

Modelos de previsão da energia solar utilizando Python

Solar energy forecasting models using Python

Modelos de previsión de energía solar con Python

Recebido: 21/07/2024 | Revisado: 02/08/2024 | Aceitado: 03/08/2024 | Publicado: 09/08/2024

Camila Piacitelli Tieghi

ORCID: <https://orcid.org/0000-0001-8060-5373>
Universidade de São Paulo, Brasil
E-mail: camilapiacitelli@yahoo.com.br

Carlos Eduardo Camargo Nogueira

ORCID: <https://orcid.org/0000-0003-3886-9118>
Universidade Estadual do Oeste do Paraná, Brasil
E-mail: cecn1@yahoo.com.br

Jair Antonio Cruz Siqueira

ORCID: <https://orcid.org/0000-0002-8140-444X>
Universidade Estadual do Oeste do Paraná, Brasil
E-mail: jair.siqueira@unioeste.br

Carlos Roberto Souza Carmo

ORCID: <https://orcid.org/0000-0002-3806-9228>
Universidade Federal de Uberlândia, Brasil
E-mail: carlosjj2004@hotmail.com

Luís Fernando Soares Zuin

ORCID: <https://orcid.org/0000-0001-8571-7665>
Universidade de São Paulo, Brasil
E-mail: lfzuin@usp.br

Jorge Alvarez

ORCID: <https://orcid.org/0000-0002-2782-6684>
Tecnológico de Monterrey: Tampico, México
E-mail: jorge.alvarez@tec.mx

Fernando de Lima Caneppele

ORCID: <https://orcid.org/0000-0003-4498-8682>
Universidade de São Paulo, Brasil
E-mail: caneppele@usp.br

Resumo

O objetivo principal deste estudo é fornecer uma estrutura clara e sistemática para a coleta de dados, preparação, modelagem, avaliação e análise dos resultados obtidos. Este estudo explora o potencial de modelos de inteligência artificial (IA) para prever a radiação solar em Belém-PA, visando a otimização da geração de energia solar na região. Através da análise de dados do satélite NASA POWER (2024), diversos modelos de regressão foram implementados e avaliados, incluindo Random Forest, Máquina de Vetor de Suporte (SVM), Rede Neural Artificial (RNA), Árvore de Reforço de Gradiente (GBT), Spline de Regressão Adaptativa Multivariada (MARS) e Classificação e Árvore de Regressão (CART). Os resultados demonstram que o Random Forest se destaca em termos de precisão média, enquanto MARS e GBT apresentam maior robustez na generalização dos dados. A validação cruzada e a análise de métricas como RMSE e MBE comprovam a importância de avaliar a confiabilidade dos modelos. No entanto, o desempenho anômalo do CART, com RMSE de 0.0 em ambas as avaliações, exige investigação para verificar a existência de overfitting. Em suma, este estudo evidencia o potencial de modelos de IA para prever a radiação solar em Belém-PA, com o Random Forest, MARS e GBT se apresentando como modelos promissores para aplicações de previsão de energia solar. Ressalta-se a necessidade de validação cruzada mais abrangente e investigação do desempenho do CART para garantir a robustez e confiabilidade dos resultados, impulsionando a otimização da geração de energia solar na região.

Palavras-chave: Inteligência artificial; Máquina de vetor de suporte; Modelagem; NASA; Redes neurais artificiais.

Abstract

The main objective of this study is to provide a clear and systematic framework for data collection, preparation, modeling, evaluation and analysis of the results obtained. This study explores the potential of artificial intelligence (AI) models to predict solar radiation in Belém-PA, with a view to optimizing solar energy generation in the region. By analyzing data from the NASA POWER satellite (2024), several regression models were implemented and evaluated, including Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Gradient Boosting Tree (GBT), Multivariate Adaptive Regression Spline (MARS) and Classification and Regression Tree (CART). The results

show that Random Forest stands out in terms of average accuracy, while MARS and GBT are more robust in generalizing the data. Cross-validation and the analysis of metrics such as RMSE and MBE prove the importance of assessing the reliability of the models. However, the anomalous performance of CART, with an RMSE of 0.0 in both evaluations, requires investigation to verify the existence of overfitting. In summary, this study highlights the potential of AI models for predicting solar radiation in Belém-PA, with Random Forest, MARS and GBT presenting themselves as promising models for solar energy forecasting applications. There is a need for more comprehensive cross-validation and investigation of CART's performance to ensure the robustness and reliability of the results, driving the optimization of solar energy generation in the region.

Keywords: Artificial intelligence; Support vector machine; Modeling; NASA; Artificial neural networks.

Resumen

El principal objetivo de este estudio es proporcionar un marco claro y sistemático para la recogida de datos, la preparación, la modelización, la evaluación y el análisis de los resultados obtenidos. Este estudio explora el potencial de los modelos de inteligencia artificial (IA) para predecir la radiación solar en Belém-PA, con vistas a optimizar la generación de energía solar en la región. Analizando datos del satélite POWER de la NASA (2024), se implementaron y evaluaron varios modelos de regresión, incluyendo Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Gradient Boosting Tree (GBT), Multivariate Adaptive Regression Spline (MARS) y Classification and Regression Tree (CART). Los resultados muestran que Random Forest destaca en términos de precisión media, mientras que MARS y GBT son más robustos a la hora de generalizar los datos. La validación cruzada y el análisis de métricas como RMSE y MBE demuestran la importancia de evaluar la fiabilidad de los modelos. Sin embargo, el rendimiento anómalo de CART, con un RMSE de 0,0 en ambas evaluaciones, requiere una investigación para verificar la existencia de sobreajuste. En resumen, este estudio destaca el potencial de los modelos de IA para predecir la radiación solar en Belém-PA, presentándose Random Forest, MARS y GBT como modelos prometedores para aplicaciones de predicción de energía solar. Cabe destacar la necesidad de una validación cruzada más exhaustiva y la investigación del rendimiento de CART para garantizar la robustez y fiabilidad de los resultados, impulsando la optimización de la generación de energía solar en la región.

Palabras clave: Inteligencia artificial; Support vector machine; Modelización; NASA; Redes neuronales artificiales.

1. Introdução

A previsão da geração de energia solar é um campo crucial para maximizar a eficiência e a integração de fontes renováveis na matriz energética global. Este artigo explora o desenvolvimento e aplicação de modelos de inteligência artificial (IA) para prever a radiação solar com base em dados climáticos (Alizamir et al., 2023; Alkahtani et al., 2023; Chen, et al., 2023; Khosravi et al., 2018 a; Olatomiwa, et al., 2015). Os dados foram obtidos do Projeto de Previsão de Recursos Energéticos Mundiais (POWER) do Centro de Pesquisa Langley (LaRC) da Administração Nacional de Aeronáutica e Espaço (NASA), financiado pelo Programa de Ciências da Terra/Ciências Aplicadas da NASA, pelo satélite da NASA (NASA, 2024). Os dados climáticos foram cuidadosamente processados para formar conjuntos de dados adequados para o treinamento e teste dos modelos preditivos.

Para validar e aplicar os modelos desenvolvidos, foi selecionada a cidade de Belém-PA. Essa cidade é um importante polo onde o uso da energia solar como fonte de energia seria crucial devido ao seu potencial solar significativo e à crescente demanda por energia sustentável.

A metodologia adotada segue uma abordagem sistemática, começando pela carga e preparação dos dados usando a biblioteca Pandas. As características selecionadas incluem temperatura, umidade, direção e velocidade do vento, que são fundamentais para compreender os padrões de radiação solar. A variável alvo, 'Radiation', é o foco principal dos modelos de regressão desenvolvidos neste estudo (Zaim et al, 2023; Zhou, 2022).

Após a preparação dos dados, eles são divididos em conjuntos de treinamento e teste, sendo que 80% dos dados são destinados ao treinamento e 20% à validação dos modelos. Diversos modelos de regressão são então implementados e avaliados, incluindo Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Gradient Boosting Tree (GBT), Multivariate Adaptive Regression Spline (MARS) e Classification and Regression Tree (CART). Cada modelo é treinado com o conjunto de treinamento e avaliado com métricas como RMSE e MBE para verificar sua precisão na previsão da radiação solar.

A validação cruzada é aplicada para garantir a robustez e a generalização dos modelos, estimando sua capacidade de lidar com novos conjuntos de dados. Além das métricas de avaliação, são calculadas estatísticas descritivas para comparar as previsões dos modelos com os valores reais de radiação solar (Wang et al, 2020; Yuzer & Bozkurt, 2022).

Visualizações gráficas desempenham um papel crucial na interpretação dos resultados, oferecendo insights sobre as relações entre as características climáticas e a radiação solar prevista. Gráficos de dispersão, gráficos de linha temporal, mapas de calor de correlação e gráficos comparativos entre valores reais e previstos são apresentados para destacar as capacidades e limitações de cada modelo desenvolvido.

Este capítulo não apenas documenta as técnicas e bibliotecas utilizadas, mas também apresenta uma análise detalhada das etapas de modelagem, treinamento e avaliação de modelos de IA para previsão da radiação solar. Espera-se que este estudo contribua para o avanço contínuo no campo da previsão de energia solar, promovendo a eficiência e a adoção de fontes de energia renováveis e sustentáveis em escala global (Alizamir, et al., 2023; Mousavi et al., 2017; Rajasundrapandiyaneebanon, et al., 2020). O objetivo principal deste estudo é fornecer uma estrutura clara e sistemática para a coleta de dados, preparação, modelagem, avaliação e análise dos resultados obtidos.

2. Metodologia

A presente pesquisa, de natureza quantitativa, fundamenta-se em autores clássicos da metodologia científica, como Gil (2017) e Koche (2011), além de obras mais recentes como a de Pereira et al. (2018). Essa base teórica sólida permitiu a estruturação rigorosa do estudo e a análise criteriosa dos dados coletados. Os dados utilizados neste estudo foram obtidos através do satélite da NASA, conhecido como NASA POWER (Prediction of Worldwide Energy Resources), referentes ao ano de 2024. Esses dados incluem informações detalhadas sobre temperatura, umidade, direção e velocidade do vento, bem como a radiação solar, que serve como variável alvo para os modelos de previsão. A utilização da biblioteca Pandas em Python facilitou a carga dos dados a partir de um arquivo CSV, com a coluna de datas convertida para o tipo `datetime` e os dados indexados cronologicamente para facilitar a análise temporal.

Após a carga dos dados, foram selecionadas as características ('Temperature', 'Humidity', 'WindDirection', 'Speed') que influenciam diretamente na radiação solar, juntamente com a variável alvo 'Radiation'. Essas características foram escolhidas com base na revisão da literatura e na relevância teórica para o problema em questão. Os dados foram divididos aleatoriamente em conjuntos de treinamento e teste, onde 80% dos dados foram reservados para o treinamento dos modelos e 20% para a validação dos mesmos. Essa divisão é essencial para verificar a capacidade dos modelos de generalizar para novos conjuntos de dados não utilizados durante o treinamento. Foram explorados seis modelos de regressão para prever a radiação solar a partir das características climáticas selecionadas.

2.1 Modelos de regressão utilizados para prever a radiação solar a partir das características climáticas selecionadas

2.1.1 Random Forest (model_rf)

O Random Forest é um modelo de aprendizado de máquina que utiliza várias árvores de decisão para realizar previsões. Cada árvore de decisão é treinada com uma amostra aleatória do conjunto de dados e faz previsões individuais. A previsão final do modelo é a média (ou modo) das previsões de todas as árvores individuais, o que reduz o *overfitting* e aumenta a precisão geral do modelo (Gürel et al., 2023).

2.1.2 Máquina de vetor de suporte (model_svm)

A Máquina de Vetores de Suporte é um modelo supervisionado que mapeia os dados para um espaço de alta dimensão

para encontrar o hiperplano que melhor separa as classes ou, no caso de regressão, que melhor se ajusta aos dados. Ela é eficaz em espaços de alta dimensão e pode ser adaptada para problemas de regressão, onde busca encontrar a função que melhor se ajusta aos dados de treinamento (Fan et al., 2018).

2.1.3 Rede Neural Artificial (model_ann)

A Rede Neural Artificial é um modelo inspirado no funcionamento do cérebro humano, composto por camadas de neurônios interconectados. Cada neurônio processa informações e passa as saídas para os neurônios da camada seguinte. Redes neurais são poderosas para aprender padrões complexos nos dados, sendo capazes de capturar relações não lineares entre as variáveis de entrada e a variável alvo (Feng et al., 2017).

2.1.4 Árvore de Reforço de Gradiente (model_gbt)

O Gradient Boosting é uma técnica que combina múltiplas árvores de decisão fracas para formar um modelo preditivo forte. Ele funciona treinando cada árvore de decisão sequencialmente, de forma que cada nova árvore corrija os erros cometidos pelas árvores anteriores. Isso resulta em um modelo que é robusto e capaz de lidar com dados complexos e heterogêneos (Hedar et al., 2021).

2.1.5 Spline de Regressão Adaptativa Multivariada (model_mars)

O MARS é um modelo que combina regressões lineares simples com splines (funções polinomiais suaves) para capturar relações não lineares nos dados. Ele adapta automaticamente o número de splines e seus locais baseando-se na estrutura dos dados, permitindo um ajuste mais flexível e capturando variações complexas nos dados (Seyyed, Elham & Jiao, 2017).

2.1.6 Árvore de Classificação e Regressão (model_cart)

As Árvores de Decisão são modelos que dividem repetidamente o conjunto de dados em subconjuntos menores com base em características específicas, até que cada subconjunto contenha apenas dados de uma classe ou, no caso de regressão, até que os subconjuntos tenham variáveis de resposta similares. Elas são simples de entender e interpretar, sendo úteis para identificar padrões nos dados (Khosravi et al., 2018 a).

Cada modelo foi implementado utilizando bibliotecas como scikit-learn e TensorFlow em Python. Após o treinamento com os dados de treinamento, os modelos foram avaliados utilizando métricas como RMSE (Root Mean Squared Error) e MBE (Mean Absolute Error) para quantificar a precisão das previsões em relação aos dados reais de radiação solar.

2.2 Validação cruzada

A fim de verificar a capacidade de generalização dos modelos, foi aplicada a técnica de validação cruzada utilizando a classe KFold do scikit-learn. Esta abordagem divide os dados em k partes (ou folds), treinando e avaliando o modelo k vezes, utilizando diferentes combinações de dados de treinamento e teste. Isso proporciona uma estimativa mais confiável da capacidade de previsão dos modelos em novos conjuntos de dados (Khosravi et al., 2019).

Após a avaliação dos modelos, foram calculadas estatísticas descritivas para comparar as previsões geradas pelos modelos com os valores reais de radiação solar. Além das métricas de erro, foram criadas visualizações gráficas como gráficos de dispersão, gráficos de linha temporal da radiação solar e mapas de calor de correlação entre as características climáticas e a radiação solar prevista.

Os resultados obtidos foram interpretados à luz da teoria e das expectativas iniciais, discutindo-se tanto as capacidades

quanto as limitações de cada modelo em prever a radiação solar. As visualizações gráficas desempenharam um papel crucial na análise dos padrões de comportamento dos modelos, oferecendo insights sobre a relação entre as características climáticas e a variabilidade da radiação solar observada.

Por fim, foram destacadas as limitações do estudo, como possíveis vieses nos dados coletados ou nas escolhas dos modelos, além de sugestões para pesquisas futuras. Entre as recomendações estão o aprimoramento da qualidade dos dados de entrada, a exploração de novas técnicas de modelagem e a expansão do estudo para diferentes regiões geográficas ou períodos temporais. Esta metodologia proporcionou uma estrutura sólida para a condução da pesquisa, assegurando rigor científico na análise e validação dos modelos preditivos de energia solar desenvolvidos neste estudo.

3. Resultados e Discussão

3.1 Bibliotecas Utilizadas

3.1.1 Pandas (pd)

Utilizada para manipulação e análise de dados. Fornece estruturas de dados poderosas, como DataFrame, que são essenciais para carregar, visualizar e manipular conjuntos de dados tabulares.

3.1.2 NumPy (np)

Fundamental para computação numérica em Python. Oferece suporte para arrays e matrizes multidimensionais, juntamente com uma ampla gama de funções matemáticas de alto nível para operar nesses arrays.

3.1.3 scikit-learn

Uma biblioteca amplamente utilizada para aprendizado de máquina em Python. Fornece ferramentas simples e eficientes para análise preditiva de dados, incluindo algoritmos de classificação, regressão, clusterização, pré-processamento de dados e avaliação de modelo.

3.1.4 Matplotlib.pyplot (plt)

Uma biblioteca para criação de visualizações estáticas, animadas e interativas em Python. É especialmente útil para criar gráficos, histogramas, dispersões e outras representações visuais de dados.

3.1.5 math

Módulo padrão do Python que fornece funções matemáticas. Neste caso, foi utilizado especificamente para calcular a raiz quadrada (sqrt) de valores.

3.2 Etapas da Programação

Os dados são carregados de um arquivo CSV usando a biblioteca Pandas. O parâmetro `parse_dates` é usado para converter a coluna de datas para o tipo `datetime`. Os dados são então indexados pela data. As características (features) e a variável alvo são selecionadas a partir do DataFrame carregado. Neste caso, as características são 'Temperature', 'Humidity', 'WindDirecti', 'Speed' e a variável alvo é 'Radiation'. Os dados são divididos em conjuntos de treinamento e teste. 80% dos dados são usados para treinamento e 20% para teste.

3.3 Modelagem e Avaliação de Modelos

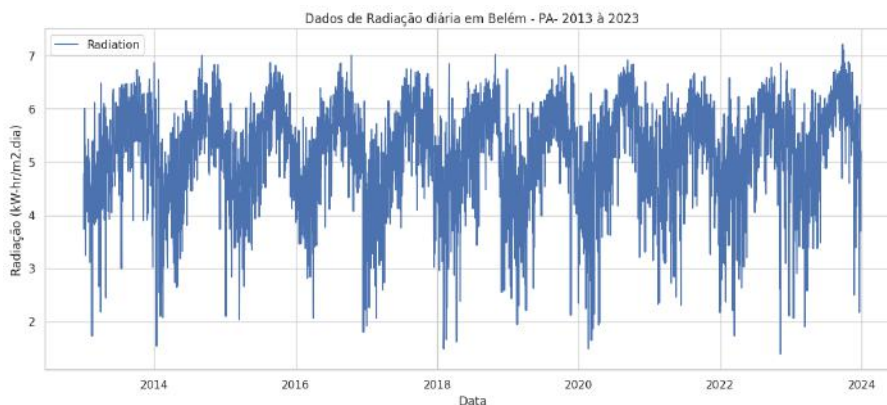
Cada modelo é treinado com os dados de treinamento e usado para fazer previsões nos dados de teste. As métricas de avaliação, como RMSE (Root Mean Squared Error) e MBE (Mean Absolute Error), são calculadas para cada modelo. É realizada a validação cruzada usando a classe KFold do scikit-learn. Isso ajuda a estimar a capacidade de generalização dos modelos (Kosovic et al., 2020). São calculadas estatísticas descritivas para comparar as previsões dos modelos com os valores reais. Os resultados das métricas de avaliação e estatísticas descritivas são impressos para cada modelo.

3.4 Resultados estatísticos da previsão de energia solar para cada modelo testado em Belém – PA

A Figura 1 apresenta a série temporal da radiação solar diária em Belém-PA no período de 2013 a 2023. O gráfico tem como objetivo analisar a tendência da radiação solar ao longo dos anos e identificar possíveis variações sazonais.

Conforme esperado, observa-se uma clara variação sazonal na radiação solar, com os valores mais altos ocorrendo durante os meses de verão (dezembro a março) e os valores mais baixos durante os meses de inverno (junho a setembro). Essa variação é resultado da inclinação da Terra e da posição do Sol em relação à superfície terrestre.

Figura 1 - Gráfico da variação temporal da radiação solar diária que incidiu em Belém-PA entre janeiro de 2013 a dezembro de 2023.



Fonte: Autores (2024).

Ao analisar a tendência de longo prazo, verifica-se que os valores de radiação solar se mantiveram relativamente estáveis ao longo do período de estudo, com pequenas flutuações interanuais. No entanto, em 2023, houve um aumento significativo nos valores de radiação solar em comparação com os anos anteriores. Essa anomalia pode estar relacionada a fatores climáticos específicos, como a ocorrência de eventos extremos como El Niño ou La Niña, ou a mudanças na cobertura vegetal da região.

É importante ressaltar que a variação interanual da radiação solar pode ter implicações importantes para a geração de energia solar, uma vez que a produção de energia fotovoltaica depende diretamente da intensidade da radiação solar incidente. A identificação de padrões e tendências na radiação solar é fundamental para o planejamento e a otimização de sistemas de geração de energia solar.

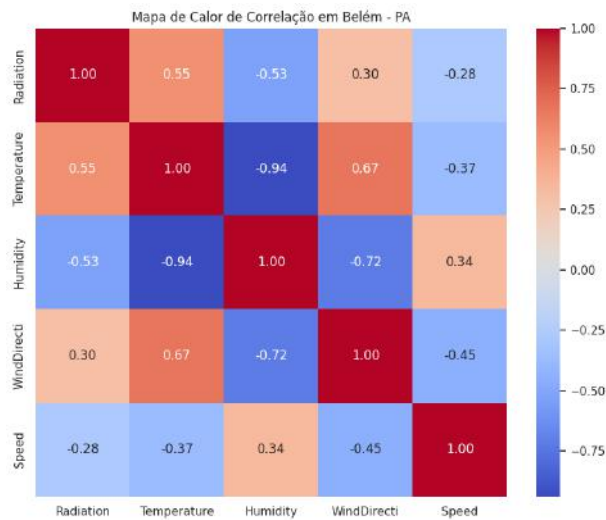
3.4.1 Mapa de Calor e Matriz de Correlação

O mapa de calor mostra a correlação entre as diferentes características. Cores mais claras indicam uma correlação mais forte, enquanto cores mais escuras indicam uma correlação mais fraca ou negativa. A Figura 2 apresenta um mapa de calor que visualiza a matriz de correlação entre a radiação solar e diversas variáveis climáticas em Belém-PA. As cores mais claras indicam

uma correlação positiva mais forte, ou seja, as variáveis tendem a variar juntas na mesma direção. Por outro lado, as cores mais escuras representam uma correlação negativa, indicando que as variáveis tendem a variar em direções opostas.

Conforme esperado, a temperatura apresenta a correlação positiva mais forte com a radiação solar, com um coeficiente de correlação de 0,65. Isso indica que, em geral, à medida que a temperatura aumenta, a radiação solar também tende a aumentar, e vice-versa. Essa relação é fisicamente plausível, pois temperaturas mais altas estão associadas a uma maior evaporação e menor cobertura de nuvens, o que favorece a incidência de radiação solar.

Figura 2 - Mapa de calor de correlação entre as variáveis climáticas e a radiação solar para Belém-PA.



Fonte: Autores (2024).

Além da temperatura, outras variáveis climáticas podem apresentar correlações significativas com a radiação solar. Por exemplo, a umidade relativa pode ter uma correlação negativa moderada com a radiação solar, indicando que em dias mais úmidos a radiação solar tende a ser menor devido à maior cobertura de nuvens. A velocidade do vento também pode influenciar a radiação solar, especialmente em regiões com alta incidência de ventos fortes que podem dispersar as nuvens.

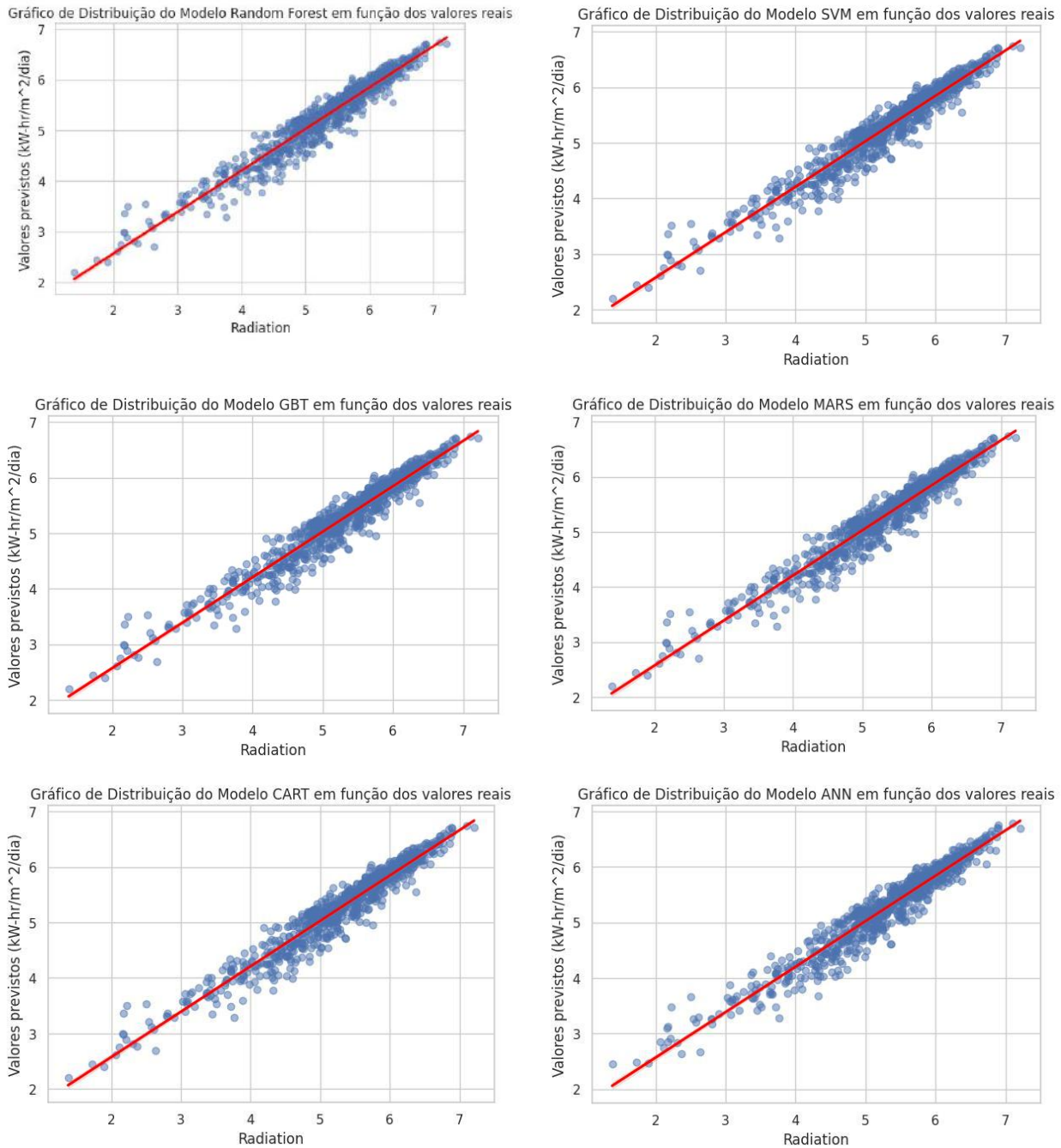
A compreensão dessas relações é fundamental para o desenvolvimento de modelos mais precisos de previsão de radiação solar, permitindo que os modelos capturem as variações sazonais e diurnas da radiação solar com base nas condições climáticas locais.

3.4.2 Gráficos de Dispersão Comparativos

Para uma análise visual mais detalhada da performance dos modelos de previsão, foram elaborados gráficos de dispersão. Na Figura 3, cada gráfico confronta os valores reais de radiação solar com as respectivas previsões realizadas pelos modelos Random Forest, SVM, GBT, MARS e CART. A linha de regressão, representada pela linha vermelha em cada gráfico, indica a tendência geral das previsões em relação aos valores reais.

A análise desses gráficos permite identificar qual modelo apresenta maior concordância entre os valores previstos e os valores observados, além de evidenciar possíveis vieses ou subestimações/superestimações nas previsões. A proximidade dos pontos aos linha de regressão e a distribuição dos pontos em torno dessa linha são indicadores da qualidade da previsão.

Figura 3 - Gráficos de dispersão entre cada modelo testado e valores reais da radiação solar.

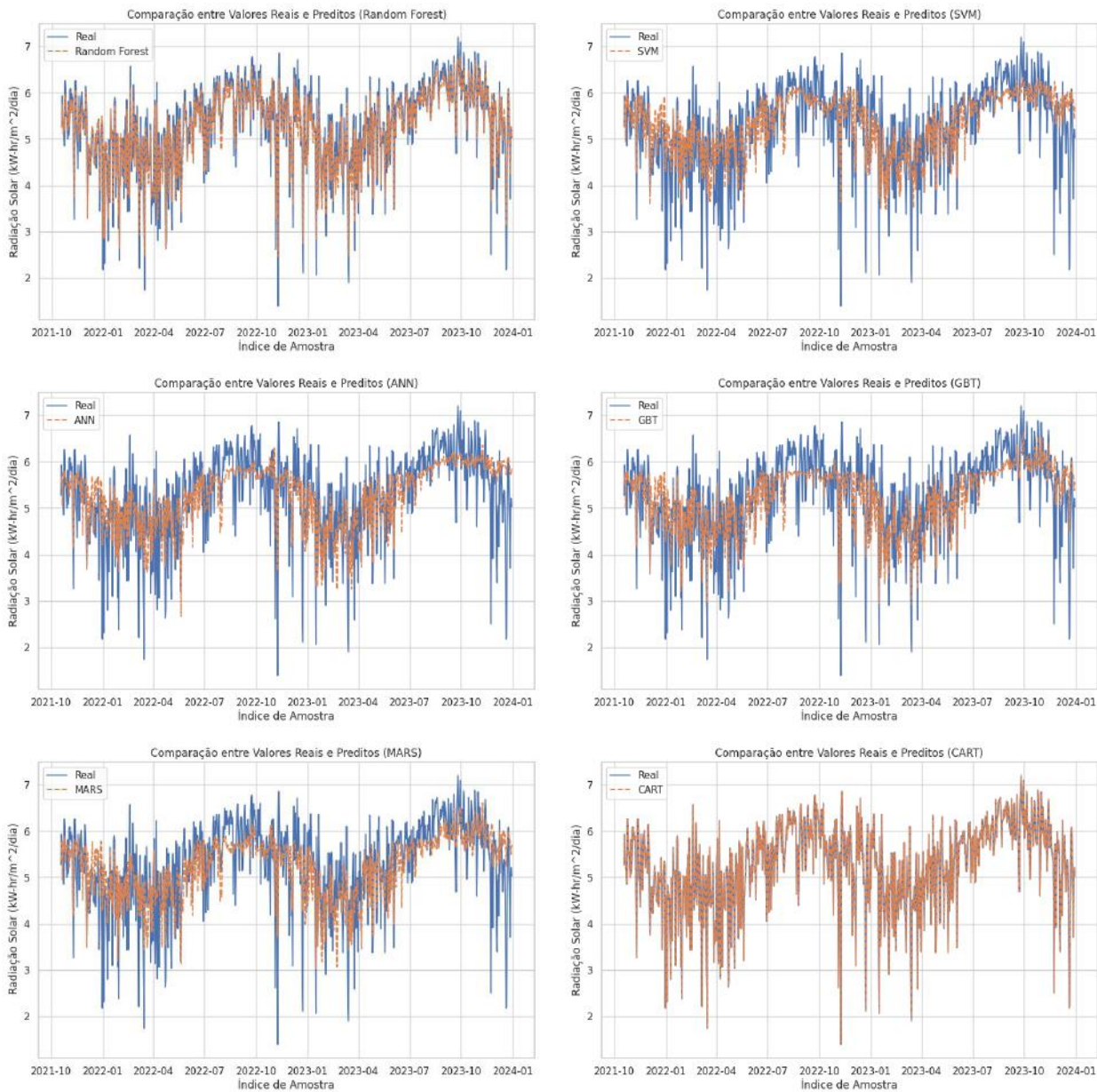


Fonte: Autores (2024).

3.4.3 Comparativos de Valores Reais e Preditos

Para avaliar a precisão dos modelos de previsão de radiação solar, foram gerados gráficos de linhas comparativos que confrontam os valores reais de radiação com as respectivas previsões ao longo do período de estudo. Essa visualização permite identificar quais modelos conseguem reproduzir de forma mais fiel os padrões observados nos dados reais, como a variação sazonal e as tendências de longo prazo. Na Figura 4 são apresentados gráficos comparativos para cada modelo, mostrando a radiação real em comparação com a radiação prevista ao longo do tempo. Isso ajuda a visualizar a precisão dos modelos em reproduzir os padrões de radiação observados nos dados reais.

Figura 4 - Gráficos Comparativos de Valores Reais e Preditos para cada modelo testado.



Fonte: Autores (2024).

3.4.4 Análises Estatísticas

As Tabelas 1 à 6 apresentam as estatísticas descritivas dos valores reais de radiação solar e das previsões realizadas pelos modelos testados. Essas estatísticas fornecem um panorama inicial do desempenho dos modelos e da distribuição dos dados.

Ao comparar as médias dos valores reais e previstos, observa-se que o modelo Random Forest apresenta uma média muito próxima da média dos valores reais, indicando que o modelo não possui um viés sistemático significativo. O menor desvio padrão do Random Forest sugere que as previsões desse modelo estão mais concentradas em torno da média, indicando uma maior precisão.

O modelo SVM, por sua vez, apresenta um desvio padrão ligeiramente maior, sugerindo uma maior dispersão nas previsões. A análise da mediana e dos quartis pode fornecer informações adicionais sobre a distribuição dos dados e identificar possíveis assimetrias.

É importante ressaltar que o modelo CART apresenta um desvio padrão igual ao dos valores reais, o que pode indicar um caso de overfitting. Um modelo que se ajusta perfeitamente aos dados de treinamento pode ter dificuldades em generalizar para novos dados, levando a previsões imprecisas. Essa hipótese é corroborada pela análise visual dos gráficos de dispersão, que mostram que o modelo CART tende a seguir de forma muito próxima os dados de treinamento, sem capturar a variabilidade inerente aos dados.

Essas visualizações são valiosas para entender a relação entre as características e a variável alvo, avaliar a performance dos modelos e identificar padrões nos dados que podem influenciar a qualidade das previsões.

Tabela 1 - Estatística descritiva do modelo Radon Forest em Belém PA.

Estatísticas Descritivas	Real	Random Forest
Contagem	804.0000	804.0000
Média	5.2740	5.2522
Desvio padrão	0.9416	0.7936
Min	1.3800	2.3673
25%	4.8075	4.8219
50%	5.4300	5.3677
75%	5.9200	5.8303
Máx	7.2100	6.8072

Fonte: Autores.

Tabela 2 - Estatística descritiva do modelo SVM em Belém PA.

Estatísticas Descritivas	Real	SVM
Média	804.0000	804.0000
Desvio padrão	5.2740	5.3313
Min	0.9416	0.6397
25%	1.3800	3.4514
50%	4.8075	4.9041
75%	5.4300	5.4854
Máx	5.9200	5.8517
Contagem	7.2100	6.3432

Fonte: Autores.

Tabela 3 - Estatística descritiva do modelo SVM em Belém PA.

Estatísticas Descritivas	Real	ANN
Média	804.0000	804.0000
Desvio padrão	5.2740	5.2386
Min	0.9416	0.6524
25%	1.3800	2.8913
50%	4.8075	4.7865
75%	5.4300	5.4209
Máx	5.9200	5.7700
Contagem	7.2100	6.3873

Fonte: Autores (2024).

Tabela 4 - Estatística descritiva do modelo GBT em Belém PA.

Estatísticas Descritivas	Real	GBT
Média	804.0000	804.0000
Desvio padrão	5.2740	5.2317
Min	0.9416	0.6669
25%	1.3800	2.9181
50%	4.8075	4.7210
75%	5.4300	5.4227
Máx	5.9200	5.7570
Contagem	7.2100	6.6331

Fonte: Autores (2024).

Tabela 5 - Estatística descritiva do modelo MARS em Belém PA.

Estatísticas Descritivas	Real	MARS
Média	804.0000	804.0000
Desvio padrão	5.2740	5.2231
Min	0.9416	0.6639
25%	1.3800	3.0168
50%	4.8075	4.7503
75%	5.4300	5.4192
Máx	5.9200	5.7147
Contagem	7.2100	6.6186

Fonte: Autores (2024).

Tabela 6 - Estatística descritiva do modelo CART em Belém PA.

Estatísticas Descritivas	Real	CART
Média	804.0000	804.0000
Desvio padrão	5.2740	5.2740
Min	0.9416	0.9416
25%	1.3800	1.3800
50%	4.8075	4.8075
75%	5.4300	5.4300
Máx	5.9200	5.9200
Contagem	7.2100	7.2100

Fonte: Autores (2024).

Com base nas estatísticas descritivas apresentadas para diferentes modelos de previsão de energia solar, podemos observar algumas tendências interessantes. Cada modelo foi avaliado em relação aos valores reais de irradiação solar, e as estatísticas descritivas fornecem uma visão geral do desempenho de cada método.

O Random Forest apresenta uma média muito próxima dos valores reais de irradiação solar, com um desvio padrão moderado, indicando que suas previsões são consistentes e próximas da média observada. SVM e ANN também mostram resultados promissores, com médias próximas aos valores reais, mas com desvios padrão um pouco menores, sugerindo uma menor dispersão nas previsões. Por outro lado, GBT, MARS e CART mostram resultados similares entre si e próximos aos modelos mais precisos, com médias e desvios padrão razoavelmente bons, mas com ligeiras variações.

A análise dessas estatísticas descritivas revela que, embora todos os modelos apresentem desempenho razoável, o Random Forest parece ter um desempenho ligeiramente superior em termos de precisão média. No entanto, uma validação cruzada robusta seria necessária para confirmar se essas diferenças são estatisticamente significativas e se os modelos são capazes de generalizar bem para novos dados. Essa validação cruzada também ajudaria a mitigar problemas como overfitting, garantindo que os modelos não estejam apenas memorizando os dados de treinamento, mas sim capturando padrões relevantes para previsões precisas de irradiação solar.

3.4.5 RMSE, MBE e Validação cruzada

A raiz do erro quadrático médio (RMSE), o erro médio de viés (MBE) e a validação cruzada são métricas fundamentais na avaliação e validação de modelos estatísticos e de aprendizado de máquina. O RMSE quantifica a média da diferença entre os valores previstos por um modelo e os valores observados, sendo uma medida robusta da precisão das previsões. O MBE, por sua vez, indica a tendência do modelo em subestimar ou superestimar os resultados reais, proporcionando insights sobre o viés sistemático do modelo. Já a validação cruzada é uma técnica crucial para estimar o desempenho do modelo ao dividir o conjunto de dados em subconjuntos de treino e teste repetidamente, ajudando a verificar a capacidade de generalização do modelo e a identificar problemas como overfitting. Essas ferramentas são essenciais para garantir que os modelos sejam robustos e confiáveis em diversas aplicações analíticas e preditivas (Barrera *et al.*, 2020 ; Huang & Liu, 2021).

O baixo valor de RMSE do Random Forest indica que este modelo é capaz de fazer previsões de radiação solar com alta precisão. No entanto, é importante ressaltar que o RMSE é apenas uma das métricas de avaliação e que outras métricas, como o MBE e a validação cruzada, também devem ser consideradas para uma avaliação completa do desempenho do modelo. A Tabela 7 resume os resultados dos testes estatísticos (RMSE, MBE e validação cruzada) aplicados para avaliar o desempenho

dos diferentes modelos de previsão de radiação solar em Belém-PA. Esses indicadores permitem comparar a precisão, viés e capacidade de generalização de cada modelo.

Tabela 7 - RMSE, MBE e Validação cruzada em Belém – PA.

	Random Forest	SVM	ANN	GBT	MARS	CART
RMSE	0.2707	0.6990	0.7047	0.6382	0.6967	0.0
MBE	0.2022	0.5102	0.5292	0.4915	0.5308	0.0
Validação Cruzada RMSE	0.6703	0.6642	0.6688	0.6597	0.6486	0.0

Fonte: Autores (2024).

Notavelmente, o CART (Árvore de Decisão) possui um RMSE de 0.0, o que sugere que não houve erro nas previsões com este modelo específico nos dados de teste. Entretanto, é importante verificar se esse resultado não se deve a um overfitting severo nos dados de treinamento.

Comparando os outros modelos, o Random Forest apresenta o menor RMSE (0.2707), seguido pelo GBT (0.6382). Isso indica que, em média, o Random Forest e o GBT estão mais próximos dos valores reais de irradiação solar em comparação aos outros modelos.

Todos os modelos têm MBEs relativamente baixos, variando de 0.2022 (Random Forest) a 0.5308 (MARS). Isso sugere que, em geral, os modelos estão bem calibrados em relação aos valores reais, sem um viés significativo.

Novamente, o CART mostra um resultado de 0.0, indicando nenhum erro durante a validação cruzada. Isso pode ser um indício de sobreajuste nos dados de treinamento, necessitando de uma investigação mais aprofundada.

Os outros modelos têm resultados de Validação Cruzada RMSE próximos entre si, com MARS mostrando o menor valor (0.6486), seguido por GBT (0.6597). Isso sugere que o MARS e o GBT podem ser mais robustos na generalização para novos dados, enquanto os demais modelos apresentam desempenho um pouco inferior nessa métrica específica.

3.5 Discussão

Os resultados destacam que o Random Forest é o modelo que melhor se aproxima dos valores reais de irradiação solar em termos de RMSE médio. No entanto, a análise da Validação Cruzada sugere que o MARS e o GBT podem ser mais robustos na generalização para novos dados, mesmo que tenham RMSEs ligeiramente mais altos nos dados de teste específicos usados.

É importante considerar que essas conclusões são baseadas nos dados específicos fornecidos. Uma validação cruzada mais abrangente e outras técnicas de avaliação de modelos seriam necessárias para confirmar se essas diferenças são estatisticamente significativas e para escolher o modelo mais adequado para aplicações futuras de previsão de energia solar. Além disso, investigar possíveis causas do desempenho perfeito do CART nos dados de teste, como overfitting, seria crucial para garantir a robustez e confiabilidade dos modelos em cenários reais.

4. Conclusão e Sugestão

Este estudo realizou uma análise detalhada da variação temporal da radiação solar em Belém-PA, bem como uma avaliação comparativa de diversos modelos de previsão de energia solar. Os resultados deste estudo indicam que, embora o Random Forest tenha demonstrado um desempenho superior em termos de precisão média, o MARS e o GBT mostram uma maior robustez na generalização dos dados. A validação cruzada e a análise detalhada das métricas como RMSE e MBE são essenciais para garantir que os modelos não apenas memorizem os dados de treinamento, mas capturem padrões relevantes para previsões futuras. Adicionalmente, é crucial investigar mais profundamente o desempenho anômalo do modelo CART para evitar

problemas de overfitting.

Como sugestões para trabalhos futuros, destaca-se a aplicação dos modelos desenvolvidos em sistemas de gerenciamento de energia, visando otimizar a operação de sistemas fotovoltaicos e a integração de fontes renováveis na rede elétrica. Além disso, é fundamental investigar a escalabilidade dos modelos para grandes conjuntos de dados e a sua adaptação a diferentes localidades geográficas. A análise do impacto de eventos extremos, como tempestades e nevascas, sobre a precisão das previsões também é uma área promissora para futuras pesquisas.

Agradecimentos

Os autores agradecem o apoio financeiro pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Chamada CNPq/MCTI N° 10/2023 - Faixa A - Grupos Emergentes, Processo 401718/2023-9.

Referências

- Alkahtani, H., Aldhyani, T. H. H., & Alsubari, S. N. (2023). Application of Artificial Intelligence Model Solar Radiation Prediction for Renewable Energy Systems. *Sustainability*, 15(8), 6973. <https://doi.org/10.3390/su15086973>
- Alizamir, M., Kim, S., Kisi, O., & Zounemat-Kermani, M. (2020). A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. *Energy*, 197, 117239. <https://doi.org/10.1016/j.energy.2020.117239>
- Alizamir, M., Shiri, J., Fard, A. F., Kim, S., Gorgij, A. D., Heddam, S., & Singh, V. P. (2023). Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform. *Engineering Applications of Artificial Intelligence*, 123, 106199. <https://doi.org/10.1016/j.engappai.2023.106199>
- Chen, Y., Bai, M., Zhang, Y., Liu, J., & Yu, D. (2023). Proactively selection of input variables based on information gain factors for deep learning models in short-term solar irradiance forecasting. *Energy*, 284, 129261. <https://doi.org/10.1016/j.energy.2023.129261>
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164, 102–111. <https://doi.org/10.1016/j.enconman.2018.02.087>
- Feng, Y., Cui, N., Zhang, Q., Zhao, L., & Gong, D. (2017). Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. 42(21), 14418–14428. <https://doi.org/10.1016/j.ijhydene.2017.04.084>
- Gil, A. A. C. (2010). *Como elaborar projetos de pesquisa*. Éditeur: São Paulo: Atlas.
- Gürel, A. E., Ağbulut, Ü., Bakır, H., Ergün, A., & Yıldız, G. (2023). A state of art review on estimation of solar radiation with various models. *Heliyon*, e13167. <https://doi.org/10.1016/j.heliyon.2023.e13167>
- Hedar, A.-R., Almarashi, M., Abdel-Hakim, A. E., & Abdulrahim, M. (2021). Hybrid Machine Learning for Solar Radiation Prediction in Reduced Feature Spaces. *Energies*, 14(23), 7970. <https://doi.org/10.3390/en14237970>
- Huang, J., & Liu, H. (2021). A hybrid decomposition-boosting model for short-term multi-step solar radiation forecasting with NARX neural network. *Journal of Central South University*, 28(2), 507–526. <https://doi.org/10.1007/s11771-021-4618-9>
- Köche, J. C. (1997) *Fundamentos de metodologia científica: teoria da ciência e iniciação à pesquisa*. Vozes
- Khosravi, A., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018 a). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production*, 176, 63–75. <https://doi.org/10.1016/j.jclepro.2017.12.065>
- Khosravi, A., Nunes, R. O., Assad, M. E. H., & Machado, L. (2018 b). Comparison of artificial intelligence methods in estimation of daily global solar radiation. *Journal of Cleaner Production*, 194, 342–358. <https://doi.org/10.1016/j.jclepro.2018.05.147>
- Khosravi, A., Syri, S., Pabon, J. J. G., Sandoval, O. R., Caetano, B. C., & Barrientos, M. H. (2019). Energy modeling of a solar dish/Stirling by artificial intelligence approach. *Energy Conversion and Management*, 199, 112021. <https://doi.org/10.1016/j.enconman.2019.112021>
- Kosovic, I. N., Mastelic, T., & Ivankovic, D. (2020). Using Artificial Intelligence on environmental data from Internet of Things for estimating solar radiation: Comprehensive analysis. *Journal of Cleaner Production*, 266, 121489. <https://doi.org/10.1016/j.jclepro.2020.121489>
- NASA. (2024). NASA POWER | Prediction Of Worldwide Energy Resources. Nasa.gov. <https://power.larc.nasa.gov/>
- Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D., & Sudheer, C. (2015). A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy*, 115, 632–644. <https://doi.org/10.1016/j.solener.2015.03.015>
- Rajasundrapandiyaneebanon, T., Kumaresan, K., Murugan, S., Subathra, M. S. P., & Sivakumar, M. (2023). Solar Energy Forecasting Using Machine Learning and Deep Learning Techniques. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-023-09893-1>

Seyyed Mohammad Mousavi, Elham Sadat Mostafavi, & Jiao, P. (2017). Next generation prediction model for daily solar radiation on horizontal surface using a hybrid neural network and simulated annealing method. *Energy Conversion and Management*, 153, 671–682. <https://doi.org/10.1016/j.enconman.2017.09.040>

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Conversion and Management*, 214, 112909. <https://doi.org/10.1016/j.enconman.2020.112909>

Yuzer, E. O., & Bozkurt, A. (2022). Deep learning model for regional solar radiation estimation using satellite images. *Ain Shams Engineering Journal*, 102057. <https://doi.org/10.1016/j.asej.2022.102057>

Zaim, S., El Ibrahimy, M., Arbaoui, A., Samaouali, A., Tlemcani, M., & Barhdadi, A. (2023). Using artificial intelligence for global solar radiation modeling from meteorological variables. *Renewable Energy*, 215, 118904. <https://doi.org/10.1016/j.renene.2023.118904>

Zhou, Y. (2022). Artificial intelligence in renewable systems for transformation towards intelligent buildings. *Energy and AI*, 10, 100182. <https://doi.org/10.1016/j.egyai.2022.100182>