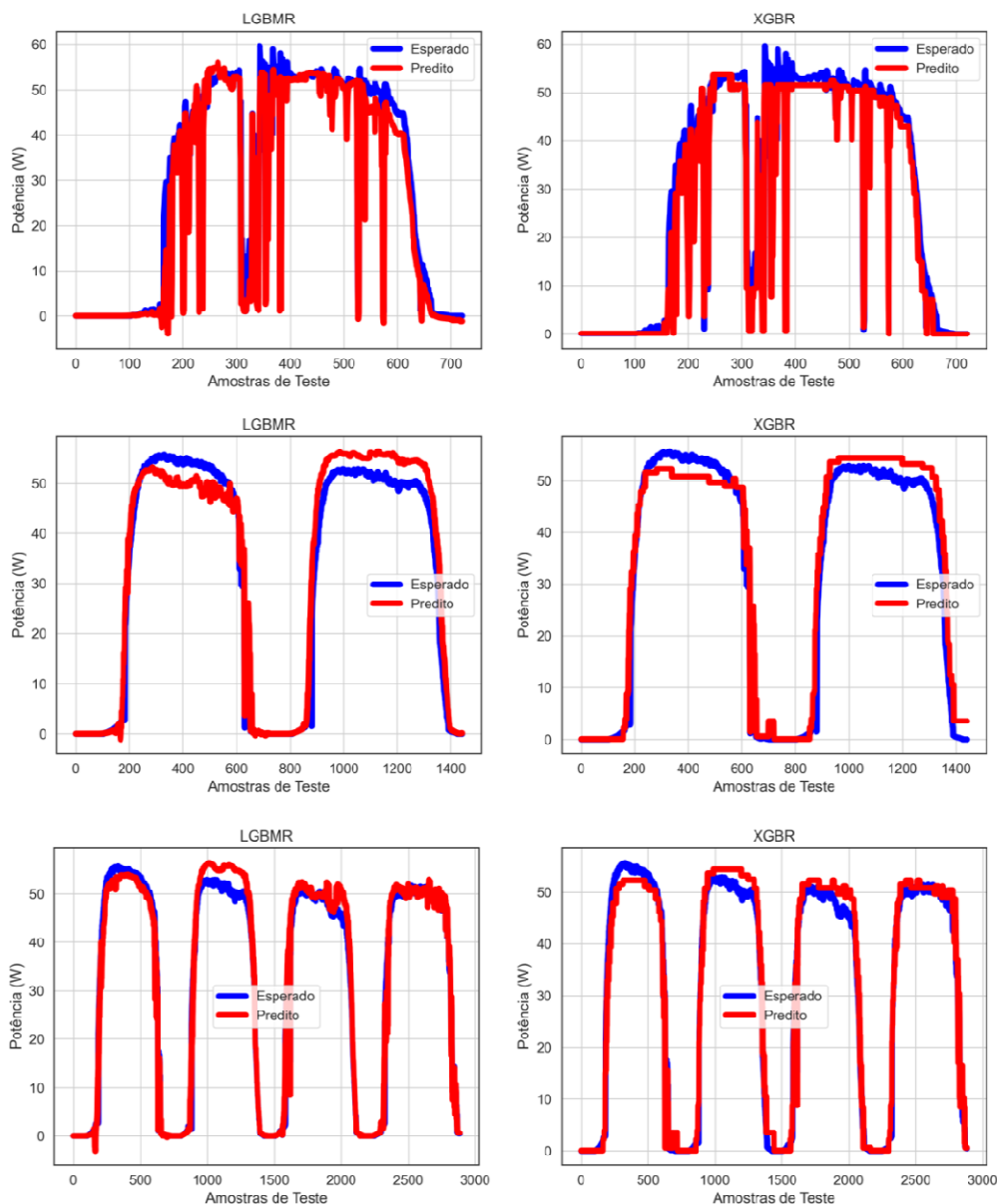
Fonte: Os autores.

4.2 Teste

Comparam-se, para testar o desempenho dos modelos XGBoost e LighGBR, os resultados obtidos nos modelos de previsão aos resultados de potência elétrica observados em campo. Para realizar essa comparação, utilizaram-se três conjuntos de teste: 1 dia (721 dados), 2 dias (1442 dados) e 4 dias (2884 dados). Observa-se que esses conjuntos de dados não

participaram da etapa de seleção dos hiperparâmetros. Na Figura 6, apresentam-se, em termos gráficos, os resultados das previsões dos valores observados e preditos da potência elétrica gerada pelo sistema fotovoltaico. Pode-se notar, nos resultados apresentados na figura, um bom ajuste entre os valores esperados e preditos pelos modelos.

Figura 6 – Curvas – Conjuntos de Teste (1 dia, 2 dias e 4 dias)



Fonte: Os autores.

Na Tabela 4 apresentam-se, para os três conjuntos de testes, os resultados das métricas de desempenho. Observa-se que as métricas foram computadas exclusivamente para o período diurno, abrangendo das 6h às 18h.

Tabela 4 – Descrição dos Parâmetros

Métricas - 1 dia			
Modelo	R²	MAE (W)	RMSE (W)
XGBoost	0,981	2,035	3,144
LighGBR	0,969	2,461	4,049
Métricas - 2 dias			
Modelo	R²	MAE (W)	RMSE (W)
XGBoost	0,969	2,872	4,146
LighGBR	0,963	3,219	4,537
Métricas - 4 dias			
Modelo	R²	MAE (W)	RMSE (W)
XGBoost	0,987	1,83	2,523
LighGBR	0,985	1,971	2,772

Fonte: Os autores.

Pode-se observar, dos resultados apresentados na Tabela 4, uma leve vantagem do modelo XGBoost sobre o Modelo LighGBR, com R² mais próximo de 1 e menores valores de MAE e RMSE.

5 CONSIDERAÇÕES FINAIS

Neste trabalho abordou-se a comparação de desempenho na previsão da potência elétrica gerada por um sistema fotovoltaico dos algoritmos XGBoost (*Xtreme Gradient Boosting Regressor*) e LighGBR (*Light Gradient Boosting Regressor*). Os modelos de previsão implementados por meio desses algoritmos passaram pelas fases de: preparação de dados, definição das estruturas, treinamento, validação e testes.

Inicialmente, na etapa de treinamento e validação, otimizaram-se, por meio do Optuna, os hiperparâmetros dos modelos XGBoost e LighGBR. Na sequência testou-se, para valores que não participaram da etapa de treinamento e validação, o desempenho dos dois modelos de previsão. Observou-se que os indicadores de desempenho (R², RMSE e MAE) obtidos pelos dois modelos, apresentaram resultados muito similares. Contudo, o modelo XGBoost apresentou, com relação ao modelo LighGBR, um melhor desempenho.

Por fim, pode-se observar que ambos os modelos são válidos e podem ser utilizados na previsão de potência elétrica de sistemas fotovoltaicos.

REFERÊNCIAS

- AGGA, A.; ABBOU, A.; LABBADI, M.; HOUM, Y. E. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. **Electric Power Systems Research**, 2022.
- AL-ALI, E.; HAJJI, Y.; SAID, Y.; HLEILI, M.; ALANZI, A. M.; LAATAR, A. H.; ATRI, M. Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model. **Mathematics**, v. 11, 2023.
- CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. 2016. Disponível em: <<https://arxiv.org/pdf/1603.02754>>. Acesso em: 20 jan. 2024.
- DENG, Z.; ZHOU, D.; KANG, Z.; DONG, H. (2024). Deep learning-based dynamic forecasting method and application for ultra-deep fractured reservoir production. **Frontiers in Energy Research**, v.12, 2024.
- EPE (Empresa de Pesquisa Energética). (2024). **EPE BEN 2024 - Relatório Síntese - Ano base 2023**. Disponível em: <<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2024>>. Acesso em: 20 jan. 2024.
- GARIP, Z.; EKINI, E.; ALAN, A. Day-ahead solar photovoltaic energy forecasting based on weather data using LSTM networks: a comparative study for photovoltaic (PV) panels in Turkey. **Electrical Engineering**, v. 105, 2023.
- JUNIOR, D. S.; OLIVEIRA, J. F.; SIQUEIRA, H. **Previsão de Carga Baseada em Ensemble de Modelos Inteligentes**. 2019. Disponível em: <<https://sbic.org.br/wp-content/uploads/2019/12/CBIC2019-114.pdf>>. Acesso em: 10 fev. 2024.
- LIM, Y. **Otimização de hiperparâmetros de aprendizado de máquina de última geração com Optuna**. 2022. Disponível em: <<https://towardsdatascience.com/state-of-the-art-machine-learning-hyperparameter-optimization-with-optuna-a315d8564de1>>. Acesso em: 15 fev. 2024.
- LU, W.; LU, J.; SUN, A. A CNN-LSTM-based model to forecast stock prices. **Complexity**, 2020.
- MICHELS, R. N. **Avaliação de um sistema de bombeamento de água alimentado por painéis fotovoltaicos**. 2007. Disponível em: <<https://tede.unioeste.br/bitstream/tede/2643/1/Roger%20Nabeyama%20Michels.pdf>>. Acesso em: 15 mar. 2024.
- OLIVEIRA, L.; RUSTICK, A.; SANTOS, J. A. A. Estimativa de vazão de um sistema de bombeamento fotovoltaico usando redes neurais artificiais. **Retec**, v. 16, 2023.
- SRIDHAR, A.; KARHUNEN, M.; HONKAPURO, S.; RUIZ, F. **Forecast or Nowcast to Predict Electricity Prices? The Role of Open Data**. 2024. Disponível em: <<https://ieeexplore.ieee.org/document/10608865>>. Acesso em: 25 jun. 2024.
- SOLANO, E. S.; DEHGHANIAN, P.; AFFONSO, C. M. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. **Energies**, v. 15, 2022.

TANG, M.; ZHAO, Q.; LONG, W.; HUANG, B. An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. **Energies**, v. 13, 2020.

XU, Y.; ZHENG, S.; ZHU, Q.; WONG, K.; WANG, X.; LIN, Q. (2023). A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting. **Energies**, v. 11, 2023.

ZHOU, M.; LI, X.; ZHONG, Z. (2023). **Research on Photovoltaic Power Generation Power Prediction Based on RF-XGBoost Model**. Disponível em: <https://ieeexplore.ieee.org/document/10393154> >. Acesso em: 5 mar. 2024.

Como Referenciar este Artigo, conforme ABNT:

A. N. B. POLO, J. A. A. SANTOS. Previsão da Potência Elétrica de um Sistema Fotovoltaico: Uma Análise Comparativa entre Modelos de Aprendizado de Máquina. **Rev. FSA**, Teresina, v. 22, n. 3, art. 9, p. 170-182, mar. 2025.

Contribuição dos Autores	A. N. B. Polo	J. A. A. Santos
1) concepção e planejamento.	X	X
2) análise e interpretação dos dados.	X	X
3) elaboração do rascunho ou na revisão crítica do conteúdo.	X	X
4) participação na aprovação da versão final do manuscrito.	X	X