

Uso de aprendizagem de máquina para estimar a energia potencial de aerogeradores

Durval Brito

Programa de Pós-Graduação em
Engenharia de Sistemas e Automação
Santos, Brasil
durval.brito@estudante.ufla.br

Arthur Públio

Engenharia de Controle e Automação
UFLA
Lavras, Brasil
arthur.souza6@estudante.ufla.br

Bruno Henrique Groenner Barbosa

Departamento de Automática
UFLA
Lavras, Brasil
brunohb@ufla.br

Abstract—A estimação da energia potencial de aerogeradores é essencial para a gestão e operação eficiente de sistemas eólicos, considerando a natureza intermitente desse recurso. Este trabalho aplica técnicas de aprendizado de máquina para essa estimativa, utilizando dados reais de um aerogerador e torres anemométricas. Foram empregados descarte de dados por inspeção e o algoritmo Floresta de Isolamento para tratamento e seleção das amostras. Na modelagem, foram testados métodos Floresta Aleatória, Máquina de Vetores de Suporte, Árvore de Decisão e Redes Neurais Perceptron Multicamadas. Foi constatado que o melhor modelo para estimação é a Floresta Aleatória, que obteve R^2 de 0,9804 e RMSE de 0,0527, apresentando qualidade na predição da energia potencial. Conclui-se que o uso de métodos de aprendizagem de máquinas é uma abordagem promissora para sensoriamento virtual em parques eólicos, contribuindo para a otimização operacional e maior aproveitamento dos recursos renováveis.

Index Terms—Machine Learning, Virtual Sensor, Wind Turbine, Torre de Medição Anemométrica, Power Curve, Digital Twin, Floresta de Isolamento, Máquina de Vetores de Suporte, Perceptron Multicamadas, Floresta Aleatória, Árvore de Decisão.

I. INTRODUÇÃO

Aerogeradores (WTG) são cada vez mais empregados na geração de energia renovável, sendo essenciais no cenário de transição energética [1]. Essas máquinas transformam a energia do vento em eletricidade, embora a intermitência desse recurso natural exija estratégias de otimização para maximizar o desempenho e a disponibilidade das turbinas. Uma gestão eficiente busca evitar paradas em períodos de alta produção potencial e otimizar a extração de energia, garantindo que o sistema opere em condições ideais e sem desperdícios. Além das turbinas, parques eólicos geralmente contam com torres de medição eólica (TME) distribuídas estrategicamente para monitorar com maior precisão o recurso eólico de uma região.

A intermitência do recurso eólico impõe a necessidade de estratégias eficazes para otimizar a extração de energia potencial, conforme discutido em [2] sobre gestão eficiente de parques eólicos. A potência disponível é influenciada por uma combinação de fatores intrínsecos ao sistema, como o ângulo de ataque das pás, a velocidade de rotação do rotor, características construtivas do aerogerador e fatores ambien-

tais, como a velocidade e direção do vento e a densidade do ar.

A curva de potência fornecida pelos fabricantes de turbina usualmente possui o formato apresentado na figura 1, onde é possível identificar 4 regiões. A região 1 apresenta um limiar mínimo de velocidade V_p (*cut-in*) para que haja produção de energia potencial. Na região 2 há crescimento de geração de energia entre V_p e V_n . Já a região 3 indica que o WTG está operando em seu máximo. Para velocidades acima de V_c (*cut-out*), o aerogerador adentra a região 4, onde há uma interrupção no funcionamento do aerogerador para que seus componentes sejam protegidos [3]. Apesar dessa curva ser utilizada, ela não costuma representar bem o desempenho operacional da turbina nas condições meteorológicas e geoespaciais do local na qual ela foi instalada [4]. Por conta disso, são criadas curvas de potência específicas para os aerogeradores, onde dois tipos de modelos são explorados: os paramétricos, baseados em condições e modelos físicos; e os não paramétricos, baseados em aprendizado de máquina e que utilizam, primordialmente, dados reais.

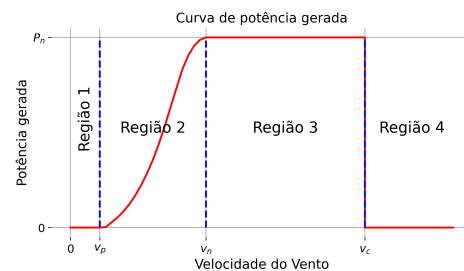


Fig. 1. Curva de Potência típica de um WTG

A curva de potência tem um papel crucial na predição de geração, monitoramento da saúde da turbina e expectativa de geração [5]. O modelo descrito na IEC 61400-12-1 é padrão na indústria e consiste em criar faixas de vento e calcular a média de potência baseada nos dados da turbina [6]. Porém, esse método é propenso a erros, devido à existência de diversos fatores externos, como turbulência, densidade do ar e influência do terreno.

Neste contexto, a aplicação de técnicas de aprendizagem de máquinas são uma abordagem promissora para aprimorar

a eficiência energética e a gestão dos parques. Essas técnicas permitem modelar as complexidades e interações não lineares presentes na geração e no aproveitamento do recurso eólico. Em [7] é apresentado o processo de desenvolvimento de uma solução de inteligência artificial (IA), baseada em *deep learning*, capaz de prever falhas no aerogerador, fazendo uso de floresta de isolamento, *one class support vector* e *autoencoders*.

Diversos trabalhos discutem modelos distintos e os respectivos dados utilizados em cada um, usualmente comparando seus desempenhos com o método de faixas de vento, conforme discutido em [8], o método de faixas apresenta a desvantagem de descartar informações relevantes, comprometendo a detecção de comportamentos anômalos da turbina eólica.

É importante observar que as diversas turbinas em um parque têm influência entre si, seja por efeito de bloqueio ou de esteira [9]. Para tornar esses fenômenos intrínsecos ao modelo, [10] compara modelos com base em dados da própria turbina, da região ao redor e do parque como um todo, evidenciando que os dados da região próxima oferecem o melhor desempenho.

O uso de TME é muito presente na exploração inicial do vento disponível, sendo importante medi-lo durante a operação de parques eólicos. Sensores virtuais têm a capacidade de estimar alguma variável desconhecida a partir de dados. No contexto de uso para WTG, [11] mostra o uso de sensores virtuais para prever danos de estresse mecânico causados pelo efeito esteira, avaliado por intensidade de turbulência. Em [12], são usados dados de TME e de WTG para estimar o vento real de um aerogerador, identificar o efeito esteira e medir seu impacto.

O objetivo deste trabalho é estudar a relação entre a geração de um WTG e as condições meteorológicas, a partir do desenvolvimento de um sensor virtual, calcular a expectativa de energia gerada a partir destas. Um diferencial do trabalho é utilizar os dados tanto de várias TMEs quanto do próprio WTG para estimar a energia esperada para o aerogerador e identificar perdas causadas por diversos fatores. Embora sensores como o de potência não apresentem falhas diretas, as características operacionais da máquina podem se alterar ao longo do tempo, impactando geração. Nesse contexto, o uso de sensores virtuais torna-se uma alternativa eficiente, permitindo estimar potência com base em outras medições do sistema, especialmente quando a máquina não desempenha como esperado. Para isso, serão utilizados dados de um sistema SCADA (Controle Supervisório e Aquisição de Dados) reais de um WTG e dados reais de TMEs distribuídas pela região onde se encontra o WTG em estudo. Assim, este estudo explora o desenvolvimento de *soft sensors* e o papel da IA na otimização da operação dos WTG, alinhando o uso avançado de tecnologia à necessidade crescente de um sistema energético sustentável e eficiente, especialmente em um momento em que a própria IA pode onerar a transição energética [13].

Este artigo está estruturado da seguinte forma: a Seção II aborda os materiais e métodos, como o banco de dados

utilizado, a filtragem de dados, assim como os algoritmos de aprendizagem de máquina utilizados. Na Seção III são apresentados e discutidos os resultados obtidos dos modelos e suas aplicações práticas. Por fim, a Seção IV apresenta as conclusões.

II. MATERIAIS E MÉTODOS

A. Banco de Dados

A turbina chamada de “L-III-01” está localizada próxima a Xique-Xique, Bahia, Brasil, em uma região de alta concentração de aerogeradores, com terreno montanhoso, onde as turbinas se encontram em diferentes elevações relativas entre si. A disposição da região pode ser vista na Figura 2, onde a região próxima está em um raio de 30 vezes o diâmetro de rotor do “L-III-01”. O WTG em questão é uma máquina da fabricante GE de 2,5 MW, com 80 metros de altura de nacela e 116 metros de diâmetro de rotor.

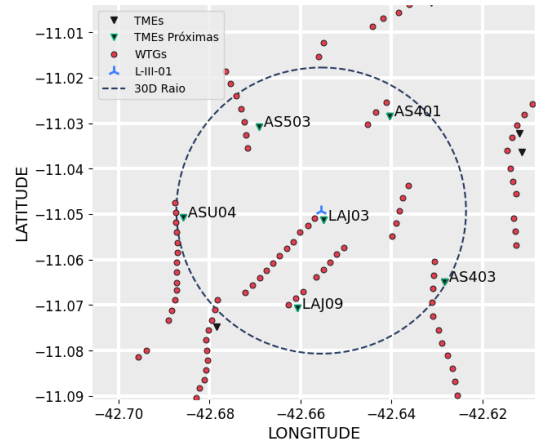


Fig. 2. Disposição de WTGs e TMEs próximas do L-III-01.

Foi concedido um conjunto de dados obtidos do sistema SCADA para o WTG “L-III-01”, tendo 55 variáveis e 54864 amostras, entre janeiro de 2024 e janeiro de 2025. Para cada variável, são registrados na amostra os valores da média, do máximo, do mínimo e do desvio padrão, a cada 10 minutos. As variáveis contidas no SCADA estão descritas na Tabela I.

Para os dados meteorológicos, foram utilizadas 6 TMEs (AS401, AS403, AS503, ASU04, LAJ03, LAJ09) à uma distância de $30D$, onde D é o diâmetro do rotor do WTG, marcadas como “TMEs Próximas” na figura 2. As variáveis também possuem período de amostragem de 10 minutos, e são padronizadas entre cada torre, como é mostrado na tabela II e III. Tais dados advêm de um sistema que já filtra os dados meteorológicos vindos do *datalogger* de cada TME, para se ter somente dados de alta confiabilidade. Cada TME tem 1 termo-higrômetro, medindo temperatura e umidade, 1 barômetro, medindo a pressão, 2 sensores de direção do vento e 3 anemômetros, medindo velocidade do vento em diferentes alturas da torre. Para cada anemômetro é calculada a sua intensidade de turbulência respectiva, T_i , que é calculada por:

TABELA I
VARIÁVEIS SCADA

Variável SCADA	Descrição	Unidade
BL1_ACT	Ângulo de ataque da pá 1	°
BL2_ACT	Ângulo de ataque da pá 2	°
BL3_ACT	Ângulo de ataque da pá 3	°
DrvTr_DrvTrOscAccel	Aceleração do drive train	m/s ²
FREQ	Frequência da linha	Hz
I_A	Corrente fase A	A
I_B	Corrente fase B	A
I_C	Corrente fase C	A
N_GEN_CCU	Velocidade do gerador	RPM
N_ROT_PLC	Velocidade do rotor	RPM
PF	Fator de potência	-
POS_NAC	Direção da Nacele	-
P_ACT	Potência ativa	kW
Q	Potência reativa	kVAR
T_AMB	Temperatura ambiente	°C
T_BEAR_A	Temperatura bearing A	°C
T_BEAR_B	Temperatura bearing B	°C
T_BEAR_SHAFT	Temperatura shaft bearing	°C
T_GEAR	Temperatura gearbox	°C
T_GEAR_BEAR	Temperatura gear box bearing	°C
T_GEN_1	Temperatura generator 1	°C
T_GEN_2	Temperatura generator 2	°C
T_GEN_COOL	Temperatura generator cooling air	°C
T_MAIN_BOX	Temperatura main box	°C
T_NAC	Temperatura Nacelle	°C
T_TOP_BOX	Temperatura top box	°C
U_A_N	Tensão Fase A-Neutro	V
U_B_N	Tensão Fase B-Neutro	V
U_C_N	Tensão Fase C-Neutro	V
V_WIN	Velocidade do Vento	m/s
WIND_DEV_1SEC	Desvio do vento incidente	°

TABELA II
DESCRIÇÃO DAS VARIÁVEIS DA TME

Variável TME	Descrição	Unidade
ATM_PRESSURE	Pressão Atmosférica	hPA
TEMP_AIR	Temperatura do ar	°C
HUMIDITY_RELATIVE	Humidade relativa	%
WIND_DIR	Direção do vento	°
WIND_SPD	Velocidade do vento	m/s

TABELA III
INFORMAÇÕES DAS VARIÁVEIS DA TME

Variável TME	Informação Disponível			
	Média	Máximo	Mínimo	Desv Pad
ATM_PRESSURE	X			
TEMP_AIR	X			
HUMIDITY_RELATIVE	X			
WIND_DIR	X			X
WIND_SPD	X	X	X	X

O X indica que a informação existe para a variável

B. Métodos

Primeiramente, foi realizada uma breve análise dos dados disponíveis, para compreender qual a relação da potência com as demais variáveis. Em seguida, foi aplicado um pré-tratamento para se obter apenas os instantes nos quais a potência do WTG se refere ao seu melhor desempenho. Também foram escolhidas as melhores variáveis a serem utilizadas, para diminuir a dimensionalidade das entradas dos modelos. Por fim, foram aplicados algoritmos diversos, com diferentes combinações de fontes de dados para se identificar o melhor modelo de estimativa de geração.

1) *Análise exploratória*: como cada TME tem múltiplos instrumentos de medição de vento em alturas distintas, foi calculada a matriz de correlação, a fim de determinar se é necessário manter as medições dos 3 anemômetros, conforme apresentado na figura 3. Como exemplo, tem-se a TME mais próxima chamada “LAJ03”, onde os instrumentos são numerados a partir do topo, ou seja, WIND_SPD_1 é o anemômetro de topo da TME. Isso permite descartar algumas informações previamente, como é o caso de todos os instrumentos secundários (em alturas mais baixas) já que estes tem alta correlação entre si como visto na figura 3. Além disso, é visível a relação entre AIR_DENSITY_01 e TEMP_AIR_01, HUMIDITY_RELATIVE_01 e ATM_PRESSURE_01 no cálculo (5), tornando possível removê-las das variáveis de entrada do modelo. Todas as TMEs obtiveram o mesmo comportamento, portanto, os dados de TME a serem utilizados para a modelagem serão apenas WIND_SPD_1, WIND_DIR_1, AIR_DENSITY_01 e TI_01.

O vento da região é considerado unidirecional, já que a maior parte do tempo o vento sopra de uma única direção. E os ventos de alta velocidade também são majoritariamente vindos da direção predominante, conforme os gráficos de direção do vento da figura 4. Isso evidencia um efeito esteira que afeta diretamente o WTG, ao considerar os outros WTGs da região, vistos na figura 2, ou seja, é essencial considerar o impacto

$$T_i = \frac{V_{std}}{V_{avg}}, \quad (1)$$

em que T_i é a intensidade da turbulência, V_{std} é o desvio padrão da velocidade do vento da amostra e V_{avg} é a média da velocidade do vento da amostra. Além disso, é calculada a densidade do ar, por meio das seguintes equações:

$$p_{sat} = 610,78 \exp\left(\frac{(17,27(T - 273,15))}{(T - 35,85)}\right), \quad (2)$$

$$p_v = H \times p_{sat}, \quad (3)$$

$$p_d = p - p_v, \quad (4)$$

$$\rho = \frac{p_d}{R_d T} + \frac{p_v}{R_v T}, \quad (5)$$

em que T é a temperatura em K, H é a umidade relativa do ar, p é a pressão atmosférica, p_{sat} é a pressão de saturação da água, p_v é a pressão de vapor da água, p_d é a pressão do ar seco, $R_d = 287,058 J/(kgK)$ é a constante específica do gás para o ar seco, $R_v = 461,495 J/(kgK)$ é a constante específica do gás para o vapor de água e ρ é a densidade do ar em kg/m³.

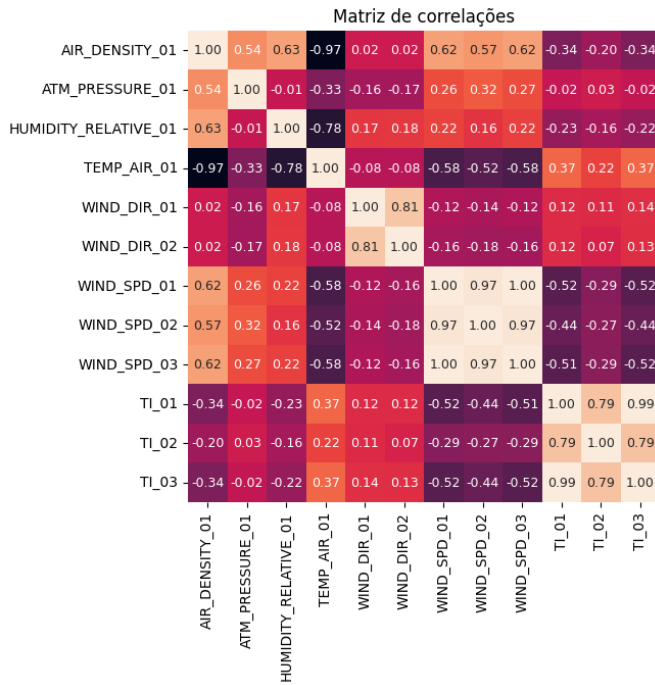


Fig. 3. Correlação entre variáveis da TME LAJ03.

da turbulência T_i na operação do aerogerador. A figura 4 também ilustra que as maiores intensidades de turbulência ocorrem em ventos de baixa velocidade; contudo, seu impacto é mais relevante em ventos de média velocidade (5 - 10m/s), pois, além dessa faixa de vento apresentar maior frequência de ocorrência, está associada a maiores níveis de geração de energia, conforme a região 2 da curva de potência esperada na figura 1.

Devido à variedade de medidas vindas do SCADA do aerogerador, foram escolhidas algumas variáveis específicas para compreender as relações entre elas e a potência gerada. Com o objetivo de filtrar as amostras nas quais a potência do aerogerador não representa o melhor desempenho, é necessário identificar os pontos que não se comportam como o esperado de uma curva de potência típica, como visto no gráfico de vento por potência da figura 5. Nela, é possível reconhecer pontos que representam paradas de máquina ou limitações durante a operação, avaliando as relações entre as variáveis e a potência. Apesar de ser plausível utilizar uma curva de potência média para filtrar os dados, este sensor apresenta recorrentes problemas de calibração, podendo resultar em filtragens ruins. A partir da relação entre o ângulo de ataque da pá e a potência na figura 5, verifica-se uma região dos dados que representam as paradas e limitações, já que a região de operação da máquina tem um formato em “C” entre, aproximadamente, -5° e 20° . Da mesma forma, a operação normal na relação entre a velocidade do gerador e a potência deveria se manter em 0kW entre 0 e 800rpm, onde se mantém até 300kW, crescendo a potência e velocidade do gerador em uma reta até atingir 1500kW e 1400rpm, a partir do

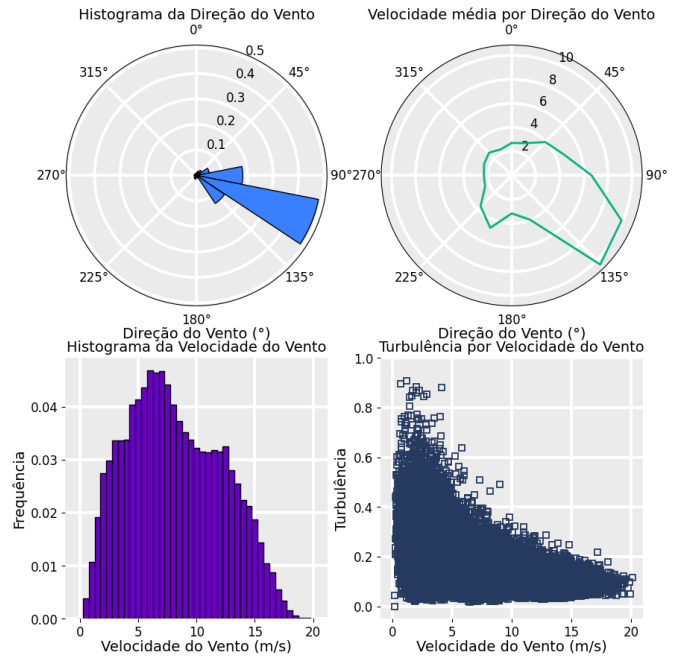


Fig. 4. Vento característico da TME LAJ03.

qual a potência segue até a nominal, mas a velocidade do gerador fica constante. Além disso, o desvio padrão do ângulo de ataque das pás deve se manter próximo a zero, já que movimentos bruscos da pá indicam problemas, os maiores valores deveriam ser observados próximos à nominal, onde a máquina aumenta o ângulo de ataque para manter a rotação do gerador constante. Pode-se observar a projeção dessas amostras a serem descartadas entre as relações de cada uma das variáveis e a potência.

2) *Pré-tratamento*: A filtragem dos dados SCADA teve início com a aplicação de regras baseadas em critérios pré-definidos de exclusão, sendo considerados inválidos os seguintes casos: (a) amostras com menos de 600 registros coletados, identificadas como '02-CNT'; (b) instantes em que alguma das variáveis apresenta ausência de dados, identificados como '03-NAN'; e (c) instantes em que a frequência da rede se encontra fora dos limites estabelecidos pela ANEEL em [14], identificados como '04-FREQ'.

Feito o descarte de dados por inspeção, implementou-se o algoritmo Floresta de Isolamento [15], na relação entre a média da potência (P_{ACT}) do WTG e o desvio padrão do ângulo de ataque das pás ($BL1_{ACT}$, $BL2_{ACT}$, $BL3_{ACT}$), chamado de "05-BL_STD_P_AVG", utilizando 5% de contaminação, 200 árvores, 1% de amostras para treinar cada árvore. E, na sequência, o mesmo algoritmo foi utilizado para uma nova filtragem na relação entre a média da potência, a média do ângulo de ataque das pás e a velocidade do gerador (N_{GEN_CCU}), chamado de "06-BL_GEN_P", utilizando 10% de contaminação, 200 árvores, 1% de amostras para treinar cada árvore.

A aplicação desses filtros resultou nos dados apresentados

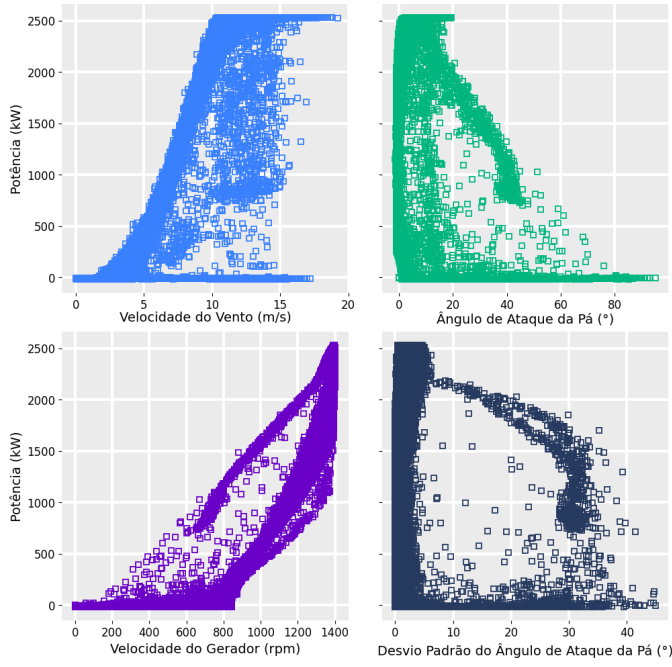


Fig. 5. Características operacionais do WTG L-III-01

na figura 6. Pode-se observar como essa filtragem foi refletida na relação da curva de vento por potência. Foram coletadas 54.864 amostras no total. Deste conjunto, 71% das amostras foram utilizadas para o treinamento dos modelos, restando 15.910 que não foram aproveitadas. Ressalta-se que esses dados não representam, necessariamente, falhas estruturais ou funcionais da turbina como um todo, mas sim amostras que não representam o melhor desempenho da turbina, como limitações e paradas de máquina. Os dados utilizados na próxima etapa de modelagem são apresentados em azul no gráfico 6.

3) *Modelagem*: com os dados filtrados, foram utilizadas técnicas de aprendizagem de máquina para gerar um modelo de previsão da geração baseado nas características atmosféricas e meteorológicas. Como visto anteriormente em II-B1, para cada TME, são escolhidas as medidas de densidade do ar, turbulência T_i , de um sensor de velocidade e de um sensor de direção. Já para o WTG, a variável T_i e o erro da direção do vento foram escolhidos. A potência foi transformada em Fator de Capacidade (FC) que representa a porcentagem da potência nominal da máquina. Os dados foram normalizados de forma específica para cada variável, onde o desvio da direção foi normalizado entre -180° e 180° , a velocidade do vento entre 0 e 32,5 m/s, a direção do vento entre 0° e 360° e a densidade do ar entre 1 e 1,225, de acordo com a equação:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (6)$$

em que X_n é o dado normalizado, X é o dado a ser normalizado, X_{min} e X_{max} são os valores mínimo e máximo

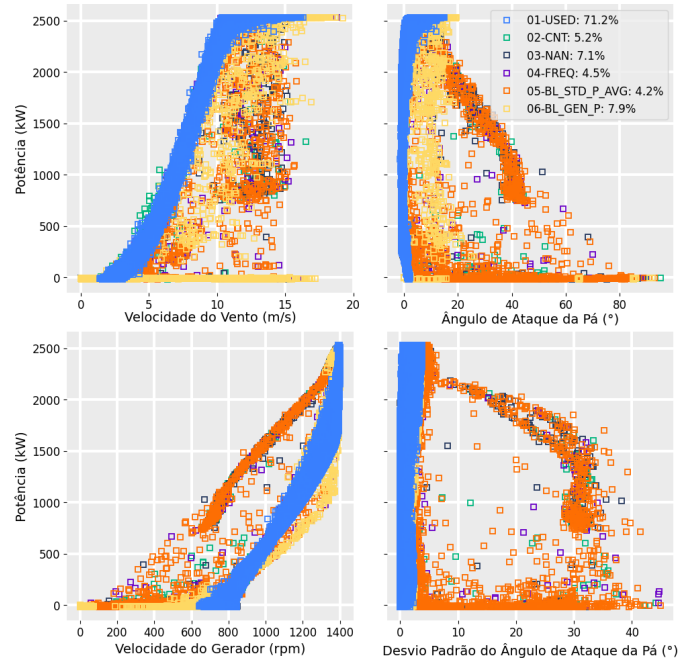


Fig. 6. Dados filtrados do WTG L-III-01, onde cada cor representa um passo de filtragem e a cor azul são os dados a serem utilizados no processo de modelagem da turbina.

respectivamente, entre os quais o dado deve ser normalizado. Para T_i não há necessidade de normalização.

Com o objetivo de prever o FC do WTG, foram aplicados os algoritmos Florestas Aleatórias (RF) e Máquina de vetor suporte (SVM) como [16]; Árvore de decisão (DT) como [17]; Rede Neural Artificial, como [1], [18], mais especificamente o *Perceptron* de Múltiplas Camadas (MLP). Os parâmetros utilizados para RF foram: 100 árvores, todas as variáveis para treinar cada árvore, amostragem por reposição, minimização do erro médio quadrático, com divisões considerando pelo menos duas amostras e folhas contendo pelo menos uma amostra, sem limite de profundidade ou número de folhas e sem pré-poda. Os parâmetros utilizados para SVM foram: função de base radial, com coeficiente do *kernel* em escala, penalidade de erro de 1, tolerância para critério de parada de 0,001, sem encolhimento heurístico. Os parâmetros utilizados para DT foram: minimização do erro quadrático médio, usando todos os recursos para tomar decisões em cada nó, com divisões considerando pelo menos duas amostras e folhas contendo pelo menos uma amostra, sem limite de profundidade ou número de folhas e sem pré-poda. Os parâmetros utilizados para MLP foram: máximo de 4000 iterações, função de ativação ReLU, otimizador Adam, taxa de aprendizado de 0,001, peso inicial aleatório, e tolerância para convergência de 0,0001. Para os modelos de MLP, foram utilizadas algumas abordagens distintas, com uma camada escondida de 60, 100 ou 200 neurônios; e duas camadas escondidas com 50 neurônios em cada camada. O nome do modelo indica o número de neurônios nas camadas escondidas separados por “_”.

Em relação ao conjunto de dados utilizados para treinar os modelos, foram utilizadas todos os dados de cada fonte de dado, seja ela TME ou o WTG. Foram feitas todas as combinações possíveis entre as possíveis fontes, variando entre utilizar apenas uma fonte, até todas as 7 (6 TMEs e o WTG). Os dados foram separados em 75% para treino e 25% para teste. Para garantir que não há interferência nos dados escolhidos para treino e teste, foi utilizado o método *K-fold* [19] para validação cruzada com 10 *folds*.

III. RESULTADOS

Na tabela IV é mostrado o melhor resultado de cada modelo e as fontes de dados utilizadas, bem como os erros apresentados por cada uma. Assim, a partir das respectivas acurácias de cada *fold*, foram obtidos a média e o desvio padrão para cada modelo e gerado o *boxplot* apresentado na figura 7.

TABELA IV
RESULTADOS DO MELHOR RESULTADO DE CADA MODELO

Modelo	Fontes	R2	RMSE
RF	AS401 AS503 ASU04 L-III-01 LAJ03 LAJ09	0,9804	0,0527
SVM	AS503 L-III-01 LAJ03 LAJ09	0,9783	0,0555
MLP_50_50	AS401 AS503 L-III-01 LAJ03	0,978	0,055
MLP_60	AS401 AS503 L-III-01 LAJ03 LAJ09	0,978	0,055
MLP_100	AS503 ASU04 L-III-01 LAJ03 LAJ09	0,978	0,055
MLP_200	AS401 AS503 L-III-01 LAJ03 LAJ09	0,978	0,055
DT	AS401 AS403 ASU04 L-III-01 LAJ03	0,959	0,076

Como pode ser observado pela tabela IV e figura 7, a DT não obteve resultados satisfatórios em comparação aos outros. Ademais, o modelo RF obteve resultados melhores que os demais. Ao analisar os modelos MLP_50_50, MLP_60, MLP_100 e MLP_200 por meio da figura 7, nota-se uma semelhança nos resultados, apesar do uso de fontes de dados diferentes. Vale ressaltar que, dentre as fontes utilizadas nos melhores modelos aqui apresentados, ao avaliar a contagem

de ocorrências de fontes na tabela V, as 3 mais comuns são AS503, L-III-01 e LAJ03. L-III-01 indica que há grande relevância dos fatores presentes no WTG; LAJ03 é uma TME imediatamente à frente do WTG, em relação ao vento incidente predominante, que é altamente relevante; AS503 é uma TME que se encontra atrás do WTG em relação ao vento incidente predominante, isso indica que a característica de como o vento se comporta ao passar pelos WTG tem relevância; ou seja, os melhores modelos são aqueles que utilizam informações sobre o caminho como um todo do vento, como entra, como passa e como sai do sistema. A presença dessas 3 fontes pode indicar que o uso de mais fontes traz benefícios, mas, em valores absolutos, os ganhos são marginais. Por exemplo, utilizando um número menor de fontes de dados em comparação com os outros algoritmos de aprendizado de máquina, o SVM apresentou uma capacidade de predição próxima do melhor modelo (RF), tornando-o um algoritmo eficiente por utilizar menos dados do processo (4 fontes apenas).

TABELA V
NÚMERO DE OCORRÊNCIAS DE USO
DE FONTE ENTRE OS MELHORES MODELOS

Fonte	Ocorrência
L-III-01	7
LAJ03	7
AS503	6
AS401	5
LAJ09	5
ASU04	3
AS403	1

Na figura 8 é mostrado o resultado do modelo de maior acurácia em um gráfico de dispersão do FC medido e o FC predito tanto na fase de treino quanto teste, mostrando sua capacidade de generalização. Na figura 9 é mostrada a predição do FC pelo modelo RF durante um dia em que não ocorreram falhas ou anomalias no processo, evidenciando a capacidade de aproximação do modelo.

Ao aplicar esses modelos ao conjunto completo de dados históricos disponíveis (sem filtragem), torna-se possível estimar o Fator de Capacidade em períodos de limitações e paradas. A Figura 10 apresenta um período com as amostras reais e com as amostras filtradas, indicadas pelas faixas coloridas; as linhas azul e laranja mostram os valores reais e preditos do Fator de Capacidade, respectivamente. Observa-se que as faixas filtradas, como a que antecede o período de “03-07 12h”, correspondem a paradas da máquina, onde os valores reais se aproximam de zero. Nesses momentos, a curva prevista continua a indicar o potencial de geração da máquina, ilustrando a aplicação prática do modelo. Assim, é possível estimar a perda de energia, expressa como a relação entre a energia efetivamente gerada e a energia total disponível, conforme indicado pela seguinte equação:

$$E_t = \frac{E_g}{E_e}, \quad (7)$$

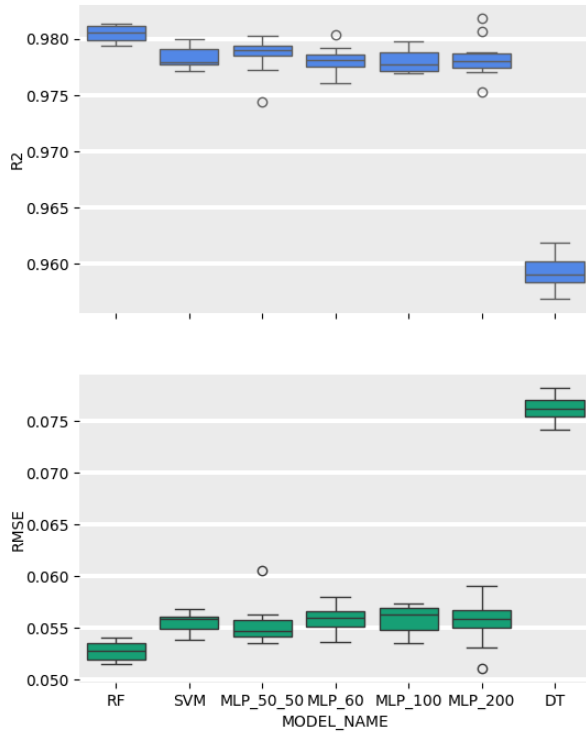


Fig. 7. Desempenho do melhor conjunto de fontes por modelo

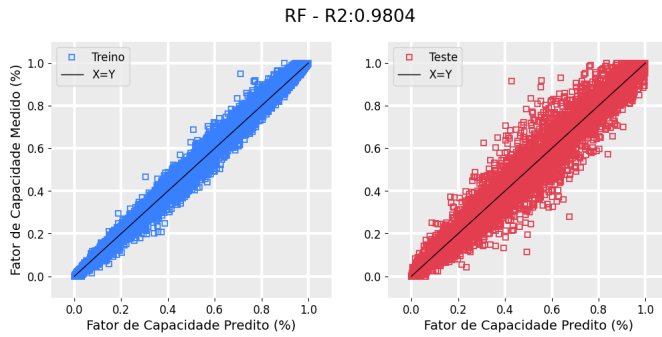


Fig. 8. Diagrama de dispersão entre Fator de Capacidade predito e medido, para dados de treino e teste da Floresta Aleatória.

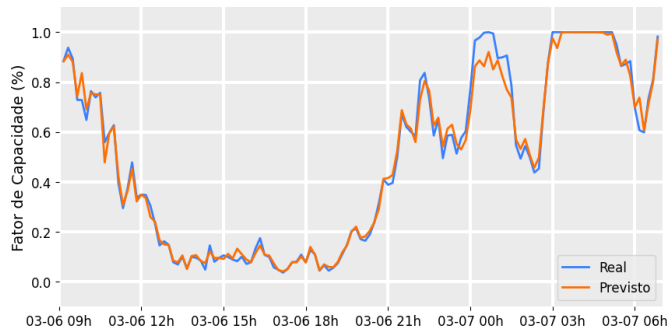


Fig. 9. Predição de Fator de Capacidade da Floresta Aleatória durante um dia de operação do WTG, em dados de teste.

em que E_l é a energia relativa, E_e é a energia total esperada e E_g é a energia total gerada no período. Aplicado a todas as amostras, seus resultados estão na tabela VI. Com base na Tabela 4, observa-se que o algoritmo RF apresentou um desempenho satisfatório, mesmo sendo o que utiliza a maior variedade de fontes distintas.

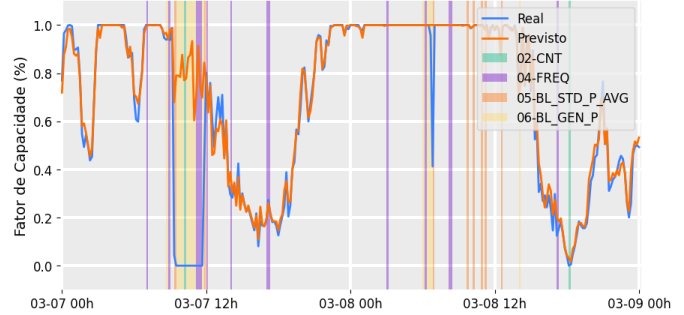


Fig. 10. Série temporal com períodos filtrados do Melhor modelo: Floresta Aleatória

TABELA VI
ENERGIA GERADA

Modelo	Fontes	Geração*
RF	AS401	95,4%
	AS503	
	ASU04	
	L-III-01	
	LAJ03	
SVM	LAJ09	94,5%
	AS503	
	L-III-01	
	LAJ03	
MLP_50_50	LAJ09	94,0%
	AS401	
	AS503	
	L-III-01	
MLP_60	LAJ03	95,7%
	LAJ09	
	AS401	
	AS503	
	L-III-01	
MLP_100	LAJ03	94,2%
	LAJ09	
	AS503	
	ASU04	
	L-III-01	
MLP_200	LAJ03	94,0%
	LAJ09	
	AS401	
	AS503	
	L-III-01	
DT	LAJ03	95,5%
	ASU04	
	L-III-01	
	AS401	

*Geração indica a porcentagem da energia total disponível

IV. CONCLUSÃO

Apesar dos avanços, há desafios remanescentes, como a transparência dos modelos de IA e a necessidade de integração de modelos físicos, que podem fornecer uma interpretação melhor dos resultados.

Os objetivos propostos neste trabalho foram alcançados com sucesso. Foram implementadas técnicas de aprendizagem de máquina que permitiram calcular a expectativa de energia gerada baseada nas condições meteorológicas, por meio da análise de dados históricos.

Idealmente, o modelo em foco não deve ser escolhido apenas com resultados de desempenho de predição, mas sim na capacidade operacional das fontes de dados. Entendendo que há chance de faltas de dados de entradas, o menor número possível de fontes leva a uma melhora da eficiência operacional, tornando o modelo mais robusto. Além disso, novos estudos são necessários para explorar e validar a operacionalidade de modelos ensemble, que utilizam diferentes abordagens e métodos, selecionando os resultados com base na confiabilidade e disponibilidade de cada um.

Os resultados discutidos neste trabalho destacam o potencial transformador da integração de inteligência artificial na operação e gestão de aerogeradores e parques eólicos. A aplicação de modelos baseados em aprendizagem de máquina fornecem previsões sobre a geração de energia, podendo aplicar cálculos de desempenho energético, oferecendo soluções para enfrentar os desafios da intermitência e da complexidade operacional dos parques eólicos, consolidando seu papel como um dos pilares da transição para uma matriz energética mais sustentável e resiliente.

AGRADECIMENTOS

Agradecimento à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e à Serena Energia pelos recursos financeiros destinados à execução do projeto

REFERENCES

- [1] J. C. d. Albuquerque, "Desenvolvimento de modelos de curvas de potência via IA para previsão de geração eólica em curto prazo," doctoralThesis, Universidade Federal de Pernambuco, 2021. [Online]. Available: <https://repositorio.ufpe.br/handle/123456789/42252>
- [2] M. V. S. Rodrigues, M. D. de Aquino, and A. C. F. Thomaz, "Análise da eficiência na produção de energia eólica nos principais estados brasileiros produtores, por meio da aplicação de análise por envoltória de dados," *Anais do Congresso Brasileiro de Iniciação Científica (CBIC)*, 2014.
- [3] V. Sohoni, S. C. Gupta, and R. K. Nema, "A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems," *Department of Electrical Engineering, Maulana Azad National Institute of Technology*, Jun. 2016, received 27 March 2016; Revised 8 June 2016; Accepted 12 June 2016. Academic Editor: Kamal Aly. [Online]. Available: <https://doi.org/10.1155/2016/8519785>
- [4] D. Astolfi, "Perspectives on SCADA data analysis methods for multivariate wind turbine power curve modeling," *Machines*, vol. 9, no. 5, p. 100, 2021. [Online]. Available: <https://www.mdpi.com/2075-1702/9/5/100>
- [5] Y. Wang, Q. Hu, L. Li, A. M. Foley, and D. Srinivasan, "Approaches to wind power curve modeling: A review and discussion," *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109422, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364032119306306>
- [6] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452–460, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364032113007296>
- [7] M. V. R. Nascimento, "Análise e predição de falhas em aerogeradores utilizando deep learning," Universidade Estadual Paulista 'Júlio de Mesquita Filho' (UNESP), Guaratinguetá, SP, Brasil, Trabalho de Conclusão de Curso, 2023, disponível em: repositorio.unesp.br/server/api/core/bitstreams/6a672330-2e23-41e2-beae-9799b900bbb3/content (Acesso em: 28/05/2025).
- [8] P. Guo and D. Infield, "Wind turbine power curve modeling and monitoring with gaussian process and sprt," *Wind Energy*, 2013. [Online]. Available: <https://scispace.com/pdf/wind-turbine-power-curve-modeling-and-monitoring-with-313gean1sp.pdf>
- [9] N. G. Nygaard, S. T. Steen, L. Poulsen, and J. G. Pedersen, "Modelling cluster wakes and wind farm blockage," *Journal of Physics: Conference Series*, vol. 1618, no. 6, p. 062072, 2020. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1618/6/062072>
- [10] J. T. Lyons and T. Göçmen, "Applied machine learning techniques for performance analysis in large wind farms," *Energies*, vol. 14, no. 13, p. 3756, 2021. [Online]. Available: <https://www.mdpi.com/1996-1073/14/13/3756>
- [11] L. D. Avendaño-Valencia, I. Abdallah, and E. Chatzi, "Virtual fatigue diagnostics of wake-affected wind turbine via gaussian process regression," *Renewable Energy*, vol. 170, pp. 539–561, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960148121001671>
- [12] J. Schreiber, C. L. Bottasso, and M. Bertelè, "Field testing of a local wind inflow estimator and wake detector," *Wind Energy Science*, vol. 5, no. 3, pp. 867–884, 2020. [Online]. Available: <https://wes.copernicus.org/articles/5/867/2020/>
- [13] C.-C. Lee and J. Yan, "Will artificial intelligence make energy cleaner? evidence of nonlinearity," *Applied Energy*, vol. 363, p. 123081, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261924004641>
- [14] ANEEL, "PROCEDIMENTOS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA NO SISTEMA ELÉTRICO NACIONAL," May 2022. [Online]. Available: <https://www.gov.br/aneel/pt-br/centrais-de-conteudos/procedimentos-regulatorios/prodist>
- [15] R. NASCIMENTO, B. Barbosa, R. Vargas, I. Santos, D. Ferreira, L. Merschmann, and M. Santos, "DETECÇÃO DE ANOMALIAS EM POÇOS DE PRODUÇÃO DE PETRÓLEO OFFSHORE COM A UTILIZAÇÃO DE autoencoders E TÉCNICAS DE RECONHECIMENTO DE padrões," *UFLA*, Mar. 2022. [Online]. Available: <http://repositorio.ufla.br/jspui/handle/1/49556>
- [16] R. K. Pandit, D. Infield, and A. Kolios, "Comparison of advanced non-parametric models for wind turbine power curves," *IET Renewable Power Generation*, vol. 13, no. 9, pp. 1503–1510, 2019. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2018.5728>
- [17] S. Barber and H. Nordborg, "Improving site-dependent power curve prediction accuracy using regression trees," *Journal of Physics: Conference Series*, vol. 1618, no. 6, p. 062003, 2020. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1618/6/062003>
- [18] J. C. de Albuquerque, R. R. B. de Aquino, O. N. Neto, M. M. S. Lira, A. A. Ferreira, and M. A. de Carvalho, "Power curve modelling for wind turbine using artificial intelligence tools and pre-established inference criteria," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 3, pp. 526–533, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9214975>
- [19] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015, available online 17 March 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320315001070>