

Aprendizado de máquina e a previsão de curto prazo para velocidade máxima do vento

Edvana da Silva da Silveira
Universidade Federal de Santa Catarina (UFSC)

 <https://orcid.org/0000-0003-2694-597X>
edvanadasilveira@gmail.com

Políbio Ícaro Moro Capo
Universidade Federal de Santa Catarina (UFSC)

 <https://orcid.org/0000-0003-3488-3000>
polibiomoro@hotmail.com

Carla de Abreu D'Aquino
Universidade Federal de Santa Catarina (UFSC)

 <https://orcid.org/0000-0002-4079-0866>
carla.daquino@ufsc.br

RESUMO

Eventos de grande intensidade de ventos como passagens de frentes frias, ocorrência de ciclones, ciclones bombas, dentre outros são frequentes no sul do Brasil. A região se destaca também pela produção de energia elétrica a partir dos ventos, a qual, depende de um recurso estocástico e, portanto, sempre haverá incerteza sobre a produção final. Considerando este cenário, o presente trabalho teve como objetivo analisar a performance de modelos de *machine learning* para a previsão da velocidade máxima do vento com antecedência de uma hora. A regressão linear se apresentou como o método com menor capacidade de previsão. A rede neural artificial e o *extreme gradient boosting* (aumento de gradiente extremo, conhecido como XGBoost) apresentaram melhor desempenho em termos de métricas. Os resultados obtidos demonstraram capacidade de previsões promissoras. Esses, representam uma possibilidade de inovação na previsão de velocidades máximas de vento, principalmente, quando

inserida em locais onde ocorrem diferentes fenômenos com ventos extremos.

Palavras-chave: previsão; velocidade máxima do vento; XGBoost; redes neurais artificiais.

Machine learning and the short-term forecast for maximum wind speed

ABSTRACT

Events of great intensity of winds, such as the passage of cold fronts, the occurrence of cyclones, bomb cyclones, among others, are frequent in southern Brazil. The region also stands out for the production of electricity from the winds, which depends on a stochastic resource and, therefore, there will always be uncertainty about the final production. Considering this scenario, the present work aimed to analyze the performance of machine learning models for predicting the maximum wind speed one hour ahead. Linear regression was the method with the lowest predictive capacity. The artificial neural network and extreme gradient boosting (XGBoost) performed better in terms of metrics. The results obtained demonstrated promising prediction capacity. These represent a possibility of innovation in the prediction of maximum wind speeds, especially when inserted in places where different phenomena with extreme winds occur.

Keywords: forecast; maximum wind speed; XGBoost; artificial neural networks.

Submissão em: 11/08/2023 | **Aprovação em:** 30/04/2024

1. INTRODUÇÃO

As mudanças climáticas têm intensificado a ocorrência de eventos com ventos extremos em todo o mundo, inclusive no Brasil. Na região Sul, verificam-se muitas frentes frias e variações bruscas de pressão, além da ocorrência de ventos intensos e eventos meteorológicos extremos, como o Furacão Catarina, ocorrido em 27 e 28 de março de 2004.

Em 2020, outro fenômeno atingiu o estado de Santa Catarina, o Ciclone Bomba, que conforme Gobato e Heidari (2020), tem como sua característica a queda repentina de pressão em seu interior, gerando ventos muito intensos. Segundo a Defesa Civil do Estado de Santa Catarina (Santa Catarina, 2021), durante o evento foram registradas velocidades acima de 130 km/h no litoral. Conforme a CELESC (Centrais Elétricas de Santa Catarina) (2020), o ciclone afetou o sistema elétrico do estado, deixando 1,5 milhão de unidades consumidoras sem energia, sendo o maior dano a rede já registrado pela companhia de energia.

Em maio de 2022, outra ocorrência de fortes rajadas de vento chamou a atenção, o Ciclone Yakecan, que conforme nota meteorológica da Defesa Civil (Santa Catarina, 2022), atingiu velocidades entre 60 a 100 km/h. Em abril de 2023, um Ciclone Bomba atinge a região do litoral sul com rajadas de vento de 79 km/h no Balneário Arroio do Silva (Santa Catarina, 2023a). No dia 13 de julho de 2023 foram emitidos diversos alertas pela Defesa Civil para vento intenso, conforme o órgão as rajadas superaram os 120 km/h em municípios do Planalto e Litoral Sul catarinense (Santa Catarina, 2023b).

As estratégias de Inteligência Artificial (IA) têm sido aplicadas em diferentes trabalhos visando a previsão da velocidade do vento, como: Peláez-Rodríguez et al. (2022); Zucatelli et al. (2019); Noorollahi, Jokar e Kalhor (2016); Samet, Reisi e Marzbani (2019); Liu et al. (2020); Navas, Prakash e Sasipraba (2020); Paula et al. (2021); Santos, Lyra e Silva Júnior (2020); Fonseca, Oliveira e Affonso (2021); Pati et al. (2023); Malakouti

(2023). Entretanto, para a previsão da velocidade máxima do vento existem lacunas na literatura, principalmente, considerando a região sul do Brasil.

A previsão do recurso eólico local, segundo Camelo et al. (2018), permite o conhecimento do regime dos ventos de uma região, e dessa forma, pode auxiliar na sua exploração para geração de energia. No entanto, as previsões não são necessárias somente para a prospecção do recurso eólico, mas também, conforme Zucatelli et al. (2019), para o planejamento, controle e monitoramento de sistemas de energia eólica. Isso permite que os operadores de parques eólicos possam tomar decisões preventivas, como desligar ou reduzir a velocidade de rotação dos aerogeradores durante condições climáticas adversas, minimizando o risco de danos e aumentando a vida útil dos equipamentos.

Com o crescimento exponencial da geração eólica e a possibilidade de geração no mar aumentam ainda mais a demanda por soluções de previsão de curto prazo das velocidades máximas de vento. Com isso, destaca-se a localização da estação meteorológica utilizada nesse estudo, que conforme Corrêa (2018) e Pires (2019), possui potencial eólico costeiro para exploração de energia eólica com futura geração *offshore*.

Nesse contexto, o presente trabalho teve por objetivo analisar o desempenho de algoritmos de *machine learning* (aprendizado de máquina) na previsão da velocidade máxima do vento com uma hora de antecedência. Assim, as investigações e os resultados contribuem para o desenvolvimento científico e tecnológico da região, bem como, podem representar uma ferramenta para auxiliar nas previsões de curto prazo, na produção de energia em um parque eólico e para a Defesa Civil na tomada de decisões sobre eventos extremos com ventos intensos.

2. MATERIAIS E MÉTODOS

Os dados utilizados foram coletados em uma estação meteorológica instalada em uma zona costeira de Santa Catarina. Essa é uma região onde ocorrem diversos fenômenos com ventos intensos. A metodologia foi dividida em três etapas:

- 1) O pré-tratamento e preparação dos conjuntos de dados,
- 2) Aplicação dos modelos de *machine learning*, e
- 3) A avaliação das métricas de desempenho.

A estação meteorológica foi construída sobre uma plataforma de pesca localizada no extremo sul catarinense, no município de Balneário Arroio do Silva, conforme apresentado na Figura 1. De acordo com Pires (2019), a plataforma possui 410 m de comprimento com aproximadamente 7 m acima do nível do mar.

Figura 1 - (A) Plataforma de pesca Entremares em Balneário Arroio Silva, em vermelho a localização do laboratório, SC. (B) Laboratório de pesquisa e a torre meteorológica.



Fonte: (A) Google Earth (2022); (B) MOVLIDAR-UFSC (2021), adaptado.

A Base de Observação do Oceano e Atmosfera (BOOA) é uma cooperação entre a Universidade Federal de Santa Catarina (UFSC), a plataforma de pesca Entremares e o Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina (CIRAM/EPAGRI), a estação é identificada pelo código 1066 - Balneário Arroio do Silva - BOOA/UFSC.

Conforme Nassif (2017), a estação localiza-se na região de transição continente-oceano, com uma distância de 250 metros da linha de praia. A BOOA é um

laboratório de pesquisa de 2 m², onde estão instalados diversos equipamentos, entre eles a torre meteorológica com anemômetro de copo (Movlidar-UFSC, 2021). Os dados utilizados no presente estudo foram medidos na altura de 17 m.

3. PRÉ-TRATAMENTO E PREPARAÇÃO DO CONJUNTO DE DADOS

Os dados utilizados foram registrados no período entre 2018 e 2021 com intervalo de 1 hora. Para o ano de 2018 foram utilizados 5419 dos 5600 observados, que compreendem 01/05/2018 até 31/12/2018. Em 2019 e 2020, o período de janeiro a dezembro, totalizando 8552 de 8724 e 8551 de 8768 dados, respectivamente. E para o ano de 2021, no período de janeiro a março, 2100 entre os 2155 dados observados. Em alguns meses não houve medições devido a manutenção dos equipamentos. Foram removidas as inconsistências no banco de dados em caso de falha dos sensores de medição, tais como, lacunas e variáveis com marcações '99999'.

Nos modelos de *machine learning*, normalmente, podem ser utilizadas duas formas de treinamento: não supervisionado, o qual compreende em utilizar um conjunto de variáveis e não se conhece as saídas, ou supervisionado, quando por meio de variáveis independentes, obtém-se uma variável dependente que é conhecida durante a fase de treinamento (Lee; Booth; Alam, 2005). Para o desenvolvimento do trabalho o treinamento supervisionado foi aplicado.

Os dados foram divididos em três blocos principais:

- 1) Treinamento: conjunto de dados utilizados para ajustar os modelos, sendo aleatoriamente selecionados 90% dos dados observados em 2018, 2019 e 2021, totalizando 14464 observações.
- 2) Teste: a amostra de dados usada para fornecer uma avaliação imparcial do modelo após ajustes de hiper parâmetros realizados durante o treinamento. Foram selecionados,

aleatoriamente, 10% dos dados observados em 2018, 2019 e 2021, contemplando 1607 observações.

3) Validação: amostra de dados utilizada para fornecer uma avaliação imparcial do modelo final após os ajustes realizados durante o treinamento. Foram utilizados os dados do ano de 2020 com o total de 8551 dados, os quais não foram incluídos nas etapas de treinamento e teste.

Os modelos e estratégias de *machine learning* são fortemente dependentes da base de dados de treinamento. Dessa forma, a mesma contempla um período maior de observações, o que pode levar a melhores resultados na validação. Com isso, a probabilidade de ocorrerem situações já treinadas, favorece o acerto. Dado que, os modelos foram capacitados para lidar com grande variedade de situações.

O treinamento supervisionado foi aplicado considerando como variável alvo a velocidade máxima do vento 1 hora depois (objeto da previsão). Assim, verificou-se a existência das medições no instante de tempo t_0 e $t+1h$ com o objetivo de garantir que não houvesse lacunas. Dessa forma, a base de dados foi tratada para ser um conjunto de variáveis de entrada com uma única saída, não sendo uma previsão de série temporal.

Além disso, optou-se por calcular a razão e a diferença entre as velocidades médias e máximas do vento para evitar o uso da variável alvo (velocidade máxima) durante as etapas de treinamento, teste e validação. Com isso, oito variáveis foram utilizadas como parâmetros de entrada: pressão atmosférica ao nível da estação (hPa), temperatura mínima do ar ($^{\circ}\text{C}$), temperatura média do ar ($^{\circ}\text{C}$), temperatura máxima do ar ($^{\circ}\text{C}$), direção vento ($^{\circ}$), umidade (%), razão (adimensional) e diferença entre velocidade média e máxima do vento (m/s).

Com o treinamento supervisionado, o conjunto de dados de entrada foi padronizado, conforme Equação (1).

$$z = \frac{(x-u)}{s} \quad (1)$$

Onde u é média das variáveis de treinamento, s é o desvio-padrão, x é a variável original e z é a variável padronizada. Os dados da variável de saída (velocidade máxima do vento) estão na mesma ordem de grandeza. Portanto, não foram padronizados para utilização nos modelos de *machine learning*.

4. APLICAÇÃO DOS MODELOS DE MACHINE LEARNING

Os modelos escolhidos foram Rede Neural Artificial (RNA) e *Extreme Gradient Boosting* (XGBoost). E, para comparação de desempenho a Regressão Linear (RL) foi adotada como algoritmo base. Todos os modelos foram desenvolvidos no ambiente *Google Colaboratory* utilizando a linguagem *Python* e as bibliotecas *scikit-learn*, *keras* e *xgboost*. Cada algoritmo possui variados parâmetros de funcionamento e otimização.

Para automatizar a busca pelos melhores parâmetros foi utilizado o método *Grid Search*. O qual tem por objetivo combinar todas as possibilidades, dentro de um conjunto de parâmetros, e retornar a melhor opção (Goodfellow et al., 2016). Esse método foi realizado para a RNA e o XGBoost com validação cruzada de duas subdivisões visando poupar recursos computacionais. Já para a RL foi mantida uma configuração padrão por ser o algoritmo base para comparação de desempenho entre os modelos.

5. REGRESSÃO LINEAR

A regressão linear calcula a relação entre duas ou mais variáveis, em que uma depende de outra ou outras. De maneira geral, o objetivo dessa abordagem é obter a variável dependente por meio das variáveis independentes. Nesse caso, a velocidade máxima do vento (dependente) é calculada conforme as variáveis independentes. A relação funcional de diferentes variáveis independentes com a variável dependente Y é dada pela Equação (2).

$$y = b_0 + b_1X_1 + \dots + b_kX_k \quad (2)$$

Onde, y é a variável alvo, b_0 é a constante quando as variáveis independentes são 0, b_1, b_2, \dots, b_k são os coeficientes e, X_1, X_2, \dots, X_k são os valores das variáveis independentes. A regressão linear foi considerada o modelo a ser superado, dada a sua simplicidade e ampla utilização.

6. REDE NEURAL ARTIFICIAL

A rede neural artificial (RNA) é uma técnica computacional semelhante à característica da rede neural biológica humana (Noorollahi; Jokar; Kalhor, 2016; Ma; Feng, 2016). Uma RNA é formada por unidades de processamento, camadas (entrada, ocultas e saída), neurônios e funções de ativação (Vafaeipour et al., 2014).

Na fase de treinamento o algoritmo *backpropagation* foi utilizado. Nele, a rede é treinada com dados que consistem em um conjunto de entrada e um conjunto de saída, no qual, os pesos (conexões neurais) são ajustados para minimizar o erro entre a saída real e a prevista (Noorollahi; Jokar; Kalhor, 2016).

A configuração de uma RNA necessita de mais parâmetros além do algoritmo de treinamento. Portanto, foram utilizados: arquitetura do tipo *Multilayer Perceptron*, uma camada de entrada, uma camada oculta com função de ativação ReLu (*Rectified Linear Unit*) que consiste na mais popular para redes neurais (Oostwal; Straat; Biehl, 2021), uma camada de saída com função de ativação linear, tamanho da amostra de dados que são utilizados nos cálculos dos pesos (*batch_size* = 512) e número de épocas de 1000.

Entretanto, muitas épocas podem levar ao *overfitting* (sobre ajustado) do conjunto de dados de treinamento, enquanto poucas podem resultar em um modelo *underfitting* (sub ajustado). A parada antecipada foi utilizada para interromper o treinamento assim que o desempenho do modelo pare de melhorar em um conjunto de dados de validação (Prechelt, 1998).

O número de neurônios na camada oculta é um fator fundamental para o desempenho da RNA. Assim, optou-se por realizar a otimização por meio do *Grid Search*,

variando a quantidade de neurônios na camada oculta de 2 a 100 em intervalos de dois, totalizando 48 redes neurais avaliadas para duas validações cruzadas.

7. O ALGORITMO XGBOOST

O XGBoost é um algoritmo de aprendizado de máquina baseado em árvore de decisão que usa uma estrutura de aumento de gradiente e seu detalhamento matemático pode ser encontrado em Chen e Guestrin (2016).

Os hiperparâmetros determinam e afetam o processo de aprendizagem do modelo de *machine learning*, para o XGBoost estão detalhados em Chen e Guestrin (2016) e em XGBoost (2022), e estes podem ser otimizados pelas mais variadas técnicas. Os intervalos dos parâmetros tiveram como base o trabalho de Dosdoğru e İpek (2022) e a documentação oficial disponível em XGBoost (2022). Os parâmetros dispostos na Tabela 1 correspondem a 98 modelos XGBoost treinados utilizando o *Grid Search*. O melhor modelo foi selecionado com base no menor erro calculado após a validação cruzada.

Tabela 1- Lista dos parâmetros analisados utilizando o método do *Grid Search*.

<i>Parâmetro</i>	<i>Modelo proposto</i>
<i>max_depth</i>	[3, 4, 5]
<i>learning_rate</i>	[0.01]
<i>n_estimators</i>	[1000, 2000]
<i>reg_alpha</i>	[0.067,1.1]
<i>reg_lambda</i>	[1, 1.2]
<i>Subsample</i>	[0.5]
<i>Gamma</i>	[0.8,1,5]
<i>min_child_weight</i>	[1.1,5]
<i>colsample_bytree</i>	[0.8]

Fonte: organizado pelos autores, 2023.

8. MÉTRICAS

O desempenho e a precisão de um determinado modelo podem ser avaliados por meio de várias métricas, as quais permitem a comparação entre diferentes modelos. Conforme Sun e Wang (2018) e Kato (2016) algumas das principais métricas são: Erro Quadrático Médio (*Mean Square Error - MSE*), Equação (3), Raiz do Erro Médio

Quadrático (*Root Mean Square Error* - RMSE), Equação (4), Erro Médio Absoluto (*Mean Absolute Error* - MAE), Equação (5). Chicco, Warrens e Jurman (2021) sugerem a adoção do coeficiente de determinação (R^2), Equação (6) para avaliação dos modelos de regressão.

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (\bar{X} - y_{i,observado})^2} \quad (6)$$

Sendo \bar{X} a média das observações e e_i definido pela Equação (7).

$$e_i = y_{i,observado} - y_{i,previsto} \quad (7)$$

Onde $y_{i,previsto}$ e $y_{i,observado}$ são os i -ésimos valores previstos e observados, respectivamente e e_i é o i -ésimo erro. Com $i = 1, \dots, N$ através de todo o conjunto de dados N . Além disso, ao utilizar o *Grid Search* determinou-se o menor MSE, encontrado para a base de dados de teste, como critério de seleção.

As métricas expressam o desvio da previsão em relação aos dados observados. Sendo que, quando menor o valor melhor os resultados. Destaca-se que a métrica MSE, e consequentemente, o RMSE, Equações (3) e (4), atribuem maiores pesos aos maiores erros de previsão. Para o R^2 , quanto mais próximo de 1, melhor o resultado. Nesse sentido, todas as métricas apresentadas, em conjunto com a distribuição de frequência dos erros obtidos pela Equação (7), foram adotadas para um melhor entendimento dos resultados.

9. RESULTADOS E DISCUSSÃO

Aplicando o método *Grid Search* determinou-se a melhor configuração com 56 neurônios na camada oculta da RNA. Já para o XGBoost o sumário dos parâmetros que otimizaram os resultados está apresentado na Tabela 2.

Tabela 2 - Resultados dos parâmetros conforme o *Grid Search* aplicado para o modelo XGBoost.

<i>Parâmetro</i>	<i>Modelo otimizado</i>
<i>max_depth</i>	4
<i>learning_rate</i>	0.01
<i>n_estimators</i>	1000
<i>reg_alpha</i>	1.1
<i>reg_lambda</i>	1
<i>Subsample</i>	0.5
<i>Gamma</i>	0.8
<i>min_child_weight</i>	1.1
<i>colsample_bytree</i>	0.8

Fonte: autores, 2023.

Com a otimização dos parâmetros foi possível realizar previsões com a base de testes e a de validação. A Tabela 3 apresenta os resultados estatísticos para ambas as bases de dados. Considerando a maior velocidade máxima medida, nenhum dos métodos reproduziu a mesma com precisão no período de testes, onde a RNA e o XGBoost subestimaram os valores e a RL superestimou (Tabela 3). Com os dados de validação, os valores da velocidade máxima previstos pela RNA e o XGBoost se aproximaram mais dos dados medidos (Linha 1), enquanto a RL não apresentou melhora.

Tabela 3 - Resultado da análise estatística para as bases de dados de teste/treino e validação (2020), comparando os dados reais e os três modelos de previsão escolhidos.

Velocidade [m/s]	Teste				Validação (2020)			
	Média	Desvio Padrão	Mínima	Máxima	Média	Desvio Padrão	Mínima	Máxima
Medidos	6,97	3,75	1,26	27,70	7,16	3,87	0,98	28,84
RL	7,04	3,06	-3,83	38,22	7,21	3,09	-4,49	41,87
RNA	7,08	3,32	2,18	18,58	7,23	3,41	1,69	23,12
XGBoost	7,03	3,37	1,90	18,08	7,19	3,44	1,85	20,49

Fonte: autores, 2023.

A Tabela 4 apresenta as métricas para os conjuntos de dados de teste e validação. Os valores de RMSE ficaram abaixo de 2 m/s para as RNA e XGBoost tanto no teste quanto para o ano de validação, isso demonstra que, mesmo a amostra extraída aleatoriamente da base denominada teste, os resultados foram próximos. Existem limitações ao comparar os resultados do presente trabalho com os obtidos por outros pesquisadores em diferentes regiões do mundo, uma vez que cada região tem sua particularidade geográfica e meteorológica. Contudo, os resultados obtidos por Yamaguchi e Ishihara (2021) para uma região no leste do Japão apresentaram RMSE entre 1,5 e 2 m/s para o horizonte de previsão de 1 hora.

Tabela 4 - Resultado das métricas calculadas em [m/s] para MAE e RMSE, R² adimensional.

Modelo	Teste			Validação		
	MAE	RMSE	R ²	MAE	RMSE	R ²
RL	1,66	2,17	0,652	1,71	2,29	0,649
RNA	1,13	1,60	0,802	1,17	1,71	0,805
XGB	1,11	1,60	0,812	1,15	1,69	0,810

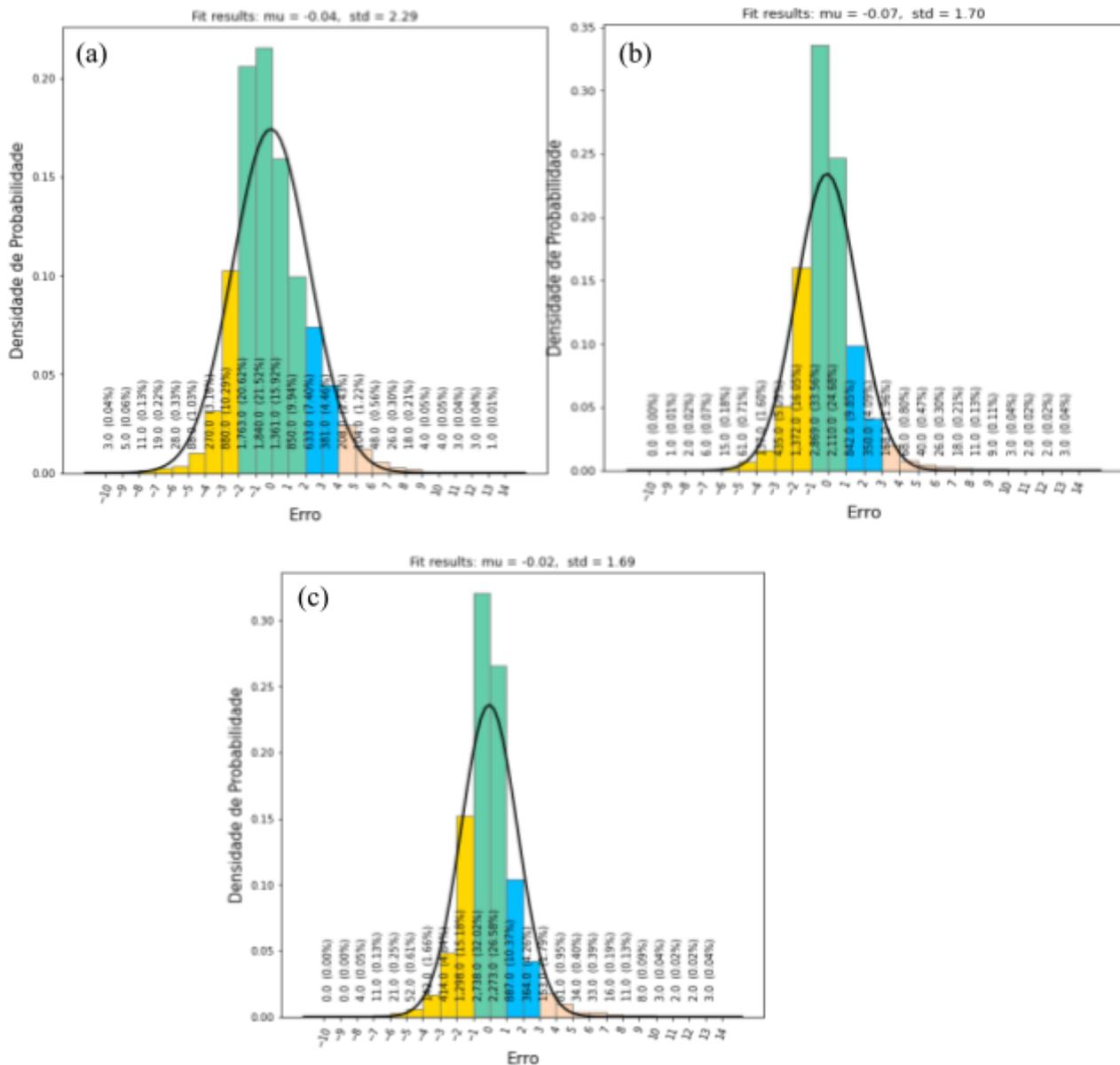
Fonte: autores, 2023.

Ao analisar as métricas, a regressão linear se apresentou como o método com menor capacidade de previsão para as variáveis selecionadas. A RNA e o XGBoost obtiveram desempenho similar em termos de métricas, ambos melhores que a RL, o que os torna mais promissores.

A Figura 2 demonstra os histogramas com a frequência dos resíduos (Equação 7) para o ano de 2020 (base de dados de validação). Para os valores positivos os modelos

realizaram previsões subestimadas em relação ao valor medido, e para os negativos a previsão foi acima do observado. Considerando ainda que, quanto maior a frequência de valores próximos de zero, melhores são os resultados das previsões.

Figura 2 - Histograma com os resíduos para o ano de 2020: (a) RL, (b) RNA, (c) XGBoost.



Fonte: autores, 2023.

Existem limitações nos modelos RNA e XGBoost quando se identifica que cerca de 16% da frequência dos erros é distribuída para os extremos (Figura 2b, c). Algumas hipóteses são levantadas e podem ter influenciado este comportamento: i) a região de

estudo apresenta grande variação de velocidades, com médias entre 5 e 7 m/s e máximas entre 19 e 28 m/s, ii) a quantidade e qualidade dos dados, devido as falhas nas medições das séries temporais. Principalmente, pela característica da velocidade máxima de ter um comportamento de bruscas mudanças conforme Santos, Lyra e Silva Júnior (2020).

Todavia, no caso da RL (Figura 2a) a frequência de acertos representa apenas 67,60%, limitando sua aplicação. A maior frequência de erros é entre -2 a 0 e entre 0 a 2 (m/s), para os modelos de RNA (Figura 2b) e XGBoost (Figura 2c), sendo esse intervalo de 84,14% e 84,15%, respectivamente. Tal comportamento demonstra os acertos das previsões indicando, em conjunto com as métricas, que para o curto prazo esses modelos podem ser recomendados. Os resultados demonstraram que o desempenho dos algoritmos de *machine learning* analisados no presente artigo foi promissor na previsão da velocidade máxima do vento com uma hora de antecedência.

10. CONSIDERAÇÕES FINAIS

Com a comparação dos algoritmos de *machine learning*, foi possível analisar o desempenho na previsão da velocidade máxima do vento, com uma hora de antecedência, para a região de estudo. Os resultados são relevantes, visto que eventos de vento com altas velocidades vêm se tornando cada vez mais comuns na região, causando estragos em sistemas elétricos e também podem vir a colocar em risco populações. Além disso, o crescimento exponencial da geração eólica e a possibilidade de geração no mar aumentam ainda mais a demanda por soluções de previsão de curto prazo das velocidades máximas de vento.

Dessa forma, destaca-se a relevância do presente trabalho o qual utilizou dados inéditos de uma estação meteorológica em zona costeira que sofre com a influência de fenômenos meteorológicos de grande intensidade de ventos. E, também, da possibilidade de futura geração *offshore* no local. Espera-se que as contribuições e o pioneirismo do estudo norteiem para a continuação e aperfeiçoamento dos algoritmos e métodos de *machine learning* focando na previsão da velocidade máxima dos ventos. Bem como, a

reprodução da metodologia em outras localidades do Brasil para comparação de desempenho.

REFERÊNCIAS

CAMELO, H. do N.; LUCIO, P. S.; LEAL JUNIOR, J. B. V.; CARVALHO, P. C. M. de. Proposta para Previsão de Velocidade do Vento Através de Modelagem Híbrida Elaborada a Partir dos Modelos ARIMAX e RNA. **Revista Brasileira de Meteorologia**, v. 33, p. 115–129, mar. 2018. <http://dx.doi.org/10.1590/0102-7786331005>.

CELESC. Centrais Elétricas de Santa Catarina. **Ciclone: Celesc restabelece sistema para cerca de 1 milhão de unidades**. 2020. Disponível em: <https://www.celesc.com.br/listagem-noticias/ciclone-celesc-restabelece-sistema-para-erca-de-1-milhao-de-unidades>. Acesso em: 23 abr. 2023.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. San Francisco California USA: ACM, 2016, p. 785–794. <http://dx.doi.org/10.1145/2939672.2939785>.

CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. **PeerJ Computer Science**, v. 7, p. e623, 5 jul. 2021. <http://dx.doi.org/10.7717/peerj-cs.623>.

CORRÊA, A. G. **Climatologia dos Ventos e Potencial Eólico Offshore de Santa Catarina**. 2018. 85 f. Dissertação (Mestrado), Curso de Oceanografia, Universidade Federal de Santa Catarina, Florianópolis, 2018.

DOSDOĞRU, A. T.; İPEK, A. B. Hybrid boosting algorithms and artificial neural network for wind speed prediction. **International Journal of Hydrogen Energy**, v. 47, n. 3, p. 1449–1460, jan. 2022. <http://dx.doi.org/10.1016/j.ijhydene.2021.10.154>.

FONSECA, S. B.; OLIVEIRA, R. C. L.; AFFONSO, C. M. Short-term Wind Speed Forecasting using Machine Learning Algorithms. 28 jun. 2021. **2021 IEEE Madrid PowerTech**, Madrid, Spain: IEEE, 28 jun. 2021. p. 1–6. <http://dx.doi.org/10.1109/powertech46648.2021.9494848>.

GOBATO, R.; HEIDARI, A. Cyclone Bomb Hits Southern Brazil in 2020. **Journal Of Atmospheric Science Research**, v. 3, n. 3, 4 ago. 2020. <http://dx.doi.org/10.30564/jasr.v3i3.2163>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge, Massachusetts: The MIT Press, 2016.

Google Earth website. **Plataforma de pesca Entremares em Balneário Arroio do Silva**. 2022. Disponível em: <https://earth.google.com/web/@-28.96237375,-49.37920088,-0.18856225a,806.36277371d,35y,0h,0t,0r>. Acesso em: 03 abr. 2023

KATO, T. Prediction of photovoltaic power generation output and network operation. **Integration Of Distributed Energy Resources In Power Systems**, [S.L.], p. 77-108, 2016. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-803212-1.00004-0>.

LEE, K; BOOTH, D; ALAM, P. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. **Expert Systems With Applications**, v. 29, n. 1, p. 1-16, jul. 2005. Elsevier BV. <http://dx.doi.org/10.1016/j.eswa.2005.01.004>.

LIU, X.; ZHANG, H.; KONG, X.; LEE, K. Y. Wind speed forecasting using deep neural network with feature selection. **Neurocomputing**, v. 397, p. 393-403, jul. 2020. <http://dx.doi.org/10.1016/j.neucom.2019.08.108>.

MA, Y.; FENG, S. Prediction and Forecast of Wind Power. **Large-Scale Wind Power Grid Integration**, [S.L.], p. 185-209, 2016. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-849895-8.00005-1>.

MALAKOUTI, S. M. Estimating the output power and wind speed with ML methods: a case study in Texas. **Case Studies in Chemical and Environmental Engineering**, v. 7, p. 100324, jun. 2023. Elsevier BV. <http://dx.doi.org/10.1016/j.cscee.2023.100324>.

MOVLIDAR-UFSC. **Base de Observação do Oceano e Atmosfera**. 2021. Disponível em: <https://movlidar.paginas.ufsc.br/estruturas/booa>. Acesso em: 13 abr. 2023

NASSIF, F. B. **A tecnologia LIDAR aplicada a medições eólicas sobre corpos hídricos e oceano**. 2017. 111 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia Mecânica, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2017.

NAVAS, R. K. B.; PRAKASH, S.; SASIPRABA, T. Artificial Neural Network based computing model for wind speed prediction: A case study of Coimbatore, Tamil Nadu, India. **Physica A: Statistical Mechanics and its Applications**, v. 542, p. 123383, mar. 2020. <http://dx.doi.org/10.1016/j.physa.2019.123383>.

NOOROLLAHI, Y.; JOKAR, M. A.; KALHOR, A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. **Energy Conversion and Management**, v. 115, p. 17-25, mai. 2016. <http://dx.doi.org/10.1016/j.enconman.2016.02.041>.

OOSTWAL, E.; STRAAT, M.; BIEHL, M. Hidden unit specialization in layered neural networks: ReLu vs. sigmoidal activation. **Physica A: Statistical Mechanics and its Applications**, v. 564, p. 125517, fev. 2021. <http://dx.doi.org/10.1016/j.physa.2020.125517>.

PATI, N.; GOURISARIA, M. K.; DAS, H.; BANIK, D. Wind Speed Prediction using Machine Learning Techniques. 28 abr. 2023. **2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP)**, Nagpur, India: IEEE, 28 abr. 2023. p. 1–6.
<http://dx.doi.org/10.1109/icetet-sip58143.2023.10151597>.

PAULA, M.; COLNAGO, M.; FIDALGO, J. N.; CASACA, W. Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine Learning Models. **IEEE Latin America Transactions**, [S. l.], v. 18, n. 11, p. 2011–2018, 2021.

PELÁEZ-RODRÍGUEZ, C.; PÉREZ-ARACIL, J.; FISTER, D.; PRIETO-GODINO, L.; DEO, R. C.; SALCEDO-SANZ, S. A hierarchical classification/regression algorithm for improving extreme wind speed events prediction. **Renewable Energy**, v. 201, p. 157–178, dez. 2022. <http://dx.doi.org/10.1016/j.renene.2022.11.042>.

PIRES, C. H. M. **Avaliação dos recursos eólicos com um LIDAR instalado em uma plataforma costeira do sul do Brasil**. 2019. 98 f. Dissertação (Mestrado) - Curso de Oceanografia, Universidade Federal de Santa Catarina, Florianópolis, 2019.

PRECHELT, L. Automatic early stopping using cross validation: quantifying the criteria. **Neural Networks**, v. 11, n. 4, p. 761–767, jun. 1998.
[http://dx.doi.org/10.1016/s0893-6080\(98\)00010-0](http://dx.doi.org/10.1016/s0893-6080(98)00010-0).

SAMET, H.; REISI, M.; MARZBANI, F. Evaluation of neural network-based methodologies for wind speed forecasting. **Computers & Electrical Engineering**, v. 78, p. 356–372, set. 2019. <http://dx.doi.org/10.1016/j.compeleceng.2019.07.024>.

SANTA CATARINA. Defesa Civil realiza “Live” sobre o “Ciclone Bomba” registrado em 2020. *Defesa Civil de Santa Catarina*, 2021. Disponível em: <https://www.defesacivil.sc.gov.br/noticias/defesa-civil-realiza-live-sobre-o-ciclone-bomba-registrado-em-2020/>. Acesso em: 06 abr. 2023.

SANTA CATARINA. Nota Meteorológica da Defesa Civil /SC - 16/05 – possível Formação da Tempestade Subtropical Yakecan nesta terça-feira (17). **Defesa Civil de Santa Catarina**, 2022. Disponível em: <http://www.defesacivil.sc.gov.br/noticias/notas-oficiais/nota-meteorologica-dc-sc-16-05-possivel-formacao-da-tempestade-subtropical-yakecan-nesta-terca-feira-17/> Acesso em: 10 abr. 2023.

SANTA CATARINA. Nota meteorológica da Defesa Civil /SC - 03/04 – Ciclone bomba provoca ressaca e alagamentos costeiros em Santa Catarina. **Defesa Civil de Santa Catarina**, 2023a. Disponível em: <https://www.defesacivil.sc.gov.br/noticias/nota-meteorologica-dc-sc-03-04-ciclone-bomba-provoca-ressaca-e-alagamentos-costeiros-em-santa-catarina/>. Acesso em: 04 out. 2023.

SANTA CATARINA. Monitoramento Meteorológico Defesa Civil /SC 13/07 10:40. **Defesa Civil de Santa Catarina, 2023b**. Disponível em: <https://www.defesacivil.sc.gov.br/previsao-do-tempo/monitoramento-meteorologico-d-c-sc-13-07-1020/>. Acesso em: 04 out. 2023.

SANTOS, I. G. DA S.; LYRA, R. F. DA F.; SILVA JÚNIOR, R. S. DA. Comparativo de Prognósticos da Velocidade do Vento Utilizando Modelo WRF e Rede Neural Artificial. **Revista Brasileira de Meteorologia**, v. 35, n. spe, p. 1017-1027, dez. 2020. <http://dx.doi.org/10.1590/0102-77863550103>.

SUN, W.; WANG, Y. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. **Energy Conversion and Management**, v. 157, p. 1-12, fev. 2018. <http://dx.doi.org/10.1016/j.enconman.2017.11.067>.

VAFAEIPOUR, M.; RAHBARI, O.; ROSEN, M. A.; FAZELPOUR, F.; ANSARIRAD, P. Application of sliding window technique for prediction of wind velocity time series. **International Journal of Energy and Environmental Engineering**, v. 5, n. 2-3, p. 105, 2014. <http://dx.doi.org/10.1007/s40095-014-0105-5>.

XGBOOST. **XGBoost Parameters**. 2022. Disponível em: <https://xgboost.readthedocs.io/en/latest/parameter.html#general-parameters>. Acesso em: 20 abr. 2023.

YAMAGUCHI, A.; ISHIHARA, T. Maximum Instantaneous Wind Speed Forecasting and Performance Evaluation by Using Numerical Weather Prediction and On-Site Measurement. **Atmosphere**, v. 12, n. 3, p. 316, 28 fev. 2021. <http://dx.doi.org/10.3390/atmos12030316>.

ZUCATELLI, P.J.; NASCIMENTO, E.G.s.; AYLAS, G.y.R.; SOUZA, N.B.P.; KITAGAWA, Y.K.L.; SANTOS, A.A.B.; ARCE, A.M.G.; MOREIRA, D.M. Short-term wind speed forecasting in Uruguay using computational intelligence. **Heliyon**, [S.L.], v. 5, n. 5, e01664, maio 2019. <http://dx.doi.org/10.1016/j.heliyon.2019.e01664>.