

Previsão da Curva de Potência de Aerogeradores Através de Redes Neurais Artificiais: Uma Abordagem Baseada em Dados SCADA

Forecasting the Wind Turbine Power Curve Using Artificial Neural Networks: A SCADA-Data-Based Approach

Vinicius de Lira Teixeira

Universidade Federal de Alfenas - UNIFAL-MG
viniciuslira.teixeira@sou.unifal-mg.edu.br

Leandro Ferreira

Universidade Federal de Alfenas - UNIFAL-MG
leandro.ferreira@unifal-mg.edu.br

Luciene Resende Gonçalves

Universidade Federal de Alfenas - UNIFAL-MG
luciene.goncalves@unifal-mg.edu.br

Resumo

A energia eólica destaca-se como uma das fontes renováveis de crescimento mais acelerado, tanto no cenário global quanto no Brasil. Um dos desafios dessa geração é a estimativa da energia produzida ao longo da vida útil de um parque eólico, considerando não apenas a variabilidade inerente à velocidade do vento, mas também outros fatores operacionais. A curva de potência, que relaciona a velocidade do vento à potência elétrica gerada, constitui uma ferramenta crucial para avaliar o desempenho dos aerogeradores. Neste trabalho, propõe-se o uso de redes neurais artificiais (RNAs) para a previsão da curva de potência com base em dados operacionais obtidos do sistema SCADA. Para cada conjunto de dados dos parques eólicos analisados, foi gerado um modelo distinto, mantendo-se os mesmos parâmetros de configuração. O modelo desenvolvido utiliza uma arquitetura composta por três camadas ocultas e emprega o otimizador Adam, visando capturar a relação não linear entre as variáveis de entrada e saída. A aplicação do modelo aos dados dos três parques eólicos localizados na região sul do Brasil demonstrou uma alta capacidade preditiva, evidenciada pelos coeficientes de determinação, mesmo diante das variações e da dispersão dos dados operacionais. Esses resultados reforçam o potencial dos métodos baseados em aprendizado de máquina para o monitoramento e a otimização da geração de energia eólica, contribuindo para o avanço das práticas de manutenção preditiva e gestão da energia renovável.

Palavras-chave

Energia eólica, curva de potência, redes neurais artificiais

Abstract

Wind energy stands out as one of the fastest-growing renewable sources, both globally and in Brazil. One of the challenges with this type of energy production is estimating the energy generated over the lifetime of a wind farm, taking into account not only the inherent variability of wind speed but also other operational factors. The power curve, which relates wind speed to the generated electrical power, is a crucial tool for evaluating the performance of wind turbines. In this work, the use of artificial neural networks (ANNs) is proposed for predicting the power curve based on operational data obtained from the SCADA system. For each dataset from the analyzed wind farms, a distinct model was generated while maintaining the same configuration parameters. The developed model utilizes an architecture composed of three hidden layers and employs the Adam optimizer, with the aim of capturing the nonlinear relationship between the input and output variables. The application of the model to data from the three wind farms located in southern Brazil demonstrated a high predictive capability, as evidenced by the determination coefficients, even in the presence of variations and dispersion in the operational data. These results reinforce the potential of machine learning-based methods for monitoring and optimizing wind energy production, thereby contributing to the advancement of predictive maintenance practices and renewable energy management.

Keywords

Wind energy, power curve, Artificial Neural Network

1 Introdução

De acordo com o [International Energy Institute \(2024\)](#), a matriz elétrica global é predominantemente composta por fontes de energia não renováveis. A demanda por eletricidade vem crescendo rapidamente em todo o mundo há décadas, com um aumento de 2,5% em 2023 em relação ao ano anterior, atingindo o recorde de 29.925 TWh. Segundo [Sohoni et al. \(2016\)](#), a crescente preocupação em atender à demanda por energia elétrica, bem como os requisitos de sustentabilidade e a redução das emissões de gases de efeito estufa na geração de energia, tem impulsionado o crescimento da participação das fontes renováveis. Em consonância, o [International Energy Institute \(2024\)](#) ratifica que a participação das fontes renováveis em 2022 era de 29% e, em 2023, alcançou 30%.

Dentre as fontes renováveis, a energia eólica tem se mostrado uma alternativa extremamente promissora. Segundo o [Global Wind Energy Council \(2024\)](#), as novas instalações de parques eólicos no mundo superaram 110 GW em 2023. A participação da América Latina foi de 5%, com o Brasil liderando as novas instalações na região, sendo responsável por mais de 3/4 dos novos parques eólicos em 2023. Além disso, o Brasil é um dos dez países com potência instalada superior a 14 GW, ocupando atualmente a sexta posição no ranking global de capacidade instalada de energia eólica *onshore*.

Segundo [Associação Brasileira de Energia Eólica e Novas Tecnologias \(2024\)](#), atualmente existem mais de 1000 parques eólicos em operação no Brasil, superando os 30 GW de potência instalada. Em consenso, [Empresa de Pesquisa Energética \(2024\)](#) reitera que a energia eólica foi responsável por 13,2% de toda a oferta de energia elétrica no Brasil em 2023, ficando atrás somente da fonte hidráulica. No entanto, a geração de energia pela fonte eólica é intermitente, tendo em vista que a potência disponível está diretamente associada à velocidade do vento local, que está sujeita a variações sazonais e diárias. Evidentemente, é fundamental prever e monitorar o desempenho dos aerogeradores, e uma ferramenta muito utilizada para verificar o desempenho de um aerogerador é a curva de potência em função da velocidade do vento ([Silva, 2020](#); [Teixeira, 2022](#)). Neste trabalho, é proposta uma abordagem simplificada para determinar a curva de potência a partir de dados experimentais, obtidos diretamente do SCADA (*Supervisory Control And Data Acquisition*) por meio de uma rede neural artificial (RNA).

2 Fundamentos teóricos

A curva de potência de um aerogerador é determinada pela relação entre a velocidade do vento que incide na altura do cubo do rotor e a potência elétrica gerada, em geral fornecida em catálogo pelo respectivo fabricante do aerogerador. No entanto, a curva de potência pode variar de acordo com o modelo do aerogerador, bem como de acordo com o fabricante ([Teixeira, 2022](#)). As curvas de potência também podem variar em função da abordagem de controle adotada pelo fabricante; na Figura 1 é exibida a curva de potência típica de um aerogerador que adota o sistema de controle por passo.

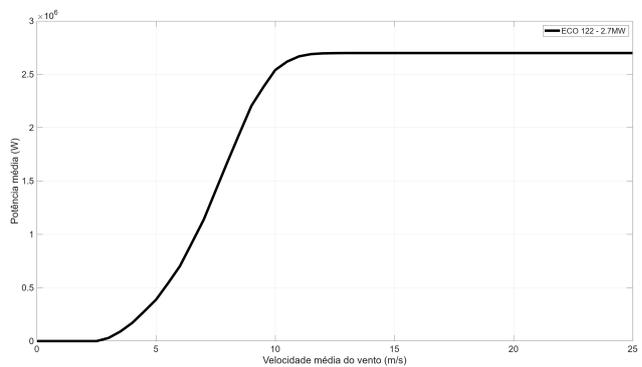


Figura 1: Curva de potência de catálogo adaptado de: ([Bauer & Matysik, 2025](#))

Com o intuito de simplificar o funcionamento do sistema de controle do aerogerador, a curva de potência ideal pode ser simplificada em 4 regiões, conforme elucidado na Figura 2.

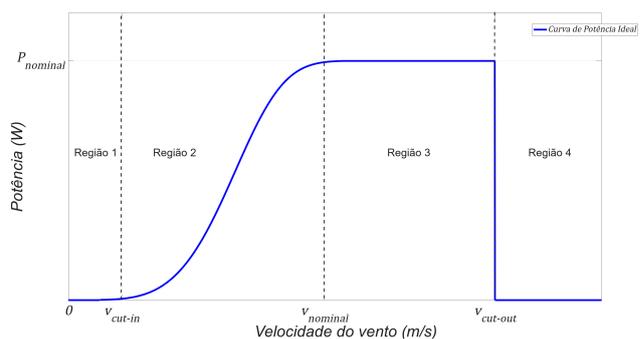


Figura 2: Curva de potência segmentada em regiões adaptado de: ([Teixeira, 2022](#))

Em certas condições, a velocidade do vento que atinge as pás da turbina eólica não é suficiente para que o sistema entre em funcionamento, seja por aspectos técnicos ou econômicos. A Região 1 da curva de potência é delimitada por ventos com velocidade maior que zero e inferior à velocidade de partida (v_{cut-in}). Formalmente, a potência extraída na Região 1 é zero. No

entanto, é importante ressaltar que, na prática, nessa região o aerogerador não gera energia, mas consome para manter todo o sistema em funcionamento. Eventualmente, quando a velocidade do vento que incide no aerogerador se torna excessivamente alta, o sistema é desligado para preservar a integridade do equipamento, travando o eixo do rotor da turbina eólica. Essa região é denominada Região 4 e, assim como ocorre na Região 1, a potência gerada é nula, sendo matematicamente expressa pela seguinte equação:

$$P(v) = 0 \quad (1)$$

A Região 2 é uma região de transição caracterizada pela maximização da conversão de energia cinética do vento, onde a potência é proporcional ao cubo da velocidade do vento e é nessa região que o sistema visa maximizar a extração dessa energia. A Região 2 é delimitada pela velocidade de partida ($v_{\text{cut-in}}$) e pela velocidade nominal (v_{nominal}). A relação entre a velocidade do vento, na altura do cubo do rotor, e a potência elétrica gerada por uma turbina eólica pode ser expressa matematicamente pela equação:

$$P(v) = \frac{1}{2} \rho A C_p v^3 \quad (2)$$

Onde

- $P(v)$ retrata a potência em W em função da velocidade do vento v ;
- ρ é a densidade do ar em kg/m^3 ;
- A simboliza a área varrida pelas pás do rotor em m^2 ;
- C_p representa o coeficiente de potência um adimensional;
- v corresponde a velocidade do vento em m/s .

A Região 3 é caracterizada pela extração da potência nominal (P_{nominal}) da turbina eólica e está delimitada pela velocidade nominal (v_{nominal}) e pela velocidade de corte ($v_{\text{cut-out}}$). Dentro desse intervalo, a turbina opera em sua capacidade máxima, formalmente representada pela equação:

$$P(v) = P_{\text{nominal}} \quad (3)$$

2.1 Modelos da curva de potência

Nas últimas décadas, pesquisadores têm se dedicado ao aprimoramento e à formulação de modelos para curvas de potência (Wang et al., 2019). Embora esses modelos possam ser classificados de diversas maneiras, a ênfase tem sido maior nos modelos determinísticos, tanto paramétricos quanto não paramétricos, sendo que o método das classes também merece destaque. Uma classificação abrangente e detalhada está ilustrada na Figura 3.

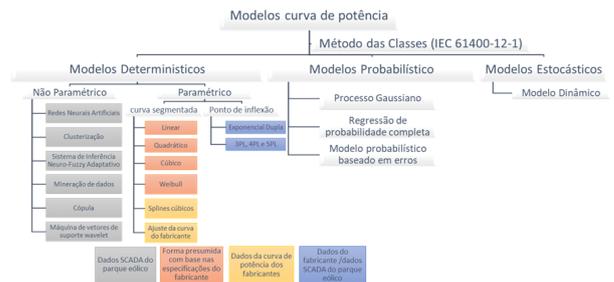


Figura 3: Modelos para se obter a curva de potência adaptado de (Sohoni et al., 2016) e (Wang et al., 2019)

2.1.1 Método das classes IEC-61400-12-1

A *International Electrotechnical Commission* (IEC) desempenha um papel central na definição de normas internacionais para as tecnologias elétricas e eletrônicas. Atualmente, o método mais amplamente aceito para medir o desempenho de aerogeradores, bem como para a obtenção de suas curvas de potência, está delineado na norma IEC 61400-12-1 (Teixeira, 2022). A abordagem baseia-se na hipótese de uma curva de potência não perturbada, denominada "curva de potência de turbulência zero", que reflete o comportamento teórico ideal de uma turbina eólica. Essa premissa implica que o aerogerador segue, continuamente, a mesma curva de potência, considerando relevantes as variações da velocidade do vento ao longo de um intervalo de 10 minutos. Com base nessa curva de referência, torna-se possível extrapolar o desempenho da turbina para diferentes níveis de intensidade de turbulência (International Electrotechnical Commission, 2022).

De forma geral, o método das classes proposto pela IEC requer, para a obtenção da curva de potência, os dados médios de 10 minutos relativos à velocidade do vento, à potência, ao desvio padrão da velocidade do vento e aos parâmetros necessários para o cálculo da densidade

dade do ar. Tais dados são agrupados em intervalos de 0,5 m/s, denominados *bins*, que organizam os valores medidos por faixas de velocidade, permitindo a análise da potência média correspondente a cada classe. Em seguida, procede-se à normalização dos dados, visando corrigir as variações na densidade do ar e assegurar a representatividade da curva de potência sob diferentes condições ambientais, utilizando a densidade do ar medida e a potência gerada pelo aerogerador. Posteriormente, realiza-se a normalização da turbulência, ajustando os valores de potência de acordo com a intensidade da turbulência local, por meio de um modelo estatístico que minimiza seu impacto na curva obtida. Por fim, determina-se a curva de turbulência zero, ou seja, a curva de potência ideal do aerogerador, livre dos efeitos da turbulência, utilizando regressões e ajustes baseados nos dados já normalizados, o que resulta em uma representação mais precisa do desempenho teórico da turbina em condições ideais (Teixeira, 2022).

2.1.2 Modelos paramétricos

As técnicas paramétricas fundamentam-se na resolução de expressões matemáticas com o objetivo de modelar a potência efetivamente gerada por turbinas eólicas (Lydia et al., 2014). Essa linha de pesquisa pode ser subdividida em duas abordagens: a primeira apresenta modelos que incorporam um ponto de inflexão, onde a curvatura da curva de potência se altera; a segunda segmenta a curva de potência em diferentes intervalos. Ambas as abordagens permitem a otimização dos parâmetros de cada segmento por meio de técnicas como algoritmos genéticos, enxame de partículas, algoritmos de polinização de flores, dentre outras metaheurísticas. Além disso, alguns modelos paramétricos oferecem interpretações físicas para cada parâmetro, enquanto outros adotam uma abordagem mais abstrata, baseando-se na forma funcional da curva de potência sem necessariamente associá-la a aspectos físicos ou de controle (Teixeira, 2022).

2.1.3 Modelos não paramétricos

Modelos não paramétricos não assumem previamente uma equação matemática para o fenômeno analisado, sendo amplamente utilizados quando há um grande volume de dados e dificuldades na aplicação de modelos paramétricos (Teixeira, 2022). Os modelos não paramétricos mais utilizados são técnicas de aprendizado de máquina não linear, pois permitem maior flexibilidade na análise sem a necessidade de presu-

mir uma equação previamente, e mesmo assim é possível captar a relação entre a velocidade do vento e a potência (Wang et al., 2019).

As redes neurais artificiais (RNAs), inspiradas no sistema nervoso biológico, replicam aspectos da inteligência humana ao aprenderem relações não lineares entre conjuntos de dados de entrada e saída, por meio da utilização de funções de ativação em seus neurônios ocultos (Haykin, 2007). Em uma análise comparativa realizada, Lydia et al. (2013) propôs um modelo de rede neural artificial de retropropagação que utiliza uma única camada oculta composta por 5 neurônios, otimizando o treinamento por meio do método de *Levenberg–Marquardt*. A abordagem proposta foi comparada com diferentes modelos paramétricos e não paramétricos aplicados a cinco conjuntos de dados distintos, demonstrando desempenho superior entre as metodologias não paramétricas. Em contrapartida, Jang et al. (1997) ressalta que as RNAs possuem uma natureza “caixa-preta”, dificultando a interpretação dos significados inerentes a cada neurônio e aos respectivos pesos, o que limita a compreensão dos mecanismos internos do modelo. Além disso, Sohoni et al. (2016) sugere que as RNAs são especialmente indicadas para a previsão da geração de energia eólica, contribuindo na fase de dimensionamento, no monitoramento online e na predição de grupos de turbinas, o que evidencia sua aplicabilidade prática, mesmo diante das limitações inerentes à sua natureza interpretativa.

A modelagem de curvas de potência por clusterização tem sido realizada, de maneira geral, por meio de técnicas baseadas em lógica *fuzzy* de centros de *clusters*, agrupamento *fuzzy c-means* e métodos de agrupamento subtrativo, os quais quantificam o grau de pertencimento das amostras aos respectivos clusters (Wang et al., 2019). Adicionalmente, métodos clássicos como *K-means* e *K-medoids* são empregados para aprimorar a construção de modelos capazes de captar o comportamento das turbinas eólicas em condições operacionais normais (Yesilbudak, 2018). Enquanto os métodos baseados em lógica *fuzzy*, *c-means* e *K-means* exigem a definição prévia do número de *clusters*, o método subtrativo permite determinar esse número de forma adaptativa, com base nos dados fornecidos (Wang et al., 2019). Uma análise comparativa apontou que a abordagem *fuzzy* de centros de *clusters* produz um RMSE significativamente inferior ao dos modelos de regressão polinomial ajustados pelo método dos mínimos quadrados (Sohoni et al., 2016).

O Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS) combina as principais características dos sistemas *fuzzy* e das redes neurais artificiais. Essa abordagem tem sido amplamente utilizada para o monitoramento online da curva de potência, por sua capacidade de incorporar variáveis como a direção do vento e a temperatura ambiente, resultando em menores erros de previsão (Sohoni et al., 2016).

A mineração de dados consiste na extração de conhecimento a partir de grandes volumes de informações, e seus avanços têm se mostrado promissores para a modelagem de curvas de potência de turbinas eólicas em monitoramento online. A seleção adequada do método e do algoritmo de mineração é essencial para a obtenção de curvas precisas, estáveis e robustas. Entre as abordagens mais empregadas estão o *Perceptron* Multicamadas (MLP), *Random Forest*, M5P, *Boosting Tree* e *k-Nearest Neighbour* (k-NN). A análise comparativa do desempenho desses modelos para monitoramento em tempo real indicou que o método k-NN se destaca por sua alta fidelidade, tornando-o adequado como curva de referência (Sohoni et al., 2016).

Uma cópula é uma função de distribuição utilizada para descrever a dependência entre variáveis aleatórias Lydia et al. (2014). Ao contrário dos modelos tradicionais de curvas de potência, que estabelecem uma relação explícita entre a velocidade do vento e a potência gerada, os modelos baseados em cópulas focam na aprendizagem da distribuição conjunta dessas variáveis Wang et al. (2019). Com base em dados SCADA de velocidade do vento e potência, Gill et al. (2012) propuseram um modelo de cópula para a obtenção de curvas de potência. O estudo demonstra que essa abordagem capta de forma eficaz a complexa dependência entre as variáveis, evidenciando sua aplicabilidade no monitoramento operacional das turbinas e na detecção precoce de falhas, propondo que diferentes modos de falha geram assinaturas estatísticas distintas, manifestadas pelas medidas do coeficiente de correlação (R^2) e do qui-quadrado (χ^2), possibilitando a identificação do tipo de falha.

A análise *wavelet* tem se mostrado uma técnica eficaz para a avaliação de sinais não estacionários, enquanto as máquinas de vetores de suporte (SVM) mapeiam os dados para um espaço de características de maior dimensão por meio de transformações não lineares. Modelos que combinam a análise *wavelet* com SVM demonstram desempenho robusto na previsão da curva de potência em curtos intervalos temporais (Sohoni et al., 2016).

3 Metodologia

Os dados provenientes do sistema SCADA representam um recurso essencial para o setor eólico, fornecendo uma base rica e multivariada de informações operacionais e ambientais. Sua aplicação vai além do monitoramento de turbinas eólicas, abrangendo também a modelagem avançada da curva de potência, a estimativa do fator de capacidade e a previsão de falhas com maior precisão. Estudos recentes demonstram que abordagens baseadas em aprendizado de máquina e estatística avançada podem reduzir as incertezas nas análises e melhorar a detecção precoce de anomalias, contribuindo diretamente para uma operação mais eficiente e para a manutenção preditiva dos parques eólicos (Pandit & Kolios, 2020; Astolfi et al., 2021; Astolfi, 2021; Pandit et al., 2022). Com o adequado aproveitamento desses dados, é possível desenvolver modelos adaptativos capazes de aprimorar a confiabilidade das turbinas e otimizar a geração de energia sob diferentes condições de operação.

Os dados do SCADA desta pesquisa referem-se ao complexo eólico Senandes, localizado na região sul do Brasil, composto por 40 aerogeradores ECO-122/2.7MW distribuídos em 4 parques eólicos Senandes II, Senandes III, Senandes IV e Vento Aragano, conforme ilustrado na Figura 4. No entanto, a análise realizada neste estudo considerou especificamente os dados do aerogerador 1 dos parques Vento Aragano, Senandes II e Senandes III, cujos períodos de observação foram: Vento Aragano (05/05/2016 até 26/08/2016), Senandes III (19/05/2016 até 14/08/2016) e Senandes II (28/09/2016 até 20/10/2016).

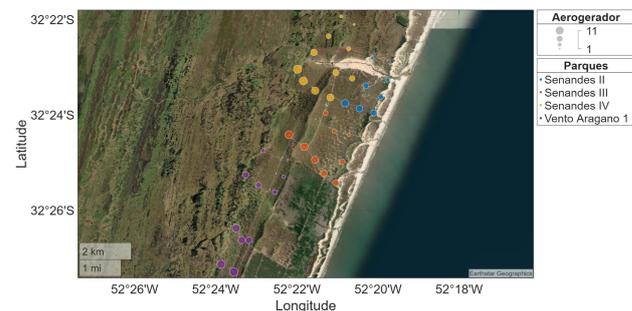


Figura 4: Complexo eólico Senandes, composto por 40 aerogeradores distribuídos em 4 parques eólicos

Em geral os dados do SCADA geralmente não são padronizados e podem apresentar pontos anômalos ou falhas de registro (Teixeira, 2022). Os dados utilizados nesta pesquisa foram filtrados por meio de uma abordagem baseada em clas-

ses construídas a partir da velocidade do vento. O fluxograma processo de filtragem de dados é ilustrado na Figura 5.

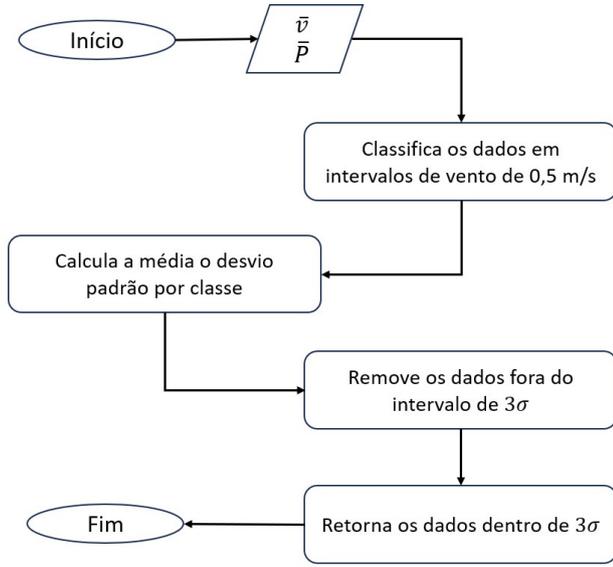


Figura 5: Fluxograma do processo de filtragem de dados

Inicialmente, os dados foram agrupados em classes de 0,5 m/s, permitindo o cálculo das médias e dos desvios padrão em cada intervalo. Em seguida, foram removidos os pontos destacados em vermelho na Figura 6, os quais não estavam localizados dentro do intervalo de 3 desvios padrão em torno da média de cada classe (representado pelas barras de erro). Esse processo reduziu a influência de outliers e ruídos nos registros, garantindo maior consistência e qualidade para as análises subsequentes. Além disso, a abordagem preserva a explicação física do comportamento da curva de potência sob a influência da intensidade da turbulência e de outros fatores.

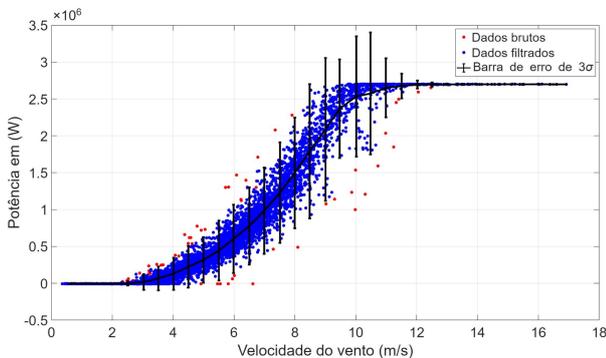


Figura 6: Curva de potência do aerogerador 1 do parque eólico Vento Aragano 01: dados brutos vs. dados filtrados

Para modelar a curva de potência a partir dos dados experimentais, foi adotada uma

abordagem baseada em Redes Neurais Artificiais (RNA). Inicialmente, os dados filtrados da velocidade do vento e da potência foram normalizados para possuir média zero e desvio padrão igual a 1, utilizando a seguinte fórmula:

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

onde

- μ é a média dos valores de treinamento;
- σ representa o desvio padrão desses valores;
- x simboliza o valor original da variável;
- z caracteriza o valor normalizado, que será utilizado como entrada para o treinamento da rede neural.

O modelo de RNA foi implementado em Python (2025) versão 3.11.2, por meio da IDE Spyder (2024) e utilizando a biblioteca Keras (2025) versão 3.8.0. A arquitetura adotada é do tipo densa, composta por três camadas ocultas e uma camada de saída, conforme ilustrado na Figura 7. Essa configuração busca equilibrar a complexidade do modelo com sua capacidade de generalização, tendo sido definida a partir de um processo iterativo de busca empírica, no qual diferentes combinações de camadas e neurônios foram testadas até se obter um arranjo satisfatório.

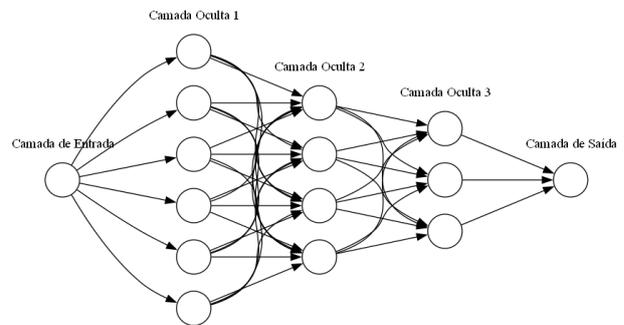


Figura 7: Arquitetura da Rede Neural Artificial proposta para estimativa da curva de potência

Na arquitetura adotada, a camada de entrada recebe os dados normalizados e encaminha as variáveis para o processamento pela RNA. O processamento ocorre por meio de três camadas ocultas, que constituem o núcleo do modelo, distribuídas da seguinte forma:

- Primeira camada oculta: 6 neurônios;
- Segunda camada oculta: 4 neurônios;
- Terceira camada oculta: 3 neurônios.

Todas as camadas ocultas utilizam a função de ativação tangente hiperbólica (\tanh), definida como

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

a qual é simétrica em torno de zero e, para valores extremos, aproxima-se de 1 ou -1, ajudando a manter os pesos da rede distribuídos de forma equilibrada.

Por fim, a camada de saída adota uma ativação linear, descrita pela seguinte equação:

$$f(x) = x \quad (6)$$

de modo que os valores provenientes da última camada oculta sejam repassados sem alteração para gerar a previsão final, que corresponde à curva de potência estimada, relacionando a velocidade do vento com a potência elétrica gerada.

O treinamento da RNA foi realizado por meio do algoritmo de retropropagação, utilizando o erro médio quadrático (MSE - *Mean Squared Error*) como função de custo para ajustar os pesos sinápticos entre os neurônios, definido por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

onde:

- n é o número total de amostras;
- y_i é o valor experimental da amostra i ;
- \hat{y}_i é o valor predito para essa amostra.

Além do MSE, também foram adotadas outras métricas para auxiliar na avaliação do desempenho do modelo, incluindo o coeficiente de determinação (R^2), a raiz do erro médio quadrático (RMSE - *Root Mean Squared Error*) e o erro médio absoluto percentual (MAPE - *Mean Absolute Percentage Error*), cujas equações são apresentadas a seguir.

O coeficiente de determinação (R^2) mede o grau de ajuste do modelo aos dados, sendo definido como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

Onde \bar{y} representa a média dos valores reais.

A raiz do erro médio quadrático (RMSE) fornece uma medida da dispersão dos erros em

relação aos valores reais, sendo dada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

O erro médio absoluto percentual (MAPE) avalia a precisão das previsões em termos percentuais:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (10)$$

Essas métricas permitem avaliar o desempenho da rede neural, identificando o grau de erro da previsão e sua capacidade de generalização. O R^2 indica o quão bem o modelo explica a variabilidade dos dados, enquanto o RMSE e o MAPE fornecem medidas de erro absoluto e percentual para analisar a precisão das estimativas (Pandit et al., 2020; Kardakaris et al., 2021).

O otimizador adotado foi o Adam (*Adaptive Moment Estimation*), um método avançado de otimização amplamente utilizado em redes neurais e aprendizado profundo. O Adam combina os conceitos de *Momentum* e *RMSprop*, ajustando dinamicamente a taxa de aprendizado de cada parâmetro durante o treinamento.

O *Momentum* acelera a descida do gradiente por meio de uma média móvel dos gradientes passados, conforme a seguinte expressão:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \frac{\partial L}{\partial w_t} \quad (11)$$

onde:

- β_1 é o parâmetro de decaimento para o *Momentum*, o valor adotado foi $\beta_1 = 0,85$;
- $\frac{\partial L}{\partial w_t}$ representa o gradiente da função de perda em relação ao peso w_t .

A atualização dos pesos é realizada pela regra:

$$w_{t+1} = w_t - \alpha \cdot m_t \quad (12)$$

onde:

- m_t representa a média móvel dos gradientes no tempo t ;
- α é a taxa de aprendizagem, o valor adotado para $\alpha = 0,0001$;
- w_t e w_{t+1} são os pesos no momento t e $t + 1$ respectivamente.

O RMSprop adapta a taxa de aprendizado utilizando uma média móvel exponencial dos gradientes quadrados, definida por:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left(\frac{\partial L}{\partial w_t} \right)^2 \quad (13)$$

e os pesos são atualizados conforme:

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{v_t + \epsilon}} \cdot \frac{\partial L}{\partial w_t} \quad (14)$$

onde:

- v_t é a média ponderada exponencialmente dos gradientes quadrados;
- ϵ é uma pequena constante adicionada para evitar a divisão por zero, neste caso foi adotado $\epsilon = 1 \cdot 10^{-7}$;
- β_2 é o fator de decaimento do segundo momento, o valor adotado foi $\beta_2 = 0,99$.

Como m_t e v_t são inicializados em zero, o algoritmo aplica correções de viés:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (15)$$

e a atualização final dos pesos é dada por:

$$w_{t+1} = w_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (16)$$

O treinamento foi conduzido ao longo de 1.000 épocas, com um tamanho de lote (*batch size*) de 10, permitindo ajustes robustos nos parâmetros do modelo. Para preservar a homogeneidade ao longo da curva de potência, optou-se por uma seleção aleatória dos pontos, evitando abordagens mais intrusivas que impusessem uma quantidade fixa de amostras com base nas regiões de atuação da curva. A estratégia de particionamento dos dados foi de 56% para treinamento, 20% para teste e 24% para validação, afim de minimizar o risco de sobreajuste e garantir uma melhor capacidade de generalização. Essa divisão foi cuidadosamente ajustada em função da distribuição desigual dos dados entre os três conjuntos disponíveis: dois deles continham menos de 2.500 pontos, enquanto apenas um ultrapassava os 5.000. Diante desse cenário, a escolha aleatória dos pontos permitiu manter uma distribuição mais equilibrada ao longo de toda a curva, sem comprometer a representatividade dos dados.

4 Resultados e Discussão

O resultado final do modelo de rede neural aplicado ao parque eólico Vento Aragano é apresentado na Figura 8. Os pontos verdes correspondem aos dados reais, e a curva amarela representa a previsão do modelo, que capturou satisfatoriamente a relação não linear entre a velocidade do vento e a potência gerada, conforme evidenciado pelo elevado grau de aderência observado no conjunto de validação, composto por 1.262 pontos. No entanto, apesar dos resultados promissores, é importante ressaltar que, no cotovelo inferior da curva de potência, o modelo tende a superestimar a potência em velocidades inferiores a 3 m/s e superiores a 2 m/s, velocidades essas que formalmente pertencem à Região 1, quando o sistema permanece em *stand-by*, e a subestimar a potência na transição entre as Regiões 2 e 3, ou seja, para ventos entre 9 m/s e 11 m/s. Em geral, essas discrepâncias são características de modelos que não consideram o efeito da turbulência, conforme identificado em Albers (2010) e discutido em Clifton et al. (2013), aprofundado em Bardal & Sætran (2017) e Teixeira (2022).

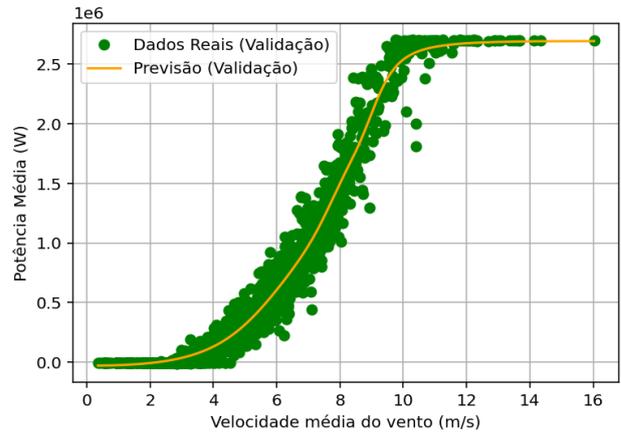


Figura 8: Curva de potência prevista para o conjunto de dados de validação do parque eólico Vento Aragano aerogerador 01

O conjunto de dados de validação do parque eólico Senandes II é composto por 573 pontos, sendo o menor dentre os conjuntos de dados disponíveis. Na Figura 9 é exibida a curva de potência prevista, evidenciando uma forte concordância entre os dados de validação e a predição do modelo. Observa-se que os pontos verdes, que representam os dados reais, acompanham de maneira consistente a curva laranja das previsões, indicando que o modelo foi capaz de reproduzir as características do sistema. Vale a pena ressaltar que neste caso também é possível

observar que para ventos maiores 3 m/s e inferiores a 5 m/s a potência é superestimada.

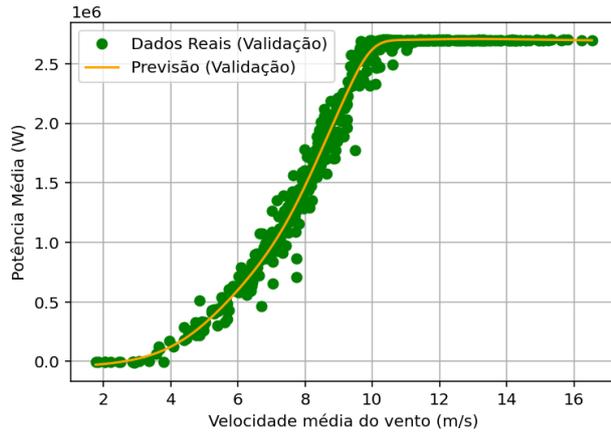


Figura 9: Curva de potência prevista para o conjunto de validação do parque eólico Senandes II aerogerador 01

A Figura 10 apresenta a predição da curva de potência para o parque eólico Senandes III, utilizando um conjunto de validação composto por 586 pontos, valor ligeiramente superior ao observado para Senandes II, porém substancialmente inferior ao de Vento Aragano. Observa-se uma predominância de pontos com velocidades superiores a 6 m/s. Entretanto, o modelo demonstrou eficácia na captura da relação não linear entre a velocidade do vento e a potência gerada. Levando em consideração o comportamento típico da curva de potência, é possível observar que a curva prevista pelo modelo se assemelha de forma consistente à curva de catálogo do aerogerador em questão, exibida na Figura 1. Um aspecto adicional a ser destacado é que, neste conjunto de dados, observa-se menor dispersão, a qual, em geral, é causada pela turbulência. Essa informação é relevante, pois o aerogerador 1 do parque eólico Senandes III está de frente ao mar, o que pode indicar que a direção do vento no período analisado estava favorável a ventos menos turbulentos, a uma menor variação na direção do vento ou ainda à ausência do efeito esteira, ou, possivelmente, trata-se apenas da baixa quantidade de dados disponíveis.

O gráfico de regressão comparativo entre os valores reais e as previsões do modelo para o conjunto de teste é ilustrado na Figura 11 para o parque eólico Vento Aragano. Os pontos pretos correspondem aos dados observados, enquanto a linha azul representa a regressão linear ajustada às previsões. A linha de identidade, representada em vermelho, serve de referência para uma predição perfeita. Analisando a linha de

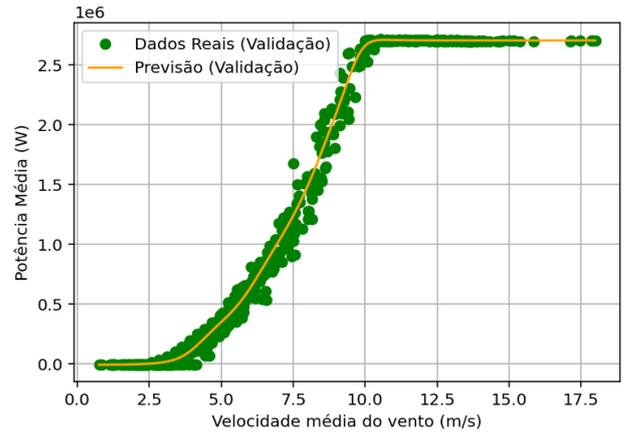


Figura 10: Curva de potência prevista para o conjunto de validação do parque eólico Senandes III aerogerador 01

regressão neste gráfico, é possível perceber que o modelo, em geral, subestima a potência para valores maiores que 1,5 MW e superestima a potência para valores menores que 0,5 MW. Em contrapartida, o coeficiente de determinação (R^2) é de 0,966, denotando um excelente ajuste do modelo para os dados do parque eólico Vento Aragano. A elevada proximidade dos pontos à linha de identidade ressalta a capacidade preditiva do modelo, embora pequenas dispersões possam ser atribuídas a variações naturais nos dados ou a limitações inerentes ao treinamento da rede neural, sem comprometer significativamente a precisão global da previsão.

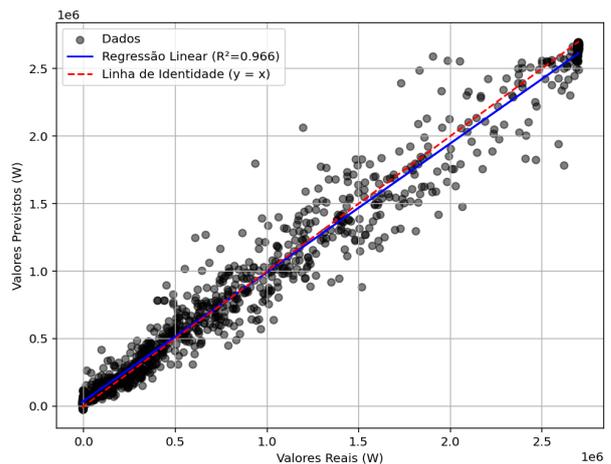


Figura 11: Regressão entre os valores reais e previstos (conjunto de teste) do parque eólico Vento Aragano aerogerador 01

A Figura 12 apresenta o gráfico de regressão do modelo para o conjunto de teste do parque eólico Senandes II. Ressalta-se que, apesar de a

amostra ser reduzida, com somente 478 pontos, o R^2 demonstrou ser extremamente elevado, sugerindo um potencial sobreajuste. Esse alto R^2 pode ser justificado parcialmente pelo tamanho menor do conjunto; no entanto, também demonstra uma robusta capacidade preditiva do modelo, sugerindo um desempenho consistente mesmo em condições de volume de dados limitado.

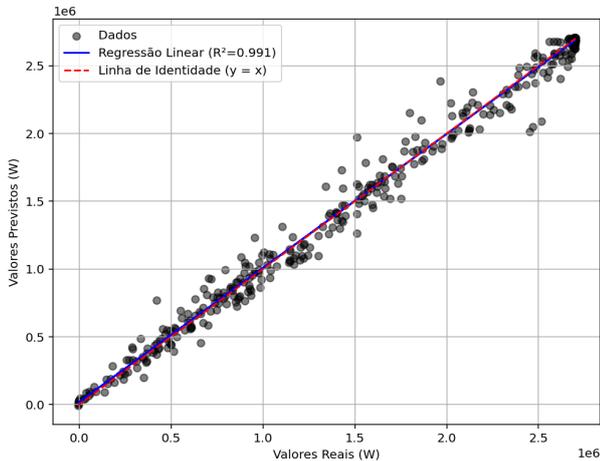


Figura 12: Regressão entre os valores reais e previstos (conjunto de teste) do parque eólico Senandes II aerogerador 01

A Figura 13 exibe a regressão linear aplicada ao parque eólico Senandes III, evidenciando um coeficiente de determinação (R^2) de 0,993, o que, à primeira vista, pode sugerir um sobreajuste ou indicar a necessidade de um volume de dados maior. De maneira geral, esse alto valor do coeficiente de determinação indica que o modelo possui elevada capacidade de reproduzir a relação entre as variáveis de estudo.

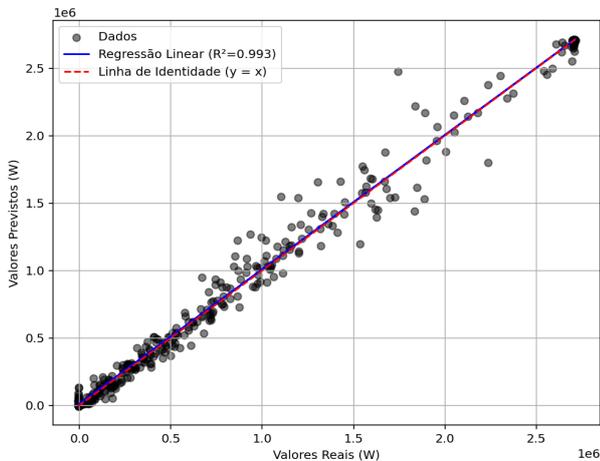


Figura 13: Regressão entre os valores reais e previstos (conjunto de teste) do parque eólico Senandes III aerogerador 01

O histograma dos erros de treinamento e de

teste para o parque eólico Vento Aragano, exibido na Figura 14, revela que a maior parte dos resíduos se concentra em torno de zero, evidenciando uma boa concordância entre as previsões e os valores observados. A distribuição, aproximadamente simétrica, mostra uma dispersão ligeiramente maior no conjunto de teste, em comparação com o conjunto de treinamento, o que sugere uma capacidade de generalização razoável do modelo.

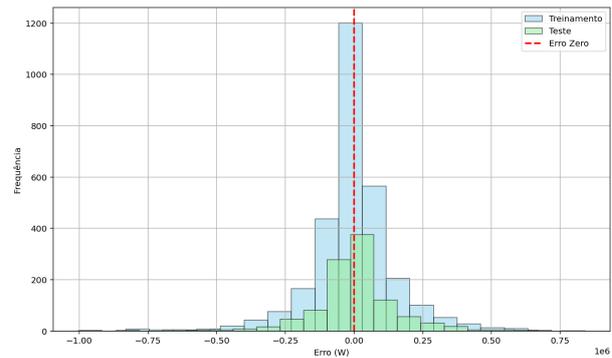


Figura 14: Histograma dos erros de treinamento e teste - Parque eólico Vento Aragano

A Figura 15 apresenta o histograma dos erros obtidos no parque eólico Senandes II. Assim como no caso do parque eólico Vento Aragano, a maior parte dos resíduos está concentrada em torno de zero, o que indica uma boa concordância entre as previsões e os dados reais. A distribuição dos erros apresenta uma dispersão moderada, sugerindo que, embora o modelo apresente um desempenho robusto, ainda possam existir pequenas variações. Essa análise evidencia, portanto, o potencial do modelo, ao mesmo tempo em que aponta para a possibilidade de ajustes adicionais visando a otimização do desempenho em faixas específicas de operação.

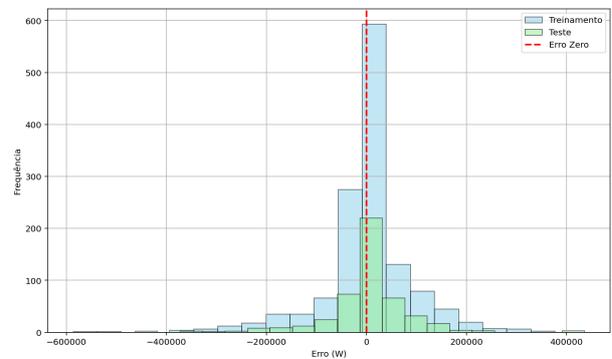


Figura 15: Histograma dos erros de treinamento e teste - Parque eólico Senandes II

A Figura 16 apresenta o histograma dos erros para o parque eólico Senandes III, evidenciando

que a maior parte dos resíduos se concentra em torno de zero, indicando uma provável ausência de viés sistemático nas previsões. No entanto, a dispersão observada nos erros sugere uma variabilidade que pode estar associada à heterogeneidade intrínseca dos dados ou a limitações do ajuste do modelo em determinadas faixas operacionais. Esses achados ressaltam a robustez geral do modelo, ao mesmo tempo em que apontam para a necessidade de análises complementares, possivelmente envolvendo o ajuste de parâmetros ou a incorporação de variáveis adicionais, a fim de aprimorar a precisão preditiva em regimes extremos.

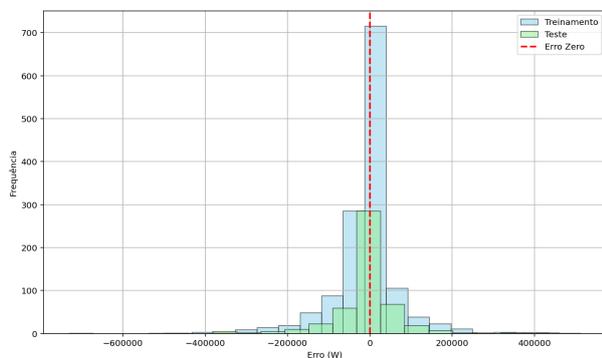


Figura 16: Histograma dos erros de treinamento e teste - Parque eólico Senandes III

A Tabela 1 apresenta a análise comparativa dos erros do modelo para os parques eólicos Vento Aragano, Senandes II e Senandes III. Os valores de R^2 superiores a 0,95 evidenciam um excelente ajuste, confirmando a capacidade do modelo de capturar a relação entre a velocidade do vento e a potência gerada. Observa-se que, em termos de erro médio quadrático (RMSE), o parque Senandes II apresenta os menores valores, o que sugere maior precisão preditiva nesse cenário. Já no caso do Senandes III o erro médio quadrático é relativamente próximo ao de Senandes II no entanto em contrapartida, o MAPE elevado pode ser explicado pela influência de valores reais muito baixos, os quais distorcem a interpretação percentual do erro. No geral, o modelo demonstra uma robusta capacidade preditiva e representa uma alternativa promissora para monitoramento e previsão, visto que utiliza exclusivamente dados disponíveis no sistema SCADA, embora sejam possíveis refinamentos adicionais por meio da inclusão de outras variáveis e do tratamento de outliers para reduzir ainda mais os erros.

Tabela 1: Erros dos modelos da curva de potência para os parques eólicos analisados

	Vento Aragano	Senandes II	Senandes III
R^2	0,966	0,991	0,993
RMSE (W)	154.173,853	89.404,983	89.607,914
MAPE (%)	128,985	24,164	85,995

5 Conclusão

A modelagem proposta neste trabalho demonstrou que o uso de redes neurais artificiais para a previsão da curva de potência de aerogeradores a partir de dados SCADA pode ser uma boa alternativa. O modelo foi capaz de capturar a complexa relação não linear entre a velocidade do vento e a potência gerada, evidenciada por coeficientes de determinação superiores a 0,95, o que atesta uma excelente capacidade preditiva, mesmo diante das variações inerentes aos dados operacionais dos parques eólicos. Embora a grande dispersão dos dados represente um desafio, a estratégia de tratamento e filtragem aplicada mostrou-se crucial para garantir a consistência das previsões. Ainda assim, vale destacar que a adoção de abordagens mais avançadas para a detecção e remoção de *outliers* pode representar uma alternativa promissora para aprimorar a qualidade dos dados de entrada e, consequentemente, os resultados do modelo. Entretanto, cabe destacar que, a fim de evitar sobreajuste, talvez seja relevante trabalhar com volumes de dados mais extensos. Observou-se que, nos casos em que o conjunto de dados possuía menos de 2.500 pontos, o modelo muito provavelmente memorizou o conjunto de treinamento, resultando em valores de R^2 extremamente elevados. Por outro lado, quando o volume de dados ultrapassou 5.000 pontos, os resultados mostraram-se mais satisfatórios. É importante ressaltar que o número mínimo ideal de pontos não deve ser estabelecido com base na duração temporal, mas sim na quantidade efetiva de registros disponíveis. Isso se deve ao fato de que a turbina pode, eventualmente, estar indisponível para geração, seja por razões técnicas, seja pela ausência de vento no período. Além disso, a disponibilidade de vento varia tanto em função da época do ano assim como pelo o período do dia, o que reforça a necessidade de avaliar a representatividade do conjunto de dados a partir do número de pontos disponíveis após a etapa de pré-processamento e não exclusivamente pela extensão temporal.

Como perspectiva para trabalhos futuros, recomenda-se a inclusão de variáveis operacionais adicionais, por exemplo, intensidade da turbulência, direção do vento e densidade do ar, a fim de aprimorar ainda mais a acurácia do mo-

delo. Dessa forma, o estudo reforça o potencial dos métodos baseados em aprendizado de máquina para a otimização e o monitoramento dos parques eólicos, além de contribuir para o avanço das práticas de manutenção preditiva e gestão da energia renovável.

Agradecimentos

O autor agradece ao Dr. Thadeu Carneiro da Silva e à ADS Energia Renováveis S.A. pela valiosa disponibilização dos dados SCADA dos parques eólicos, os quais foram fundamentais para o desenvolvimento deste trabalho.

Referências

- Albers, Axel. 2010. Turbulence and shear normalisation of wind turbine power curve. *Deutsche WindGuard Consulting GmbH*.
- Associação Brasileira de Energia Eólica e Novas Tecnologias. 2024. Boletim anual 2024.
- Astolfi, D. 2021. Perspectives on scada data analysis methods for multivariate wind turbine power curve modeling. *Machines* 9. 100. doi: 10.3390/machines9050100.
- Astolfi, D., F. Castellani, A. Lombardi & L. Terzi. 2021. Multivariate scada data analysis methods for real-world wind turbine power curve monitoring. *Energies* 14. 1105. doi: 10.3390/en14041105.
- Bardal, Lars Morten & Lars Roar Sætran. 2017. Influence of turbulence intensity on wind turbine power curves. *Energy Procedia* 137. 553–558. doi:10.1016/j.egypro.2017.10.384554.
- Bauer, Lucas & Silvio Matysik. 2025. Alstom eco 122/2700 - wind turbine. Acesso em: 24 maio 2025. <https://en.wind-turbine-models.com/turbines/629-alstom-eco-122-2700>.
- Clifton, Andrew, Levi Kilcher, Julie K. Lundquist & Paul Fleming. 2013. Using machine learning to predict wind turbine power output. *Environmental Research Letters* 8. 024009. doi: 10.1088/1748-9326/8/2/024009.
- Empresa de Pesquisa Energética. 2024. Anuário estatístico de energia elétrica 2024.
- Gill, Simon, Bruce Stephen & Stuart Galloway. 2012. Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy* 3(1). 94–101. doi:10.1109/TSTE.2011.2167164.
- Global Wind Energy Council. 2024. Global wind report 2024.
- Haykin, Simon. 2007. *Redes neurais: princípios e prática*. Porto Alegre: Bookman 2nd edn.
- International Electrotechnical Commission. 2022. *IEC 61400-12-1: Wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines*. Geneva: International Electrotechnical Commission 3rd edn.
- International Energy Institute. 2024. *Statistical review of world energy 2024*.
- Jang, Jyh-Shing R., Eiji Mizutani & Chuen-Tsai Sun. 1997. Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. *IEEE Transactions on Automatic Control* 42(10). 1482–1484. doi: 10.1109/TAC.1997.633847.
- Kardakaris, Kimon, Ifigeneia Boufidi & Takvor Soukissian. 2021. Offshore wind and wave energy complementarity in the greek seas based on era5 data. *Atmosphere* 12(10). 1360. doi:10.3390/atmos12101360. <https://www.mdpi.com/2073-4433/12/10/1360>. Acesso em: 28 ago. 2025.
- Keras. 2025. Keras api documentation. Acesso em: 3 jul. 2025. <https://keras.io/api/>.
- Lydia, M., S. Suresh Kumar, A. Immanuel Selvakumar & G. Edwin Prem Kumar. 2014. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews* 30. 452–460. doi: 10.1016/j.rser.2013.10.030.
- Lydia, M., A. Immanuel Selvakumar, S. Suresh Kumar & G. Edwin Prem Kumar. 2013. Advanced algorithms for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy* doi:10.1109/TSTE.2013.2247641.
- Pandit, Ravi, Davide Astolfi, Jiarong Hong, David Infield & Matilde Santos. 2022. Scada data for wind turbine data-driven condition performance monitoring: A review on state-of-art, challenges, and future trends. *Wind Engineering* 46. 1–20. doi:10.1177/0309524X221124031.
- Pandit, Ravi & Athanasios Kolios. 2020. Scada data-based support vector machine wind turbine power curve uncertainty estimation and its comparative studies. *Applied Sciences* 10. 8685. doi:10.3390/app10238685.

- Pandit, Ravi Kumar, David Infield & Athanasios Kolios. 2020. Gaussian process power curve models incorporating wind turbine operational variables. *Energy Reports* 6. 1658–1669. doi:10.1016/j.egy.2020.06.018. <https://doi.org/10.1016/j.egy.2020.06.018>.
- Python. 2025. Python release python 3.11.2. Acesso em: 27 ago. 2025. <https://www.python.org>.
- Silva, Thadeu Carneiro da. 2020. *Proposta de planejamento de manutenção de aerogeradores, com base nas condições ambientais locais*. Santo André: Universidade Federal do ABC. Dissertação de mestrado.
- Sohoni, V., S.C. Gupta & R.K. Nema. 2016. A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy* 2016. 1–20. doi:10.1155/2016/8519785.
- Spyder. 2024. Spyder: The python ide that scientists and data analysts deserve. Acesso em: 27 ago. 2025. <https://www.spyder-ide.org>.
- Teixeira, V. L. 2022. *Modelo para a curva de potência de geradores eólicos a partir de dados experimentais*. Santo André: Universidade Federal do ABC. Dissertação de mestrado.
- Wang, Yun, Qinghua Hu, Linhao Li, Aoife M. Foley & Dipti Srinivasan. 2019. Approaches to wind power curve modeling: A review and discussion. *Renewable and Sustainable Energy Reviews* 116. 109422. doi:10.1016/j.rser.2019.109422.
- Yesilbudak, Mehmet. 2018. Implementation of novel hybrid approaches for power curve modeling of wind turbines. *Energy Conversion and Management* 171. 156–169. doi:10.1016/j.enconman.2018.05.092.