Revista Brasileira de Climatologia
Brazilian Journal of Climatology
OPEN ACCESS

PREVISÃO DE TEMPERATURA DE UM MÓDULO FOTOVOLTAICO: UM ESTUDO DE CASO UTILIZANDO DADOS CLIMÁTICOS DA REGIÃO OESTE PARANAENSE

Photovoltaic module temperature forecasting: a case study using climate data from the Western Paraná Region

Predicción de la temperatura de un módulo fotovoltaico: un estudio de caso utilizando datos climáticos de la Región Oeste Paranaense

José Airton Azevedo dos Santos
Universidade Tecnológica Federal do Paraná (UTFPR)
Medianeira, Paraná, Brasil
airton@utfpr.edu.br

Resumo: A temperatura de operação dos módulos fotovoltaicos é um fator crítico que impacta diretamente a eficiência da conversão de energia. Seu comportamento térmico é dinâmico e governado pelas condições meteorológicas locais, uma vez que a irradiância solar, a temperatura ambiente e a velocidade do vento determinam, em conjunto, o balanço térmico do sistema. Neste contexto, este trabalho analisa, utilizando dados climáticos locais, o desempenho dos algoritmos XGBoost (*eXtreme Gradient Boosting*) e ElasticNet (*Elastic Net Regression*) na previsão da temperatura de um módulo fotovoltaico. Esta análise de desempenho baseou-se em dados medidos, no período entre janeiro de 2024 e janeiro de 2025, por uma estação solarimétrica instalada no campus da Universidade Tecnológica Federal do Paraná (UTFPR). A base de dados, com 307308 observações, é formada por 6 variáveis (Tempo (*Time*), Irradiação Solar (RS), Velocidade do Vento (VV), Temperatura Ambiente (TA), Umidade Relativa (UA) e Temperatura do Módulo (TP). A biblioteca Optuna foi utilizada para os ajustes dos hiperparâmetros dos modelos. As métricas R^2, MAE e RMSE foram utilizadas para avaliar o desempenho dos dois modelos. Resultados experimentais demonstraram o melhor desempenho estatístico do modelo XGBoost ($p < 0,001$ - Teste de Wilcoxon (*two-sided*)). A integração com a técnica SHAP (*Shapley Additive exPlanations*)

revelou que a temperatura ambiente e a irradiância solar, na região Oeste do Paraná, são os principais condicionantes climáticos da temperatura operacional dos módulos fotovoltaicos.
Palavras-chave: Aprendizado de máquina. Climatologia aplicada. Shap. XGBoost.

Abstract: The operating temperature of photovoltaic modules is a critical factor that directly affects energy conversion efficiency. Their thermal behavior is dynamic and governed by local meteorological conditions, since solar irradiance, ambient temperature, and wind speed jointly determine the system's thermal balance. In this context, we analyzed the performance of the XGBoost (eXtreme Gradient Boosting) and ElasticNet (Elastic Net Regression) algorithms in predicting module temperature using local climatic data. This performance assessment was based on measurements collected between January 2024 and January 2025 by a solarimetric station installed on the campus of the Federal University of Technology – Paraná (UTFPR). The dataset, comprising 307,308 observations, included six variables: Time (Time), Solar Irradiance (RS), Wind Speed (VV), Ambient Temperature (TA), Relative Humidity (UA), and Module Temperature (TP). Hyperparameter tuning was conducted using the Optuna library. Model performance was evaluated using R^2, MAE, and RMSE metrics. Experimental results demonstrated the superior statistical performance of the XGBoost model ($p < 0.001$, Wilcoxon two-sided test). Integration with the SHAP (Shapley Additive exPlanations) technique revealed that ambient temperature and solar irradiance in Western Paraná are the primary climatic determinants of photovoltaic module operational temperature.
Keywords: Machine learning. Applied climatology. Shap. XGBoost.

Resumen: La temperatura de operación de los módulos fotovoltaicos constituye un factor crítico que impacta directamente en la eficiencia de conversión de energía. Su comportamiento térmico es dinámico y está gobernado por las condiciones meteorológicas locales, dado que la irradiancia solar, la temperatura ambiente y la velocidad del viento determinan conjuntamente el balance térmico del sistema. En este contexto, analizamos el desempeño de los algoritmos XGBoost (eXtreme Gradient Boosting) y ElasticNet (Elastic Net Regression) en la predicción de la temperatura de un módulo fotovoltaico utilizando datos climáticos locales. Esta evaluación de desempeño se basó en mediciones realizadas entre enero de 2024 y enero de 2025 por una estación solarimétrica instalada en el campus de la Universidad Tecnológica Federal de Paraná (UTFPR). El conjunto de datos, compuesto por 307.308 observaciones, incluyó seis variables: Tiempo (Time), Irradiancia Solar (RS), Velocidad del Viento (VV), Temperatura Ambiente (TA), Humedad Relativa (UA) y Temperatura del Módulo (TP). La biblioteca Optuna fue empleada para el ajuste de los hiperparámetros de los modelos. Las métricas R^2, MAE y RMSE se utilizaron para evaluar el desempeño de ambos modelos. Los resultados experimentales demostraron un mejor rendimiento estadístico del modelo XGBoost ($p < 0,001$; prueba de Wilcoxon bilateral). La integración con la técnica SHAP (Shapley Additive exPlanations) reveló que la temperatura ambiente y la irradiancia solar en la región Oeste de Paraná son los principales condicionantes climáticos de la temperatura operacional de los módulos fotovoltaicos.
Palabras clave: Aprendizaje automático. Climatología aplicada. Shap. XGBoost.

Submetido em: 16/02/2026
Aceito para publicação em: 25/08/2026
Publicado em: 15/11/2026

1. INTRODUÇÃO

A conversão da energia solar em energia elétrica, por meio de sistemas fotovoltaicos, destaca-se como uma das técnicas mais utilizadas para geração de energia a partir de fontes renováveis. Contudo, a eficácia desse processo de conversão é influenciada por diversos fatores ambientais e climáticos, como: radiação solar, temperatura, velocidade dos ventos, umidade relativa do ar e acúmulo de poeira. Além de questões operacionais e de manutenção, como degradação dos painéis, limpeza e técnicas de resfriamento. Todos esses elementos impactam direta ou indiretamente no desempenho e na potência dos módulos fotovoltaicos (IHADDADENE et al., 2022; KEDDOUDA et al., 2024).

A eficiência e a durabilidade dos painéis fotovoltaicos são diretamente impactadas pela temperatura em que operam. A eficiência de conversão de energia tende a cair à medida que a temperatura aumenta, o que eleva o risco de degradação acelerada dos materiais. Dessa forma, entender e prever com precisão a temperatura operacional é essencial para: maximizar a geração de energia (evitando perdas por superaquecimento), prolongar a vida útil dos módulos (minimizando danos térmicos) e otimizar o projeto de usinas solares (escolha de local, inclinação e ventilação) (BAMISILE et al., 2025; EBHOTA; TABAKOV, 2023).

Diversos estudos têm empregado modelos de aprendizado de máquina para a previsão da temperatura de módulos fotovoltaicos. Entre os principais trabalhos destacam-se: Coskun et al. (2017) que utilizaram redes neurais artificiais para estimar a temperatura da superfície de módulos fotovoltaicos. Os autores adotaram como variáveis de entrada a temperatura ambiente, a irradiância solar e a velocidade do vento e como variável de saída a temperatura do módulo fotovoltaico. Diferentes algoritmos de treinamento foram testados e comparados neste trabalho. Os resultados indicaram que o algoritmo de Levenberg-Marquardt (LM) apresentou o melhor desempenho. Keddouda et al. (2024) realizaram a previsão, utilizando 12 algoritmos de aprendizado de máquina, da temperatura de módulos fotovoltaicos. Como entrada dos modelos selecionaram as variáveis: temperatura ambiente, irradiação solar, velocidade do vento e umidade relativa. Concluíram, neste trabalho, que o algoritmo MLP (Multilayer Preceptron) apresentou o melhor desempenho. Este estudo reforça a capacidade das redes neurais em capturar padrões complexos entre variáveis como

irradiação solar, temperatura ambiente, umidade relativa e velocidade do vento. Li et al. (2024) realizaram, por meio de algoritmos de aprendizado de máquina, a previsão de temperatura de painéis fotovoltaicos instalados na terra e na água. Observaram que os modelos Random Forest e ResNet apresentaram, para ambos os ambientes, o melhor desempenho. Já Campos et al. (2020) utilizaram métodos de inteligência artificial e machine learning para estimar, a partir de variáveis exógenas como irradiância, temperatura ambiente e velocidade do vento, os valores de temperatura de módulos solares fotovoltaicos. Os dados foram obtidos de um conjunto de sistemas fotovoltaicos instalados em Brotas de Macaúbas (BA). Observaram que o modelo KNN (K-Nearest Neighbors), dentre os modelos testados, apresentou o melhor desempenho.

Além destes trabalhos, estudos recentes reforçam a importância da climatologia aplicada na compreensão dos padrões meteorológicos que afetam diretamente o desempenho dos sistemas solares (PEREIRA; MOURA; LUCENA, 2020). A relação entre clima e energia solar tem sido amplamente discutida em análises nacionais, como no Atlas Brasileiro de Energia Solar (PEREIRA; MARTINS; RUTHER, 2020), que evidencia o potencial solar e suas variações regionais. A variabilidade atmosférica e sua influência na eficiência energética também têm sido objeto de investigação, mostrando que oscilações de vento, temperatura e irradiância impactam significativamente a conversão fotovoltaica (TIEGUI et al., 2025). Por fim, os controles climáticos da irradiância solar foram detalhados por Rodríguez Gómez et al. (2018). Os autores, neste artigo, discutem os conceitos físicos e a variabilidade da irradiância como condicionantes fundamentais do desempenho dos módulos fotovoltaicos.

A escolha, neste trabalho, dos algoritmos ElasticNet e XGBoost, para estimar a temperatura de um módulo fotovoltaico, fundamenta-se na necessidade de avaliar, na modelagem térmica dos módulos, o compromisso entre a regularização linear e a capacidade de captura de padrões não lineares. O ElasticNet foi selecionado por sua robustez no tratamento da multicolinearidade inerente às variáveis meteorológicas. Ele atua como um baseline linear que penaliza coeficientes irrelevantes. Já o XGBoost foi adotado por representar o estado da arte em gradient boosting (ZHANG et al., 2022). Este algoritmo permite identificar se a complexidade adicional do modelo é capaz de mapear as interações termodinâmicas sutis que escapam às abordagens lineares clássicas.

Neste contexto, este trabalho analisa, utilizando dados climáticos locais, o desempenho dos algoritmos XGBoost (eXtreme Gradient Boosting) e ElasticNet (Elastic Net Regression) na previsão da temperatura de um módulo fotovoltaico. As principais contribuições deste estudo residem na avaliação comparativa, utilizando a técnica de validação cruzada TimeSeriesSplit, entre métodos de regularização linear (ElasticNet) e de gradient boosting (XGBoost) para a previsão da temperatura de um módulo fotovoltaico. A técnica TimeSeriesSplit respeita a cronologia dos dados e mitiga o vazamento de informações. Adicionalmente, o trabalho incorpora, com foco na importância global das variáveis, a análise de interpretabilidade via valores SHAP (SHapley Additive exPlanations). Essa abordagem permite, ao calcular a média dos valores absolutos das contribuições de cada variável, hierarquizar os preditores de acordo com seu impacto no modelo.

2. METODOLOGIA

2.1. Algoritmos

XGBoost (Extreme Gradient Boosting): O XGBoost, desenvolvido com foco em alto desempenho, é um algoritmo de machine learning baseado em árvores de decisão. Constrói, por meio da técnica de Gradient Boosting, modelos de forma sequencial. Cada novo modelo corrige os erros cometidos pelos anteriores. O XGBoost, muito utilizado em tarefas de regressão e classificação, se destaca por sua velocidade de treinamento, capacidade de generalização e por incorporar recursos de regularização (LI et al., 2022; LIEW et al., 2021; NIAZKAR et al., 2024; SOLANO et al., 2022).

ElasticNet (Elastic Net Regression): O algoritmo ElasticNet é uma técnica de regressão linear regularizada que combina linearmente, buscando superar as limitações individuais de cada método, as penalidades L1 (Lasso) e L2 (Ridge) na função de otimização. A componente Lasso promove a esparsidade do modelo ao zerar coeficientes de variáveis irrelevantes, realizando uma seleção automática de features. Já a componente Ridge, lidando eficazmente com a multicolinearidade, estabiliza a solução através da contração dos coeficientes. Essa abordagem híbrida torna o ElasticNet particularmente robusto em cenários onde existem variáveis preditoras altamente correlacionadas. Isto permite que o algoritmo selecione grupos de variáveis simultaneamente (efeito de agrupamento) em vez

de escolher arbitrariamente apenas uma, garantindo assim um modelo mais generalizável e estável. O algoritmo ElasticNet é particularmente vantajoso quando há necessidade de lidar simultaneamente com seleção automática de variáveis e com a colinearidade presente em datasets reais (AMINI; HU, 2021; CHAMLAL et al., 2024; MAHMOOD; KADIR, 2025).

2.2. Base de dados e Pré-processamento

O sistema fotovoltaico, localizado no campus Medianeira da UTFPR, é composto por 543 painéis fotovoltaicos da ZNSHINE SOLAR ZXM7-SHLDD144 Série de 555 Wp (Figura 1) e 5 Inversores da GROWATT. A estação solarimétrica, localizada nas proximidades do sistema fotovoltaico, possui os seguintes equipamentos: Piranômetro SR05 (Hukseflux), Sensor Pluviométrico SS-RG, Anemômetro FA013, Anemoscópio FA023, Sensor de Temperatura Ambiente e Sensor de Temperatura para módulo fotovoltaico SS-PT1000 (Medida de temperatura do módulo).

Figura 1 – Sistema fotovoltaico.



Fonte: Elaborado pelos autores (2025).

A área de estudo compreende o município de Medianeira, localizado na região Oeste do Estado do Paraná (latitude 25°17'43" S, longitude 54°05'38" W e altitude média de 415 m). A região, segundo a classificação climática de Köppen-Geiger, é caracterizada pelo tipo Cfa (clima subtropical úmido), apresentando verões quentes, geadas pouco frequentes e uma distribuição pluviométrica uniforme ao longo de todo o ano, sem a ocorrência de uma estação seca definida (LIMA et al., 2023).

Além dessas características gerais, o clima local sofre forte influência das massas de ar tropicais e polares, que modulam a variabilidade térmica e hídrica ao longo das estações. A sazonalidade térmica é marcada por verões de elevada temperatura e invernos amenos, com episódios ocasionais de frio intenso associados à penetração de massas polares. A atuação dos sistemas frontais é frequente, especialmente durante o outono e inverno, contribuindo para a ocorrência de chuvas frontais e quedas bruscas de temperatura. A variabilidade da nebulosidade também desempenha papel relevante, afetando diretamente os níveis de irradiância solar disponíveis para conversão fotovoltaica. Por fim, as características da radiação solar regional refletem a posição geográfica do Oeste Paranaense, com elevada disponibilidade anual de energia solar, mas sujeita a oscilações diárias e sazonais em função da cobertura de nuvens e da dinâmica atmosférica.

Essa caracterização é fundamental para a análise do comportamento térmico dos módulos fotovoltaicos, dado que a interação entre temperatura, umidade relativa, nebulosidade e radiação solar influencia diretamente as trocas térmicas entre o sistema e o meio ambiente.

2.3. Pré-processamento

Inicialmente, na etapa de pré-processamento, observou-se que a base de dados continha valores ausentes. Optou-se, para garantir a integridade dos dados, por remover todas as instâncias com valores ausentes. Na sequência, dividiu-se o conjunto de dados (307308 observações) nos Conjuntos de Treinamento (245846 observações – 80%) e Teste (61462 observações – 20%). Os dados do conjunto de treinamento foram normalizados por meio da função MinMaxScaler. Observa-se que a normalização, a fim de evitar vazamento de informações, ocorreu somente nos dados do conjunto de treinamento (BOBADILLA; SCHMITT, 2024; BOUKE; ABUDULLAH, 2023).

2.4. Método SHAP

O método SHAP (SHapley Additive exPlanations) é uma técnica de interpretabilidade, em modelos de machine learning, que utiliza conceitos da teoria dos jogos para quantificar a contribuição de cada variável nas predições do modelo. O SHAP, com relação à importância

global das variáveis, agrega, em todas as instâncias do conjunto de dados, os valores absolutos de contribuição (valores SHAP) de cada feature. Ele calcula, para determinar o impacto geral de cada variável no comportamento do modelo, a média desses valores absolutos. Isto permite identificar as características mais influentes de forma holística, independentemente da direção (positiva ou negativa) de seu efeito, o que facilita a compreensão da relevância relativa das variáveis em escala global (BOBADILLA-PONCE et al., 2024; DAI; YANG, 2025; HAMILTON; PAPADOPOULOS, 2024; SOMALA et al., 2024).

2.5. Métricas

Os modelos, implementados neste trabalho, foram avaliados pelos parâmetros: RMSE (Raiz Quadrada do Erro Quadrático Médio: mede o erro médio, mas penaliza mais os erros grandes), MAE (Erro Médio Absoluto: média dos erros absolutos entre os valores reais e os previstos) e R^2 (Coeficiente de Determinação: mede o quanto da variação da variável dependente é explicada pelo modelo). Tem-se uma previsão perfeita com RMSE=MAE=0 e R^2=1 (SANTOS; POLO, 2024; OLIVEIRA; SANTOS, 2023).

Tabela 1- Métricas R^2, MAE e RMSE.

R^2	MAE	RMSE
$R^2 = 1 - \frac{\sum_{i=1}^{n}((y_i - \hat{y})^2)}{\sum_{i=1}^{n}((y_i - \bar{y})^2)}$	$MAE = \frac{1}{n}\sum_{1}^{n}\lvert(y_i - \hat{y}_i)\rvert$	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((y_i - \hat{y}_i)^2)}$

Fonte: Elaborado pelos autores (2025).

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

2.6. O Teste de Wilcoxon (*two-sided*)

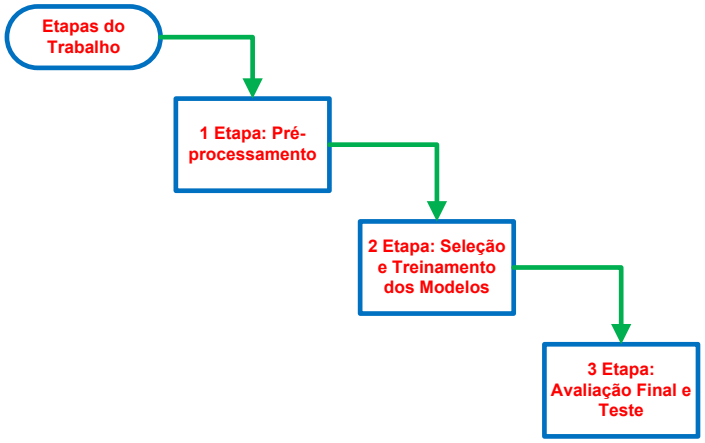
O teste de Wilcoxon (*two-sided*) é um teste estatístico não paramétrico usado para comparar duas amostras relacionadas ou pares de observações. Ele faz a avaliação, sem assumir normalidade dos dados, se as distribuições diferem de forma significativa. O teste, na versão bilateral (*two-sided*), verifica se há qualquer diferença entre os grupos, baseando-se nos postos (*ranks*) das diferenças em vez das médias. Este teste é uma alternativa,

quando os pressupostos paramétricos não são atendidos, ao teste *t* pareado (OKEH; ONYEAGU, 2020).

2.7. Etapas do trabalho

A estrutura proposta, para prever a temperatura do módulo fotovoltaico, foi dividida em três etapas (Figura 2).

Figura 2 – Etapas do trabalho.



Fonte: Elaborado pelos autores (2025).

1. **Pré-processamento e Preparação dos Dados**

- Obtenção da base de dados da estação solarimétrica da UTFPR;
- Tratamento de dados (limpeza, normalização e preparação);
- Divisão do conjunto de dados em Conjunto de Treinamento e Conjunto de Teste por meio da função train_test_split(X, y, test_size= 0.2, shuffle=False). A divisão dos dados reservou as últimas 61462 observações para teste;

2. **Seleção e Treinamento dos Modelos**

- Implementação dos algoritmos XGBoost e ElasticNet;
- Otimização de Hiperparâmetros com o Optuna;
- Avaliação de desempenho via validação cruzada (TimeSeriesSplit(*n_splits*=3)). Esta técnica cria 3 divisões (*splits*) de teste diferentes, usando uma abordagem de "Janela em Expansão" (*Expanding Window*). O conjunto de treino cresce a cada iteração, enquanto o teste se move para frente no tempo (HAYAT et al., 2025; TORRES, 2024).
- Seleção dos melhores modelos.

3. **Avaliação Final e Teste**

- Validação dos modelos no conjunto de teste (dados não vistos durante o treinamento);
- Análise comparativa do desempenho, dos modelos XGBoost e ElasticNet, na previsão de temperaturas de um módulo fotovoltaico.

3. RESULTADOS

3.1. Treinamento

A otimização de hiperparâmetros desempenha um papel muito importante no aprimoramento da capacidade de generalização de modelos de previsão. Neste contexto, o *framework* Optuna destaca-se como uma ferramenta avançada para automação desse processo, oferecendo flexibilidade e eficiência na busca pelos parâmetros ideais. O Optuna, fundamentado em técnicas de otimização bayesiana, emprega métodos probabilísticos para explorar de forma inteligente o espaço de hiperparâmetros. Ele garante, quando comparado a abordagens tradicionais, uma convergência mais rápida para soluções ótimas (DADA et al., 2025; PRAVIN et al., 2022; SRINIVAS; KATARYA, 2022). Os melhores modelos XGBoost e ElasticNet, encontrados pelo Optuna, apresentam os seguintes hiperparâmetros (Tabela 2).

Tabela 2- Parâmetros dos modelos

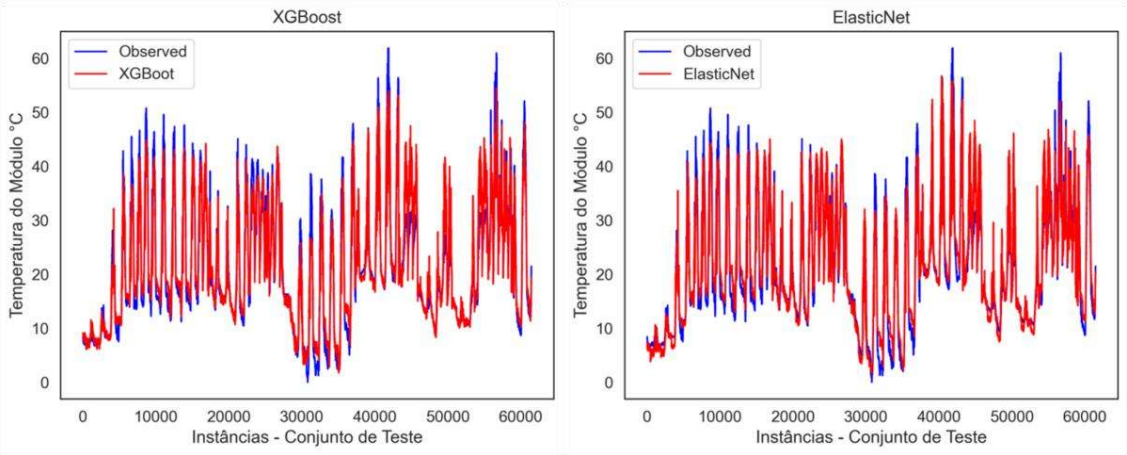
Algoritmo	Hiperparâmetro	Intervalo de Busca	Hiperp. Ótimo
XGBoost	*n_estimators*	[50,300]	233
	learning_rate	[0.00001,0.1]	0.019
	max_depth	[1,15]	5
	min_child_weight	[1,20]	8
	subsample	[0.0,1.0]	0.514
	colsample_bytree	[0.0,1.0]	0.867
	reg_alpha	[0.0,1.0]	0.421
	reg_lambda	[0.0,1.0]	0.905
ElasticNet	*alpha*	[0.00001,10]	0.00232
	l1_ratio	[0,1]	0,00018
	max_iter	5000	5000

Fonte: Elaborado pelos autores (2025).

3.2. Teste

Inicialmente, para testar o desempenho dos dois algoritmos, compararam-se, ao longo do Conjunto de Teste (61.462 observações), os valores da Temperatura do Módulo previsto e observado (Figura 3).

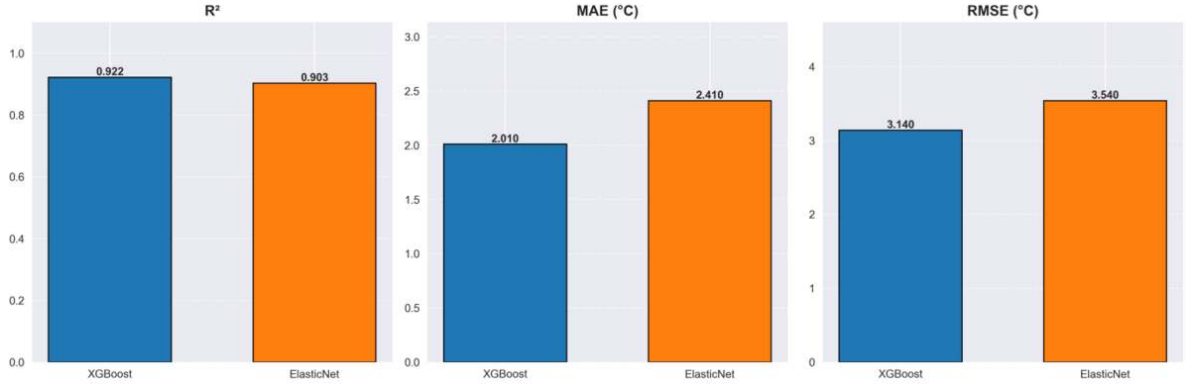
Figura 3 – Resultados de previsão – Conjunto de Teste.



Fonte: Elaborado pelos autores (2025).

Pode-se observar, por meio desta figura, que as temperaturas previstas, pelos modelos XGBoost e ElasticNet, têm um bom ajuste com os valores de temperaturas observados. Contudo, a análise visual não garante precisão na eficiência dos resultados. Portanto, optou-se, para maior confiabilidade, pelo cálculo das métricas R2, MAE e RMSE. Tais indicadores avaliam o ajuste dos modelos (Figura 4).

Figura 4 – Resultados das métricas R^2, MAE (°C) e RMSE (°C) – Conjunto de Teste**.**



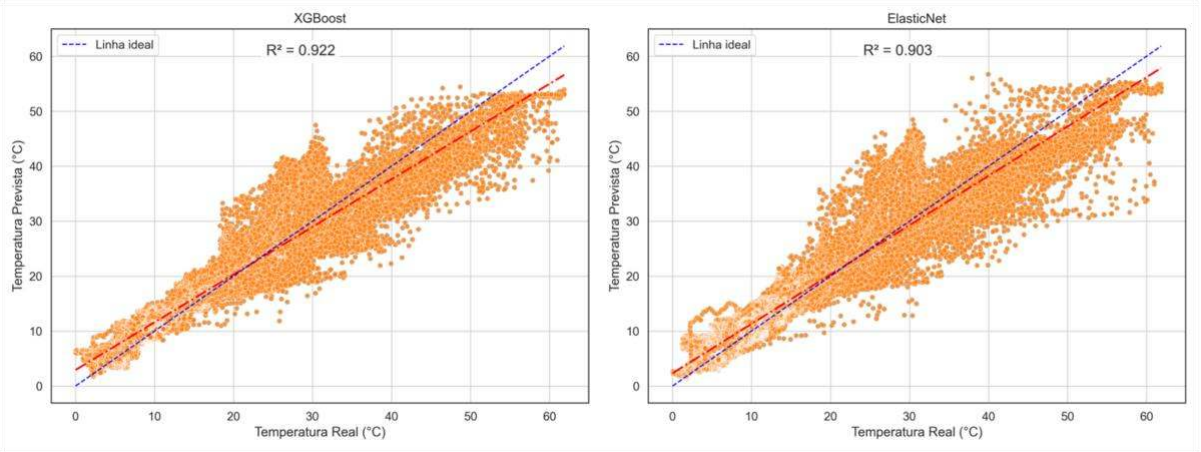
Fonte: Elaborado pelos autores (2025).

Observa-se, dos resultados apresentados na Figura 4, que o modelo XGBoost apresentou, com relação ao modelo ElasticNet, o melhor desempenho, com R^2 de 0,922, indicando que o modelo explica 92,2% da variância dos dados de temperatura do módulo. Além disso, obteve os menores valores de erro (MAE=2,01 °C e RMSE=3,14 °C). Isto demonstra sua muito boa precisão e capacidade de generalização. O melhor desempenho do XGBoost pode ser atribuído à sua habilidade de lidar com dados complexos e à aplicação de regularização, que ajuda a evitar o overfitting. Já o modelo ElasticNet obteve também um bom desempenho, mas inferior ao XGBoost (R^2=0,903, MAE=2,41 °C e RMSE=3,54 °C). O bom desempenho do modelo ElasticNet (R^2=0,903) se deve à forte correlação linear entre as variáveis de entrada e a temperatura do módulo.

Esses valores refletem não apenas a robustez estatística dos modelos, mas também a influência da elevada variabilidade da irradiância solar na região Oeste do Paraná. Essa variabilidade aumenta a complexidade da modelagem e explica por que o XGBoost, ao capturar relações não lineares, apresentou desempenho superior.

Na Figura 5 apresentam-se, para o Conjunto de Teste, os resultados de dispersão dos valores reais e preditos para os dois modelos.

Figura 5– Dispersão entre valores reais e previstos – Conjunto de Teste**.**



Fonte: Elaborado pelos autores (2025).

A análise comparativa, entre os gráficos de dispersão dos modelos XGBoost e ElasticNet, evidencia um melhor desempenho do XGBoost. Ele alcançou um coeficiente de determinação (R2) de 0,922, bem maior que o valor obtido pelo modelo ElasticNet (0,903). O gráfico de dispersão do XGBoost apresenta, ao longo de todo o eixo, uma nuvem de pontos compacta e aderente à linha ideal. Esta maior compactação, da nuvem de pontos no XGBoost, indica que o modelo conseguiu representar melhor os efeitos da sazonalidade climática típica da região, marcada por verões quentes e invernos amenos. Essa sazonalidade impacta diretamente os extremos de temperatura dos módulos, que foram previstos com maior precisão pelo modelo não-linear. Já o modelo ElasticNet exibe maior dispersão dos dados. Isso indica que o modelo ElasticNet tem mais dificuldade em prever valores extremos ou que há maior variabilidade nos erros de previsão.

3.3. Análise Estatística Comparativa

A análise estatística comparativa, realizada via teste de Wilcoxon (two-sided), confirmou o melhor desempenho do modelo XGBoost, evidenciando uma diferença estatisticamente significativa ($p < 0,001$) na precisão preditiva. Embora o tamanho do efeito de Rosenthal ($r=0,231$) seja considerado pequeno, a redução do erro médio absoluto de 2,41 °C para 2,01 °C representa um ganho prático relevante, equivalente a cerca de 16,6%. Esse avanço demonstra que o XGBoost, por ser um modelo não-linear, consegue capturar padrões mais complexos da série temporal. Isto resulta em estimativas de temperatura mais próximas da realidade operacional do sistema fotovoltaico e maior confiabilidade para aplicações que exigem previsões consistentes.

A superioridade estatística do XGBoost sobre o ElasticNet, observada neste estudo, alinha-se aos resultados de Keddouda et al. (2024). Esses autores demonstraram que algoritmos não-lineares, incluindo o Gradient Boosting, superaram, na predição de temperatura de módulos fotovoltaicos, o desempenho de modelos de regressão linear.

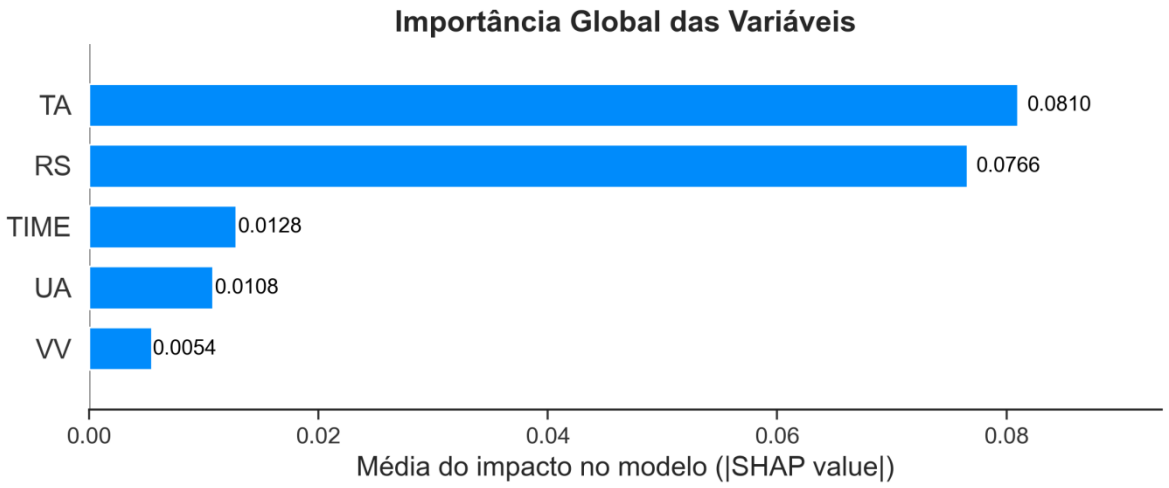
A superioridade estatística do XGBoost também se relaciona às condições subtropicais úmidas do Oeste Paranaense, caracterizadas por elevada umidade relativa e frequente nebulosidade. Esses fatores modulam a disponibilidade de radiação solar e, consequentemente, a temperatura operacional dos módulos, exigindo modelos capazes de lidar com padrões atmosféricos complexos.

3.4. SHAP

A análise de interpretabilidade via SHAP values (Figura 6) confirma que o modelo XGBoost capturou corretamente a fenomenologia física do sistema. A Temperatura Ambiente (TA), que estabelece o patamar térmico base para as trocas de calor, foi identificada como a variável de maior impacto preditivo. Esse resultado está em consonância com os mecanismos climáticos regionais: a temperatura do ar estabelece o patamar térmico de fundo para as trocas de calor entre o módulo e o ambiente. Em um clima subtropical úmido, como o do Oeste Paranaense, esse controle atmosférico é dominante, pois regula tanto o fluxo de calor sensível quanto a eficiência da dissipação térmica. A Temperatura Ambiente é seguida pela Irradiação Solar (RS), que atua como fonte primária de energia térmica. A irradiância solar atua como a principal fonte de aquecimento superficial, modulando diretamente o balanço energético do módulo. A variabilidade diária e sazonal da radiação, influenciada pela nebulosidade e pela posição solar, explica os picos de temperatura observados e reforça a importância dessa variável como determinante da carga térmica. Já as variáveis Tempo (TIME), Velocidade do Vento (VV) e Umidade Relativa (UA) assumiram papéis secundários.

Observa-se que a velocidade do vento na região, para este período de estudo, apresentou uma média de 2,74 ± 2,10 m/s, superior às condições padrão de teste (NOCT: 1 m/s). Isto teoricamente favoreceria a troca térmica por convecção. Contudo, a análise de importância de atributos (SHAP) indicou que esta variável teve impacto secundário na predição do modelo XGBoost. Este comportamento é explicado pela dominância dos fatores de carga térmica. Enquanto o vento atua como um termo de dissipação instável (alta volatilidade, CV de 76,6%), a temperatura ambiente e a irradiância solar atuam em conjunto para ditar o equilíbrio térmico macroscópico do módulo. O algoritmo, ao buscar a minimização global do erro, identificou que as flutuações do vento, apesar de fisicamente relevantes para o resfriamento instantâneo, possuem menor poder preditivo para a variância total da temperatura do que a carga térmica imposta pelo sol e a temperatura ambiente.

Figura 6 – Importância global das variáveis (SHAP)**.**



Fonte: Elaborado pelos autores (2025).

3.5. Limitações do Estudo

O estudo realizado possui as seguintes limitações:

Especificidade geográfica: O modelo foi validado para as condições de clima subtropical de transição (Cfa/Cfb). Portanto, a extrapolação direta destes resultados para regiões com dinâmicas atmosféricas distintas, como o semiárido nordestino (baixa umidade constante) ou a região amazônica (alta nebulosidade convectiva), carece de validação adicional. Visto que as variáveis preditoras podem apresentar pesos de importância (SHAP *values*) divergentes nesses cenários.

Variáveis relacionadas ao estado físico da superfície dos módulos: Por fim, o modelo não incorporou variáveis relacionadas ao estado físico da superfície dos módulos, especificamente a acumulação de poeira (*soiling*). O acúmulo de material particulado, comum em áreas agrícolas como a região do estudo, pode alterar a emissividade térmica e a absorção de radiação do painel. A ausência dessa variável implica que o modelo assume uma condição ideal de limpeza ou uma taxa de sujidade constante, o que pode introduzir resíduos no erro de previsão em períodos de estiagem prolongada.

Ausência de variáveis climatológicas complementares: O modelo não considerou variáveis de nebulosidade, cobertura de nuvens e radiação difusa, que são determinantes para o balanço de energia à superfície e modulam a disponibilidade de irradiância direta. Além disso, não foram avaliadas as diferenças de temperatura da superfície dos módulos em

diferentes orientações, o que poderia fornecer uma visão mais abrangente sobre os efeitos microclimáticos e geométricos na conversão fotovoltaica.

3.6. Diferencial Metodológico

Uma distinção metodológica importante deste trabalho, em relação à literatura similar, é o rigor na validação. Enquanto muitos estudos utilizam validação cruzada aleatória (K-Fold tradicional), que pode subestimar o erro ao misturar dados futuros e passados, o uso de TimeSeriesSplit neste estudo oferece uma estimativa de erro mais conservadora. Porém mais realista para a operação em tempo real. Portanto, o MAE de 2,01 °C aqui reportado representa uma métrica de operação real, e não apenas um ajuste teórico idealizado. Adicionalmente, para assegurar a integridade do modelo e evitar o vazamento de informações (data leakage), a normalização foi parametrizada exclusivamente a partir do conjunto de treinamento.

4. CONSIDERAÇÕES FINAIS

Este estudo comparou o desempenho, na previsão de temperatura de um módulo fotovoltaico, dos modelos *Xtreme Gradient Boosting* (XGBoost) e *Elastic Net Regression* (ElasticNet). Os resultados demonstraram que o XGBoost superou, em relação ás métricas, o algoritmo ElasticNet, alcançando o maior R^2 (0,922), o menor MAE (2,01 °C) e RMSE (3,14 °C). Em contraste, o ElasticNet mostrou-se menos preciso (R^2 = 0,903, MAE = 2,41 °C e RMSE = 3,54 °C). A análise via SHAP validou a coerência do modelo, identificando a Temperatura Ambiente e a Irradiação Solar como os preditores determinantes do aquecimento do módulo. A baixa influência da variável Velocidade do Vento, na análise SHAP, indica que, para as condições climáticas locais, o efeito de arrefecimento convectivo é menos importante do que a combinação da Temperatura Ambiente e da Irradiância Solar, que acabam determinando de forma mais forte o aquecimento do módulo.

A validação estatística, conduzida via teste de Wilcoxon, confirmou que, na previsão da temperatura do módulo fotovoltaico, o XGBoost oferece maior confiabilidade. Apesar do efeito estatístico identificado ser classificado como pequeno, a redução consistente do erro médio absoluto demonstra, em relação ao ElasticNet, um avanço prático relevante. Esse

ganho evidencia que o modelo não-linear consegue representar melhor as dinâmicas presentes nos dados, proporcionando estimativas mais próximas da realidade operacional do sistema.

Portanto, conclui-se que o algoritmo XGBoost se adapta melhor à dinâmica climática de transição (Cfa/Cfb) do Oeste do Paraná. A análise via SHAP confirmou a coerência física do modelo, em que a temperatura ambiente e a irradiância solar atuam como as principais forçantes térmicas do sistema. A superioridade estatística do XGBoost, validada por testes robustos e validação cruzada temporal, sugere que métodos não lineares são importantes para capturar a variabilidade imposta por climas subtropicais na geração fotovoltaica.

Os resultados, além da contribuição metodológica, também evidenciam a importância das condições climáticas regionais na determinação do comportamento térmico de sistemas fotovoltaicos, reforçando a necessidade de modelos calibrados para diferentes contextos climáticos brasileiros.

REFERÊNCIAS

AMINI, F.; HU, G. A two-layer feature selection method using Genetic Algorithm and Elastic Net. **Expert Systems with Applications**, v. 166, n. 1, 2021. https://doi.org/10.1016/j.eswa.2020.114072

BAMISILE, O.; ACEN, C.; CAI, D.; HUANG, Q.; STAFFELL, I. The environmental factors affecting solar photovoltaic output. **Renewable and Sustainable Energy Reviews**, v. 208, n. 1, 2025. https://doi.org/10.1016/j.rser.2024.115073

BOUKE, M. A.; ABUDULLAH, A. An empirical study of pattern leakage impact during data preprocessing on machine learning-based intrusion detection models reliability. **Expert Systems with Applications**, v. 230, n. 1, 2023. https://doi.org/10.1016/j.eswa.2023.120715

BOBADILLA, A. V.; SCHMITT, V.; MALER, C. S.; MENSING, S.; STODTMANN, S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. **Clinical and Translational S***cience*, v. 17, n. 11, 2024. https://doi.org/10.1111/cts.70056

CAMPOS, R. A.; MARTINS, G. L.; RUTHER, R. Aplicação de machine learning para previsão de temperatura de módulos solares fotovoltaicos. 2020. Disponível em: <https://anaiscbens.emnuvens.com.br/cbens/article/view/877>. Acesso em: 26 jan. 2026.

CHAMLAL, H., BENZMANE, A. & OUADERHMAN, T. Elastic net-based high dimensional data selection for regression. **Expert Systems with Applications**, V. 244, N. 1, 2024. https://doi.org/10.1016/j.eswa.2023.122958

COSKUN, C.; TOYGAR, U.; SARPDAG, O. Sensitivity analysis of implicit correlations for photovoltaic module temperature: A review. **Journal of Cleaner Pro***duction*, v. 264, n. 1, 2017. https://doi.org/10.1016/j.jclepro.2017.07.080

DADA, B. A.; NWULU, N.; OLUKANI, O. Bayesian optimization with Optuna for enhanced soil nutrient prediction: a comparative study with genetic algorithm and particle swarm optimization. **Smart Agricultural Techn***ology*, v. 12, n. 1, 2025. https://doi.org/10.1016/j.atech.2025.101136

DAI, X.; YANG, Y. Photovoltaic power prediction and SHAP interpretability analysis based on multi-model comparative learning. **Highlights in Science, Engineering and Technology**, v. 142, n. 1, 2025. https://doi.org/10.54097/ktr7gn53

EBHOTA, W. S.; TABAKOV, P. Y. Influence of photovoltaic cell technologies and elevated temperature on photovoltaic system performance. **Ain Shams Engineering Journal**, v. 14, n. 1, 2023. https://doi.org/10.1016/j.asej.2022.101984

HAMILTON, R. I.; PAPADOPOULOS P. N. Using SHAP Values and Machine Learning to Understand Trends in the Transient Stability Limit. **IEEE Transactions on Power Sys***tems*, v. 39, n. 1, 2024. https://doi.org/10.1109/TPWRS.2023.3248941

HAYAT, S.; NAWAZ, A.; ALMANI, A. A.; MUSTAFA, E.; JAVID, Z.; HOLDERBAUM, W. Enhancing Power System Flexibility Using AI-Based Forecasting Techniques: A Comparative Study. **Battery Energy**, 2025. https://doi.org/10.1002/bte2.70079

IHADDADENE, R.; JED, M. E.; IHADDADENE, N.; SOUZA, A. Analytical assessment of Ain Skhouna PV plant performance connected to the grid under a semi-arid climate in Algeria. **Solar Energy**, v. 223, n. 1, p. 56-62, 2022. https:/doi.org/10.1016/j.solener.2021.12.055

KEDDOUDA, A.; IHADDADENE, R.; BOUKHARI, A.; ATIA, A.; ARICI, M.; LEBBIHIAT, A.; IHADDADENE, N. Photovoltaic module temperature prediction using various machine learning algorithms: Performance evaluation. **Applied Energy**, v. 363, n. 1, 2024. https://doi.org/10.1016/j.apenergy.2024.123064

LI, J.; AN, X.; LI, Q.; WANG, C.; YU, H.; ZHOU, X.; GENG, Y. Application of XGBoost algorithm in the optimization of pollutant concentration. **Atmospheric Research**, v. 276, n. 1, 2022. https://doi.org/10.1016/j.atmosres.2022.106238

LI, P.; LOU, Y.; XIA, X.; SHI, W.; ZHENG, J.; LIAO, Z.; GAO, X.; CHANG, R. Advancing photovoltaic panel temperature forecasting: A comparative study of numerical simulation and machine learning in two types of PV power plant. **Renewable Energy**, v. 237, n. 1, 2024. https://doi.org/10.1016/j.renene.2024.121602

LIEW, X. Y.; HAMEED, N.; CLOS, J. An investigation of XGBoost-based algorithm for breast cancer classification. **Machine Learning with Ap***plications*, v. 16, n. 1, 2021. https://doi.org/10.1016/j.mlwa.2021.100154

LIMA, R. F.; APARECIDO, L. E. O.; TORSONI, G. B.; ROLIM, G. S. Climate Change Assessment in Brazil: Utilizing the Köppen-Geiger (1936) Climate Classification. **Revista Brasileira de Meteorologia**, v. 38, n. 1, 2023. https://doi.org/10.1590/0102-77863810001

MAHMOOD, N.; KADIR, D. H. An Elastic Net Approach to Logistic Regression for Genetic Selection in High-Dimensional Brain Cancer Data. **Cihan University-Erbil Scientific Journal (CUESJ)**, v. 9, n. 1, 2025. https://doi.org/10.24086/cuesj.v9n1y2025.

NIAZKAR, M.; MENACPE, A.; BRETAN, B.; PIRAREI, R.; JIMENEZ, P.; RIGHETTE, M. Applications of XGBoost in water resources engineering: A systematic literature review. **Environmental Modelling & Software**, v. 174, n. 1, 2024. https://doi.org/10.1016/j.envsoft.2024.105971

OKEH, U. M.; ONYEAGU, I. S.. Comparison of Two Diagnostic Test Procedures Using Modified Wilcoxon Signed Rank Test. **Asian Journal of Mathematics & St***atistics*, v. 13, n. 1, 2020. https://doi.org/10.3923/ajms.2020.14.20

OLIVEIRA, L.; SANTOS, J. A. A. Um estudo comparativo da aplicação de redes Neurais artificiais na previsão de geração eólica. **CEREUS**, v. 15, n. 1, 2023. https:/doi.org/10.18605/2175-7275/cereus.v15n1p105-119

PEREIRA, M. D. B.; MOURA, M. O.; LUCENA, D. B. (2021). Análise da variabilidade pluviométrica interanual da zona da mata nordestina e a identificação de anos padrão. **Revista Brasileira De Climatologia**,v. 26, 2020. https://doi.org/10.5380/abclima.v26i0.61276

PEREIRA, E. B.; MARTINS, F. R.; RÜTHER, R. Atlas Brasileiro de Energia Solar. São José dos Campos: INPE, 2020. Disponível em: http://labren.ccst.inpe.br/atlas/ (labren.ccst.inpe.br in Bing). Acesso em: 26 jun. 2026.

PRAVIN, P. S.; ZHI, J.; TAN, M.; YAP, S.; WU, Z. Hyperparameter optimization strategies for machine learning-based stochastic energy efficient scheduling in cyber-physical production systems. **Digital Chemical Engineering**, v. 4, n. 1, 2022. https://doi.org/10.1016/j.dche.2022.100047

RODRÍGUEZ GÓMEZ, J. M.; CARLESSO, F.; VIEIRA, L. E. A.; SILVA, L. A irradiância solar: conceitos básicos. **Revista Brasileira de Ensino de Física**, v. 40, n. 3, 2018. DOI: 10.1590/1806-9126-RBEF-2017-0342.

SANTOS, J. A. A.; POLO, A. N. B. Previsão de custos de produção de frangos: uma aplicação do modelo CNN-LSTM. **Revista ABCustos**, v. 19, n. 3, 2024. https://doi.org/10.47179/abcustos.v19i3.751

SRINIVAS, P.; KATARYA, R. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. **Biomedical Signal Processing and Control**, v.73, n. 1, 2022. https://doi.org/10.1016/j.bspc.2021.103456

SOLANO, E. S., DEHGHANIAN, P.; AFFONSO, C. M. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. **Energies**, 2022. https://doi.org/10.3390/en15197049

SOMALA, S. N.; CHANDA, S.; ALHAMAYDEH, M.; MANGALATHU, S. Explainable XGBoost–SHAP Machine-Learning Model for Prediction of Ground Motion Duration in New Zealand. **Natural Hazards Review**, v. 25, n. 1, 2024. https://doi.org/10.1061/NHREFO.NHENG-1837

TIEGHI, C. P.; CANEPPELE, F. L.; DAL PAI, A.; GODINHO, E. Z.; ALMEIDA, C. F. M.; MALAGHETA D. C.; FRIGO, M. M. Aplicações e desafios da inteligência artificial na previsão da radiação solar: uma revisão sistemática. **Revista Brasileira de Climatologia**. v. 36, n. 21, 2025. https://doi.org/10.55761/abclima.v36i21.18872

TORRES, A. Machine Learning for Predicting Stock Market Movements. **American Journal of Machine Learning**, v. 5, n. 2, 2024. https://doi.org/10.71465/ajml3047

ZHANG, P.; JIA, Y.; SHANG, Y. Research and application of XGBoost in imbalanced data. **International Journal of Distributed Sensor Networks**, 2022. https://doi.org/10.1177/1550132922110.