

Modelos Perceptron Multicamadas Específicos e Generalista para Estimação de Irradiação Solar

1st Matheus Henrique da Silva
Engenharia de Controle e Automação
Universidade Tecnológica Federal do Paraná
Cornélio Procópio, Brasil
matheussilva.2019@alunos.utfpr.edu.br

2nd Wesley Angelino de Souza
Departamento Acadêmico de Elétrica (DAELE)
Universidade Tecnológica Federal do Paraná
Cornélio Procópio, Brasil
wesleyangelino@utfpr.edu.br

Resumo—A energia solar é essencial para a transição energética, especialmente para sistemas fotovoltaicos, cuja eficiência depende da precisão na estimativa da irradiação solar. Apesar de existirem vários métodos de previsão, muitos não consideram variações climáticas regionais, comprometendo a acurácia. Este estudo desenvolveu e comparou modelos Perceptron Multicamadas (PMC) específicos e generalistas para estimar a irradiação solar diária em Cornélio Procópio, Paraná. Dados históricos (2010-2021) foram analisados e clusterizados com K-Means para diferentes condições climáticas. Modelos PMC foram treinados e testados com dados coletados por um protótipo baseado no sensor DHT22 e no microcontrolador ESP32. O modelo generalista teve melhor desempenho, apresentando menor erro absoluto médio (EAM) e maior coeficiente de determinação (R^2), sendo mais robusto e preciso. Conclui-se que o modelo generalista é preferível para previsões práticas, melhorando o planejamento e a otimização de sistemas fotovoltaicos.

Palavras-chaves—Aprendizado de máquina, Clusterização, Irradiação solar

I. INTRODUÇÃO

A irradiação solar se destaca como o principal parâmetro utilizado no projeto de implementação e eficiência de sistemas de energia elétrica fotovoltaica [4]. O século 21 é marcado por uma crescente conscientização sobre os impactos do uso de combustíveis fósseis nas mudanças climáticas, nesse cenário, crescem os investimentos em pesquisa e desenvolvimento focados na realização de uma transição energética para métodos de geração via recursos renováveis, como o Sol [2].

Dada a forte perspectiva futura de implementação de sistemas de energia fotovoltaica, destacado pelo contínuo aumento da capacidade produtiva com recorde de 266 GW em conjunto com energia eólica, detendo 72% desses acréscimos, em 2022 [3]. Embora diversos métodos existam para prever a irradiação solar, muitos não consideram as variações climáticas específicas das regiões, o que pode comprometer a acurácia das previsões. Métodos como redes neurais e regressão linear são eficientes, mas a abordagem de clusterização para aumentar a precisão pode ser explorada [1]. Este trabalho contribui propondo analisar dados provenientes de estações meteorológicas, realizar a clusterização em diferentes condições climáticas, aplicar modelos de aprendizado de máquina (AM) do tipo Perceptron Multicamadas (PMC) para estimar a irradiação solar ao longo do dia e comparar o desempenho dos diferentes modelos, visando encontrar o com maior acurácia.

II. METODOLOGIA

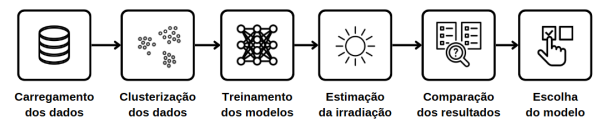


Fig. 1. Diagrama da metodologia utilizada (Autoria própria, 2023).

Carregamento dos dados: Os dados históricos foram obtidos de estações meteorológicas públicas fornecidas pelo Instituto Nacional de Meteorologia (INMET), abrangendo o período de 2010 a 2021. As variáveis climáticas selecionadas foram temperatura e umidade, consideradas essenciais para a estimativa de irradiação solar. Os dados foram analisados para identificar e tratar possíveis valores faltantes, garantindo a qualidade das informações para as etapas seguintes.

Clusterização dos dados: Utilizou-se a técnica de K-Means com a biblioteca `sklearn.cluster.KMeans` do Python para agrupar os dados em diferentes condições climáticas. Foram definidos dois clusters, um com maior umidade e outro com menor umidade, permitindo a identificação de padrões distintos nas condições climáticas da região estudada. A separação dos clusters foi visualizada em gráficos que relacionavam temperatura, umidade, irradiação solar e precipitação.

Treinamento dos modelos: Para o treinamento, foi utilizado o modelo de regressão Perceptron Multicamadas (MLP), uma rede neural artificial. A otimização dos hiperparâmetros foi realizada por meio de busca em grade (`sklearn.model_selection.GridSearchCV`), variando as configurações das camadas escondidas, função de ativação `relu` e otimizador `adam`. Os dados foram divididos em 80% para treinamento e 20% para teste, e as métricas de desempenho foram utilizadas para avaliar a precisão dos modelos.

Estimação da irradiação: A estimativa da irradiação solar foi realizada com dados coletados por um protótipo desenvolvido com o sensor de temperatura e umidade DHT22 e o microcontrolador ESP32. A coleta de dados ocorreu durante 8 semanas na primavera de 2024, permitindo uma análise detalhada das condições climáticas da estação. Os dados foram

enviados para um banco de dados na nuvem para posterior análise.

Comparação dos resultados: Os resultados dos modelos foram comparados utilizando métricas de erro e o coeficiente de determinação (R²). Esta comparação permitiu avaliar a precisão e a acurácia dos diferentes modelos de previsão de irradiação solar, destacando as capacidades dos modelos específicos e do modelo generalista em lidar com as variações climáticas da região.

Escolha do modelo: A escolha do modelo foi baseada na análise das métricas de desempenho, optando-se pelo modelo que demonstrou maior precisão e robustez na estimativa dos valores diários de irradiação solar. Este modelo foi considerado o mais adequado para aplicações práticas, proporcionando estimativas confiáveis para o planejamento e otimização de sistemas de energia fotovoltaica.

III. DESENVOLVIMENTO

A. Carregamento dos dados

As amostras históricas utilizadas foram aglutinadas a partir de dados públicos disponibilizados pelo Instituto Nacional de Meteorologia (INMET), entre os anos de 2010 e 2021 de mais de 600 estações no Brasil. Focando na região de Cornélio Procópio, no Paraná, 100 estações meteorológicas ao redor foram selecionadas para esse trabalho, como mostra a Figura 2. [6] destaca que as variáveis meteorológica mais importante para estimar a irradiação solar é a temperatura e umidade, portanto essas serão as variáveis selecionadas para a clusterização.

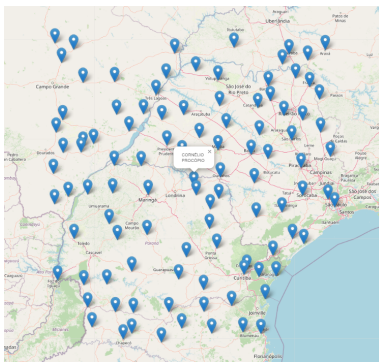


Fig. 2. Estações meteorológicas selecionadas (Autoria própria, 2023).

B. Clusterização dos dados

A clusterização é uma técnica de mineração de dados utilizada para realizar o agrupamento de amostras similares em cluster, as quais não possuem uma classificação prévia definida, o que facilita o processo de identificação de padrões [7]. No contexto de estudos meteorológicos, técnicas de clusterização têm sido amplamente aplicadas para melhorar a precisão das estimações ou previsões [4]. No presente trabalho, o método será realizado com o uso do técnica de K-Médias que utiliza o ponto de dados com base na similaridade com seus vizinhos mais próximos, calculando a distância entre o

ponto de dados e seus K vizinhos mais próximos no espaço multidimensional.

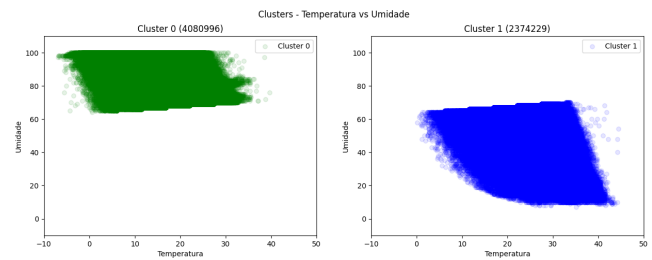


Fig. 3. Resultado da clusterização - Temperatura vs Umidade (Autoria própria, 2023).

A biblioteca em Python `sklearn.cluster.KMeans` oferece funções completas para a utilização da técnica K-Médias, no caso com os parâmetros de número de clusters com 2 unidades e estados aleatórios de 42, utilizando as variáveis de entrada a temperatura ($^{\circ}\text{C}$) e a umidade relativa do ar (%). Na Figura 3 é possível visualizar a separação linear entre as variáveis de clusterização, sendo o Cluster 0, verde, os dados com maior umidade e 4.080.996 amostras e o Cluster 1, azul, os dados de menor umidade e 2.374.229 amostras.

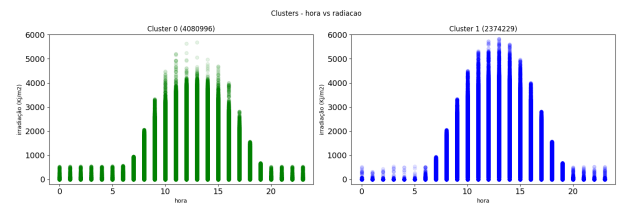


Fig. 4. Resultado da clusterização - Irradiação vs Hora (Autoria própria, 2023).

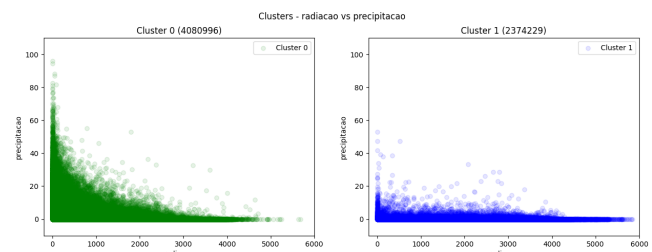


Fig. 5. Resultado da clusterização - Irradiação vs Precipitação (Autoria própria, 2023).

A Figura 4 mostra uma distribuição semelhante na amostragem horária de irradiação solar ao longo das horas do dia e a Figura 5 mostra a maior concentração de elevados valores de precipitação no Cluster 0, de maior umidade relativa do ar.

C. Treinamento dos modelos

A estimação dos valores de irradiação solar, o modelo de regressão perceptron multicamadas (MLP) foi selecionado ??.

Trata-se de uma rede neural artificial direta composta por várias redes neurais básicas, denominadas perceptron, uma unidade básica com uma camada de entradas, na qual recebe os dados, uma camada escondida onde há a presença dos pesos sinápticos de cada entrada e suas dinâmicas internas como bias e limiar de ativação, resultando em um valor na camada de saída a ser verificado pela função de ativação [6].

Para a maior acurácia da rede neural, a determinação dos hiper-parâmetros foi buscada por meio da busca em grade com valores variados das camadas escondidas, por meio da biblioteca `sklearn.model_selection.GridSearchCV`, mostrados na Tabela I.

TABLE I
PARÂMETROS DA BUSCA EM GRADE PARA PERCEPTRON MULTICAMADAS

| Parâmetro | Valores |
|--------------------|--|
| Camadas escondidas | (10, 10), (10, 20), (10, 30), (10, 40), (10, 50), (50, 50), (50, 75), (50, 100), (10, 10, 10), (10, 20, 20), (10, 20, 30), (10, 30, 30), (10, 25, 50), (10, 50, 10), (10, 100, 10) |
| Função de ativação | relu |
| Otimizador | adam |
| Estados aleatórios | 42 |

O modelo MLP para o Cluster 0, com maior umidade, com menor erro quadrático médio negativo possui a configuração de 10, 30 e 30 neurônios na camada escondida, para o Cluster 1, com menor umidade, a configuração de 10, 50 e 10, e para todos os dados a configuração de 10, 25 e 50 neurônios.

TABLE II
RESULTADOS DE ERROS E R2 NO TREINAMENTO

| Métrica | Todos os Dados | Cluster 0 | Cluster 1 |
|---------|----------------|-----------|-----------|
| EAM | 0.1943 | 0.1906 | 0.2794 |
| EQM | 0.1276 | 0.1925 | 0.1820 |
| REQM | 0.3572 | 0.4388 | 0.4266 |
| R2 | 0.8723 | 0.8072 | 0.8179 |

TABLE III
RESULTADOS DE ERROS E R2 NO TESTE

| Métrica | Todos os Dados | Cluster 0 | Cluster 1 |
|---------|----------------|-----------|-----------|
| EAM | 0.1939 | 0.1907 | 0.2798 |
| EQM | 0.1272 | 0.1921 | 0.1828 |
| REQM | 0.3566 | 0.4383 | 0.4275 |
| R2 | 0.8728 | 0.8083 | 0.8171 |

Os dados selecionados foram divididos para treinamento e teste em, respectivamente, 80% e 20% das amostras. Nas Tabelas II e III são mostrados os valores de Erro Absoluto Médio (EAM) mede a média das diferenças absolutas entre os valores previstos e reais, indicando a proximidade das previsões com os valores observados, o Erro Quadrático Médio (EQM) calcula a média dos quadrados das diferenças, penalizando erros maiores mais severamente, a Raiz do Erro Quadrático Médio (REQM), sendo a raiz quadrada do MSE, fornece uma medida de erro na mesma unidade dos dados

originais e o Coeficiente de Determinação (R2), varia de 0 a 1 e indica a proporção da variância dos dados explicita a acurácia do modelo. No geral, os modelos mostram boa precisão, com menor erro nos dados gerais e no Cluster 0, e maior erro no Cluster 1. No treinamento, o EAM para todos os dados é 0.1943, o EQM é 0.1276, o REQM é 0.3572 e o R2 é 0.8723. No teste, esses valores são semelhantes, indicando consistência na performance. O Cluster 1 apresenta maiores valores de erro EAM de 0.2798 no teste, sugerindo maior variabilidade nos dados desse cluster. Os altos valores de R2, maiores que 0.8, indicam que os modelos seguem bem a variabilidade dos dados em todas as horas analisadas.

D. Estimação da irradiação

Após a realização dos treinamentos dos modelos, a estimação da irradiação solar diária é realizada com dados coletados por um protótipo desenvolvido com o sensor de temperatura e umidade DHT22 e o microcontrolador ESP32, responsável pela coleta e envio dos dados para o banco de dados na nuvem. A coleta fora realizada entre as datas de 09/03/2024 e 05/05/2024, totalizando 8 semanas de dados na estação da Primavera, entre o Verão e o Inverno. Para a verificação da acuraria da estimação da irradiação solar, os dados meteorológicos coletados pelo INMET no mesmo período foram coletados.

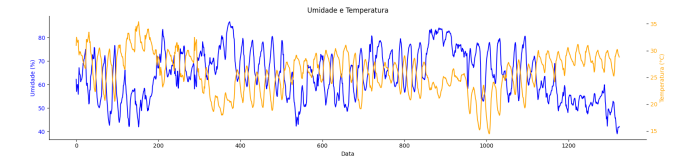


Fig. 6. Dados de temperatura e umidade coletados (Autoria própria, 2023).

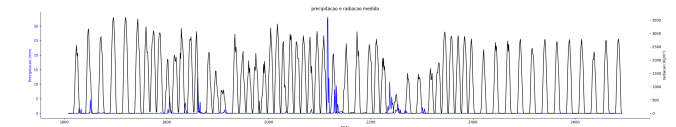


Fig. 7. Dados de irradiação solar e precipitação do INMET (Autoria própria, 2023).

E. Comparação dos resultados

Conforme mostrado na Tabela IV, os resultados das métricas de erro e do R2 indicam a precisão e a robustez dos modelos na estimação da irradiação solar. Para todos os dados, o EAM foi de 0.2207, o EQM de 0.1293, o REQM de 0.3596 e o R2 de 0.8707, demonstrando boa precisão geral do modelo, veja na Figura 8 nas linhas vermelhas. No Cluster 0, os valores foram 0.4031 para o EAM, 0.2746 para o EQM, 0.5240 para o REQM e 0.7254 para o R2, indicando uma precisão moderada, mas menor em comparação com os dados gerais, veja na Figura 8 nas linhas verdes. No Cluster 1, os valores foram 0.6923 para o EAM, 0.6690 para o EQM, 0.8179 para o

REQM e 0.3310 para o R2, sugerindo maior variabilidade e menor precisão na estimação para este grupo de dados, veja na Figura 8 nas linhas azuis.

TABLE IV
RESULTADOS DE ERROS E R2 DA ESTIMAÇÃO

| Métrica | Todos os Dados | Cluster 0 | Cluster 1 |
|-------------|----------------|-----------|-----------|
| EAM | 0.2207 | 0.4031 | 0.6923 |
| EQM | 0.1293 | 0.2746 | 0.6690 |
| REQM | 0.3596 | 0.5240 | 0.8179 |
| R2 | 0.8707 | 0.7254 | 0.3310 |

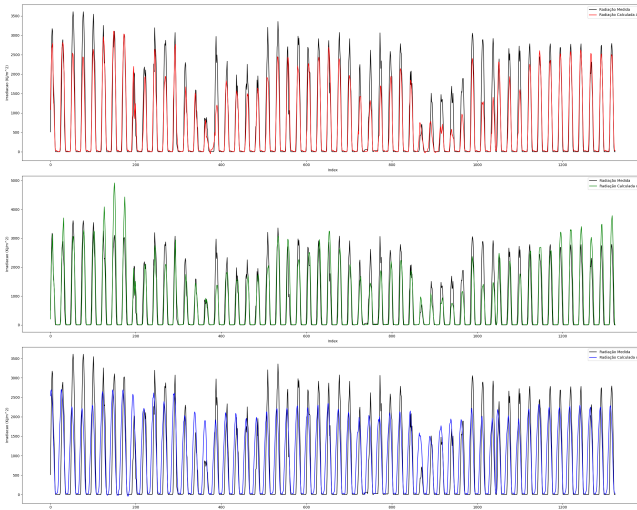


Fig. 8. Estimação da irradiação solar pelos modelos (Autoria própria, 2023).

F. Escolha do modelo

A partir da comparação, o modelo geral foi escolhido pois, embora os modelos dos clusters específicos apresentem variações nos valores de erro, o modelo geral demonstra uma maior capacidade de estimar os valores diários de irradiação solar de forma consistente, tornando-se a escolha preferida para aplicações práticas, dados os valores de erros similares aos de treinamento e teste. [8] Destaca que os modelos combinados com k-means apresentaram maior precisão e estabilidade em comparação com modelos únicos, com apenas modelos de Redes Neurais Artificiais e Aprendizado de Máquina, por exemplo, a combinação de TD-NN com k-means para estimativas diárias e mensais mostrou melhor desempenho em termos de rRMSE em comparação com modelos ARMA e TD-NN isolados, o que mostra que a abordagem para séries temporais de previsão da irradiação solar pode ser mais eficaz do que a estimação.

IV. CONCLUSÃO

Este trabalho apresentou uma análise detalhada de dados históricos de irradiação solar e aplicou modelos de aprendizado de máquina do tipo Perceptron Multicamadas (PMC) para a estimação dos valores diários de irradiação solar na

região de Cornélio Procópio, Paraná. Utilizando dados de estações meteorológicas públicas e técnicas de clusterização por K-Médias, foram desenvolvidos e comparados modelos específicos para clusters e um modelo generalista. A partir dos resultados obtidos, verificou-se que o modelo generalista apresentou melhor desempenho, com menor erro absoluto médio (EAM) e maior coeficiente de determinação (R2), indicando maior precisão e robustez na estimação dos valores de irradiação solar. Esses achados demonstram a eficácia do modelo generalista em captar as variabilidades climáticas e fornecer estimativas precisas, sendo a escolha preferida para aplicações práticas na previsão de irradiação solar. As contribuições deste estudo são relevantes para o desenvolvimento de sistemas de energia fotovoltaica, proporcionando dados mais precisos para o planejamento e otimização da geração de energia solar.

REFERÊNCIAS

- [1] Chauhan, D.; Thakur, J., Data Mining Techniques for Weather Prediction: A Review, International Journal on Recent and Innovation Trends in Computing and Communication, v. 2, n. 8, p. 2184-2189, 2014.
- [2] Frei, C. and Kim, Y.D., World Energy Scenarios 2019, Technical report, World Energy Council, London, United Kingdom, 2019.
- [3] Institute, E., Statistical Review of World Energy 2023, Technical Report 72, 2023.
- [4] Mehdizadeh, S., Behmanesh, J., and Khalili, K., Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation, The Journal of Atmospheric and Solar-Terrestrial Physics (JASTP), vol. 146, pp. 215–227, 2016.
- [5] Silva, I. N. d.; Spatti, D. H.; Flauzino, R. A. Redes neurais artificiais para engenharia e ciências aplicadas: curso prático. São Paulo: Editora, Artliber, 2010
- [6] Silva, M. H. d.; Souza, W. A. d., Análise de Métodos de Aprendizado de Máquina para Estimativa da Irradiação Solar, XXVIII Seminário de Iniciação Científica e Tecnológica da UTFPR, Ponta Grossa, PR, 2023.
- [7] Xu, J.; Lange, K., Power k-Means Clustering, 36th International Conference on Machine Learning, PMLR 97:6921-6931, 2019.
- [8] Zhou, Y., Liu, Y., Wang, D., Liu, X., & Wang, Y. (2021). A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Conversion and Management, 235, 113960.