Analysis of Machine Learning Methods for Estimating Solar Irradiation

Conference Paper · November 2023

CITATIONS
0

READS
8

2 authors:

Matheus Henrique da Silva
Federal University of Technology - Paraná/Brazil (UTFPR)
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Wesley Angelino Souza
Federal University of Technology - Paraná/Brazil (UTFPR)
86 PUBLICATIONS **316** CITATIONS
SEE PROFILE

Análise de Métodos de Aprendizado de Máquina para Estimativa da Irradiação Solar

Analysis of Machine Learning Methods for Estimating Solar Irradiation

Matheus Henrique da Silva[1], Wesley Angelino de Souza[2]

RESUMO

Os efeitos das mudanças climáticas e seus impactos ambientais estão impulsionando o desenvolvimento de ações, como a transição energética para geração com fontes renováveis. A energia solar se destaca por sua ampla aplicabilidade para a geração elétrica, e a irradiação é o parâmetro essencial em sua implementação. No entanto, há desafios associados à aquisição e precisão dos dados dos sensores, devido à sua complexidade, elevados custos, falta de políticas de incentivo e por conta de variações climáticas. Diante deste cenário, dada a importância estratégica e econômica da irradiação solar, este trabalho visa analisar diferentes métodos de aprendizado de máquina (AM) para estimar o valor da irradiação solar ao longo do dia utilizando como parâmetros dados sensoriais meteorológicos mais acessíveis. Por meio da análise das características climáticas da região de interesse, bem como o agrupamento, tratamento e exploração dos dados, e aplicando de forma otimizada os algoritmos por meio de testes de desempenho e precisão, destacaram-se dois modelos: Perceptron Multicamadas (MLP) e Árvore de Decisão (DT). Em testes de estimação realizados para diferentes anos e estações, e utilizando dados de hora, temperatura e umidade, o MLP alcançou precisão média de 92,99%, sendo o mais adequado para previsão de irradiação solar para o cenário estudado.
PALAVRAS-CHAVE: análise de dados; aprendizado de máquina; irradiação solar.

ABSTRACT

The effects of climate change and its environmental impacts drive the development of actions such as the transition to renewable energy generation. Solar energy is widely applicable for electricity generation, and irradiation is an essential parameter in its implementation. However, there are challenges associated with the acquisition and accuracy of sensor data due to its complexity, high costs, lack of incentive policies, and climate variations. Given the strategic and economic importance of solar irradiation, this work aims to analyze different machine learning (ML) methods to estimate the value of solar irradiation throughout the day using easily accessible meteorological sensory data as parameters. Through the analysis of the climatic characteristics of the region of interest, as well as the grouping, processing, and exploration of the data, and by applying the algorithms in an optimized way through performance and accuracy tests, two models stood out: Multilayer Perceptron (MLP) and Decision Tree (DT). In estimation tests conducted for different years and seasons, using data of the hour, temperature, and humidity, MLP achieved an average accuracy of 92.99%, making it the most suitable model for solar irradiation forecasting for the studied scenario.
KEYWORDS: data analytics; solar irradiation; machine learning.

INTRODUÇÃO

Nas últimas décadas, há uma crescente conscientização sobre as mudanças climáticas e os impactos das atividades humanas no meio ambiente, o que vem proporcionando atividades de transição energética para métodos de geração de energia renovável (FREI; KIM, 2019). Nesse cenário, sistemas de energia fotovoltaica são destaques por sua ampla aplicabilidade e geração elétrica distribuída, porém sua eficiência depende da irradiação solar, grandeza de difícil aquisição por elevados custos instrumentais e operacionais (TAVARES; ALONSO; SOUZA, 2023).

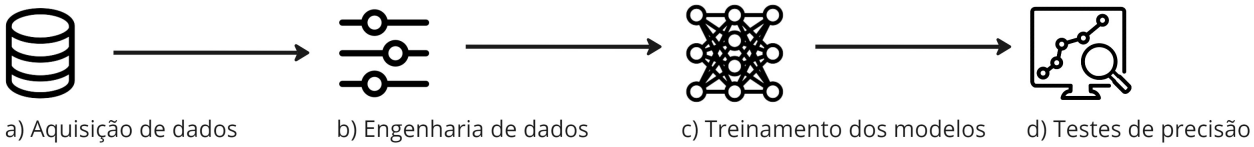
[1] Discente de Iniciação Científica. Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil. E-mail: matheussilva.2019@alunos.utfpr.edu.br. ID Lattes: 5450995625966991.
[2] Docente do Departamento Acadêmico de Elétrica. Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil. E-mail: wesleyangelino@utfpr.edu.br. ID Lattes: 8594457321079718.

No entanto, existem sensores de outras grandezas, como temperatura e umidade, que são mais acessíveis e podem ser usados para auxiliar na estimativa da irradiação. Portanto, dada a importância econômica e a gestão energética, este trabalho propõe analisar dados provenientes de estações meteorológicas e a aplicação de métodos de aprendizado de máquina (AM) para estimar a irradiação solar ao longo do dia. Com a extração de grandezas meteorológicas, são realizados estudos das características climáticas da região de interesse, bem como o agrupamento, tratamento e exploração dos dados. Por meio da aplicação e otimização de parâmetros de algoritmos regressores, é validada a metodologia de estimativa de irradiação solar utilizando sensores mais acessíveis por meio de análise de desempenho e precisão de métodos de AM.

MATERIAIS E MÉTODOS

A Figura 1 apresenta as etapas de desenvolvimento do método proposto neste trabalho.

Figura 1 – Diagrama da metodologia utilizada



Fonte: Elaborado pelos autores (2023).

a) **Aquisição dos dados:** Primeiramente, é necessária a seleção de uma área geográfica de interesse, dada as dimensões continentais do Brasil. Em seguida, é realizado o estudo climático da região para compreender as características ao longo das estações do ano e possíveis padrões. A obtenção dos dados é realizada pela exploração de bases públicas das agências nacionais e estaduais de meteorologia, com ações de manuseio e agrupamento em uma única base. É importante destacar que a coleta de dados também é viável, desde que realizada com sensores de elevada precisão.

b) **Engenharia de dados:** A análise exploratória é essencial para a compreensão inicial dos dados, possibilitada pelo uso de diferentes métodos estatísticos de estimativas de posição e variabilidade, além da relação entre as variáveis. Na sequência, os dados são limpos e tratados, para que possam ser preparados para melhor expor os padrões intrínsecos aos modelos e seus algoritmos, evitando que escalas diferentes influenciem na estimação do valor alvo.

c) **Treinamento dos modelos:** A partir dos conhecimentos sobre a base e as especificações das variáveis, é realizado o estudo e seleção dos modelos de aprendizado de máquina (AM). Os treinamentos para aprendizado e testes dos algoritmos com hiper-parâmetros pré-estabelecidos ocorrem, seguindo para ajustes e otimizações da seleção de variáveis e validação cruzada de diferentes hiper-parâmetros, analisando os resultados segundo pontuações estabelecidas.

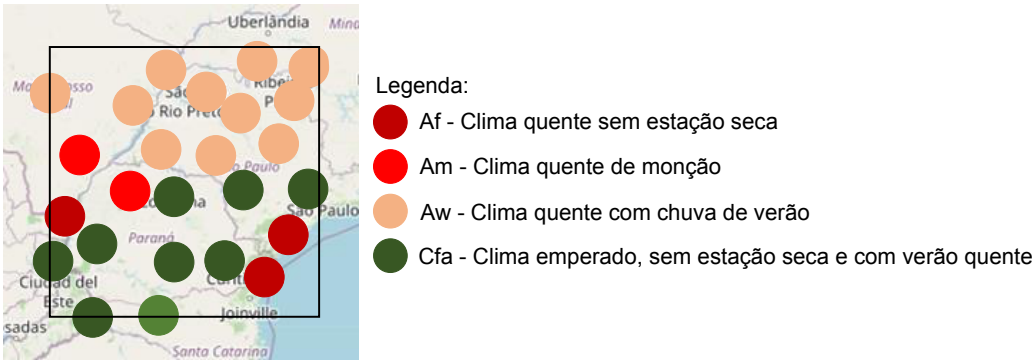
d) **Testes de precisão:** Com os algoritmos otimizados, a validação é realizada por meio da avaliação do desempenho dos modelos treinados de AM para dados ainda não observados. Com tais resultados, é possível elencar o melhor modelo de AM para o problema e objetivo proposto.

Considerando as etapas supracitadas, a seção de resultados apresenta o estudo de caso considerando a região de Cornélio Procópio, Paraná, Brasil.

RESULTADOS

O Brasil possui diversos climas e uma das classificações é a de Wilhelm Köppen (1846-1940), que utiliza temperatura e precipitação, principalmente, para classificar segundo a característica geral, as particularidades dos regimes de chuvas e a temperatura característica (DUBREUIL et al., 2017). A Figura 2 mostra as classificações na área de interesse de ±4º em torno do município de Cornélio Procópio. No Paraná, entre maio e agosto, é registrada uma menor temperatura, capacidade de armazenamento de umidade e evapotranspiração na atmosfera. Tais períodos apresentam menor atuação das massas de ar responsáveis pela pluviosidade, o que acarreta em um menor índice de irradiação solar (SALTON; MORAIS; LOHMANN, 2020).

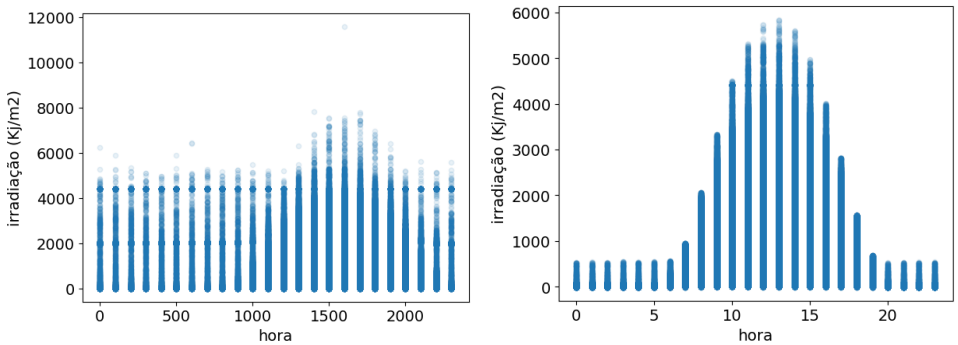
Figura 2 – Classificações climáticas de Köppen na área



Fonte: Elaborado pelos autores com base em Dubreuil et al. (2017).

A base de dados das estações meteorológicas do Instituto Nacional de Meteorologia (INMET), contém 100 estações e 6.470.986 amostras. A análise estatística exploratória da irradiação (Kj/m^2), de medições horárias, é apresentada na Figura 3, com correlação de Pearson de 0,3983 em relação com a hora. No entanto, há valores maiores que 4.000 Kj/m^2 em períodos noturnos, um *outlier* acima de 11.000 Kj/m^2, além de um deslocamento de medições entre as 09h00 e 22h00, fora do período de luz solar, devido ao padrão de fuso horário GMT0. Portanto, foi realizado o processo de tratamento dos dados, selecionando as que ficaram no intervalo médio com ±3,5 desvios-padrão e as amostras foram deslocadas para o GMT -3, resultando na distribuição exemplificada na Figura 3, com 0,4022 de correlação e 6.456.111 amostras.

Figura 3 – Distribuição da irradiação horária antes e depois do tratamento de dados



Fonte: Elaborado pelos autores (2023).

Entre as variáveis preditoras hora, precipitação, pressão atmosférica, temperatura, temperatura de orvalho, umidade, direção do vento, vento máximo e velocidade, é feita a análise de colinearidade

(CHAN et al., 2022) por meio de filtro de correlação em 0,8, eliminando a variável velocidade. Posteriormente, os dados foram normalizados através do método *StandardScaler* (STD) e min-max.

Como etapa posterior, foram utilizados quatro técnicas de regressão: *multilayer perceptron* (MLP), *decision tree* (DT), *random forest* (RF) e k-*nearest neighbors* (k-NN). O MLP é composto por várias redes neurais perceptron com camada de entrada, uma ou mais escondidas com pesos e dinâmicas internas, e saída, permitindo a extração de padrões não lineares e gerando maior precisão (ZHOU et al., 2021). A DT é composta por um conjunto de regras, com habilidade de descobrir padrões ocultos nos dados pelas iterações complexas das amostras, apresentando desempenho satisfatório na previsão de irradiação solar (RAHUL et al., 2021). A RF é composta por várias árvores de decisão no treinamento, de melhor eficiência que DT ao reduzir o viés e a variância das previsões para as diversas condições e generalidades (SHETTY et al., 2021). O k-NN calcula distâncias entre os dados de acordo com a função estabelecida, reconhecendo os comportamentos e atribuindo pesos, segundo os dados da vizinhança (LIU; ZHANG, 2016).

Os dados das 100 estações foram divididos aleatoriamente na proporção de 70, 20 e 10% para treinamento, testes e validação, respectivamente. A análise dos resultados utilizam as métricas Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Pontuação R2 (R2). Foi feita a análise da qualidade dos atributos para os modelos de regressão, tendo o MLP e k-NN utilizando como dados de entrada a hora, umidade e temperatura, o DT utilizando todos, exceto precipitação e o RF utilizando como dados de entrada a hora, temperatura, umidade e pressão atmosférica. Para a determinação dos hiper-parâmetros dos modelos, foi feita uma busca em grade com valores variados para cada modelo. Os melhores valores encontrados estão apresentados na Tabela 1. Para a avaliação dos resultados, foi feita a validação cruzada com k-fold (k=10). A Tabela 2 apresenta o desempenho dos modelos otimizados, com a porcentagem (%) de ganho ou perda de desempenho em comparação aos modelos base, conforme expostos no código do trabalho, disponibilizados no final do documento.

Tabela 1 – Parâmetros dos modelos otimizados.

Algoritmo	Parâmetros
MLP	Camada escondida: (60,30), Função de ativação: ReLU, Otimizador: adam
DT	Profundidade máxima: 17, Número de folhas: 18
RF	Número de árvores: 70, Profundidade máxima: 25; Mínimo de folhas = 10
k-NN	k (grau de vizinhança): 27; métrica: euclidiana

Fonte: Elaborado pelos autores (2023).

Tabela 2 – Desempenho dos modelos otimizados.

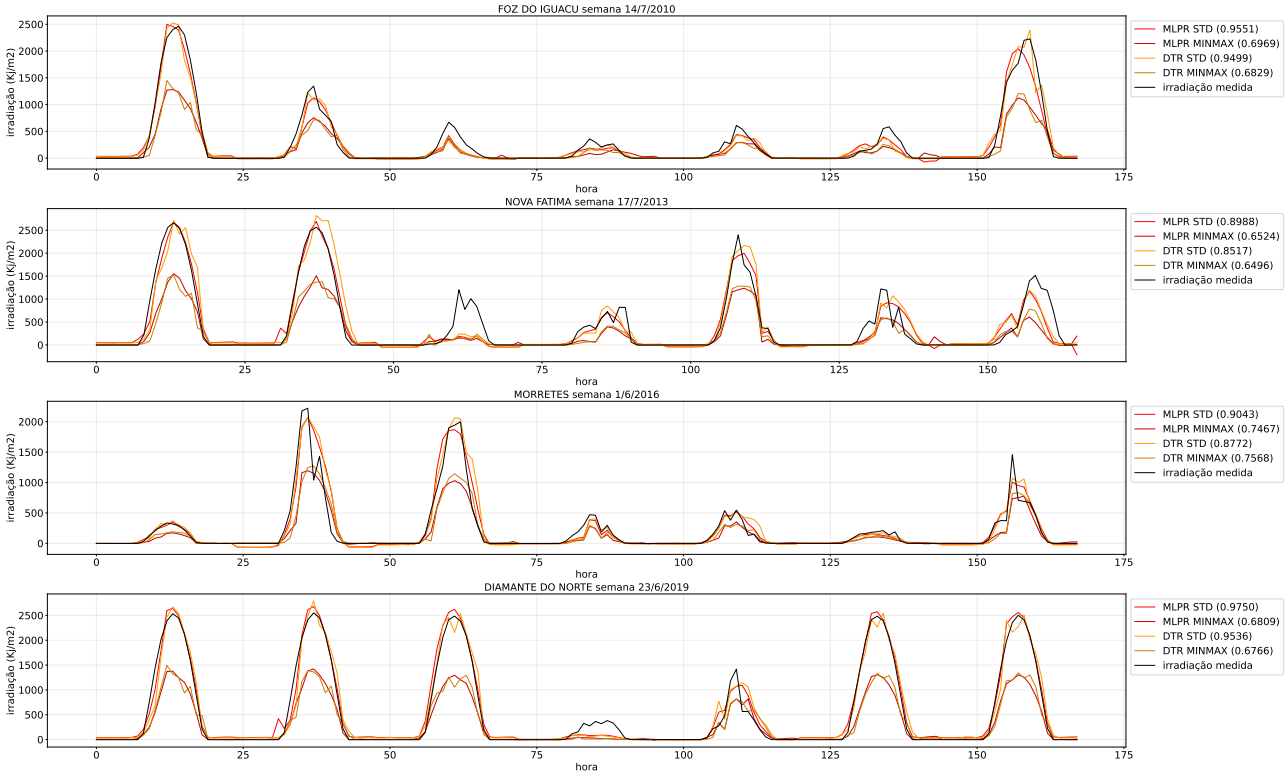
Variável	MLP		DT		RF		k-NN	
MAE	0,1869	($\uparrow 5{,}77\%$)	0,1747	($\downarrow 23{,}44\%$)	0,1766	($\uparrow 4{,}37\%$)	0,1910	($\downarrow 4{,}45\%$)
MSE	0,1268	($\uparrow 14{,}34\%$)	0,1171	($\downarrow 42{,}46\%$)	0,1187	($\uparrow 6{,}84\%$)	0,1360	($\downarrow 8{,}66\%$)
RMSE	0,3562	($\uparrow 6{,}93\%$)	0,3422	($\downarrow 24{,}16\%$)	0,3446	($\uparrow 3{,}36\%$)	0,3687	($\downarrow 4{,}45\%$)
R2	0,8730	($\downarrow 1{,}74\%$)	0,8828	($\uparrow 10{,}86\%$)	0,8812	($\downarrow 0{,}84\%$)	0,8639	($\uparrow 1{,}52\%$)
t_{treino}	441,2	($\downarrow 25{,}52\%$)	30,9	($\downarrow 36{,}02\%$)	1011,4	($\downarrow 72{,}71\%$)	-	-
t_{teste}	3,2	($\downarrow 80{,}33\%$)	0,3	($\downarrow 70{,}00\%$)	37,2	($\downarrow 79{,}80\%$)	80,4	($\downarrow 75{,}91\%$)

Fonte: Elaborado pelos autores (2023).

DISCUSSÕES

A análise de validação mostra que o MLP possui uma média semelhante aos demais, aliado ao menor desvio-padrão o torna preciso, semelhante ao DT. O RF é o mais preciso, porém com maiores tempos, e o k-NN é o menos preciso. Ademais, outro fator importante na escolha são as variáveis preditoras. Portanto, os modelos MLP e DT foram selecionados para testes específicos pela precisão, número de variáveis preditoras e tempos. O k-NN e o RF foram descartados, pois apresentaram baixa precisão e maior complexidade computacional, sendo inviáveis em futuras implementações com sistemas embarcados. Os municípios selecionados do Paraná são: Morretes - Cfa e Af; Diamante do Norte - Am; Foz do Iguaçu - Cfa; Nova Fátima - Cfa. A Figura 4 mostra o desempenho durante o inverno dos anos de 2010, 2013, 2016 e 2019, período com menor irradiação solar e alguns dias com precipitação. Estudos e melhoras com relação aos dias com baixo índice de irradiação, nos quais os modelos não são capazes de prever com boa precisão dada a nebulosidade que interfere nas coletas. Nas quatro cidades o MLP teve desempenho superior com relação ao DT. Nos testes semanais em diferentes estações e cidades, o MLP obteve precisão R2 média de 92,99%.

Figura 4 – Previsões de irradiação solar durante a estação inverno



Fonte: Elaborado pelos autores (2023).

CONCLUSÃO

O objetivo deste trabalho foi analisar métodos AM para estimação de irradiação solar ao longo do dia, baseados em dados climatológicos mais acessíveis, como hora, temperatura, umidade e demais. Para isso, foi utilizada uma base do INMET entre 2010 e 2021 em torno de Cornélio Procópio, Paraná. O treinamento dos modelos, segundo hiper-parâmetros específicos, revelou a possibilidade de estimar a irradiação com precisão. Isso estimulou a realização de otimizações de seleção de

variáveis e hiper-parâmetros. Em seguida, testes de validação com amostras desconhecidas e análise de desempenho e precisão demonstraram que os modelos MLP e DT são mais adequados ao objetivo em comparação ao RF e k-NN, considerando precisão e custo computacional por meio do tempo. Em uma segunda etapa de testes semanais ao longo dos anos, estações e condições climáticas, o MLP com variáveis hora, temperatura e umidade apresentou baixos valores de erro, elevados valores de precisão e tempo de teste otimizado. Para futuros trabalhos, os modelos de AM serão otimizados visando a melhor estimação da irradiação solar em períodos de baixa indecência e nebulosidade. Ademais, o desenvolvimento de um sistema embarcado com sensores para a estimação instantânea é uma etapa crucial no auxílio ao progresso eficiente da transição energética.

Agradecimentos

Os autores agradecem à Universidade Tecnológica Federal do Paraná - Campus Cornélio Procópio pela possibilidade do desenvolvimento da presente pesquisa.

Disponibilidade de Código

O código do trabalho desenvolvido está disponível no seguinte link: Acesso ao código.

Conflito de interesse

Os autores declaram que não há conflito de interesse.

REFERÊNCIAS

CHAN, J.Y. et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. **Mathematics**, v. 10, n. 8, 2022.

DUBREUIL, V. et al. Les types de climats annuels au Brésil: une application de la classification de Köppen de 1961 à 2015. **Pôle de recherche pour l'organisation et la diffusion de l'information géographique**, v. 41, p. 1–26, 2017.

FREI, C.; KIM, Y.D. **World Energy Scenarios**. [S.l.], 2019. P. 152.

LIU, Z.; ZHANG, Z. Solar forecasting by K-Nearest Neighbors method with weather classification and physical model. In: NORTH-AMERICAN Power Symposium. Denver, CO, USA: IEEE, 2016. P. 1–6.

RAHUL et al. Solar Energy Prediction using Decision Tree Regressor. In: 5. INT. Conference on Intelligent Computing and Control Systems. Madurai, India: IEEE, 2021. v. 5, p. 489–495.

SALTON, F.G.; MORAIS, H.; LOHMANN, M. Períodos Secos no Estado do Paraná. **Revista Brasileira de Meteorologia**, v. 36, n. 2, p. 295–303, 2020.

SHETTY, D.G. et al. Prediction of Solar Energy Using ML. **International Journal of Scientific Research & Engineering Trends**, v. 7, n. 3, p. 2069–2074, 2021.

TAVARES, K.; ALONSO, A.M.S.; SOUZA, W.A. Sistema de estimação de irradiação solar utilizando grandezas meteorológicas e redes neurais profundas. **Peer Review**, v. 5, n. 14, p. 225–238, jul. 2023.

ZHOU, Y. et al. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. **Energy Conversion and Management**, v. 235, 2021.