



Comparação do desempenho dos algoritmos RF e KNN na previsão da energia gerada por quatro tecnologias fotovoltaicas *Comparison of RF and KNN algorithm performance in forecasting energy generated by four photovoltaic technologies*

José Airton Azevedo dos Santos^{1,*}, Aldino Normelio Brun Polo²

¹ Professor do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio, Universidade Tecnológica Federal do Paraná – UTFPR, campus Medianeira, PR, Brasil

² Discente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio, Universidade Tecnológica Federal do Paraná – UTFPR, campus Medianeira, PR, Brasil

*Autor para correspondência, E-mail: airton@utfpr.edu.br

Received: 03 June 2025 | Accepted: 25 July 2025 | Published online: 03 August 2025

Resumo: Esta pesquisa tem como objetivo avaliar e comparar o desempenho de previsão, de geração de energia fotovoltaica, dos algoritmos *Random Forest* (RF) e *K-Nearest Neighbors* (KNN), utilizando uma base de dados com poucas observações. A base de dados, obtida de uma estação solarimétrica localizada na Universidade Tecnológica Federal do Paraná, apresenta, no período entre jan/2020 e dez/2023, séries históricas de geração de energia elétrica das tecnologias de silício monocristalino (m-Si), silício policristalino (p-Si), disseleneto de cobre, índio e gálio (CIGS) e de telureto de cádmio (CdTe). Para avaliar o desempenho, dos modelos RF e KNN, foi utilizada a métrica *Mean Absolute Percentage Error* (MAPE). Os resultados mostraram, para um horizonte de 4 meses, que o modelo *Random Forest* apresentou, para todas as tecnologias, o melhor desempenho.

Palavras-chave: aprendizado de máquina; energia solar; modelos preditivos; previsão de geração; tecnologias fotovoltaicas.

Abstract: This research aims to evaluate and compare the forecasting performance of photovoltaic energy generation using the *Random Forest* (RF) and *K-Nearest Neighbors* (KNN) algorithms, based on a dataset with limited observations. The dataset, obtained from a solarimetric station located at the Federal Technological University of Paraná, includes historical series of electricity generation from monocrystalline silicon (m-Si), polycrystalline silicon (p-Si), copper indium gallium diselenide (CIGS), and cadmium telluride (CdTe) technologies, covering the period from January 2020 to December 2023. To assess the performance of the RF and KNN models, the *Mean Absolute Percentage Error* (MAPE) metric was used. The results showed that, for a 4-month forecasting horizon, the *Random Forest* model outperformed the KNN model across all technologies.

Keywords: machine learning, solar energy; predictive models; generation forecasting; photovoltaic technologies.

1 Introdução

A crescente necessidade mundial de energia, aliada aos efeitos das mudanças climáticas, colocou as fontes de energia renováveis em evidência. A energia solar, entre as fontes de energia renováveis, se destaca como uma excelente opção, graças à sua vasta disponibilidade e capacidade de produção de energia em longo prazo. Esta forma de energia, reconhecida como um recurso sustentável e inesgotável, é um pilar fundamental para futuras estratégias energéticas (Flores et al., 2024; Abdelsattar et al, 2025).

A energia solar fotovoltaica, impulsionada pela redução de custos e pelo foco na diminuição das emissões de gases de efeito estufa, têm se destacado no setor energético mundial. No Brasil, a capacidade instalada de energia solar fotovoltaica, até dezembro de 2023, alcançou 37,3 GW, refletindo um

crescimento exponencial nos últimos anos. Desse total, 25,8 GW correspondem à micro e minigeração distribuída e 11,5 GW à geração centralizada (Nakahata, 2023).

As células fotovoltaicas de silício cristalino e de filmes finos, dentre as diversas tecnologias disponíveis para células fotovoltaicas, são as mais dominantes no mercado. A tecnologia com maior participação é a de silício cristalino, abrangendo tanto o silício monocristalino (m-Si) quanto o silício policristalino (p-Si), responsáveis, desde 2012, por mais de 90% da produção global de módulos fotovoltaicos. A principal diferença entre essas duas tecnologias, além dos aspectos visuais, reside na eficiência de conversão, que influencia diretamente a área requerida para os módulos, e nos coeficientes de temperatura, responsáveis pela redução da eficiência quando os módulos operam sob altas temperaturas (Frauhofer Ise, 2023; Nakahata, 2023).

As células de silício monocristalino são as mais amplamente utilizadas e, geralmente, apresentam maior eficiência. No entanto, seu processo de fabricação é mais complexo e custoso. Por outro lado, as células de silício policristalino possuem eficiência inferior, em comparação com as monocristalinas, mas seu processo de fabricação é menos exigente, resultando em um custo mais baixo. Já as células de telureto de cádmio (CdTe) têm um processo de fabricação mais simples e de menor custo. No entanto, o uso de telureto, um material raro, eleva o preço dessas células, apesar do cádmio ser um elemento abundante. As células de disseleneto de cobre, índio e gálio (CIGS) oferecem eficiência superior às de CdTe, mas inferior às de silício. Quanto ao custo de produção, elas se posicionam entre as células de silício e as de CdTe, sendo mais baratas que as primeiras e mais caras que as últimas (Frauhofer Ise, 2023; Villalva; Gazoli, 2012).

A integração em larga escala de energia fotovoltaica (FV), nas redes elétricas, devido à variabilidade e incerteza da radiação solar, enfrenta grandes desafios. No entanto, os modelos preditivos, ao oferecer previsões confiáveis da geração de energia fotovoltaica, em diferentes escalas de tempo, possibilitam, de maneira mais eficiente, o planejamento, a programação e a operação da rede (Leo et al., 2025).

Vários estudos, apresentados na literatura, utilizaram, para preverem variáveis relacionadas com sistemas fotovoltaicos, modelos de aprendizado de máquina. Vrablecova et al. (2018) propõem o algoritmo *Support Vector Regression* (SVR) para previsão de energia fotovoltaica de curto prazo. Concluíram que o modelo SVR alcançou a mesma precisão que modelos baseados em árvores. Suyambu et al. (2025) utilizaram, na previsão da energia solar no Marrocos, os algoritmos *Artificial Neural Networks* (ANN), *Support Vector Machine* (SVM) e *K-Nearest Neighbours* (KNN). Observaram, por meio dos resultados obtidos deste trabalho, que o algoritmo ANN apresentou o melhor desempenho de previsão. Konstantinou et al. (2021) utilizaram, para previsão da geração de energia solar fotovoltaica, a rede neural recorrente *Long Short-Term Memory* (LSTM). Dados históricos de uma usina fotovoltaica em Nicósia, Chipre, foram usados na previsão da geração de energia fotovoltaica. Observaram, neste trabalho, que o modelo LSTM apresentou um bom desempenho de previsão. Polo e Santos (2025) compararam, na previsão da potência elétrica gerada por um sistema fotovoltaico, os desempenhos dos modelos *Xtreme Gradient Boosting* (XGBoost) e *Light Gradient Boosting Machine* (LightGBM). Observaram, dos resultados de previsão de um, dois e quatro dias, que o modelo XGBoost apresentou, com relação ao modelo LightGBM, um melhor desempenho. Já Zhou et al. (2023) propõem, para previsão de potência de saída de um sistema fotovoltaico, a utilização de um modelo híbrido RF–XGBoost. Resultados experimentais demonstraram que o modelo RF-XGBoost obteve um desempenho de previsão superior aos modelos RF e XGBoost.

Neste contexto, este trabalho tem como objetivo avaliar e comparar o desempenho de previsão, de geração de energia fotovoltaica, dos algoritmos *Random Forest* (RF) e *K-Nearest Neighbors* (KNN). A escolha dos algoritmos *Random Forest* (RF) e *K-Nearest Neighbors* (KNN), em um contexto de poucos dados, são fundamentadas em suas abordagens robustas, que lidam com a escassez de informações e minimizam o risco de *overfitting*. O *Random Forest* combate o superajuste de forma intrínseca, utilizando um método de *ensemble learning* que agrega previsões de múltiplas árvores de decisão, treinadas com subconjuntos aleatórios dos dados. Essa estratégia garante que o modelo não se prenda a particularidades do pequeno conjunto de treinamento, resultando em uma maior capacidade de generalização. Por outro lado, o *K-Nearest Neighbors* (KNN) se destaca por sua simplicidade, focando na similaridade local dos pontos de dados sem construir um modelo complexo. Em um conjunto de dados limitado, a proximidade entre os pontos é uma informação valiosa, e o KNN a utiliza diretamente para fazer previsões, evitando a criação de funções complexas que poderiam levar ao *overfitting*. A comparação entre esses dois algoritmos permite avaliar se a complexidade do *Random Forest* oferece uma vantagem significativa em relação à simplicidade do KNN, proporcionando um *insight* claro sobre a abordagem mais eficaz para a previsão em cenários com dados limitados.

Este estudo, estruturado em diferentes etapas, inclui: a metodologia utilizada na previsão de energia fotovoltaica, os resultados obtidos, nas diferentes tecnologias fotovoltaicas, da aplicação desta metodologia e a análise destes resultados.

2 Introdução

Nesta seção apresentam-se os algoritmos de previsão, a estação solarimétrica, a base de dados, bem como as métricas utilizadas.

2.1 Algoritmos

K-Nearest Neighbors (KNN): O KNN, método simples e não paramétrico, é um algoritmo de aprendizado supervisionado usado para classificação e regressão. Este algoritmo emprega métricas, como a distância Euclidiana, para classificar dados. Ele identifica, para prever uma nova amostra no conjunto de treino, os K vizinhos mais próximos e decide pela classe majoritária (classificação) ou pela média dos valores (regressão). É um algoritmo simples, não paramétrico, mas sensível à escala das variáveis e a dados irrelevantes. Sua principal vantagem é a simplicidade, enquanto sua maior limitação é o custo computacional em grandes bases (Shams, 2024).

Random Forest (RF): O *Random Forest* é um algoritmo, de aprendizado supervisionado, que combina, para melhorar a precisão e reduzir o *overfitting*, múltiplas árvores de decisão. Cada árvore é treinada com uma amostra aleatória dos dados (*bootstrapping*) e usa um subconjunto aleatório de *features* em cada divisão. A previsão final é obtida por votação majoritária (classificação) ou média (regressão) dos resultados individuais das árvores. É um algoritmo robusto que funciona bem com amostras pequenas de dados. O RF fornece boa generalização, atenua o *overfitting* e é menos propenso a *outliers* (Villar et al., 2024).

2.2 Estação EPESOL - Medianeira

A estação EPESOL, localizada no campus Medianeira da UTFPR (Figura 1), possui quatro tecnologias: silício monocristalino (m-Si) de tecnologia PERC com 5,11 kWp, silício policristalino (p-Si) com 4,69 kWp, disseleneto de cobre, índio e gálio (CIGS) com 1,68 kWp e o telureto de cádmio (CdTe) com 1,53 kWp. Este sistema fotovoltaico possui uma estação solarimétrica de padrão SONDA (Figura 2) (Nakahata, 2023).

Observa-se que esta estação de pesquisa foi criada a partir de uma parceria entre a Companhia Paranaense de Energia (COPEL) e o Laboratório de Energia Solar (LABENS) da Universidade Tecnológica Federal do Paraná (UTFPR). A rede de estações faz parte do Projeto P&D ANEEL/COPEL Distribuição - PD 2866-0464/2017 - Metodologia para Análise, Monitoramento e Gerenciamento da Geração Distribuída por Fontes Incentivadas (LABENS, 2025).



Figura 1. Sistema fotovoltaico.



Figura 2. Estação Sonda.

2.3 Base de dados

Os dados gerados, pelo sistema fotovoltaico, são registrados por meio de um sistema de aquisição da Campbell Scientific e enviados, em arquivos no formato CSV, pela internet para um servidor FTP. A base de dados, utilizada para previsão da energia gerada pelas quatro tecnologias, foi formada pela energia gerada por cada tecnologia e a Irradiação Solar (IR). Com 48 observações de cada uma das variáveis (12 meses x 4 anos). A Figura 3 apresenta a energia gerada por cada tecnologia (m-Si, p-Si, CdTe e CIGS).

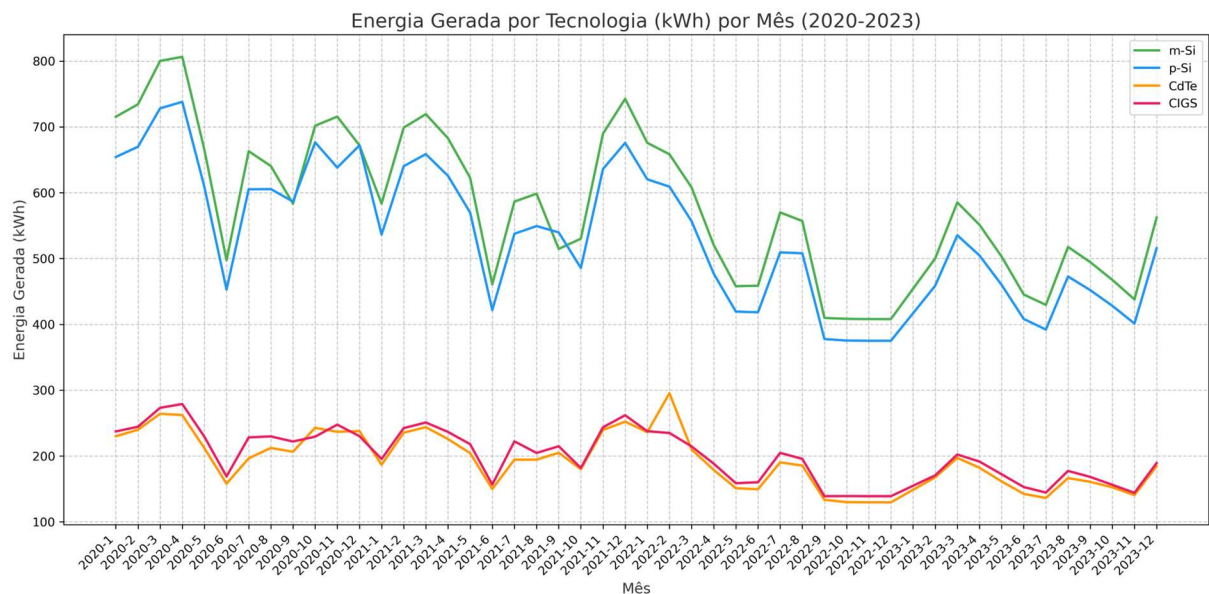


Figura 3. Energia gerada por tecnologia de 2020 a 2023.

Inicialmente, na etapa de pré-processamento, observou-se que a base de dados não continha valores ausentes e *outliers* (Figura 4).

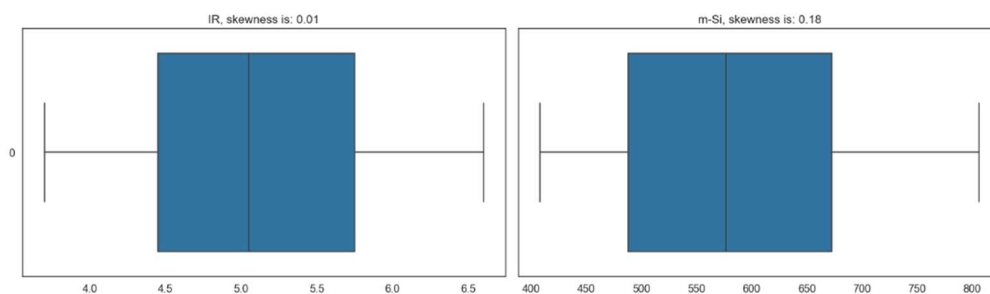


Figura 4. Base de dados sem outliers (IR e m-Si).

Neste trabalho também se estudou a correlação da variável IR com a energia gerada por cada uma das tecnologias. Na Figura 5 apresenta-se, como exemplo, o gráfico de correlação da variável IR com a energia gerada pela tecnologia p-Si.

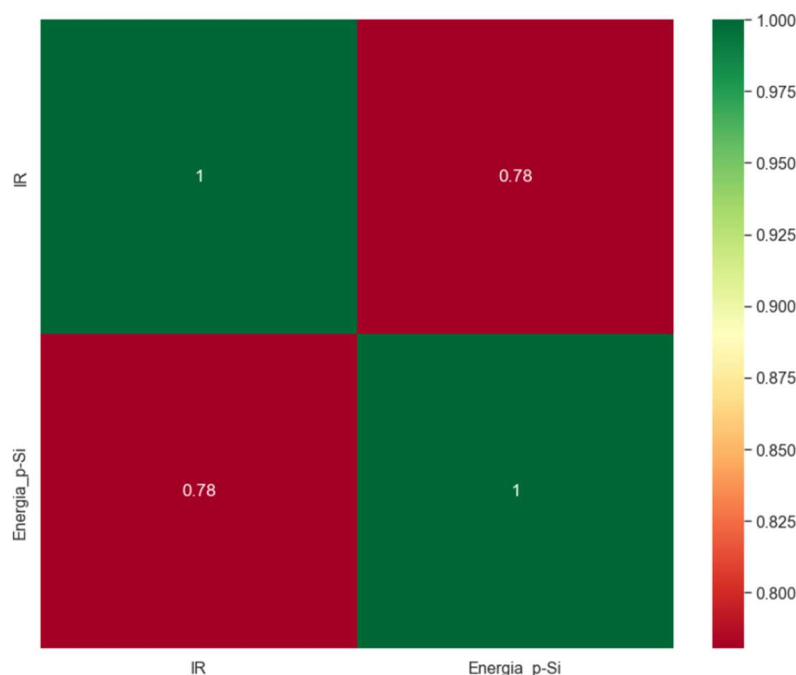


Figura 5. Correlação entre as variáveis IR e Energia_p-Si.

O valor de 0,78 sugere que há uma relação positiva significativa entre IR e Energia_p-Si. Isso significa que, à medida que uma variável aumenta, a outra tende a aumentar também, e vice-versa. Um valor de 0,78 sugere que a irradiação solar (IR) tem um impacto considerável na energia produzida pelo material.

Na sequência, realizou-se uma análise descritiva dos dados de geração (Tabela 1).

Tabela 1. Análise descritiva de dados (kWh).

Medida	m-Si	p-Si	CdTe	CIGS
Média	577.79	529.10	191.47	195.83
Mediana	583.35	536.40	194.50	195.50
Desvio Padrão	112.94	105.96	40.97	40.93
Mínimo	408.00	375.00	129.70	139.00
Máximo	806.40	738.00	295.50	279.00

Observou-se, dos dados apresentados na Tabela 1, que as tecnologias m-si (média - 577.79 kWh, máximo - 806.40 kWh) e p-si (média - 529.10 kWh, máximo - 738.00 kWh) superam, em termos de médias e máximos de energia gerada, as tecnologias CdTe (média - 191.47 kWh, máximo - 295.50 kWh) e CIGS (média - 195.83 kWh, máximo - 279.00 kWh). Isso é um resultado esperado, dado que os painéis monocristalinos e policristalinos têm eficiências de conversão geralmente entre 15-22%, enquanto CdTe e CIGS (filmes fino) ficam na faixa de 10-14%. Embora as tecnologias CdTe e CIGS gerem menos energia,

sua menor variabilidade (desvio padrão baixo - 40.97 kWh e 40.93 kWh) sugere maior consistência em diferentes condições de Irradiação Solar (IR). As tecnologias de filmes finos são menos afetadas por essas variações.

2.4 Métrica

Os modelos, implementados neste trabalho, foram avaliados pela seguinte métrica (Santos, 2021; Santos; Chaucoski, 2020).

Erro Médio Absoluto Percentual (MAPE): esta métrica mede o tamanho do erro em termos percentuais (Eq. 1). O MAPE é uma métrica confiável para comparar, em uma base de dados com poucas observações, o desempenho dos modelos RF e KNN, pois é menos afetada por valores discrepantes. Valor perfeito – MAPE=0.

$$MAPE = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i)/y_i| \times 100 (\%) \quad (1)$$

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

3 Resultados e discussões

3.1 Treinamento

A capacidade de um algoritmo de aprendizado de máquina, de generalizar para novos dados, está diretamente ligada à escolha adequada de seus hiperparâmetros. Esses parâmetros controlam aspectos essenciais do aprendizado, como, por exemplo, o número de árvores de decisão e a profundidade das árvores. Quando mal ajustados, podem resultar em *overfitting* (quando o modelo se adapta excessivamente aos dados de treinamento e perde desempenho em novos dados) ou *underfitting* (quando a aprendizagem é insuficiente e prejudica a acurácia).

Nesse cenário, o Optuna se destaca como uma ferramenta poderosa para otimização de hiperparâmetros. Essa biblioteca automatiza e aprimora o processo de ajuste, com métodos baseados em Otimização Bayesiana e técnicas adaptativas de busca, como o *Tree-structured Parzen Estimator* (TPE). O Optuna, com uma abordagem inteligente e automatizada, facilita, para maximizar o desempenho de modelos de aprendizado de máquina, a identificação das melhores combinações de hiperparâmetros (Srinivas e Katarya, 2022).

O melhor modelo RF, encontrado pelo Optuna, apresenta, para as 4 tecnologias, os seguintes hiperparâmetros (Tabela 2).

Tabela 2. Parâmetros dos modelos - RF.

Hiperparâmetro	Intervalo de Busca	M-Si	P-Si	CdTe	CIGS
<i>n_estimators</i>	[50,1000]	63	63	60	61
<i>max_depth</i>	[2,10]	2	9	8	4
<i>min_samples_leaf</i>	[1,10]	5	6	5	7
<i>min_samples_split</i>	[2,10]	9	6	4	3
<i>max_features</i>	['sqrt', 'log2', 0.5, 0.7, 1.0]	1	1	1	1
<i>ccp_alpha</i>	[1e ⁻⁵ , 0.1]	1.84e ⁻⁴	4.98e ⁻⁴	9.19e ⁻⁵	2.54e ⁻⁴

Já para o modelo KNN obteve-se, para todas as tecnologias, os seguintes hiperparâmetros: *n_neighbors*=5, *weights*=distance, *metric*=Manhattan, *p*=33, *algorithm*=brute e *leaf_size*=14.

Aplicou-se também, para avaliar a consistência dos modelos, a função *RepeatedKfold* da biblioteca *scikit-learn*. Esta função é um gerador de validação cruzada que combina a ideia de *K-Fold* com repetições. Ao repetir o processo de *K-Fold* várias vezes, com diferentes embaralhamentos, se reduz a variância dos resultados da validação cruzada. Isso é particularmente útil para um conjunto de dados pequeno (Palacios et al., 2022).

3.2 Teste

Para testar os desempenhos dos modelos RF e KNN, compararam-se, ao longo de 4 observações mensais (Conjunto de Teste – setembro, outubro, novembro e dezembro de 2023), os valores de energia prevista e observada (Figura 6). Observa-se que as 4 observações, do conjunto de teste, não participaram da etapa de treinamento dos modelos.

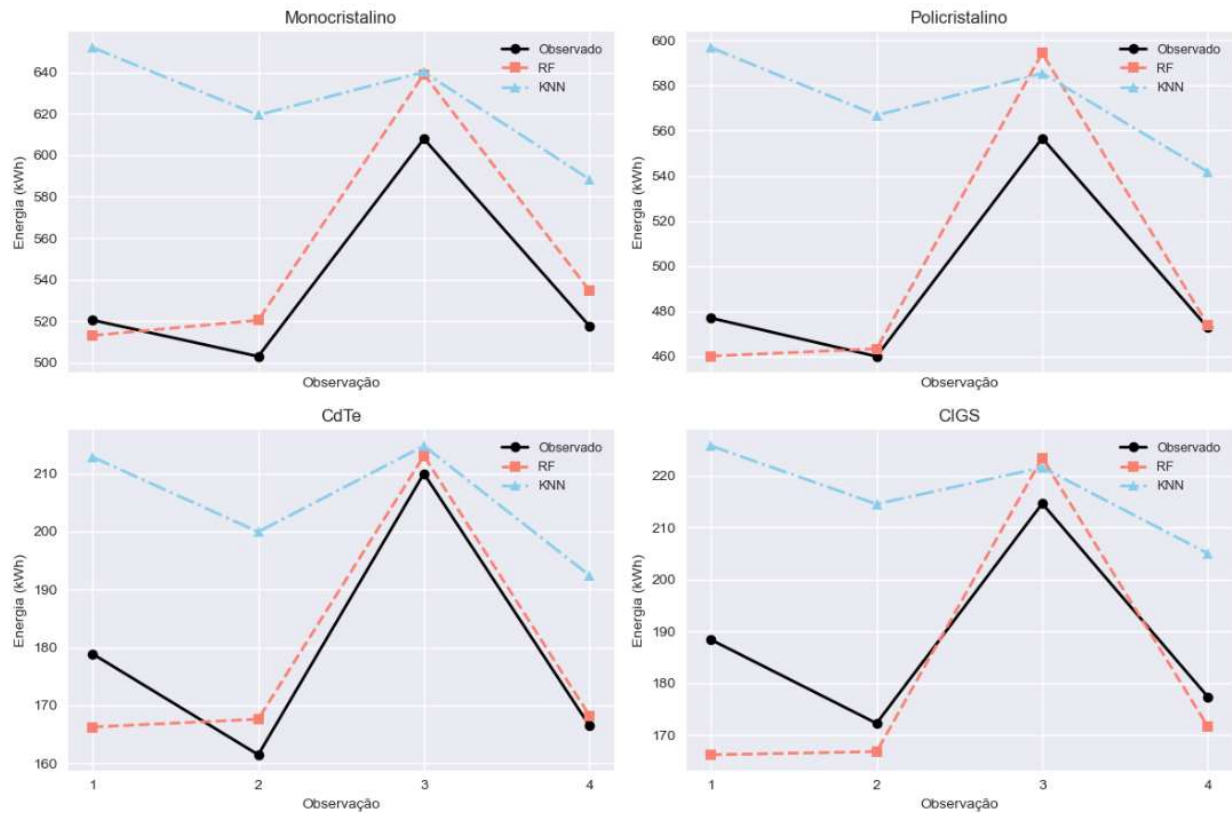


Figura 6. Comparação das curvas por tecnologia: Observado vs. Predições (RF e KNN).

Os gráficos mostram, para tecnologias fotovoltaicas (m-Si, p-Si, CdTe e CIGS), a energia prevista pelos algoritmos (RF e KNN) *versus* a energia observada (real). Observa-se, por meio dos gráficos da Figura 6, que os valores previstos pelo modelo RF estão mais ajustados aos valores observados, indicando maior precisão. Já o modelo KNN apresenta maiores desvios em relação aos valores observados. A discrepância do modelo KNN mostra sua dificuldade em modelar relações complexas.

A análise visual por si só não permite avaliar com precisão a eficiência dos resultados. Portanto, foi calculado, para maior confiabilidade, os valores da métrica MAPE para as quatro tecnologias. Estas métricas avaliam a qualidade do ajuste dos modelos (Tabela 3).

Tabela 3. Resultados da métrica - MAPE.

Tecnologia	RF - MAPE (%)	KNN - MAPE (%)
m-Si	3.21	14.0
p-Si	2.74	14.14
CdTe	3.42	12.72
CIGS	5.96	13.20

Pode-se notar, por meio dos valores apresentados na Tabela 3, que o modelo *Random Forest* (RF) apresentou, para todas as tecnologias, valores de MAPE menores do que os valores do *k-Nearest Neighbors* (KNN). Isso indica previsões mais próximas dos valores reais, ou seja, maior acurácia do modelo RF. O RF, devido à sua capacidade de modelar interações complexas e não lineares entre variáveis, tem desempenho superior ao KNN. O algoritmo KNN, por ser baseado em distâncias, sofre com a baixa densidade de dados (48 observações por tecnologia).

A aleatoriedade na seleção de variáveis e a construção de múltiplas árvores (com *bagging*) permitem que o RF explore diferentes subespaços da base de dados. Essa diversidade de árvores ajuda a mitigar o risco de que uma única árvore suprajuste o ruído. A agregação das previsões suaviza as particularidades de cada árvore, resultando em um modelo mais estável e com menor variância, mesmo com poucos dados.

Na Figura 7 apresentam-se, em termos gráficos, os resultados da métrica MAPE, para as quatro tecnologias.

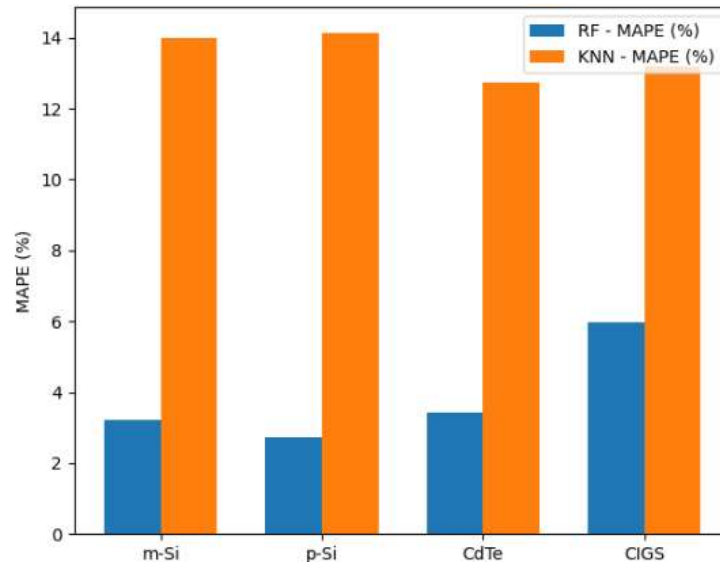


Figura 7. Comparação do MAPE por tecnologia (RF e KNN).

4 Conclusões

Este estudo comparou o desempenho, dos modelos *Random Forest* (RF) e *K-Nearest Neighbors* (KNN), na previsão da geração de energia, para quatro distintas tecnologias fotovoltaicas: m-Si, p-Si, CdTe e CIGS, utilizando a métrica MAPE (*Mean Absolute Percentage Error*).

Os resultados, obtidos neste trabalho, indicam que o modelo *Random Forest*, em comparação com o KNN, é mais adequado para prever, para todas as tecnologias avaliadas, a geração de energia fotovoltaica. Os erros (MAPEs), para o modelo RF, variaram de 2,74% (p-Si) a 5,96% (CIGS), demonstrando, mesmo com pouco dados, sua boa capacidade preditiva. Portanto, para esta base de dados, o modelo RF foi robusto o suficiente para extrair padrões relevantes. Provavelmente devido à sua natureza de *ensemble* que reduz o *overfitting*. Já o modelo KNN apresentou um desempenho inferior, com MAPEs que oscilaram entre 12,72% (CdTe) e 14,14% (p-Si).

Observou-se também, neste trabalho, que a metodologia de otimização de hiperparâmetros, com o Optuna, e a validação cruzada, com o *RepeatedKfold*, forneceram, na obtenção destes resultados, uma base sólida. No entanto, é sempre importante considerar, para aprimorar ainda mais os modelos de previsão, o tamanho do conjunto de dados e a potencial inclusão de outras variáveis.

Agradecimentos

Os autores agradecem a UTFPR pelo apoio e infraestrutura disponibilizada para o desenvolvimento desta pesquisa e ao LABENS pelos dados fornecidos no âmbito do projeto de P&D ANEEL / COPEL-DIS / UTFPR - PD 2866-0464/2017 - Metodologia para Análise, Monitoramento e Gerenciamento da GD por Fontes Incentivadas".

Referências bibliográficas

Abdelsattar M, AbdelMoety A and Eldeen AE (2025) 'Advanced machine learning techniques for predicting power generation and fault detection in solar photovoltaic systems', *Neural Computing and Applications*, vol. 37. <http://doi.org/10.1007/s00521-025-11035-6>.

- Flores JL, Márques CR, Castro ER and Ortega JMP (2024) ‘Solar photovoltaic panel production in Mexico. A novel machine learning approach’, *Environmental Research*, vol. 29. <http://doi.org/10.1016/j.envres.2023.118047>.
- Fraunhofer I (2023) ‘Photovoltaics Report’, Disponível em: www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/Photovoltaics-Report.pdf. (Acesso em: 18 mai. 2025).
- Konstantinou M, Peratikou S and Charalambide AG (2021) Solar photovoltaic forecasting of power output using LSTM networks’, *Atmosphere*, vol. 21, <http://doi.org/10.3390/atmos12010124>.
- LABENS. ‘Energia gerada pelas estações EPESOL’, Disponível em: <https://lookerstudio.google.com/reporting/d4eeef94-0967-45c488f2087e92bf81e3/page/OnwhB?s=rIzDee2c8X0>. (Acesso em: 18 mai. 2025).
- Leo DP, Ciocia A, Malgaroli G and Spertino F (2025) ‘Advancements and challenges in photovoltaic power forecasting: a comprehensive review’, *Energies*, vol. 18. <http://doi.org/10.3390/en18082108>.
- Nakahata BY (2023) *Análise de desempenho de quatro tecnologias aplicadas a sistemas fotovoltaicos conectados a rede elétrica instalados em diferentes campus da UTFPR no Paraná*. Dissertação de Mestrado, Universidade Tecnológica Federal do Paraná, Curitiba.
- Palacios LEO, Guerrero VB and Ordones H (2022) ‘Machine Learning for Solar Resource Assessment Using Satellite Images’, *Energies*, vol. 15. <https://doi.org/10.3390/en15113985>.
- Polo ANB and Santos JAA (2025) ‘Previsão da potência elétrica de um sistema fotovoltaico: uma análise comparativa entre modelos de aprendizado de máquina’, *Revista FSA*, vol. 22. <http://doi.org/10.12819/2025.22.3.9>.
- Santos JA and Chaucoski, Y (2020) ‘Previsão do consumo de energia elétrica na região sudeste: um estudo de caso usando SARIMA e LSTM’, *Revista CEREUS*, vol. 12. <https://10.18605/21757275/cereus.v12n4p93-104>.
- Santos JAA (2021) ‘Aplicação de Redes Neurais na Previsão de Custos de Produção de Frangos de Corte no Estado do Paraná’, *Revista FSA*, vol. 18. <http://doi.org/10.12819/2021.18.6.9>.
- Shams MY, Elshewey AM, El-Kenawy EM, Ibrahim A, Talaat FM and Tarek, Z (2024) ‘Water quality prediction using machine learning models based on grid search method’, *Multimedia Tools and Applications*, vol. 83. <http://doi.org/10.1007/s11042-023-16737-4>.
- Srinivas P and Katarya R (2022) ‘hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost’, *Biomedical Signal Processing and Control*, vol. 73, <http://doi.org/10.1016/j.bspc.2021.103456>.
- Suyambu MR, Vishwakarma PK and Gupta S (2025) ‘Predicting Solar Power Based on Machine Learning Approaches and Their Efficacy’, Disponível em: <https://ieeexplore.ieee.org/document/9917846>. (Acesso em: 20 jun. 2025).
- Villar A, Robledo C and Andrade V ‘Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study’, *Discover Artificial Intelligence*, vol. 4, <http://doi.org/10.1007/s44163-023-00079-z>.
- Villalva MG and Gazoli, JR (2012) ‘Energia solar fotovoltaica - sistemas conectados à rede elétrica’. *Revista o Setor Elétrico*, Disponível em: <https://www.osetoreletrico.com.br/capitulo-x-energia-solar-fotovoltaica-sistemas-conectados-a-rede-eletrica/>. (Acesso em: 20 jun. 2025).
- Vrablecova P, Ezzeddine AB and Rozinajová V (2018) ‘Smart grid load forecasting using online support vector regression’, *Comput. Electr. Eng.*, vol. 65. <https://doi.org/10.1016/j.compeleceng.2017.07.006>.
- Zhou M, Li X and Zhong Z (2023) ‘Research on Photovoltaic Power Generation Power Prediction Based on RF-XGBoost Model’, Disponível em: <https://ieeexplore.ieee.org/document/10393154>. (Acesso em: 20 jun. 2025).