



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS PEDRO II
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

VINÍCIUS MENDES DE MESQUITA

**MODELAGEM PREDITIVA DO CRESCIMENTO DO SETOR DE ENERGIA SOLAR
NO PIAUÍ**

PEDRO II
2026

VINÍCIUS MENDES DE MESQUITA

MODELAGEM PREDITIVA DO CRESCIMENTO DO SETOR DE ENERGIA SOLAR NO
PIAUÍ

Trabalho de Conclusão de Curso (artigo científico) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Pedro II.

Orientador: Prof^o. M^e. Antônio de Alencar Costa Silva.

Coorientador: Prof^a. Maria Hellem Teixeira Abreu.

PEDRO II
2026

MODELAGEM PREDITIVA DO CRESCIMENTO DO SETOR DE ENERGIA SOLAR NO PIAUÍ

Vinícius Mendes de Mesquita¹

Prof^o. M^e. Antônio de Alencar Costa Silva²

Coorientador: Prof^a. Maria Hellem Teixeira Abreu³

RESUMO

As mudanças climáticas e o aumento da emissão dos gases responsáveis pelo efeito estufa têm impulsionado a busca por fontes de energia sustentáveis, a fim de amenizar as mudanças climáticas. No Brasil, a energia solar tem ganhado mais espaço como uma das principais fontes de expansão da matriz elétrica nacional. Nesse contexto, soluções baseadas em *machine learning* vêm sendo desenvolvidas com a finalidade de auxiliar a tomada de decisão sobre as tendências de crescimento do setor de energia solar. O presente trabalho aborda o desenvolvimento de um modelo de *machine learning* capaz de estimar a potência instalada (medida em kW) utilizando dados de geração de energia solar do Piauí. Esta pesquisa pode ser um complemento a outras pesquisas que buscam desenvolver soluções utilizando modelos de *machine learning* para setor de energia solar, como também, servir de referência para pesquisas que buscam implementar *machine learning* em outros contextos de geração de energias renováveis.

Palavras-chave: machine learning; energia solar; modelos de machine learning; support vector machines.

ABSTRACT

Climate change and the rising emission of greenhouse gases have driven the search for sustainable energy sources in order to mitigate climate change. In Brazil, solar energy has gained increasing prominence as one of the main sources driving the expansion of the national electricity grid. Within this context, machine learning-based solutions have been developed to support decision-making regarding growth trends in the solar energy sector. This study addresses the development of a machine learning model capable of estimating installed power capacity (measured in kW) using solar energy generation data from the state of Piauí. This research may serve as a complement to other studies seeking to develop machine learning-based solutions for the solar energy sector, and may also serve as a reference for research aiming to apply machine learning in other contexts of renewable energy generation, thereby contributing to a broader understanding of sustainable energy planning across the country overall.

keywords: machine learning; solar energy; machine learning models; support vector machines.

Data de aprovação: 18/07/2026 (data de apresentação do TCC)

¹ Graduando em Análise e Desenvolvimento de Sistemas. IFPI. caped.2024123tads0038@aluno.ifpi.edu.br

² Graduado em Administração e Matemática; Esp. em Matemática Para o 2º Grau; Mestre em Matemática. IFPI. antonio.alencar@ifpi.edu.br

³ Graduada em Matemática; Esp. em Docência do Ensino Superior; Esp. em Estatística. UFPI. hellemmaria.mhta@gmail.com

1 INTRODUÇÃO

A preocupação crescente com as mudanças climáticas e a necessidade de redução das emissões de gases responsáveis pelo efeito estufa têm impulsionado a busca por novas fontes de energia limpas e sustentáveis. Dentre as fontes de energias sustentáveis, a energia solar se destaca como uma das principais fontes de geração de energia limpa amplamente distribuída pelo mundo (Pourasl; Barenji; Khojastehnezhad, 2023).

No Brasil, a energia solar tem ganhado mais espaço como uma das principais fontes de expansão da matriz elétrica nacional (Bezerra, 2026). As condições geográficas do país favorecem o aproveitamento da energia solar na maior parte de seu território, com as regiões que mais se destacam sendo o Nordeste e Centro-Oeste (Pereira *et al.*, 2017, p. 58-59). Esse potencial tem sido impulsionado em razão dos marcos legais de geração distribuída e da queda dos preços dos equipamentos fotovoltaicos (Bezerra, 2026).

Entre os estados do Nordeste, o Piauí é considerado um dos estados brasileiros com maior potencial para o aproveitamento de energia solar, principalmente a energia solar - devido à intensidade da irradiação do sol em diversas áreas de seu território (Pereira *et al.*, 2017, p. 58-59).

No contexto de geração de energia solar, o principal indicador de crescimento do setor é a potência instalada, medida em *kilowatts* (kW) e suas potências. Essa métrica se refere à capacidade líquida de geração de energia em usinas conectadas à rede elétrica (International Renewable Energy Agency, 2023). No Brasil, a geração de energia solar chegou a atingir 65 GW no ano de 2025, com estados do nordeste brasileiro concentrando 52% dos projetos fotovoltaicos centralizados (Bezerra, 2026).

No contexto de *machine learning*, existem na literatura trabalhos que discutem a aplicação dessas tecnologias na área de fontes de energia renováveis (Munir *et al.*, 2024). Qiu, Liu e Zheng (2024) avaliou a performance dos modelos de machine learning *Random Forest*, *Adaboost*, *ExtraTrees*, *CatBoost* e *XGBoost* para estimativas de geração de energia solar em painéis solares; Sharma *et al.* (2011) aplicaram modelos de machine learning para estimar a geração de energia solar para casas inteligentes, com base em dados meteorológicos; Shaw e Prakash (2019) aplicaram o modelo de *machine learning* SVR para estimar a potência de saída em sistemas fotovoltaicos.

Apesar da variedade de trabalhos que aplicam *machine learning* no setor de energia solar, a maioria dos trabalhos focam em criar modelos para estimar a geração de energia em painéis solares ou a irradiação solar (Nguyen *et al.*, 2025). Sendo assim, existe uma carência de trabalhos

que aplicam *machine learning* para estimar o crescimento do setor energia fotovoltaica.

Dessa forma, este trabalho terá como contribuição o desenvolvimento de um modelo preditivo capaz de estimar o crescimento do setor de energia solar através da medida da potência instalada, utilizando dados geração de energia solar do Piauí.

Neste trabalho, é utilizado dados públicos de geração de energia solar disponíveis nos acervos da ANEEL. Em seguida, foram aplicados processos de pré-processamento no conjunto de dados levantado para o tratamento de dados faltosos. Após a etapa de pré-processamento, são treinados quatro modelos de *machine learning* baseados em SVR (*Support Vector Regression*), sendo cada modelo correspondente aos *kernels Linear, Quadratic, Cubic e RBF*. A escolha do SVR como modelo de regressão justifica-se pela sua utilização no trabalho de Zazoum (2022), que demonstrou bons resultados ao realizar previsões de geração de energia solar em painéis solares, não podendo ser abordado também o GPR por limitações de tempo na realização deste trabalho. O trabalho finaliza com a aplicação de métricas estatísticas R^2 , RMSE e MAE para avaliação de performance dos modelos.

Este trabalho se organiza da seguinte forma: a Seção 2 apresenta os temas que esclarece este estudo, além dos trabalhos que fundamentam essa pesquisa; a Seção 3 apresenta os processos realizados para a realização deste estudo; a Seção 4 apresenta os resultados obtidos ao fim da realização deste estudo; a Seção 5 apresenta as conclusões e trabalhos futuros, além das principais discussões sobre este trabalho.

1.1 OBJETIVOS:

Este trabalho objetiva o desenvolvimento de um modelo preditivo capaz de prever o crescimento do setor de energia solar no estado do Piauí através da medida da potência instalada. Para atingir o objetivo geral, foram elencados os seguintes objetivos específicos:

- Selecionar fontes de dados relevantes para o treinamento dos modelos.
- Realizar tratamento dos dados selecionados.
- Selecionar as *features* mais relevantes para o treinamento dos modelos.
- Treinar diferentes modelos de *machine learning* com os dados adquiridos.
- Aplicar métricas de acurácia para selecionar o modelo mais preciso.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 GERAÇÃO DE ENERGIA SOLAR NO CENÁRIO NACIONAL

A utilização de energia solar no cenário nacional tem sido motivada por ela ser uma das principais fontes de energia sustentáveis no quesito de geração de energia elétrica em larga escala, e isso se deve ao fato de que ela se utiliza da irradiação emitida pelo sol como fonte energética principal, sendo adequada para muitas regiões do país (Guangul; Chala, 2019).

No Brasil, a adoção dessa fonte de energia tem crescido de forma acelerada, no ano de 2024, a matriz de produção elétrica teve um aumento de 10,9 GW (Gigawatts) em potência instalada, sendo 51,87% desse total proveniente de fontes fotovoltaicas (Brasil, 2025). Entre os estados da federação, o estado do Piauí destaca-se como tendo um grande potencial para a geração de energia fotovoltaica por se situar em uma área com grande índice de irradiação solar (Pereira *et al.*, 2017). Nas estimativas da expansão da matriz elétrica nacional, o Piauí contribuiu com a geração de 1.299 MW (Megawatts) apenas de fontes de energia fotovoltaica (Brasil, 2025).

A instalação de sistemas fotovoltaicos no Brasil é atualmente regida pela norma ABNT NBR 16690:2019 que regulamenta os procedimentos de instalação elétricas de arranjos fotovoltaicos, incluindo também normas de segurança próprias para esses sistemas. A norma define diretrizes como a aplicação de medidas de proteção contra choques elétricos, para segurança em operação de arranjos fotovoltaicos, além de isolamento dupla ou reforçada para componentes do lado em corrente contínua de sistemas fotovoltaicos (Associação Brasileira de Normas Técnicas, 2019). Essas orientações normativas contribuem para a padronização de sistemas fotovoltaicos no país.

Nesse contexto, surgem soluções que aplicam modelos de *machine learning* para resolver problemas relacionados ao setor energético solar. Sharma *et al.* (2011) desenvolveu modelos preditivos para estimar a geração de energia solar utilizando dados climáticos da National Weather Service (NWS).

2.2 MODELOS DE MACHINE LEARNING NO GERENCIAMENTO ENERGÉTICO SOLAR

A adesão de *machine learning* no setor de energia fotovoltaica tem ganhado destaque pela sua capacidade de processar grandes quantidades de dados e realizar previsões de curto a longo prazo (Zazoum, 2022). No ramo da energia solar, modelos de *machine learning* vêm

sendo empregados para solucionar problemas relacionados à geração de energia e predição de irradiação solar. Obando, Carvajal e Agudelo (2019) realiza uma revisão da literatura sobre o uso de *machine learning* para estimativa de irradiação solar, outros trabalhos como o de Sharma *et al.* (2011) aplicam modelos para realizar predições da geração de energia solar com base em dados climáticos.

Na literatura, as abordagens mais utilizadas para previsão no contexto de energia solar em séries temporais são o ARIMA (Autoregressive Integrated Moving Average) e o SVR (Obando; Carvajal; Agudelo, 2019). A primeira abordagem, ARIMA, é um modelo estatístico que costuma ser aplicado em séries temporais para descobrir padrões sazonais, tendência e ruídos, apresentando melhor desempenho quando os dados possuem comportamento de variação linear, também possuindo menor custo computacional comparados a algoritmos de redes neurais (Obando; Carvajal; Agudelo, 2019). A segunda abordagem, SVR, é aplicado para solucionar problemas de regressão. Considerado computacionalmente mais demandante que o ARIMA, ele demonstra melhor desempenho em problemas que envolvem regressão não linear (Obando; Carvajal; Agudelo, 2019).

2.3 TRABALHOS RELACIONADOS

Sharma *et al.* (2011) desenvolveu um modelo preditivo que permite estimar a geração de energia solar para casas inteligentes, permitindo aos moradores realizarem planejamentos avançados de geração e consumo de energias em suas residências. Eles utilizaram dados meteorológicos disponibilizados pela NWS (*National Service of Meteorology*) com fonte de dados principal, correlacionando as métricas meteorológicas adquiridas das previsões com a intensidade solar. Na etapa do treinamento dos modelos, eles utilizaram linear regression of minimum squares e SVR.

Zazoum (2022) abordou o treinamento dos modelos de *machine learning* SVR e GPR (*Gaussian Process Regression*) para a predição de geração de energia solar. Ele considerou os parâmetros como temperatura do painel, temperatura ambiente, fluxo solar, hora do dia e umidade relativa como *features* para o treinamento do modelo. A acurácia dos modelos treinados foi estimada utilizando a métrica RMSE (*Root Mean Squared Error*), com o modelo GPR demonstrando melhor performance dentre os dois.

Shaw e Prakash (2019) aplicaram o modelo de *machine learning* SVR para estimar a potência de saída em sistemas fotovoltaicos. O autor considerou fatores como pressão, temperatura

e umidade. O modelo foi treinado utilizando o modelo SVR com os *kernels* *Linear*, *Quadratic*, *Cubic* e *RBF*. Por fim, a acurácia dos modelos treinados foi estimada utilizando as métricas R^2 e RMSE.

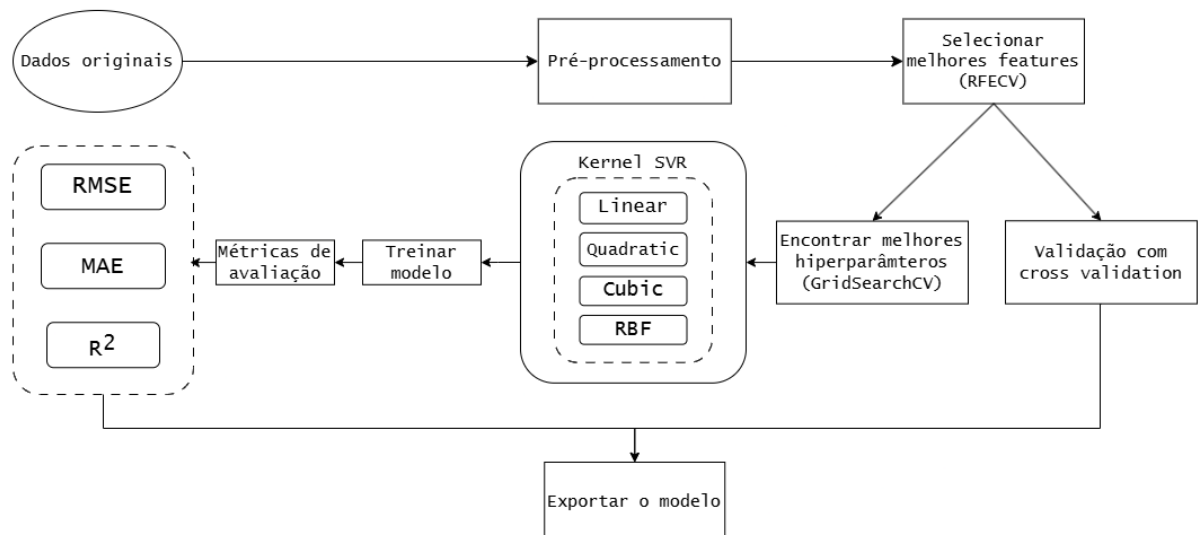
3 METODOLOGIA

Este trabalho de pesquisa caracteriza-se como uma pesquisa exploratória e de estudo de caso que aplica *machine learning* em séries temporais. Especificamente, foi desenvolvido um modelo preditivo capaz de estimar o crescimento do setor de energia solar no estado do Piauí através da medida da potência instalada utilizando dados históricos de geração de energia solar do estado.

3.1 ABORDAGENS DE MACHINE LEARNING:

Este trabalho aplica o modelo de *machine learning* SVR para estimar a potência instalada em uma série temporal. Zazoum (2022) propõe um fluxo de trabalho para treinamento de modelos de *machine learning* por regressão, sendo utilizado neste trabalho uma versão adaptada deste fluxo exibido na Figura 1.

Figura 1 – Fluxo de desenvolvimento



Fonte: produzido pelo autor (2026)

O fluxo do trabalho inicia com a aplicação de técnicas de pré-processamento nos dados, como tratamento de dados faltosos, remoção de inconsistências e seleção de *features* mais relevantes para resolução do problema. Em seguida, é realizada uma busca das melhores *features*

para o treinamento utilizando RFECV (*Recursive Feature Elimination with Cross-Validation*). Após isso, o fluxo se divide em duas etapas paralelas: a primeira realiza a otimização dos hiperparâmetros para cada *kernel* com o *GridSearchCV*; a segunda procede com a aplicação da validação cruzada a fim de obter um diagnóstico de estabilidade dos modelos. Após o treinamento, os modelos serão avaliados utilizando métricas estatísticas RMSE, MAE e R^2 para avaliar as suas capacidades preditivas. Por fim, será feita uma comparação dos resultados obtidos comparando a eficácia de cada um dos modelos gerados, permitindo identificar qual modelo apresenta melhor desempenho geral e, por essa razão, o mais adequado para utilização em casos reais (Zazoum, 2022). Esse fluxo de trabalho foi apresentado por Zazoum (2022), e adaptado para as finalidades específicas deste trabalho.

O SVR é uma variação do modelo SVM, que é utilizado para resolver problemas de classificação (Zazoum, 2022). O SVR tem como principal ideia determinar uma função que tenha uma variação n dos vetores reais de treinamento de dados e requer que seja o mais suave possível (Zazoum, 2022). Neste trabalho, será abordado o modelo SVR com os *kernels Linear, Quadratic, Cubic* e RBF.

3.2 FERRAMENTAS DE TRABALHO:

Neste trabalho, os experimentos foram realizados em uma plataforma *Windows 11*, AMD, *Ryzen(TM) 5 7600 3.8 GHz* (6 CPUs), 16 GB RAM. Foram utilizados também *Python 3.14.1* e *Scikit-learn 1.7.2* para a implementação e avaliação dos modelos de regressão. Para importação e tratamento dos dados, foi empregado a biblioteca *Pandas 2.3.3*. Também foi empregada a biblioteca *Matplotlib 3.10.9* para criar gráficos estatísticos. O *Scikit-learn* é uma biblioteca *open-source* de *machine learning* para a linguagem de programação *Python* que oferece um vasto conjunto de algoritmos de ML para classificação, regressão, clusterização, dentre outros (Flexa; Gomes; Viademonte, 2019). A biblioteca *Pandas* pode se integrar com o *Scikit-learn* para consumir dados de *datasets* e realizar pré-processamento (Flexa; Gomes; Viademonte, 2019).

3.3 COLETA DOS DADOS:

Os dados utilizados no desenvolvimento do projeto correspondem aos dados históricos de geração de energia solar do estado do Piauí entre 2015 e 2025, que correspondem aos dados mais recentes de geração de energia solar do estado. Esses dados serão coletados nos acervos da Agência Nacional de Energia Elétrica (ANEEL) que disponibiliza *datasets* com dados

históricos nacionais de geração de energia solar (Costa *et al.*, 2023). Será utilizado o *dataset* de microgeração e minigeração distribuída (MMGD) da ANEEL, disponível em um único arquivo CSV.

3.4 PRÉ-PROCESSAMENTO:

O tratamento do *dataset* utilizado iniciou-se com a seleção de *features*. Do total 33 (trinta e três) das *features* disponíveis no *dataset* original, foram selecionadas 7 (sete) com valor preditivo para o modelo, além de um atributo alvo para regressão. Após isso, foi aplicado um filtro para selecionar apenas as linhas que correspondam, simultaneamente, à geração fotovoltaica (UFV), à Piauí (PI) e ao intervalo de 2015 a 2025. Procedeu-se realizando a conversão de todos as *features* com tipos não numéricos em representações numéricas. O tratamento de valores ausentes foi realizado seguindo os seguintes critérios de acordo com a natureza e importância de cada *feature*:

- **Colunas obrigatórias:** MdaPotenciaInstaladaKW, SigUF e AnmPeriodoReferencia que possuam valores NaN foram removidas.
- **Colunas opcionais:** QtdUCRecebeCredito, DscPorte, DscClasseConsumo, SigTipoConsumido com valores NaN foram substituídos por, respectivamente, "0", "Desconhecido", "Não Informado" e "NI".

Após o tratamento de valores ausentes, foram criadas as *features* temporais *mes*, *trimestre* e *tendencia* com a finalidade de capturar sazonalidade na série temporal e a tendência de crescimento da série. Também foi criada uma *feature* chamada *crescpct* que representa a taxa de variação percentual mensal da potência instalada acumulada. Na sequência, realizou-se a agregação das instâncias individuais por período de referência, transformando o conjunto de dados originais em um série temporal de granularidade mensal. O *dataset* antes dessa etapa possuía 85985 linhas de dados brutos, sendo reduzido após o agregamento para 126 linhas. Por fim, foram criadas 12 (doze) *lags features* de defasagem temporal, referentes aos 12 meses anteriores de potência instalada. Estas *features* possuem a finalidade de fornecer ao modelo SVR o histórico recente da série, tornando possível aprendizado autorregressivo.

3.5 TREINAMENTO DO MODELO:

Neste projeto, foi utilizado o modelo SVR com os *kernels* *Linear*, *Quadratic*, *Cubic* e *RBF*. A primeira etapa do processo de treinamento do modelo envolveu a divisão dos conjuntos

de teste e treino, sendo optada uma escala 80/20 (80% dos dados destinados para treino e 20% para testes). As *features* passaram por um processo de normalização com RobustScaler. Após isso, foi realizada a seleção das *features* com maior relevância utilizando o RFECV. Por fim, foi aplicado o algoritmo *GridSearchCV* para buscar os melhores hiperparâmetros para o conjunto de treino, finalizando com treinamento do modelo. Ainda foi aplicado uma validação cruzada com o *3-fold TimeSeriesSplit*, a fim de validar a estabilidade dos modelos treinados.

3.6 AVALIAÇÃO DE PERFORMANCE:

Para avaliar a acurácia dos modelos de *machine learning* utilizados, foram aplicadas as seguintes métricas de avaliação: *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) e o coeficiente de determinação (R^2). Zazoum (2022) define matematicamente essas métricas da seguinte maneira:

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (z_j - \hat{z}_j)^2}{N}} \quad (1)$$

$$MAE = \frac{\sum_{j=1}^N |z_j - \hat{z}_j|}{N} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{j=1}^N (\hat{z}_j - z_j)^2}{\sum_{j=1}^N (\hat{z}_j - \hat{z}_{ave})^2} \quad (3)$$

onde z_j , $\hat{z}_j$ e $\hat{z}_{ave}$ representam o dado previsto, o dado alvo e a média do dado alvo, respectivamente, sendo N o número de dados.

4 RESULTADOS E DISCUSSÕES

4.1 PRÉ-PROCESSAMENTO:

O Quadro 1 mostra as *features* com valor preditivo selecionadas na etapa de tratamento do *dataset*.

Na última etapa de tratamento, foi realizado uma transformação no *dataset* para adaptá-lo em uma série temporal. O Quadro 2 mostra as *features* do *dataset* após essa etapa.

Quadro 1 – Features do dataset após pré-processamento

Feature	Tipo	Descrição
DthAtualizaCadastralEmpreend	numérico	Data do cadastro do empreendimento no sistema MMDG da ANEEL.
SigUF	categórico	Sigla da Unidade Federativa do empreendimento.
DscPorte	categórico	Porte do empreendimento conforme a ANEEL.
DscClasseConsumo	categórico	Classe de consumo da unidade ao empreendimento.
SigTipoConsumido	categórico	Tipo jurídico do titular do empreendimento.
QtdUCRecebeCredito	numérico	Quantidade de unidades consumidoras beneficiária dos excedentes de energia gerados por essa UC (Unidade Consumidora).
SigTipoGeracao	categórico	Abreviação do tipo de geração.
MdaPotenciaInstaladaKW	numérico	Potência elétrica ativa nominal instalada no empreendimento, expressa em kilowatts (kW).

Fonte: produzido pelo autor (2026)

Quadro 2 – Features do dataset após a etapa final do pré-processamento

Feature	Tipo	Descrição
QtdEmpreendimentos	numérico	Quantidade de unidades consumidoras beneficiárias dos créditos de energia gerada pelo empreendimento.
PctResidencial	numérico	Proporção de empreendimentos classificados como consumidores residenciais em relação ao total de instalações do mês.
PctMicrogeracao	numérico	Proporção de empreendimentos classificados como microgeração em relação ao total de instalações do mês.
PctPessoaFisica	numérico	Proporção de empreendimentos cujo titular é pessoa física em relação ao total de instalações do mês.
lag_1 ... lag_12	numérico	Valores defasados da variável target (MdaPotenciaInstaladaKW) para os 12 meses anteriores ao período de referência.

continua...

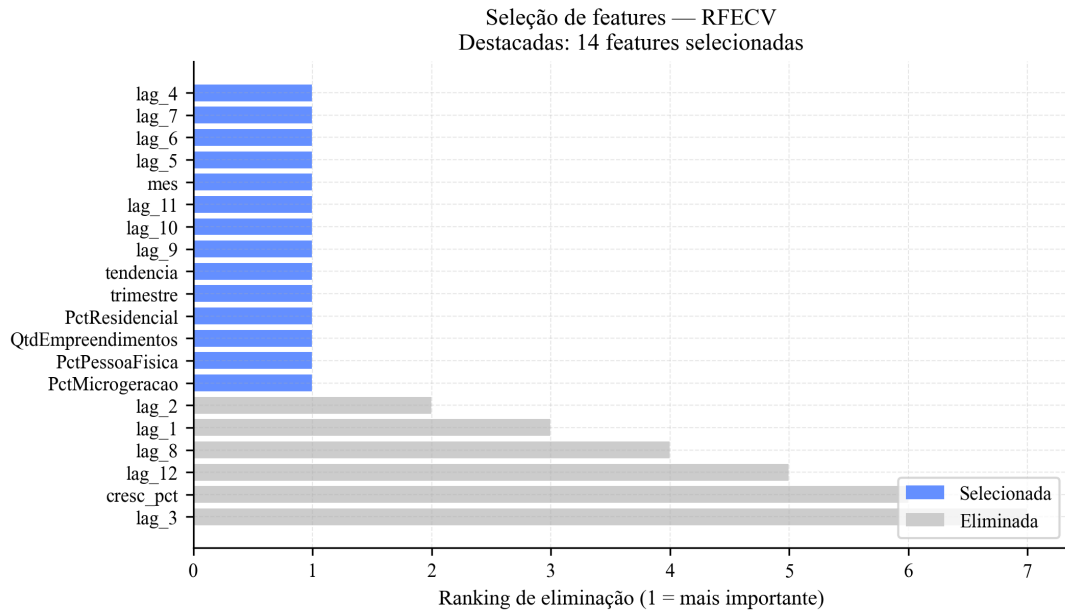
(continuação)

Feature	Tipo	Descrição
mes	numérico	Número do mês do ano correspondente ao período de referência.
trimestre	numérico	Trimestre do ano correspondente ao período de referência.
tendencia	numérico	Índice sequencial crescente representando a posição temporal de cada observação na série histórica.
cres_pct	numérico	Taxa de variação percentual da potência instalada em relação ao mês anterior, calculada como a variação relativa defasada em um período.
MdaPotenciaInstaladaKW	numérico	Target do modelo. Somatório da potência elétrica ativa nominal instalada em kW de todos os empreendimentos solares fotovoltaicos do Piauí em cada mês de referência.

Fonte: produzido pelo autor (2026)

Na etapa da seleção de *features*, a importância de cada *feature* foi determinada utilizando RFECV junto ao SVR com *kernel Linear*, sendo decidido manter somente as *features* com importância igual a 1. A Figura 2 mostra as 14 *features* selecionadas.

Figura 2 – Importância de cada feature



Fonte: produzido pelo autor (2026)

4.2 RESULTADOS DOS TESTES:

A Tabela 1 mostra o resultados das métricas de avaliação do modelo SVR para cada *kernel* no conjunto de testes. O *kernel* que demonstrou maior acurácia foi o RBF com $R^2 = 0.8632$, RMSE = 1,436 kW e MAE = 1,355 kW.

Tabela 1 – Métricas de acurácia por kernel

Kernel	R^2	RMSE(kW)	MAE(kW)
Linear	0.6149	2,409	2,129
Quadratic	0.8629	1,437	1,261
Cubic	0.7765	1,835	1,584
RBF	0.8632	1,436	1,355

Fonte: produzido pelo autor (2026)

Os *kernels* RBF e Quadratic demonstraram resultados de desempenho muitos próximos, com pequenas diferenças nos coeficientes de determinação ($R^2 = 0.8632$ e $R^2 = 0.8629$, respectivamente). Na métrica MAE, o *kernel* Quadratic demonstrou ser ligeiramente superior (MAE = 1.261 kW), em comparação ao RBF (MAE = 1.355 kW). No entanto, após a aplicação de uma validação cruzada temporal com 3-fold *TimeSeriesSplit*, notou-se uma diferença significativa em questão de estabilidade entre os dois resultados. O *kernel* RBF obteve R^2 médio de 0.7093, enquanto o *kernel* Quadratic apresentou R^2 médio de 0.4678, demonstrando maior instabilidade.

O *kernel Linear*, por sua vez, obteve R^2 médio de 0.6867, enquanto o *kernel Cubic* apresentou o pior desempenho entre todos os avaliados, obtendo R^2 médio de 0.0799.

Os resultados obtidos demonstram a superioridade da performance do *kernel RBF* dentre as demais configurações. Dentre os modelos treinados, o modelo com o *kernel RBF* demonstrou maior R^2 (0,8632) em comparação aos demais *kernels*; Nas outras métricas, o *kernel* apresentou o menor RMSE (1,436 kW) e segundo menor MAE (1,261 kW), sendo superado apenas pelo *kernel Quadratic*, que também apresentou resultados muito próximos ao RBF em todas as métricas. Essa diferença marginal foi contestada durante a aplicação da validação cruzada com o *3-fold TimeSeriesSplit*, com os novos resultados apresentando um R^2 médio de 0,7093 para o *kernel RBF* e 0,4678 para o *kernel Quadratic*, representando uma diferença de 0,2415 entre os resultados. Por fim, o *kernel Linear* também se demonstrou ser uma alternativa viável ao apresentar um R^2 médio de 0,6867 na validação cruzada, demonstrando uma consistência e capacidade de generalização semelhante ao *kernel RBF*.

Os resultados indicam que o *kernel RBF* demonstrou maior acurácia e melhor capacidade de generalização que os outros *kernels* testados. Essa afirmação é coerente com os resultados obtidos no trabalho de Sharma *et al.* (2011) que constatou o desempenho superior do *kernel RBF* na predição de irradiação solar em painéis solares.

4.3 AMEAÇAS A VALIDAÇÃO:

O trabalho foi desenvolvido considerando um único modelo de *machine learning* (SVR) utilizando quatro *kernels* diferentes. Além disso, foi utilizado somente o *dataset* microgeração e minigeração distribuída (MMGD) da ANEEL, que não cobre usinas de geração centralizadas cadastradas no SIGA (Sistema de Informações de Geração), além de instalações ainda não regularizadas. Sendo assim, o seguinte trabalho limita-se ao arcabouço regulatório vigente. Outras questões que podem ameaçar a validade dessa pesquisa são abordadas a seguir:

Durante o desenvolvimento, foram tomadas um conjunto de decisões com a finalidade de melhorar a acurácia do modelo final, decisões como: número de folds no *TimeSeriesSplit* (*3-fold TimeSeriesSplit*), o número de *lags features* (12), a escolha do escalador (*RobustScaler*). Todas essas decisões foram tomadas com base em testes empíricos observando os resultados de testes anteriores. Essa abordagem, segundo Cawley e Talbot (2010), introduz um viés na seleção nas métricas finais, fazendo que os resultados possam ser considerados irreais.

O modelo foi treinado utilizando um número consideravelmente reduzido de amostras

(126) em proporção ao número de *features* (14 após a seleção). Essa desproporção afeta diretamente o desempenho de modelos baseados em regressão, como o SVR, que performam melhor quando o número de dados aumenta. Estudos como o de Cerqueira, Torgo e Soares (2022) demonstram essa relação entre o crescimento do número de dados e o aumento da precisão de modelos de machine learning.

5 CONCLUSÕES E TRABALHOS FUTUROS:

Este trabalho desenvolveu um modelo preditivo capaz de estimar o crescimento do setor de energia solar através da medida da potência instalada. Foram treinados quatro modelos SVR com os *kernels* *Linear*, *Quadratic*, *Cubic* e *RBF*, seguindo com a aplicação de métricas estatísticas robustas para avaliar a performance dos modelos treinados. Com a finalidade de validar com mais precisão os resultados anteriores, foi aplicada uma validação cruzada com *3-fold TimeSeriesSplit*.

Conclui-se a partir dos resultados que, dentre os modelos avaliados, o SVR com o *kernel* RBF demonstrou melhor desempenho em relação aos outros *kernels*, superando o *kernel Quadratic* nas métricas R^2 e RMSE. Apesar do *kernel Quadratic* ter apresentado resultado ligeiramente superior na métrica MAE, a aplicação da validação cruzada com o *3-fold TimeSeriesSplit* resultou em um R^2 médio ligeiramente inferior ao *kernel* RBF, que demonstrou maior estabilidade. O *kernel Cubic* apresentou uma precisão muito baixa, enquanto o *kernel Linear* apresentou resultados muito próximos ao *kernel* RBF na validação cruzada. A análise por validação cruzada reforçou a confiabilidade na estabilidade do *kernel* RBF, demonstrando que o desempenho no conjunto de testes trata-se de uma métrica mais otimista para o desempenho real do modelo.

Acredita-se que os resultados desse trabalho podem contribuir com a literatura em pesquisas que visam aplicar *machine learning* para auxiliar na tomada de decisões sobre as tendências de crescimento do setor de energia solar, ao demonstrar a viabilidade da aplicação do modelo SVR para essa finalidade. Como trabalhos futuros, propõe-se a expansão do escopo geográfico para inclusão de outros estados do Nordeste brasileiro e a aplicação de outros modelos de regressão como *Random Forest Regressor*, *XGBoost Regressor*, *Adaboost Regressor*.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 16690:2019**: Instalações elétricas de arranjos fotovoltaicos — requisitos de projeto. Rio de Janeiro: Associação Brasileira de Normas Técnicas (ABNT), 2019.

BEZERRA, F. D. ENERGIA SOLAR: v. 11, n. 430, maio de 2026. **Caderno Setorial ETENE**, Fortaleza, v. 11, p. 5–6, jun. 2026. Disponível em: <https://www.bnb.gov.br/revista/cse/article/view/3392>. Acesso em: 19 ago. 2026.

BRASIL. Agência Nacional de Energia Elétrica. **RALIE – Acompanhamento da Expansão da Oferta de Geração**. Brasília, DF: Agência Nacional de Energia Elétrica, 2025. Disponível em: <https://portalrelatorios.aneel.gov.br/Ralie>. Acesso em: 18 ago. 2026.

CAWLEY, G. C.; TALBOT, N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. **Journal of Machine Learning Research**, Norwich, United Kingdom, v. 11, p. 2079–2107, 2010. Disponível em: <https://jmlr.org/papers/volume11/cawley10a/cawley10a.pdf>. Acesso em: 18 ago. 2026.

CERQUEIRA, V.; TORGO, L.; SOARES, C. A case study comparing machine learning with statistical methods for time series forecasting: size matters. **Journal of Intelligent Information Systems**, Porto, Portugal, v. 59, p. 415–433, 2022. DOI: <https://doi.org/10.1007/s10844-022-00713-9>. Disponível em: <https://link.springer.com/article/10.1007/s10844-022-00713-9>. Acesso em: 18 ago. 2026.

COSTA, V. B. F. *et al.* Assessment of the status and trends of photovoltaic distributed generation in brazil: An in-depth approach based on big data processing. **Solar Energy: The Official Journal of International Solar Energy Society**, Itajubá, MG, v. 249, p. 694–711, 2023. DOI: <https://doi.org/10.1016/j.solener.2022.12.018>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0038092X22008891?via%3Dihub>. Acesso em: 18 ago. 2026.

FLEXA, C.; GOMES, W.; VIADEMONTTE, S. An exploratory study on machine learning frameworks. In: WORKSHOP EM DESEMPENHO DE SISTEMAS COMPUTACIONAIS E DE COMUNICAÇÃO, 2019., 2019, Belém. **Anais [...]**. Porto Alegre, RS: Sociedade Brasileira de Computação, 2019. p. 1–5. DOI: <https://doi.org/10.5753/wperformance.2019.6476>. Disponível em: <https://sol.sbc.org.br/index.php/wperformance/article/view/6476>. Acesso em: 19 ago. 2026.

GUANGUL, F. M.; CHALA, G. T. Solar energy as renewable energy source: Swot analysis. In: INTERNATIONAL CONFERENCE ON BIG DATA AND SMART CITY (ICBDSC), 4., 2019, Bali, Indonesia. **Proceedings[...]**. Piscataway, NJ: IEEE, 2019. p. 1–5. DOI: <https://doi.org/10.1109/ICBDSC.2019.8645580>. Disponível em: <https://ieeexplore.ieee.org/document/8645580>. Acesso em: 18 ago. 2026.

INTERNATIONAL RENEWABLE ENERGY AGENCY. **Renewable Capacity Statistics 2023**. Abu Dhabi: International Renewable Energy Agency, 2023. p. 21–29. Disponível em: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Mar/IRENA_RE_Capacity_Statistics_2023.pdf. Acesso em: 18 ago. 2026.

MUNIR, M. S. *et al.* Machine learning in renewable energy systems. In: INTERNATIONAL CONFERENCE ON ELECTRICAL, COMPUTER, COMMUNICATIONS AND MECHATRONICS ENGINEERING (ICECCME), 4., 2024, Malé, Maldives. **Proceedings [...]**. Piscataway, NJ: IEEE, 2024. p. 1–5. DOI: <https://doi.org/10.1109/ICECCME62383.2024.10796132>. Disponível em: <https://ieeexplore.ieee.org/document/10796132>. Acesso em: 18 ago. 2026.

NGUYEN, H. N. *et al.* Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms. **PLOS ONE**, Kuala Lumpur, v. 20, n. 1, p. 1–23, 2025. DOI: <https://doi.org/10.1371/journal.pone.0315955>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0315955>. Acesso em: 18 ago. 2026.

OBANDO, E. D.; CARVAJAL, S. X.; AGUDELO, J. P. Solar radiation prediction using machine learning techniques: A review. **IEEE Latin America Transactions**, v. 17, n. 4, p. 684–697, 2019. DOI: <https://doi.org/10.1109/TLA.2019.8891934>. Disponível em: <https://ieeexplore.ieee.org/document/8891934>. Acesso em: 18 ago. 2026.

PEREIRA, E. B. *et al.* **Atlas Brasileiro de Energia Solar**. 2. ed. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. 80 p. Disponível em: <https://repositorio.unifesp.br/items/3a0c7b2a-a107-4cc9-9449-f08db316e47e>. Acesso em: 18 ago. 2026.

POURASL, H. H.; BARENJI, R. V.; KHOJASTEHNEZHAD, V. M. Solar energy status in the world: A comprehensive review. **Energy Reports**, v. 10, p. 3474–3493, 2023. DOI: <https://doi.org/10.1016/j.egyr.2023.10.022>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352484723014579?via%3Dihub>. Acesso em: 18 ago. 2026.

QIU, Y.; LIU, Y.; ZHENG, X. Research on solar photovoltaic panel power generation prediction and comparative analysis based on multiple machine learning algorithms. In: INTERNATIONAL CONFERENCE ON ENERGY, POWER AND ELECTRICAL TECHNOLOGY (ICEPET), 3., 2024, Chengdu, China. **Proceedings [...]**. China: IEEE, 2024. p. 83–86. DOI: <https://doi.org/10.1109/ICEPET61938.2024.10625860>. Disponível em: <https://ieeexplore.ieee.org/document/10625860>. Acesso em: 19 ago. 2026.

SHARMA, N. *et al.* Predicting solar generation from weather forecasts using machine learning. In: IEEE INTERNATIONAL CONFERENCE ON SMART GRID COMMUNICATIONS, 2., 2011, Brussels, Belgium. **Proceedings [...]**. Piscataway, NJ: IEEE, 2011. p. 528–533. DOI: <https://doi.org/10.1109/SmartGridComm.2011.6102379>. Disponível em: <https://ieeexplore.ieee.org/document/6102379>. Acesso em: 19 ago. 2026.

SHAW, S.; PRAKASH, M. Solar radiation forecasting using support vector regression. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING AND COMMUNICATION ENGINEERING, 6., 2019, Sathyamangalam, India. **Proceedings [...]**. Piscataway, NJ: IEEE, 2019. p. 1–4. DOI: <https://doi.org/10.1109/ICACCE46606.2019.9080008>. Disponível em: <https://ieeexplore.ieee.org/document/9080008>. Acesso em: 19 ago. 2026.

ZAZOUM, B. Solar photovoltaic power prediction using different machine learning methods. In: INTERNATIONAL CONFERENCE ON POWER AND ENERGY SYSTEMS ENGINEERING, 8., 2022, Fukuoka, Japan. **Proceedings [...]**. Al Khobar, Saudi Arabia: Elsevier, 2022. p. 19–25. DOI: <https://doi.org/10.1016/j.egyr.2021.11.183>. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S2352484721013287?via%3Dihub>. Acesso em: 19 ago. 2026.