



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS ARARANGUÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENERGIA E SUSTENTABILIDADE

Edvana da Silva da Silveira

Machine Learning e a Previsão de Curto Prazo da Velocidade do Vento Costeiro

Araranguá

2022

Edvana da Silva da Silveira

***Machine Learning* e a Previsão de Curto Prazo da Velocidade do Vento Costeiro**

Dissertação submetida ao Programa de Pós-Graduação em Energia e Sustentabilidade da Universidade Federal de Santa Catarina para a obtenção do Título de Mestre em Energia e Sustentabilidade.

Orientadora: Profa. Dra. Carla de Abreu D'Aquino

Araranguá

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Silveira, Edvana da Silva da
Machine Learning e a Previsão de Curto Prazo da
Velocidade do Vento Costeiro / Edvana da Silva da Silveira
; orientadora, Carla de Abreu D'Aquino, 2022.
89 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Campus Araranguá, Programa de Pós-Graduação em
Energia e Sustentabilidade, Araranguá, 2022.

Inclui referências.

1. Energia e Sustentabilidade. 2. Energia Eólica. 3.
Machine Learning. 4. Previsão. 5. Vento Costeiro. I.
D'Aquino, Carla de Abreu. II. Universidade Federal de
Santa Catarina. Programa de Pós-Graduação em Energia e
Sustentabilidade. III. Título.

Edvana da Silva da Silveira

***Machine Learning* e a Previsão de Curto Prazo da Velocidade do Vento Costeiro**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Dr. Giuliano Arns Rampinelli
Universidade Federal de Santa Catarina

Prof. Dr. Felipe Mendonça Pimenta
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Energia e Sustentabilidade.

Prof. Dr. Tiago Elias Allievi Frizon
Coordenação do Programa de Pós-Graduação

Profa. Dra. Carla de Abreu D'Aquino
Orientadora

Araranguá, 2022.

AGRADECIMENTOS

Presto meus agradecimentos a todos que ajudaram, de alguma forma, na realização e conclusão deste trabalho.

Ao meu companheiro de pesquisa e vida, Políbio. Obrigada por sempre incentivar, e, proporcionar todo o suporte e força que eu precisava para este momento. Agradeço por todos os abraços apertados, pela partilha, e, por me aturar nos momentos de ansiedade. Sei que na nossa vida nunca foi nada fácil, mas, se eu consegui escrever e desenvolver essa dissertação foi porque você esteve ao meu lado!

À Profa Carla D'Aquino, por acreditar nesse trabalho e aprender sobre o mundo de *machine learning* junto comigo. Não poderia deixar de reconhecer sua importância para que eu encarasse esse desafio. Sou grata pela sua amizade e ensinamentos que levarei para o resto da vida. Obrigada por sua compreensão, motivação e por suas excelentes contribuições para o desenvolvimento desta Dissertação.

À Universidade Federal de Santa Catarina - Campus Araranguá, em especial ao Programa de Pós-graduação em Energia e Sustentabilidade (PPGES), pela oportunidade de estudo e realização deste trabalho.

Aos professores e técnicos administrativos do PPGES.

À equipe da EPAGRI e BOOA/UFSC pelo fornecimento dos dados, com menção especial ao João Casemiro (UFSC), Prof. Felipe Pimenta (UFSC) e Andrey Abreu (EPAGRI), pelo apoio na realização deste trabalho.

Aos responsáveis pela plataforma de pesca Entremares - Balneário Arroio do Silva/SC.

Aos colegas do que ingressam comigo no PPGES, Aline e Marcos.

Agradeço aos membros da banca, prof. Dr. Giuliano Arns Rampinelli e prof. Dr. Felipe Mendonça Pimenta, pela imensa contribuição dada para este trabalho desde a qualificação. Aos professores Luciano Lopes Pfitscher, Arcilan Trevenzoli Assireu e Tiago Elias Allievi Frizon, pela disponibilidade e participação.

Agradeço à FAPESC pela bolsa de estudos concedida.

"Não espere o futuro mudar sua vida. Porque o futuro será a consequência do presente."
(Racionais MC's)

RESUMO

As concessionárias e o operador do sistema precisam de previsões e estimativas quanto ao recurso eólico para assim, gerenciar o despacho antecipado de eletricidade das usinas tradicionais, como por exemplo, as termelétricas ou hidrelétricas com reservatórios de acumulação. Dessa forma, o objetivo geral desse trabalho é analisar o desempenho de algoritmos de *Machine Learning* na previsão da velocidade do vento costeiro, com uma hora de antecedência, utilizando três algoritmos: *Extreme Gradient Boosting* (XGBoost), Redes Neurais Artificiais e Regressão Linear Multivariável. Para o desenvolvimento do estudo, foram utilizados dados coletados por um anemômetro de copo instalado na Base de Observação do Oceano e Atmosfera (BOOA), construída sobre uma plataforma de pesca localizada no Balneário Arroio do Silva, Santa Catarina. Foram utilizados cerca de 25000 dados horários correspondentes ao período de 2018 a 2021. A metodologia foi dividida nas seguintes etapas: tratamento e preparação do conjunto de dados, treinamento, aplicação e avaliação dos modelos e, seleção do melhor modelo e sua validação para o ano de 2020. O modelo utilizando XGBoost apresentou melhores resultados para a região de estudo. Sua acurácia foi avaliada considerando métricas de desempenho, análise estatística e a investigação de sua capacidade frente a um evento extremo, o Ciclone Bomba. Destacam-se o valor de MAE (*Mean Absolute Error*) menores que 1,00 m/s, MBE (*Mean Bias Error*) próximos de zero, e, também valores de RMSE (*Root Mean Square Error*) próximos de 1,00 m/s. Assim, o modelo apresentou capacidade de realizar previsões promissoras com antecedência de 1 hora. Essas, são muito importantes para a setor de energia eólica, principalmente quando inserida em locais onde ocorrem diferentes fenômenos extremos, especialmente no Sul do Brasil, e, considerando a presente mudança climática global.

Palavras-chave: energia eólica; velocidade do vento; previsão de curto prazo; *machine learning*; XGBoost; redes neurais artificiais; regressão linear multivariável; Python.

ABSTRACT

The concessionaires and the system operator need forecasts and estimates regarding the wind resource in order to manage the anticipated dispatch of electricity from traditional plants, such as thermoelectric plants or hydroelectric plants with accumulation reservoirs. Thus, the general objective of this work is to analyze the ability to predict the coastal wind speed, one hour ahead, using three Machine Learning algorithms: Extreme Gradient Boosting (XGBoost), Artificial Neural Networks and Multivariable Linear Regression. For the development of the study, were used data collected by a cup anemometer installed at the Ocean and Atmosphere Observation Base (BOOA), built on a fishing platform located in Balneário Arroio do Silva, Santa Catarina. Near 25,000 hourly data corresponding to the period from 2018 to 2021 were used. The methodology was divided into the following steps: cleaning and preparation of the data set, training, application and evaluation of the models, and selection of the best model and its validation for the year 2020. The model using XGBoost showed better results for the study case. Its accuracy was evaluated considering performance metrics, statistical analysis and the investigation of its capacity in the face of an extreme event, Bomb Cyclone. Some highlights are the value of MAE (Mean Absolute Error) lower than 1.00 m/s, MBE (Mean Bias Error) close to zero, and also values of RMSE (Root Mean Square Error) close to 1.00 m/ s. Thus, the model was able to make promising predictions 1 hour ahead. These are very important for the wind energy sector, especially when inserted in places where different extreme phenomena occur, especially in southern region of Brazil, and considering the present global climate change.

Keywords: wind power; wind speed; short term forecast; machine learning; XGBoost; artificial neural networks; multivariable linear regression; Python.

LISTA DE FIGURAS

Figura 1 - Distribuição dos empreendimentos eólicos em operação no Sul do Brasil.	15
Figura 2 - Balanço de energia da Região Sul do Brasil.	15
Figura 3 - Estimativa de velocidade do vento costeiro e offshore no Sul do Brasil.	16
Figura 4 - Representação esquemática do neurônio artificial.	25
Figura 5 - Estrutura típica de um modelo de rede neural artificial.	25
Figura 6 - (A) RNA feedforward e (B) arquitetura recorrente RNA.	27
Figura 7 - Processamento do algoritmo backpropagation.	28
Figura 8 - Exemplo de uma árvore de decisão.	30
Figura 9 - Exemplo XGBoost.	31
Figura 10 - Local de estudo.	36
Figura 11 - (A) Plataforma de Pesca (B) Base de Observação do Oceano e Atmosfera.	37
Figura 12 - Etapas do desenvolvimento da metodologia da pesquisa.	37
Figura 13 - Fluxograma Cross Validation.	40
Figura 14 - Gráfico de dispersão base de treino.	45
Figura 15 - Previsão da velocidade do vento para Março de 2020.	48
Figura 16 - Resíduos XGBoost: Março de 2020.	49
Figura 17 - Previsão da velocidade do vento para Agosto de 2020.	50
Figura 18 - Resíduos XGBoost: Agosto de 2020.	50
Figura 19 - Previsão da velocidade do vento para Setembro de 2020.	51
Figura 20 - Resíduos XGBoost: Setembro de 2020.	51
Figura 21 - Previsão com 50 dados horários considerando a base de dados de validação para o modelo selecionado (2020).	53
Figura 22 - Observações e previsões do Ciclone Bomba.	54
Figura 23 - Gráfico de dispersão (R^2) para o melhor modelo (M4) utilizando XGBoost.	54
Figura 24 - Resíduos para o banco de dados de validação com o algoritmo XGBoost.	55
Figura 25 - Curva de potência típica de um aerogerador.	56
Figura 26 - Resíduos para o banco de dados de validação utilizando o algoritmo XGBoost para a faixa de operação de uma turbina.	57

LISTA DE QUADROS

Quadro 1 - Classificação com base na escala de tempo.	17
Quadro 2 - Parâmetros de entrada para previsão eólica utilizando machine learning....	22
Quadro 3 - Parâmetros XGBoost.....	32
Quadro 4 - Métricas de avaliação de desempenho da previsão eólica.	33
Quadro 5 - Conjuntos de variáveis de entrada para os modelos de machine learning. ...	39
Quadro 6 - Quantidade de dias utilizados da base de dados de validação - mensal.....	44

LISTA DE TABELAS

Tabela 1 - Exemplo de entrada e saída para o treino dos modelos com horizonte de previsão de 1 hora.	38
Tabela 2 - Lista dos parâmetros analisados utilizando o método do GridSearchCV.	42
Tabela 3 - Análise exploratória das variáveis referentes ao banco de dados horários de treinamento.	43
Tabela 4 - Análise exploratória do conjunto de dados de 2020.	43
Tabela 5 - Resultados das métricas para a base de dados de teste com diferentes modelos de machine learning e variáveis de entrada.	45
Tabela 6 - Resultados dos parâmetros conforme o GridSearchCV aplicado para o modelo selecionado (M4) XGBoost.	46
Tabela 7 - Resultados da análise estatística mensal com base no ano de validação (2020).	46
Tabela 8 - Resultados de métricas mensais – validação (2020).	47
Tabela 9 - Resultados da análise estatística para a base de dado de validação (2020), comparando os dados reais para o modelo de previsão selecionado.	52
Tabela 10 - Resultados das métricas anuais para 2020 calculadas para o modelo selecionado.	52
Tabela 11 - Faixa de velocidades baseada na curva de potência - ano de 2020.	56

LISTA DE ABREVIATURAS E SIGLAS

ABEEólica - Associação Brasileira de Energia Eólica
AR - *Auto Regressive*
ARIMA - *Auto Regressive Integrated Moving Average*
ARMA - *Auto Regressive Moving Average*
BOOA - Base de Observação do Oceano e Atmosfera
CIRAM - Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina
EPAGRI - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina
EPE - Empresa de Pesquisa Energética
GWEC - *Global Wind Energy Council*
IA - Inteligência Artificial
IBGE - Instituto Brasileiro de Geografia e Estatística
IEA - *International Energy Agency*
LIDAR - *Light Detection and Ranging*
MAAPE - *Mean Arctangent Absolute Percentage Error*
MAE – *Mean Absolute Error*
MAPE - *Mean Absolute Percentage Error*
MBE – *Mean Bias Error*
ML - *Machine Learning*
MLP- *Multilayer Perceptron*
ONS - Operador Nacional do Sistema
ReLU - *Rectified Linear Unit*
RLM - Regressão Linear Multivariável
RMSE – *Root Mean Square Error*
RNA – Rede Neural Artificial
XGBOOST - *Extreme Gradient Boosting*
UFSC – Universidade Federal de Santa Catarina

SUMÁRIO

1 INTRODUÇÃO	11
1.1 OBJETIVOS	13
1.1.1 Objetivo Geral	13
1.1.2 Objetivos específicos	13
2 REFERENCIAL TEÓRICO	14
2.1 ENERGIA EÓLICA	14
2.2 MÉTODOS DE PREVISÃO DA VELOCIDADE DO VENTO	17
2.3 MÉTODOS DE APRENDIZADO DE MÁQUINA.....	21
2.3.1 Regressão Linear Multivariável (RLM)	23
2.3.2 Redes Neurais Artificiais	24
2.3.2.1 Estrutura e treinamento da rede neural artificial.....	26
2.3.3 Algoritmo XGBoost	30
2.4 MÉTRICAS PARA AVALIAÇÃO DO DESEMPENHO DAS PREVISÕES	33
3 METODOLOGIA	36
3.1 TRATAMENTO E PREPARAÇÃO DO CONJUNTO DE DADOS	38
3.2 TREINAMENTO, APLICAÇÃO E AVALIAÇÃO DOS MODELOS	40
3.2.1 Regressão Linear Multivariável	41
3.2.2 Parâmetros da Rede Neural Artificial	41
3.2.3 Parâmetros do XGBoost	42
3.3 VALIDAÇÃO E AVALIAÇÃO DO MODELO SELECIONADO.....	42
4 RESULTADOS E DISCUSSÃO	43
5 CONSIDERAÇÕES FINAIS	58
REFERÊNCIAS	60
ANEXO A - Descrição matemática da função objetivo -XGBoost com base nos trabalhos de Chen (2014) e Chen e Guestin (2016)	74

APÊNDICE A - Previsão da velocidade do vento e resíduos para Janeiro de 2020	
.....	77
APÊNDICE B - Previsão da velocidade do vento e resíduos para Fevereiro de 2020	
.....	78
APÊNDICE C - Previsão da velocidade do vento e resíduos para Abril de 2020	79
APÊNDICE D - Previsão da velocidade do vento e resíduos para Maio de 2020	80
APÊNDICE E - Previsão da velocidade do vento e resíduos para Junho de 2020	
.....	81
APÊNDICE F - Previsão da velocidade do vento e resíduos para Julho de 2020	82
APÊNDICE G - Previsão da velocidade do vento e resíduos para Outubro de 2020	
.....	83
APÊNDICE H - Previsão da velocidade do vento e resíduos para Novembro de 2020	
.....	84
APÊNDICE I - Previsão da velocidade do vento e resíduos para Dezembro de 2020	
.....	85

1 INTRODUÇÃO

De acordo com a publicação dos Indicadores de Desenvolvimento Sustentável, divulgada pelo Instituto Brasileiro de Geografia e Estatística - IBGE (2015), a produção, o consumo e os subprodutos decorrentes da oferta de energia impactam o meio ambiente e os recursos naturais. Conforme a publicação, para um desenvolvimento sustentável, é necessário atender às demandas energéticas com o uso de fontes renováveis, conciliando a oferta de energia com a conservação do meio ambiente, além de atender a demanda energética de maneira confiável.

O Balanço Energético Nacional, publicado e elaborado pela Empresa de Pesquisa Energética - EPE (2022) - Ano Base 2021, afirma que o Brasil tem 78,1% de sua energia elétrica gerada por fontes renováveis. A energia eólica aumentou sua participação na matriz elétrica brasileira que em 2020 era 8,7% para 10,6% em 2021, e, a geração de energia oriunda dessa fonte teve um crescimento de 26,7%, correspondendo a 72,29 TWh (EPE, 2022).

Conforme dados da IEA (*International Energy Agency*) (2021), no ano de 2020, a geração de eletricidade proveniente da energia eólica *onshore* (continental) aumentou 144 TWh em relação ao ano anterior (2019). O incremento da capacidade eólica foi de 108 GW, principalmente devido ao comissionamento de diversos parques na China e nos Estados Unidos. No ano de 2020 foram gerados 1592 TWh de eletricidade a partir de instalações eólicas *onshore/offshore* (continental/marinha), sendo a principal fonte renovável não hídrica no mundo.

No Brasil, os dados do “Infovento” nº 25 – 2022, elaborados e publicados pela ABEEólica (Associação Brasileira de Energia Eólica), apontam que o país terá uma capacidade eólica instalada de aproximadamente 34,69 GW até 2026. O fator de capacidade médio brasileiro, ou seja, a proporção de geração efetiva em um intervalo de tempo e capacidade total no mesmo período, foi de 40,6% no ano de 2020 (ABEEÓLICA, 2020).

Conforme o trabalho de Corrêa (2018), a região de maior potencial e viabilidade de implantação de um parque eólico em Santa Catarina está localizada na região sul oceânica, próxima de Araranguá. Conforme resultados apresentados por Pires (2019), a densidade de potência média para a zona costeira na cidade de Balneário Arroio do Silva, foi de 336,8 W/m² em 2017 e de 278,5 W/m² em 2018.

A geração de energia eólica pressupõe a existência de condições naturais específicas e favoráveis. Conforme Trentini (2016), o vento é variável no tempo e espaço, sendo influenciado por diferentes fatores físicos e climáticos, que afetam sua intensidade, direção e velocidade, tais

como: relevo da região, presença de obstáculos, rugosidade, variação da velocidade com a altura.

Nesse contexto, apesar da energia eólica possuir ativos atraentes para as demandas crescentes de energia, ela possui comportamento intermitente. De acordo com Samet, Reisi e Marzbani (2019), a intermitência drástica do vento ocasiona dificuldades para os operadores do sistema elétrico quanto ao despacho de energia e planejamento do mercado. Conforme Zucatelli et al. (2019), a previsão do recurso eólico de forma eficaz é importante no planejamento, controle e monitoramento dos sistemas de energia eólica. Samet, Reisi e Marzbani (2019) apontam que uma vantagem da previsão da velocidade do vento é a possibilidade de comercialização no mercado de energia em tempo real.

Assim, com o crescimento da geração eólica no Brasil, é relevante e importante a previsão em diferentes horizontes de previsão para conhecer os impactos na operação, manutenção e interligação com o sistema elétrico existente visto à natureza intermitente da energia eólica (BULHÕES, 2015). Nos estudos para implementação de empreendimentos eólicos, sabe-se que a geração de energia depende da velocidade do vento incidente no aerogerador e sua potência é proporcional ao cubo da velocidade, e conseqüentemente, pequenas alterações na estimativa do vento provocam grandes mudanças na previsão de geração de energia (PICOLO; BÜHLER; RAMPINELLI, 2014; NOOROLLAHI; JOKAR; KALHOR, 2016).

Nesse sentido, Alencar (2018) afirma que é crucial o desenvolvimento de métodos para realizar a previsão da velocidade do vento para a geração de energia, de forma que possibilite sua inserção na matriz elétrica com segurança e confiabilidade. Conforme as características climáticas e físicas particulares de cada região, que afetam os parâmetros de intensidade, direção e velocidade do vento, é relevante o desenvolvimento de ferramentas que considerem as características locais e que realizem a previsão de velocidade do vento com elevado nível de confiança no período de interesse.

Dentre os horizontes de previsão, a previsão de curto prazo, no qual o horizonte de 1 hora está inserido, é destaque devido às operações de mercado e despacho de energia (DHIMAN; DEB; BALAS, 2020a). Bulhões (2015) afirma que com a previsão do recurso eólico, as usinas de energia convencionais como por exemplo, as termelétricas ou hidrelétricas com reservatórios de acumulação, poderão agendar a produção das centrais elétricas para atender à demanda energética solicitada. O autor aponta que é necessário prever a parcela de contribuição da geração eólica, em um determinado horizonte, para saber quanto de energia será gerada com as usinas convencionais.

Além disso, de maneira geral, horizontes de previsão de curtíssimo e curto prazo podem fornecer resultados precisos e com mais detalhes, enquanto que um horizonte de previsão maior fornece informações de longo prazo sobre a previsão de energia eólica, entretanto, sua capacidade de produzir resultados precisos é relativamente baixa (SILVA et al., 2021).

Diante do cenário eólico nacional e da crescente necessidade do desenvolvimento de técnicas cada vez mais precisas de previsão de velocidades do vento, do favorável cenário para prospecção eólica no litoral sul catarinense, o presente trabalho pretende contribuir com essas demandas, utilizando ferramentas computacionais de aprendizado de máquina (*machine learning*) para prever as velocidades do vento. Além disso, a presente pesquisa é pioneira na região do Sul de Santa Catarina, que conforme Pires (2019), possui potencial eólico costeiro para exploração de energia eólica. A seguir serão descritos os objetivos de forma mais detalhada. Essa dissertação está organizada nos tópicos: referencial teórico, metodologia, resultados e discussão e conclusões.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Analisar o desempenho de algoritmos de *machine learning* na previsão da velocidade do vento costeiro, com uma hora de antecedência, utilizando dados de uma estação meteorológica localizada no Extremo Sul de Santa Catarina.

1.1.2 Objetivos específicos

1. Identificar na literatura os diferentes modelos de previsão da velocidade do vento.
2. Analisar o conjunto de dados coletados por uma estação meteorológica costeira.
3. Aplicar diferentes modelos de *machine learning* no banco de dados disponível.
4. Avaliar a performance dos modelos.
5. Selecionar o melhor modelo de *machine learning* e analisar seu desempenho.

2 REFERENCIAL TEÓRICO

2.1 ENERGIA EÓLICA

Os ventos são produzidos pelo aquecimento desigual da superfície terrestre através do Sol, que produz variações na pressão atmosférica, provocando o deslocamento das camadas de ar, as quais se movem de região de alta para baixa pressão (LYNN, 2012). A atmosfera terrestre é aquecida de maneira não uniforme devido ao fato de que a radiação solar incidente sobre a superfície da Terra não se distribui igualmente e, em decorrência dos movimentos de rotação do planeta (PICOLO; BÜHLER; RAMPINELLI, 2014). Assim, o vento é o movimento horizontal das parcelas de ar em resposta a força do gradiente de pressão horizontal.

Estima-se que cerca de 2% da energia solar absorvida pela Terra é transformada em energia cinética dos ventos (CRESESB, 2008). Dessa forma, a energia eólica é uma fonte sustentável para a geração de energia elétrica, pois aproveita um recurso natural (vento), de grande disponibilidade e utiliza-o de maneira renovável (CRUZ, 2016).

De acordo com o *Global Wind Report 2022* publicado pelo Global Wind Energy Council (GWEC) (ano base - 2021), o Brasil ocupa a 6ª posição no *ranking* internacional de capacidade instalada de energia eólica *onshore*. Em relação a capacidade eólica instalada no território brasileiro, 46% dos parques eólicos estão localizados a cerca de 25km da costa e 26%, na faixa de até 5 km do litoral (BRANNSTROM et al., 2018).

Nesse contexto, conforme a Associação Brasileira de Energia Eólica (ABEEólica), no boletim “Infovento” 2022, a geração de energia eólica do Brasil apresentava, em março de 2022, 21,80 GW de capacidade instalada com 805 parques eólicos distribuídos em 12 Estados. De acordo com a publicação, o Sul do País contém 96 parques eólicos, sendo esses: 80 no Rio Grande do Sul, 15 parques em Santa Catarina e 1 parque no Paraná, Figura 1.

Conforme os dados do Sistema de Informações de Geração da ANEEL (SIGA) (2022), especificamente no estado de Santa Catarina as usinas eólicas em operação estão localizadas nas seguintes cidades: Água Doce, Bom Jardim da Serra, Laguna, São José do Cedro e Tubarão. E, ainda, os dados do SIGA apontam que na região Sul do país 3 empreendimentos estão em construção, ambos em Rio Grande - Rio Grande do Sul, e, 108 empreendimentos em estudo (Requerimento de Outorga).

Figura 1 - Distribuição dos empreendimentos eólicos em operação no Sul do Brasil.



Fonte: ANEEL (2022).

Segundo o ONS (Operador Nacional do Sistema Elétrico) (2022a), nos últimos anos, houve um forte crescimento no número de instalações eólicas, principalmente nas regiões Nordeste e Sul do Brasil, aumentando a importância dessa geração para o mercado de energia. Ainda, conforme os dados de Balanço de Energia atualizados em um curto período de tempo no site oficial do ONS (2022b), do dia 11 de maio de 2021, às 19 h e 22 min, a região Sul do Brasil, gerou 14537,0 MW, sendo a geração proveniente da fonte eólica teve uma contribuição de 948,3 MW, conforme Figura 2.

Figura 2 - Balanço de energia da Região Sul do Brasil.



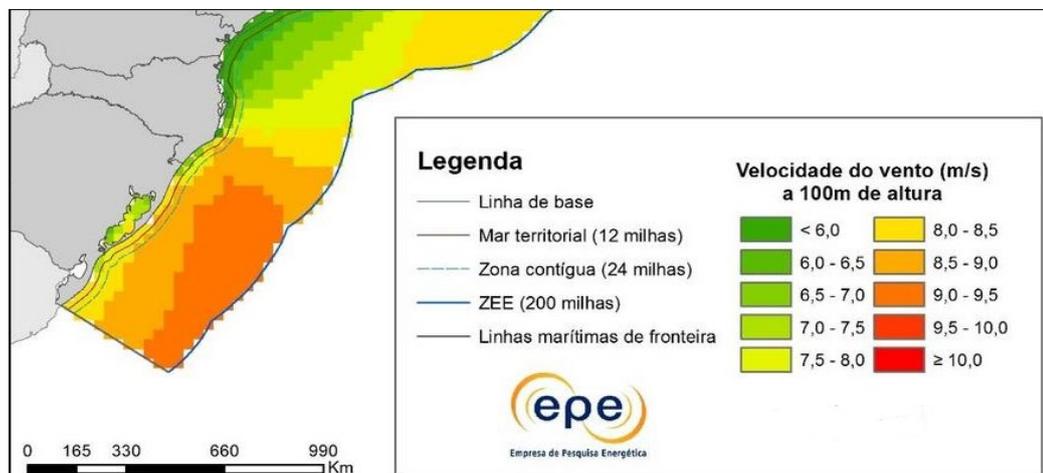
Fonte: ONS (2022b).

Entretanto, existem poucos estudos relativos ao setor eólico no estado de Santa Catarina, em específico na região do extremo sul. Pires et al. (2020) realizaram um estudo do potencial

eólico na zona costeira no Extremo Sul catarinense com o uso de dados de um perfilador de vento LIDAR instalado sobre a Base de Observação do Oceano e Atmosfera no Balneário Arroio do Silva. Os autores utilizaram três modelos de turbinas eólicas para a simulação da produção de energia no local. Corrêa (2018), identificou as áreas de maior potencial para exploração da energia eólica *offshore* e a viabilidade utilizando análise batimétrica em Santa Catarina.

Conforme o primeiro *Roadmap* da Energia Eólica *Offshore* (EPE, 2020), as regiões com velocidade acima de 7 m/s são consideradas atrativas para o setor. Cabe destacar que o relatório abrange toda a região costeira/*offshore*. Observa-se que no Sul do Brasil, em específico no Extremo Sul Catarinense, Figura 3, o recurso eólico é uma das opções promissoras para geração de energia eólica, a região atinge velocidade média entre 7,5 e 8,0 m/s para a zona costeira e 8,5 e 9,0 m/s para *offshore* na altura de 100 m. Segundo Pires (2019), o potencial eólico é mais abundante nas regiões costeiras e oceânicas, também devido a fatores como: ausência de obstáculos e baixa rugosidade da superfície.

Figura 3 - Estimativa de velocidade do vento costeiro e *offshore* no Sul do Brasil.



Fonte: EPE (2020).

O conhecimento da velocidade do vento é requisito indispensável para a localização e instalação de turbinas eólicas, conseqüentemente, o uso da energia eólica requer um detalhamento das características do vento local (VELO; LÓPEZ; MASEDA, 2014). Um desafio tecnológico para expansão desse setor é estimar o recurso eólico, principalmente devido a sua natureza sazonal e intermitente. A previsão da velocidade do vento desempenha um papel importante no planejamento e gerenciamento da rede elétrica, auxiliando nas operações de despacho (WANG; YANG, 2021).

2.2 MÉTODOS DE PREVISÃO DA VELOCIDADE DO VENTO

Os modelos matemáticos e computacionais são utilizados para simulação teórica, avaliação de recursos exploratórios e previsão de ventos. Conforme Kumar, Pal e Tripathi (2020), a previsão é um método para prever um valor futuro utilizando dados históricos. Os autores dividem a previsão nas seguintes categorias:

- Baseada na escala temporal;
- Baseada em dados de observações históricas / meteorológicos;
- Baseada em técnicas / modelos físicos e matemáticos.

Conforme o levantamento bibliográfico realizado, não existe um consenso na literatura sobre a classificação dos métodos de previsão com relação a escala temporal. Nesse sentido, foi utilizada a classificação proposta por Letcher (2017), a qual, a previsão de energia eólica baseada na escala temporal, também chamada de horizonte de previsão, é mostrada no Quadro 1.

Quadro 1 - Classificação com base na escala temporal.

Tipo de previsão	Intervalo de tempo	Aplicações
Curtíssimo prazo	Entre alguns segundos e antes de 30 minutos a frente.	Compensação do mercado de eletricidade; regulação do setor elétrico.
Curto prazo	Entre 30 minutos e antes de 6 horas a frente	Despacho econômico de carga, Decisões de incremento / decrémento de carga.
Médio prazo	Entre 6 horas e antes de um dia	Decisões relacionadas ao gerador e segurança operacional no mercado do dia seguinte.
Longo prazo	Entre um dia e uma semana ou mais.	Auxiliar nos processos de tomada de decisão, programação de manutenção.

Fonte: Letcher (2017).

A previsão de energia eólica com horizonte de curto prazo é um campo de pesquisa importante para o setor de energia. Nesse contexto, Rodríguez et al. (2020) afirmam que o controle da microrrede de energia elétrica é melhorado com a previsão da densidade de energia/velocidade do vento. Visto que os operadores do sistema devem lidar com uma quantidade importante de energia intermitente, decorrente do aumento da capacidade instalada de energia eólica.

A categoria de previsão baseada em modelos/técnicas contempla diversas abordagens. Conforme Kumar, Pal e Tripathi (2020), diferentes técnicas podem ser adotadas no campo de previsão de energia eólica: persistência, estatística, modelo ARMA (*Auto Regressive Moving Average*), ARIMA (*Auto Regressive Integrated Moving Average*), modelo de previsão

numérica e métodos computacionais (Redes Neurais Artificiais (RNA), *Support Vector Machine*, lógica *Fuzzy*, *Support Vector Regression*), entre outros.

Nos métodos baseados em persistência, o valor futuro permanecerá inalterado quando comparado ao medido no presente (COIMBRA; PEDRO, 2013). O desempenho e a precisão utilizando essa abordagem diminui rapidamente com o incremento da escala temporal (JUNG; BROADWATER, 2014; HANIFI et al., 2020).

Samet, Reisi e Marzbani (2019) classificam os modelos de previsão eólica existentes em cinco principais grupos: correlação espacial, modelos matemáticos, métodos estatísticos, inteligência artificial e métodos híbridos. Conforme Chang (2014), existem diferentes abordagens estatísticas empregadas para a previsão da velocidade do vento: autorregressiva (AR, do inglês *Auto Regressive*), média móvel autorregressiva (ARMA, do inglês *Auto Regressive Moving Average*), média móvel integrada autoregressiva (ARIMA, do inglês *AutoRegressive Integrated Moving Average*), dentre outras.

Os modelos estatísticos são adequados para implementação quando existe uma grande quantidade de dados de séries históricas (ZHANG et al., 2020). Torres et al. (2005), por exemplo, aplicaram o modelo ARMA para prever a velocidade do vento em Navarra (Espanha) e verificaram que o modelo ARMA é superior ao método de persistência, principalmente quando o horizonte de previsão é maior. Os modelos estatísticos são adequados para previsões de curto prazo, pois o processamento é rápido e sua precisão diminui rapidamente com o incremento do horizonte de previsão (SILVA, 2017).

No entanto, Chen et al. (2019) afirmam que esses modelos não conseguem lidar de maneira efetiva com a série temporal da velocidade do vento com características não lineares devido à limitação de suas estruturas de correlação linear. Nesse sentido, Zhang et al. (2020), apontam que o método estatístico não consegue prever a velocidade do vento com flutuações instáveis, e essa metodologia é sensível a mudança de dados.

Os modelos matemáticos, a exemplo do algoritmo de previsão numérica do tempo (NWP), utilizam dados atmosféricos (temperatura, velocidade e direção do vento, pressão, etc.), topográficos e outros fatores para resolver problemas complexos (SEWDIEN et al., 2020). As técnicas de simulação numérica podem ser usadas para previsão do recurso eólico utilizando cálculos numéricos e imagens (JIANG et al., 2021).

Entretanto, alguns pontos devem ser levados em consideração nesse modelo: dificuldade de implementação, informações do ambiente atualizadas continuamente, alto poder computacional e custos elevados (ZHANG et al., 2020; CHANG, 2014). Além disso, Letcher

(2017) afirma que este modelo apresenta um bom desempenho para o horizonte de longo prazo, mas as previsões de curto prazo são inferiores em comparação com outras técnicas.

As abordagens de correlação espacial são destinadas para caracterizar a relação espacial entre diversos locais vizinhos, utilizando dados históricos disponíveis, e os correlacionam com o local destinado para a previsão (CHEN et al., 2019; CARTA, VELÁZQUEZ; CABRERA, 2013). Porém, os modelos só conseguem prever quando as medições da velocidade do vento são feitas de vários locais espacialmente correlacionados (JUNG; BROADWATER, 2014). Nesse sentido, a aplicação dos métodos de correlação espacial apresenta dificuldade devido às limitações de medição e atraso (JIANG et al., 2021).

Os métodos híbridos são compostos por duas ou mais abordagens de previsão (CHANG, 2014). A finalidade dos modelos híbridos é aumentar a eficiência da previsão quando comparada com as técnicas individuais (SAMET; REISI; MARZBANI, 2019). O modelo é baseado na combinação de técnicas auxiliares, como remoção de ruído de dados, otimização de parâmetros e análise de erros de previsão para melhorar o seu desempenho (WANG; LI; ZENG, 2021).

Com o rápido desenvolvimento de estratégias de Inteligência Artificial (IA), vários algoritmos foram desenvolvidos para previsão da velocidade do vento. Segundo Camelo (2018), alguns dos métodos de IA mais utilizados para previsão no setor de energia eólica são: Redes Neurais Artificiais (RNA), *Adaptative Neural Fuzzy Inference System* (ANFIS), métodos de lógica difusa (*Fuzzy Logic Methods*), dentre outros. Nesse sentido, as RNAs e as Máquina de Vetores de Suporte são os dois métodos não lineares mais utilizados para a previsão da velocidade do vento (FENG et al., 2017). As Redes Neurais Artificiais (RNA) são amplamente utilizadas na pesquisa em energia renováveis.

Geralmente, devido à sua capacidade de previsão não linear, as tecnologias de previsão com IA são amplamente utilizadas para tratar de relações complicadas, principalmente, para previsões de curto prazo (DHIMAN; DEB, 2020b; SEWDIEN et al. 2020). Conforme Zucatelli et al. (2019), as redes neurais artificiais apresentam custos computacionais mais baixos quando comparado com outras técnicas, como a modelagem matemática.

No trabalho de Bilgili e Sahin (2010), para prever a velocidade do vento foram utilizadas as variáveis meteorológicas mensais como temperatura, pressão atmosférica, umidade relativa do ar, precipitação e velocidade do vento. Os autores estabeleceram uma correlação entre as mesmas. As melhores variáveis independentes foram utilizadas nos modelos de regressão linear e não linear, e também, como parâmetros de entrada da RNA. No estudo adotou-se a estrutura de rede neural artificial *feedforward* (alimentada a frente) de três camadas e um algoritmo de

backpropagation (propagação dos erros da saída para entrada) foi usado para o treinamento de RNAs. A RNA apresentou melhor desempenho quando comparada aos métodos de regressão linear e regressão não linear.

De forma semelhante, Barhmi, Elfatni e Belhaj (2019) e Verma, Ghritlahre e Chandrakar (2021), a previsão da velocidade do vento utilizando RNA obteve melhores métricas de desempenho quando comparadas ao modelo de Regressão Linear Multivariável (RLM). A relação complexa não linear entre a velocidade do vento e as variáveis meteorológicas, foi apontada por Barhmi, Elfatni e Belhaj (2019), como um fator que contribuiu para o melhor desempenho do modelo RNA (não linear) com relação ao modelo de RLM.

Velo, López e Maseda (2014) adotaram a rede *Perceptron* Multi-camadas (*Multilayer Perceptron* - MLP) para prever a velocidade média anual do vento em um local onde não tinha uma estação meteorológica disponível. Foram utilizados dados de estações próximas. A arquitetura da RNA para o estudo foi a seguinte: dois parâmetros na camada de entrada, 15 neurônios na camada oculta e um neurônio na camada de saída. O algoritmo de aprendizado supervisionado utilizado no estudo foi o *backpropagation*. Os resultados encontrados no estudo obtiveram erros menores que 6%.

Em Navas, Prakash e Sasipraba (2020), os autores realizaram um estudo para prever a velocidade do vento no estado de Tamil Nadu (Índia). No trabalho foram utilizados e comparados o desempenho de diferentes modelos: Rede Neural *Multilayer Perceptron* (MLP), Rede Neural de Base Radial e Regressão Categórica. Conforme os resultados das validações dos modelos, a Rede Neural de Múltiplas Camadas apresentou melhores resultados, sendo mais adequado para prever a velocidade do vento quando comparado com o modelo de base radial e de Regressão Categórica.

Em Silveira, Capo e D'Aquino (2021), os autores analisaram a capacidade de previsão da velocidade do vento costeiro utilizando Redes Neurais Artificiais com um horizonte de previsão de 30 minutos em uma cidade do Extremo Sul Catarinense. Para o estudo, os melhores resultados foram obtidos com 15 neurônios na camada oculta. Os resíduos/erros entre os valores reais e previstos da velocidade do vento tiveram uma maior frequência entre -1 a 0 e 0 a 1, onde combinados tiveram 71,46% do total.

Philippopoulos e Deligiorgi (2012) utilizaram Redes Neurais *feedforward* para prever a média horária da velocidade do vento, em diferentes locais, com base em dados locais próximos. A precisão da previsão das redes neurais treinadas foi superior às metodologias tradicionais de interpolação espacial para a área de estudo (planície de Chania, localizada na parte noroeste da ilha de Creta, na Grécia).

Zucатели et al. (2019) realizaram a previsão da velocidade do vento de curto prazo em Colonia Eulacio no Uruguai com Redes Neurais Artificiais. O objetivo do estudo foi identificar a configuração de RNA mais eficiente. Os autores utilizaram *Multilayer Perceptrons* com o algoritmo de treinamento *Backpropagation Levenberg-Marquardt* para previsão de velocidade do vento entre 1 h e 12 h à frente. Conforme Fadare (2010), a RNA é uma ferramenta que auxilia na resolução de diferentes problemas e, também, pode aprender com exemplos. Além disso, de acordo com o autor, uma vez treinada, a rede pode realizar previsão e generalização de maneira rápida.

No trabalho de Purohit, Ng e Kabir (2022), os algoritmos de *Support Vector Regression*, Redes Neurais Artificiais e XGBoost (*Extreme Gradient Boosting*) foram utilizados para prever a velocidade e a intensidade de turbulência de uma esteira de turbina eólica. Os modelos de aprendizado de máquina ofereceram um desempenho satisfatório e semelhante na previsão, porém, o XGBoost apresentou um treinamento mais rápido quando comparado com os outros dois modelos.

Cai et al. (2020), aplicaram três algoritmos para prever a velocidade do vento: XGBoost, Redes Neurais Artificiais com *backpropagation* e Regressão Linear. Entre os modelos, o XGBoost apresentou melhores métricas de desempenho em todo o conjunto de dados. Assim como no trabalho de Phan, Wu e Phan (2020) o XGBoost superou os demais modelos testados.

Dentro desse contexto, conforme Alkhayat e Mehmood (2021), métodos de inteligência artificial, incluindo os algoritmos de *Machine Learning* vêm sendo estudados por diversos pesquisadores devido à sua capacidade de identificar relações não lineares e apresentarem um bom desempenho.

2.3 MÉTODOS DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina (*Machine Learning* – ML) é uma área da IA em que métodos e algoritmos permitem que o sistema aprenda a partir de experiências/dados, sendo utilizados para modelar recursos da entrada em relação à saída esperada e relacionar a saída de previsão com os registros históricos (SINGH et al., 2021). Cabe destacar que, a previsão do recurso eólico consiste em obter resultados por meio de um algoritmo de *machine learning* a partir de valores não vistos anteriormente. Ou seja, dado um conjunto de dados qualquer, os modelos possuem a capacidade de aprender com a experiência do passado e conseguem prever valores futuros (DHIMAN; DEB, 2020b).

Além disso, a quantidade de dados e qualidade dos dados são fatores determinantes para o desempenho das técnicas de aprendizado de máquina (CHU et al., 2022). Ainda, o desempenho do modelo de previsão pode ser afetado se dados irrelevantes/incoerentes forem adotados para sua construção (ZHA et al., 2022). Conforme Siefert et al. (2021), no Brasil, a principal limitação/dificuldade para estudos climatológicos regionais é a disponibilidade de dados meteorológicos quanto à quantidade, qualidade e sua distribuição espacial.

Assim, selecionar um banco de dados adequado é o passo mais importante em qualquer projeto de *machine learning*. Essa etapa pode afetar o desempenho e convergência dos modelos. Nesse sentido, para garantir a acurácia do projeto de previsão, é necessário que o banco de dados utilizado na etapa de treinamento tenha confiabilidade (RODRÍGUEZ et al., 2020). Conforme Zha et al., (2022), os dados de entrada precisam ser removidos e/ou selecionados antes da construção do modelo de ML e, dessa forma, reduzir o risco de *overfitting* (sobreajuste). Contudo, essa etapa pode ser refeita após o treinamento, teste e validação dos modelos.

O desempenho de um modelo de *machine learning* depende das variáveis do local de estudo. Ou seja, um modelo adequado para uma determinada região pode apresentar imprecisão em outra. Diante disso, não é possível generalizar as configurações, a exemplo dos parâmetros de entrada. Assim, não há um consenso sobre o assunto, o que se identifica é que em cada região de estudo existe sua própria correlação entre variáveis de entrada e saída. O Quadro 2 apresenta alguns exemplos de parâmetros de entrada encontrados na literatura.

Quadro 2 - Parâmetros de entrada para previsão eólica utilizando *machine learning*.

Autores	Parâmetros de entrada	Modelo(s)
ZUCATELLI et al. (2019)	Hora, dia, mês, ano, valores médios horários da velocidade do vento, direção do vento e temperatura.	Redes Neurais Artificiais
RODRÍGUEZ et al. (2020)	Mês, hora e medições de densidade de energia eólica (W/m ²).	Redes Neurais Artificiais
PELLETIER; MASSON; TAHAN (2016)	Velocidade do vento, densidade do ar, intensidade da turbulência, cisalhamento do vento, direção do vento e erro de guinada.	Redes Neurais Artificiais
DEMOLLI et al. (2019)	Velocidade diária do vento, desvio padrão diário e potência eólica diária	XGBoost, Floresta Aleatória, K-vizinhos mais próximos, <i>Least Absolute Shrinkage and Selection Operator</i> , <i>Support Vector Regression</i>
VELO; LÓPEZ; MASEDA (2014)	Velocidade e direção do vento	Redes Neurais Artificiais
RANGANAYAKI e DEEPA (2016)	Temperatura, direção e velocidade do vento, umidade relativa	Redes Neurais Artificiais
CAI et al. (2020)	Radiação, temperatura, pressão, umidade, direção e velocidade do vento.	XGBoost

Fonte: Organizado pela autora, 2022.

Os algoritmos de aprendizado de máquina têm diversos parâmetros denominados como hiperparâmetros, que são ajustáveis e permitem o controle no processo de treinamento e, além disso, afetam o tempo de execução, capacidade de generalização e o desempenho da previsão (SINGH et al., 2021). Conforme Demolli et al. (2019), os algoritmos de ML têm a capacidade de identificar e se adaptar às mudanças no conjunto de dados. Além disso, cada técnica de aprendizado de máquina é adotada para uma aplicação específica, por exemplo, algoritmos de regressão e classificação são amplamente utilizados em problemas de previsão (SINGH et al., 2021).

Nos modelos de *machine learning*, os conjuntos de dados de entrada e saída são subdivididos para uso em duas etapas: treinamento e validação das estimativas (ZUCATELLI et al., 2019; ZHA et. al., 2022). Existem dois principais tipos de aprendizagem: supervisionada e não supervisionada. Se a saída desejada for conhecida no início do processo, o treinamento é denominado supervisionado; ao contrário, será denominado não-supervisionado (LEE; BOOTH; ALAM, 2005).

Segundo Braga, Carvalho e Ludermir (2007), no aprendizado não-supervisionado, não há o acompanhamento de um supervisor durante o processo de treinamento e somente as variáveis de entrada estão disponíveis para o modelo. Os autores afirmam que, no caso do aprendizado supervisionado, a entrada e saída alvo são fornecidas por um supervisor, ou seja, o conjunto de treinamento possui pares de entrada e saída, e, no processo de aprendizagem compara-se a saída desejada com a saída obtida pelo modelo.

Os modelos de aprendizado de máquina possuem diferentes aplicações em sistemas de energia, por exemplo: custo de eletricidade, detecção de falhas, despacho econômico e previsão de recursos energéticos/energia (MOSAVI et al., 2019). Assim, os algoritmos de aprendizado de máquina são ferramentas importantes para previsão do recurso eólico, a próxima seção aborda diferentes métodos para esse fim que foram aplicados na presente pesquisa: Regressão Linear Multivariável, Redes Neurais Artificiais e XGBoost.

2.3.1 Regressão Linear Multivariável (RLM)

A RLM é uma das ferramentas mais conhecidas e compreendidas em estatística e aprendizado de máquina, sendo utilizada em diferentes estudos direcionados para o setor eólico, a exemplo dos trabalhos de Barhmi, Elfatni e Belhaj (2019), Verma, Ghritlahre e Chandrakar (2021), Bilgili e Sahin (2010).

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 (AZADI; KARIMI-JASHNI, 2016). As variáveis independentes do modelo de RLM são as variáveis de entrada e as variáveis dependentes são as variáveis de saída (VERMA; GHRITLAHRE; CHANDRAKAR, 2021).

Assim, para a previsão do recurso eólico, a velocidade média do vento é calculada (prevista) conforme as variáveis de entrada do modelo. A relação funcional de diferentes variáveis independentes com a variável dependente Y é dada pela Equação (1) (BARHMI; ELFATNI; BELHAJ, 2019).

$$y = b_0 + b_1X_1 + \dots + b_kX_k \quad (1)$$

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.

2.3.2 Redes Neurais Artificiais

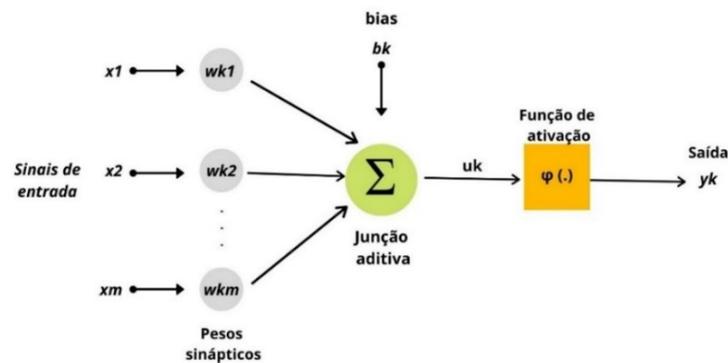
A rede neural artificial (RNA) é uma técnica computacional semelhante à característica da rede neural biológica humana (NOOROLLAHI; JOKAR; KALHOR, 2016; KALOGIROU, 2014; MA; FENG, 2016). É uma das técnicas mais populares e eficazes do sistema de aprendizagem utilizando o meio computacional (RUSSEL et al., 2016). Além disso, o neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural e atua em conjunto com outros neurônios da rede (CHOUDHURY et al., 2018; SAIRAMYA et al., 2019).

A Figura 4 mostra a representação básica de um neurônio artificial, no qual, cada neurônio apresenta m terminais de entrada $x_1, x_2, x_3 \dots x_m$ e um terminal de saída y_k . Nesse sentido, cada terminal de entrada possui um valor de entrada x_j , que é conectado a um neurônio k e multiplicado pelo peso sináptico correspondente (W_{kj}), podendo apresentar valores positivos ou negativos. Em seguida, as entradas ponderadas são somadas (BRAGA; CARVALHO; LUDERMIR, 2007). O *bias* (viés) aumenta ou diminui a entrada da função de ativação, dependendo se este é positivo ou negativo, respectivamente (HAYKIN, 2001).

O resultado do somatório passa por uma função de ativação, também denominada de função de transferência, que com base nas entradas e pesos sinápticos, definirá a saída do neurônio artificial (BRAGA; CARVALHO; LUDERMIR, 2007; HAYKIN, 2001). Sendo

assim, o objetivo geral de uma rede neural é identificar a relação entre a entrada e a saída (KHOSRAVI; MACHADO; NUNES, 2018).

Figura 4 - Representação esquemática do neurônio artificial.

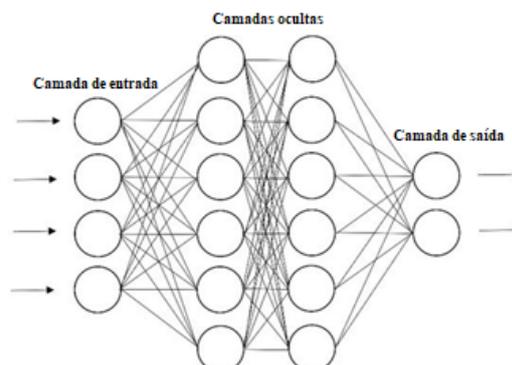


Fonte: HAYKIN (2001), adaptado.

As principais características das redes neurais incluem: aprendizagem, propriedade não linear, capacidade de se adaptar e lidar com grande quantidade de dados, além da generalização a partir do conhecimento adquirido durante a etapa de treinamento (RANGANAYAKI; DEEPA, 2016; JUNG; BROADWATER, 2014). Uma RNA é formada por unidades de processamento, camadas (entrada, ocultas e saída), neurônios e funções de ativação (VAFAEIPOUR et al., 2014). São necessárias pelo menos três camadas: uma camada de entrada, uma camada oculta e uma camada de saída (RUSSEL et al., 2016; HANIFI et al., 2020).

A camada de entrada da rede recebe as variáveis de entrada do modelo, que consistem em observações históricas dos dados, enquanto a camada de saída representa o resultado final da etapa de processamento. Nas camadas intermediárias/ocultas há o processamento dos dados, cálculos e a formação de relações não lineares (HAYKIN, 2001). Na Figura 5 é apresentada a estrutura típica de um modelo de rede neural artificial (RNA), denominada de *perceptron* multicamadas.

Figura 5 - Estrutura típica de um modelo de rede neural artificial.



Fonte: HANIFI et al. (2020), adaptado.

As camadas ocultas armazenam as funções de transferência que são utilizadas para processar os dados de entrada (PANDEY et al., 2016). A função de ativação é utilizada para controlar a amplitude da saída de um neurônio (RIOS et al., 2019). Assim, uma função de ativação é anexada em cada neurônio da rede neural e decide se ele deve ser ativado caso sua entrada for importante para o modelo de previsão (IOSIFIDIS, TEFAS, 2022).

As funções de ativação são divididas em dois grupos: linear e não linear (LONDRES, 2018; HAYKIN, 2001). Conforme Lederer (2021), existem diversas funções de ativação para a rede neural, tais como: sigmoide (logística/tangente hiperbólica), linear, ReLu. A função de ativação ReLu (*Rectified Linear Unit*) vem sendo amplamente utilizada para projetos de redes neurais artificiais, a exemplo dos trabalhos de Gandhi (2020); Nana et al. (2019); Silvera, Chamorro e Ochoa (2021). A função ReLu apresenta rapidez no processamento e característica não linear (PAJANKAR, JOSHI, 2022).

Dessa forma, o desempenho e a capacidade dos modelos de RNAs dependem diversos fatores, dentre eles: preparação e tratamento dos dados, treinamento, conexões entre entrada e saída, estrutura do modelo, parâmetros de entrada e estrutura dos dados (MARUGÁN et al., 2018; JUNG; BROADWATER, 2014).

2.3.2.1 Estrutura e treinamento da rede neural artificial

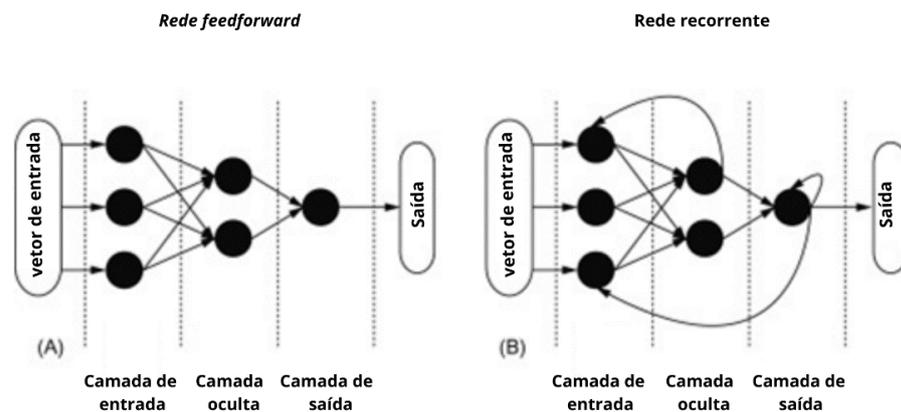
As redes neurais artificiais são, geralmente, utilizadas para previsão de velocidade ou potência, com um horizonte de previsão variando entre 30 minutos e 6 horas (DHIMAN; DEB, 2020b). Diferentes estudos sobre previsão da velocidade do vento de curto prazo, por exemplo: Sewdien et al. (2020), Zucatelli et al. (2019) e Marugán et al. (2018), afirmam que o aumento dos horizontes de previsão ocasiona maiores erros quando se utiliza RNA.

Em Zucatelli et al. (2019), os resultados para o horizonte de 12 horas apresentaram erros superiores quando comparados com o horizonte de 1 hora. Conforme os resultados apresentados no estudo de Hiruma (2016), com dados da Usina Eólica Palmas no estado do Paraná, a Rede Neural Artificial *Multilayer Perceptron* produziu bons resultados de previsão de geração para horizontes de previsão de até seis horas. Segundo Hanifi et al. (2020), existem mais de cinquenta tipos de RNAs, dentre elas: *multilayer perceptron* (MLP), rede neural *wavelet*, Rede Neural de função de base Radial, Rede Neural Elman, *Long-Short-Term Memory*, Rede Neural Convolutacional.

A primeira etapa do projeto da rede neural artificial contempla a seleção da estrutura da rede e a especificação da direção das informações, de acordo a topologia. Existem duas

topologias principais para a rede neural artificial: redes não recorrentes (*feedforward*) que transmite os dados em uma direção, ou seja, das camadas de entrada para a saída, não existindo ciclos e as redes recorrentes (*feedback*) que transmite os dados em direções mútuas, apresentando laços de realimentação em uma ou mais camadas do modelo (HIRUMA, 2016; HAYKIN, 2001; HANIFI et al. 2020; RUSSEL et al., 2016). A representação esquemática das topologias RNA *feedforward* e recorrente está apresentada na Figura 6.

Figura 6 - (A) RNA *feedforward* e (B) arquitetura recorrente RNA.



Fonte: IMRAN; ALSUHAIBANI (2019), adaptado.

A arquitetura da RNA pode ser dividida em três tipos principais: (1) redes *feedforward* de camada simples ou única, (2) arquitetura *feedforward* de múltiplas camadas e (3) rede recorrente (HAYKIN, 2001). A diferença entre a *feedforward* simples e de múltiplas camadas é que na primeira existe somente um nó entre qualquer entrada e qualquer saída da rede, ou seja, sem camada intermediária, e a segunda apresenta uma ou mais camadas de neurônios ocultas (BRAGA; CARVALHO; LUDERMIR, 2007).

Entre as RNAs a topologia mais comumente utilizada para problemas de previsão do recurso eólico é a *perceptron* multicamadas (MLP), a exemplo dos trabalhos de Pelletier, Masson e Tahan (2016), Zucатели et al. (2019), Manero, Béjar e Cortés (2019), Hiruma (2016), Masrur et al. (2016). Conforme Velo, López e Maseda (2014), a MLP modela funções complexas e apresenta boa capacidade em desconsiderar dados irrelevantes e ruídos.

A segunda etapa para implementação do projeto de rede neural é a escolha do algoritmo de treinamento. O aprendizado da rede é realizado por meio do treinamento, e as forças de conexão entre os neurônios (pesos sinápticos) guardam o conhecimento adquirido durante o processo (ZUCATELLI et al., 2019).

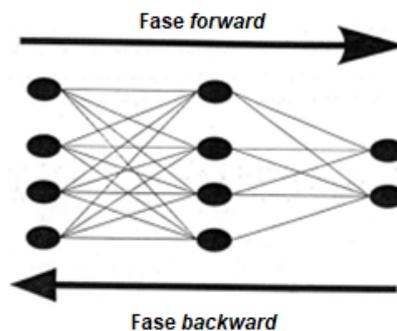
O algoritmo *backpropagation* (BP) é consolidado e amplamente utilizado em RNA, sendo o método mais adequado para o treinamento de redes do tipo *multilayer feedforward*,

com base na propagação reversa de erros (REN, H.; CHEN, X.; CHEN, Y., 2017; ERMIS, 2018; WONG, ZENG, AU, 2013). O método *backpropagation* é uma rede de aprendizagem supervisionada, no qual é percorrida em um sentido, que utiliza pares (entrada e saída alvo) para ajustar os pesos da RNA, por meio do seu mecanismo de correção (OZSAHIN, I.; OZSAHIN, D., 2020).

No algoritmo *backpropagation*, a rede neural é treinada com dados que são divididos em um conjunto de entrada e um conjunto de saída, onde, durante a etapa de treinamento, os pesos são ajustados para minimizar o erro entre a saída esperada/observada e a prevista (NOOROLLAHI; JOKAR; KALHOR, 2016).

Segundo Haykin (2001), o treinamento *backpropagation* ocorre em duas fases: *forward* e *backward*. Na primeira fase do treinamento, os vetores de entrada são transferidos na direção *forward* (para frente) da camada de entrada para a saída, ou seja, essa etapa determina a saída a partir de um dado de entrada. Na segunda fase, o erro é propagado na direção *backward* (para trás) para atualizar os pesos e reduzir os erros, assim, essa fase utiliza a saída alvo e a saída obtida pela rede neural para atualizar os pesos das conexões (BRAGA; CARVALHO; LUDERMIR, 2007; NOOROLLAHI; JOKAR; KALHOR, 2016). A Figura 7 mostra a representação esquemática dessas duas fases.

Figura 7 - Processamento do algoritmo *backpropagation*.



Fonte: Braga, Carvalho e Ludermir, (2007), adaptado.

O algoritmo de *backpropagation* é um procedimento que utiliza o gradiente descendente e, dessa forma, o erro está ligado aos pesos e limites que são ajustados por meio das fases *forward* e *backward*, até que se atenda ao requisito de minimização de erros, ou seja, o processo de treinamento finalizará quando um determinado nível de precisão ou critério for atingido (REN, H.; CHEN, X.; CHEN, Y., 2017; VEIT, 2012; MENG e PIAN, 2016). A descrição matemática sobre as Redes Neurais Artificiais pode ser encontrada nos livros “Redes Neurais: Princípios e prática” de Haykin (2001) e “Redes neurais artificiais: teoria e aplicações” de Braga, Carvalho e Ludermir (2007).

Segundo Multazam et al. (2017), o algoritmo *Levenberg-Marquardt Backpropagation* é adequado para a previsão de velocidade do vento de curto prazo porque apresenta as vantagens de coeficiente de correlação mais alto, menores erros e baixo custo computacional relativo ao treino. O método de treinamento *Levenberg-Marquardt* é um algoritmo de mínimos quadrados que tem como base o método de Gauss-Newton (RODRÍGUEZ et al., 2020).

Pelletier, Masson e Tahan (2016), Masrur et al. (2016), Zucatelli et al. (2019), Bilal et al. (2018), Vafaeipour et al. (2014), entre outros, obtiveram bons resultados para previsão do recurso eólico de curto prazo usando técnicas de inteligência computacional, especialmente com redes neurais do tipo *perceptron* multicamadas com algoritmo de treinamento *Backpropagation de Levenberg-Marquardt*.

Li e Shi (2010) utilizaram três tipos de RNA (elemento linear adaptativo, *feed forward back-propagation* e função de base radial) para previsão da velocidade do vento de curto prazo na altura de 10 m, com horizonte de previsão de 1 h à frente. Segundo os autores, a seleção da rede com a melhor performance depende fortemente dos dados utilizados. Eles descobriram que diferentes dados de entrada, taxas de aprendizagem e as estruturas do modelo, influenciam no desempenho das previsões do recurso eólico.

O aprendizado do modelo de rede neural artificial também depende de outros parâmetros, a citar: Número de Épocas e do *Batch Size*. O *Batch Size* (tamanho do lote) é o número de exemplos/dados usados durante o treino da rede neural (VALERIO et al., 2022). Assim, o lote de treinamento é particionado dentro da rede neural. O parâmetro influencia no processo de aprendizado de uma rede neural, visto que as amostras são utilizadas para estimar o erro antes da atualização dos pesos do modelo (BROWNLEE, 2018). Geralmente o valor do *Batch Size* fica entre 64 e 1024, variando de acordo com o modelo (HUANG et al., 2018).

A época (*Epoch*) é definida como a passagem completa pelo conjunto de dados de treino (IOSIFIDIS; TEFAS, 2022). A taxa de perda é reduzida conforme o modelo é treinado com um número maior de épocas e converge depois de um determinado ponto (PAJANKAR; JOSHI, 2022). A quantidade de épocas pode afetar o desempenho das previsões, se o seu número for pequeno a RNA pode subajustar (*underfitting*) os dados, caso for muito grande, pode ocorrer o sobreajuste (*overfitting*) no modelo (IOSIFIDIS; TEFAS, 2022).

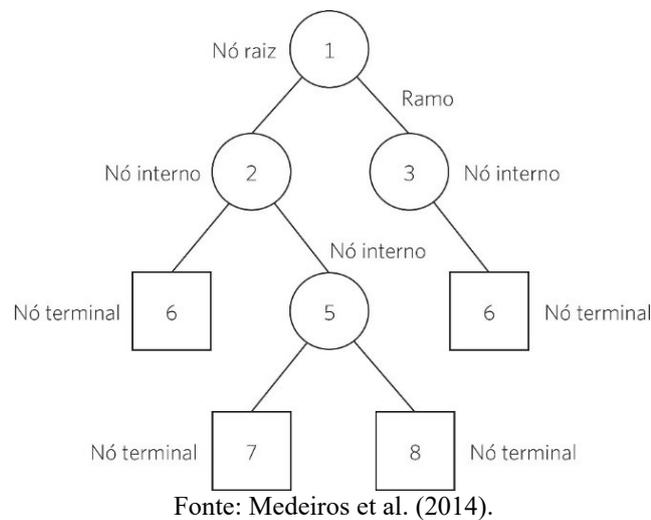
A terceira etapa do desenvolvimento de uma rede neural envolve a validação do modelo por meio de métricas. Entretanto, os modelos de RNA também têm alguns pontos fracos; por exemplo, eles podem apresentar *overfitting*, local ótimo e baixas taxas de convergência (ZHANG et al., 2020). Diante disso, o treinamento ideal ainda é um desafio.

2.3.3 Algoritmo XGBoost

O XGBoost (eXtreme Gradient Boosting) é uma técnica de *machine learning* baseado em árvore de decisão que usa uma estrutura de aumento de gradiente (CHEN; GUESTRIN, 2016). Conforme Mayrink (2016), as árvores de decisão são modelos representados em uma estrutura hierárquica semelhante ao processo de tomada de decisão humana, sendo formada por dois tipos de nós: nós de decisão e nós terminais. Conforme o autor, cada nó de decisão executa um teste, e, o resultado define a direção para o nó posterior. E, assim, os testes são executados até atingir o nó terminal (folha), que representa um valor da variável de destino.

Um exemplo de aplicação de árvore de decisão é dado na Figura 8: 1 é o nó raiz que contém o atributo de previsão; 2, 3 e 5 são os nós internos ou filhos que abrangem as variáveis de decisão; 4, 6, 7 e 8 são denominados de folhas ou nó terminal, que apresentam a resposta da variável de decisão. Os ramos conectam os nós da árvore e possuem os valores de cada variável (MEDEIROS et al., 2014).

Figura 8 - Exemplo de uma árvore de decisão.

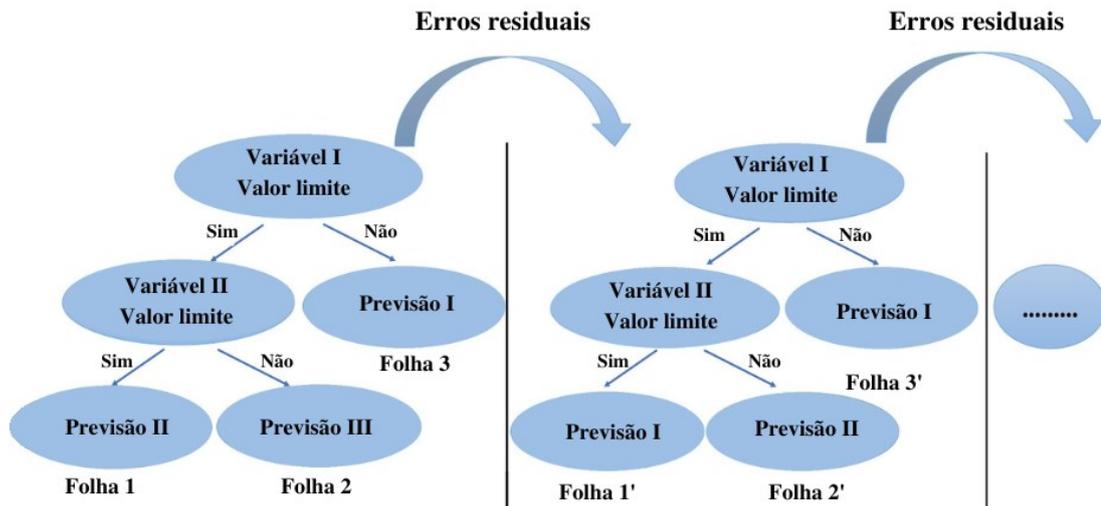


O XGBoost foi desenvolvido como um projeto de pesquisa na Universidade de Washington e publicado em 2016 (CHEN, GUESTRIN, 2016). Graf, Kolerski e Zhu (2022) afirmam que o algoritmo de aumento de gradiente em cada etapa tem por objetivo a redução do erro de previsão da etapa anterior. Conforme os autores, em cada etapa o algoritmo faz uma estimativa dos parâmetros para prever os resíduos do modelo previsto na etapa anterior.

Resumidamente, durante o processo de construção do modelo, o XGBoost gera uma nova árvore de decisão em cada iteração para ajustar o resíduo da previsão anterior, e, posteriormente, o resultado final é obtido adicionando os valores de saída previstos por cada

árvore (ZHA et al., 2022). Assim, de acordo com Jin et al. (2020), o XGBoost gera uma árvore de regressão forte combinando diversas árvores de regressão fracas. Um esquema de árvore adotando o algoritmo XGBoost é apresentado na Figura 9.

Figura 9 - Exemplo XGBoost.



Fonte: Dong et al. (2020), adaptado.

A finalidade do treino com o algoritmo XGBoost é minimizar a função objetivo (ZHA et al., 2022). Dessa maneira, tendo como base a função objetivo, o XGBoost aplica o algoritmo *greedy* (guloso) para construir árvores de decisão de maneira constante, e, posteriormente, um modelo XGBoost completo é criado (ZHENG; WU, 2019).

A execução paralela e distribuída do algoritmo XGBoost proporciona rapidez na etapa de aprendizagem (LI; ZHANG, 2018). Desta forma, o algoritmo XGBoost é um algoritmo rápido e escalável, construindo diferentes árvores constantemente para minimizar a função objetivo, e, assim, pode encontrar soluções rapidamente (CHEN, GUESTRIN, 2016).

O XGBoost tem os seguintes objetivos: evitar *overfitting*; reduzir custos computacionais; minimizar a função objetivo utilizando seus termos de regularização (JIN et al., 2020). Além disso, pode ser utilizado em problemas de regressão e classificação (OSMAN et al., 2021). Cabe destacar que o algoritmo XGBoost é adotado em diversos desafios de aprendizado de máquina e mineração de dados. Nas competições de aprendizado de máquina na comunidade Kaggle, entre as 29 soluções vencedoras dos desafios publicados no site durante o ano 2015, 17 soluções utilizaram o XGBoost no seu desenvolvimento (CHEN; GUESTRIN, 2016).

O método tem sido aplicado em vários estudos, tais como: classificação de imagens (JIAO; HAO; QIN, 2021; REN et al., 2017), estudos biomédicos (TAHMASSEBI et al., 2019), previsão de carga (ZHENG; YUAN; CHEN, 2017; AGUILAR MADRID; ANTONIO, 2021),

intensidade de ciclones (JIN et al., 2020), previsão de energia eólica (PUROHIT; NG; KABIR, 2022; JIADING et al., 2022; CAI et al., 2020; PHAN; WU; PHAN, 2020), dentre outros. O equacionamento matemático da função objetivo e alguns detalhes de como é realizada a pontuação e ramificação do algoritmo XGBoost é apresentado no trabalho de Chen (2014) e Chen e Guestrin (2016), descrito brevemente no Anexo A.

No algoritmo XGBoost existe um conjunto de diversos parâmetros que podem ser ajustados durante o treinamento para melhorar o modelo que está sendo construído (GUPTA; GUSAIN; POLI, 2016). Os parâmetros podem ser vistos na página da documentação oficial do algoritmo (XGBOOST, 2021), e, estes podem ser otimizados pelas mais variadas técnicas. O Quadro 3 apresenta alguns parâmetros e suas respectivas descrições.

Quadro 3 - Parâmetros XGBoost.

Parâmetro	Descrição
<i>max_depth</i>	Profundidade máxima de uma árvore. Utilizar valores altos pode acarretar em <i>overfitting</i> do modelo e, consumirá mais recursos computacionais. O valor deve ser um número inteiro maior que 0, e, o valor padrão é 6.
<i>learning_rate</i>	Controla o aprendizado do modelo por meio da redução do tamanho da etapa utilizado na atualização para evitar o <i>overfitting</i> . Após cada etapa de <i>boosting</i> , o peso dos recursos é atualizado. O valor deve estar entre 0 e 1, e o padrão é 0,3.
<i>n_estimators</i>	Número de árvores a serem consideradas, ou seja, é o número de iterações que o modelo realizará, sendo que uma nova árvore é criada por iteração. O valor deve ser um número inteiro maior que 0, e, o padrão é 100.
<i>reg_alpha</i>	Prazo de regularização L1 sobre os pesos. Valor padrão: 0.
<i>reg_lambda</i>	Prazo de regularização L2 sobre os pesos. Valor padrão: 1.
<i>Subsample</i>	Fração dos dados de treinamento considerada a cada iteração. O intervalo deve estar entre 0 e 1, e o valor padrão é 1.
<i>Gamma</i>	Redução de perda mínima para fazer uma divisão em um nó folha da árvore. Quanto maior o valor de gamma, mais conservador será o modelo. Intervalo: $[0, \infty]$, valor padrão: 0.
<i>min_child_weight</i>	Somatório mínimo dos pesos de um nó folha. Se a etapa de divisão da árvore resultar em um nó folha com a soma do peso menor que <i>min_child_weight</i> , o particionamento é interrompido. Intervalo: $[0, \infty]$, valor padrão: 1.
<i>colsample_bytree</i>	Fração das <i>features</i> (variáveis) a serem subamostradas ao construir cada árvore. A subamostragem é aleatória e ocorre para cada árvore construída. Intervalo: $[0, 1]$, valor padrão: 1.

Fonte: XGBoost (2021).

Uma descrição detalhada do XGBoost pode ser encontrada na página oficial do algoritmo (xgboost.readthedocs.io). Dessa forma, estudos com diferentes algoritmos, configurações e estruturas de dados são necessários. Portanto, é fundamental a avaliação dos resultados, por meio de métricas adequadas.

2.4 MÉTRICAS PARA AVALIAÇÃO DO DESEMPENHO DAS PREVISÕES

O processo de avaliação do modelo de previsão é uma etapa essencial para que se determine sua validade quanto à estimativa dos valores futuros do recurso eólico (GONZÁLEZ-SOPENA; PAKRASHI; GHOSH, 2021). Nesse sentido, existem diversas métricas estatísticas, que mostram os desvios da velocidade do vento prevista em relação à real. No Quadro 4, estão apresentados os critérios de avaliação de diferentes pesquisas destinadas à previsão eólica.

Quadro 4 - Métricas de avaliação de desempenho da previsão eólica.

Autores	RMSE	MAE	MSE	MBE	MAPE	R ²	R	Erro _{máx}
Pelletier, Masson e Tahan (2016)		X		X				
Xu e Mao (2016)	X	X						X
Rodríguez et al. (2020)	X							
Zucатели et al. (2019)	X	X	X		X	X	X	
Ranganayaki e Deepa (2016)			X					
Noorollahi, Jokar e Kalhor(2016)	X	X			X			
Vafaeipour et al. (2014)	X		X			X		
Li e Shi (2010)	X	X			X			
Azad, Mekhilef e Ganapathy(2014)		X						
Filik e Filik (2017)	X	X						
Hiruma (2016)	X	X	X		X	X		
Fadare (2010)			X		X		X	
Samet, Reisi e Marzbani (2019)	X	X	X		X			
Liu e Wang (2021)	X	X			X			
Begam e Deepa (2019)		X	X		X			
Liu, Mi e Li (2018)	X	X			X			
Camelo (2018)	X	X			X			
Liu et al. (2020)	X	X			X			
Alencar (2018)	X	X			X			
Masrur et al. (2016)					X			
Total	14	15	7	1	13	3	2	1

Fonte: Organizado pela autora, 2022.

Conforme demonstrado no Quadro 4, a maioria dos estudos envolveu um número seletivo de critérios para a avaliação das previsões, sendo os mais citados: erro médio absoluto (*Mean Absolute Error*, MAE), Raiz Quadrada do Erro Quadrático Médio (*Root Mean Square Error*, RMSE), Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*, MAPE). Os mesmos critérios de desempenho dos modelos de previsão foram adotados por Sun e Wang (2018). As equações das métricas mais citadas são vistas em (2), (3) e (4).

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_p - p| \quad (2)$$

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (p_p - p)^2 \right)^{1/2} \quad (3)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|p_p - p|}{p} \quad (4)$$

Sendo que p_p representa o valor previsto e p o valor observado. Com $i=1, \dots, N$ através de todo o conjunto de dados até o N -ésimo valor.

O erro médio absoluto é um valor não negativo e seu melhor resultado é zero (HANIFI et al., 2020). Essa métrica é definida como o valor médio dos erros de previsão em valores absolutos. A raiz do erro quadrático médio representa o desvio padrão dos erros de previsão, quanto menor o valor do critério, melhor é o desempenho (KHOSRAVI; MACHADO; NUNES, 2018).

A média do erro absoluto percentual quantifica a precisão do modelo na forma de porcentagem do erro. Entretanto, conforme González-Sopeña, Pakrashi e Ghosh (2021), o MAPE resulta em valores elevados quando os valores reais/observados estão próximos de zero, e, é indefinido quando o valor real é igual a zero. Kim e Kim (2016) propõem uma nova medida de previsão denominada de erro médio percentual absoluto do arco tangente (*Mean arctangent Absolute Percentage Error*, MAAPE) como uma opção alternativa ao MAPE. Conforme os autores, essa métrica transforma o MAPE utilizando a função arco tangente. Sua vantagem é a preservação das características do MAPE ao mesmo tempo em que supera suas limitações (GONZÁLEZ-SOPEÑA; PAKRASHI; GHOSH, 2021). A Equação (5) desenvolvida por Kim e Kim (2016) é definida como:

$$MAAPE = \frac{1}{N} \sum_{i=1}^N (AAPE_i) \quad t = 1, \dots, N. \quad (5)$$

Onde,

$$AAPE_i = \arctg \left(\left| \frac{p - p_p}{p} \right| \right) \quad (6)$$

Além disso, conforme Kato (2016), a métrica MBE (*Mean Bias Error*), Equação 7, é o erro médio da previsão, no qual representa sua tendência de superestimar (valores acima) e subestimar (valores abaixo) a saída da RNA. Para o MBE valores próximos a zero são desejáveis.

$$MBE = \frac{1}{N} \sum_{i=1}^N (p_p - p) \quad (7)$$

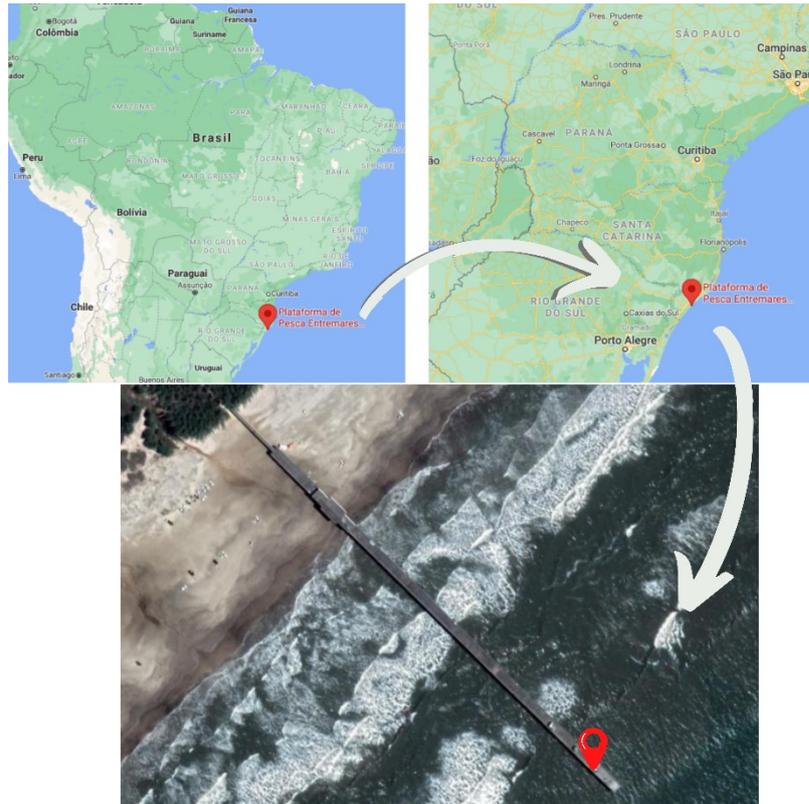
Sendo que p_p representa o valor previsto e p o valor observado. Com $i=1, \dots, N$ através de todo o conjunto de dados até o N-ésimo valor. Chai e Draxler (2014) afirmam que, muitas vezes é necessário a combinação de diferentes métricas para avaliar o desempenho dos modelos visto que não existe uma única métrica.

O referencial teórico abordou os principais aspectos relacionados ao tema de previsão da velocidade do vento: métodos e suas características/limitações, horizontes de previsão e respectivas aplicações, e métricas para avaliação de desempenho das previsões. Além disso, conforme visto, os modelos de aprendizado de máquina possuem diversas aplicações nos sistemas de energia. A Rede Neural é amplamente pesquisada em energia renováveis. O XGBoost, surgiu nos últimos anos e vem ganhando destaque na resolução de problemas. E, a Regressão Linear devido a sua simplicidade é um modelo adotado como base para diferentes estudos, sendo um método a ser superado por algoritmos mais complexos. Dessa forma, a revisão bibliográfica forneceu embasamento teórico para a metodologia, que será apresentada em detalhes a seguir.

3 METODOLOGIA

Para o desenvolvimento do trabalho foram utilizados dados coletados de uma estação meteorológica construída sobre uma plataforma de pesca localizada no Extremo Sul Catarinense, no município de Balneário Arroio do Silva (-28,9612817, -49,3821256), Figura 10.

Figura 10 - Local de estudo.



Fonte: Google Maps (2022); Google Earth (2022).

Conforme Pires (2019), a plataforma de pesca Entremares possui 410 m de comprimento com aproximadamente 7 m acima do nível do mar, Figura 11 (a). A Base de Observação do Oceano e Atmosfera (BOOA), Figura 11 (b), é uma base avançada da UFSC (Universidade Federal de Santa Catarina) em cooperação com a plataforma de pesca Entremares e os dados meteorológicos são disponibilizados *on-line* pelo Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina (CIRAM/EPAGRI). Conforme Nassif (2017), a estação localiza-se na região de transição continental-oceânica, com uma distância de 250 metros da linha de praia. A BOOA consiste em um laboratório de pesquisa de 2 m², que no seu telhado estão instalados equipamentos diversos, entre eles a torre meteorológica com anemômetro de copo (MOVLIDAR-UFSC, 2022). Dessa forma, os dados utilizados no presente estudo são medidos na altura de 17 m.

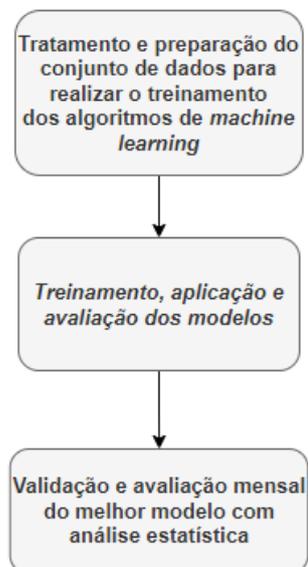
Figura 11 - (A) Plataforma de Pesca (B) Base de Observação do Oceano e Atmosfera.



Fonte: MOVLIDAR-UFSC (2022).

A metodologia foi dividida em três principais etapas (Figura 12): tratamento e preparação do conjunto de dados para realizar o treinamento dos algoritmos de *machine learning* (ML): treinamento, aplicação e avaliação dos modelos e, validação e a avaliação das métricas mensais do melhor modelo com análise estatística.

Figura 12 - Etapas do desenvolvimento da metodologia da pesquisa.



Fonte: Elaborado pela autora (2022).

Conforme a Figura 12, todos os modelos de ML foram executados a partir do pré-tratamento e preparação da base de dados. Destaca-se que todas as etapas da metodologia foram realizadas através de programação na linguagem Python no ambiente Google Colab (*Colaboratory*).

3.1 TRATAMENTO E PREPARAÇÃO DO CONJUNTO DE DADOS

Os dados utilizados foram registrados no período entre 2018 e 2021 com intervalo de 1 hora e altura de 17 m. Para o ano de 2018 foram utilizados 6057 dados que compreendem os meses entre maio a dezembro. Em 2019 e 2020, o período de janeiro a dezembro, totalizando 8724 e 8551 dados, respectivamente. E para o ano de 2021, 1289 dados para o período de janeiro a março. Em alguns meses não houveram medições devido a manutenção dos equipamentos.

As variáveis utilizadas para desenvolvimento dos modelos foram: pressão atmosférica ao nível da estação (hPa), temperatura mínima do ar (°C), temperatura média do ar (°C), temperatura máxima do ar (°C), direção do vento (°), umidade (%), velocidade média e máxima do vento (m/s). Foi realizado o tratamento de dados para remover possíveis inconsistências no banco de dados em caso de falha dos sensores de medição. Uma vez que o desempenho dos modelos de *machine learning* é, fortemente, afetado caso existam dados de entrada e saída incorretos.

Os dados foram divididos em três blocos principais: treinamento, testes e validação. O conjunto de treinamento (2018, 2019 e 2021) foi subdividido, aleatoriamente, em 90% para treino e 10% para testes. Para realizar uma validação com um período maior de tempo, e que contemplasse todos os meses, utilizou-se o ano de 2020.

Para o desenvolvimento do trabalho o treinamento supervisionado foi aplicado, sendo a variável alvo a velocidade média do vento 1 hora depois (objeto da previsão). Dessa maneira, verificou-se a existência das medições no instante de tempo t_0 e t_{+1h} a fim de garantir que não houvessem lacunas na base de dados. Assim, com os parâmetros de entrada (t_0), verifica-se a ocorrência da saída esperada (velocidade do vento no instante $t+1h$). Na Tabela 1 é apresentado um exemplo de aplicação utilizando o horizonte de 1 hora com a velocidade média do vento como parâmetro de entrada.

Tabela 1 - Exemplo de entrada e saída para o treino dos modelos com horizonte de previsão de 1 hora.

Entrada t_0		Saída ($t + 1h$)	
Horário (h)	Velocidade [m/s]	Horário (h)	Velocidade [m/s]
0	9,80	1	9,19
1	9,19	2	7,77
2	7,77	3	3,66
3	3,66	4	7,82

Fonte: Elaborado pela autora (2022).

A padronização do banco de dados foi realizada através da confirmação da existência dos pares de entrada e saída. Dessa forma, o treinamento e a previsão ocorrem quando os dados

estão completos, caso algum componente do par de dados não seja encontrado, é removido antes da utilização nos modelos.

Optou-se por calcular a razão (Velocidade Média/Velocidade Máxima) e a diferença entre as velocidades médias e máximas do vento (Velocidade Máxima – Velocidade Média) e inseri-las como variáveis de entrada, de forma que os modelos não utilizaram a variável alvo (velocidade média) durante as fases de treinamento, teste e validação.

Cabe destacar que, as variáveis de entrada selecionadas foram as que se apresentaram em maior quantidade e qualidade no banco de dados. Dessa forma, foram analisados quatro modelos com diferentes conjuntos de variáveis de entrada para cada algoritmo de *Machine Learning*, conforme Quadro 5.

Quadro 5 - Conjuntos de variáveis de entrada para os modelos de *machine learning*.

Modelo	Variáveis de entrada	Variável de saída (1 hora)
1	Hora, Pressão atmosférica, temperatura mínima do ar, temperatura média do ar, temperatura máxima do ar, direção vento, umidade e velocidade	Velocidade Média T+1 hora
2	Hora, Pressão atmosférica, temperatura mínima do ar, temperatura média do ar, temperatura máxima do ar, direção vento, umidade e diferença de velocidade	Velocidade Média T+1 hora
3	Hora, Pressão atmosférica, temperatura mínima do ar, temperatura média do ar, temperatura máxima do ar, direção vento, umidade e razão de velocidade	Velocidade Média T+1 hora
4	Hora, Pressão atmosférica, temperatura mínima do ar, temperatura média do ar, temperatura máxima do ar, direção vento, umidade, diferença de velocidade e razão de velocidade	Velocidade Média T+1 hora

A normalização dos dados é uma parte importante no desenvolvimento de modelos de previsão com ML, pois, as variáveis utilizadas no estudo são de unidades e ordem de grandeza diferentes (MABEL; FERNANDEZ, 2008; ZHENG; WU, 2019). Com o treinamento supervisionado, o conjunto de dados de entrada foi padronizado por meio das funções disponíveis na biblioteca *scikit-learn StandardScaler* em Python, conforme Equação (8).

$$z = \frac{(x - u)}{s} \quad (8)$$

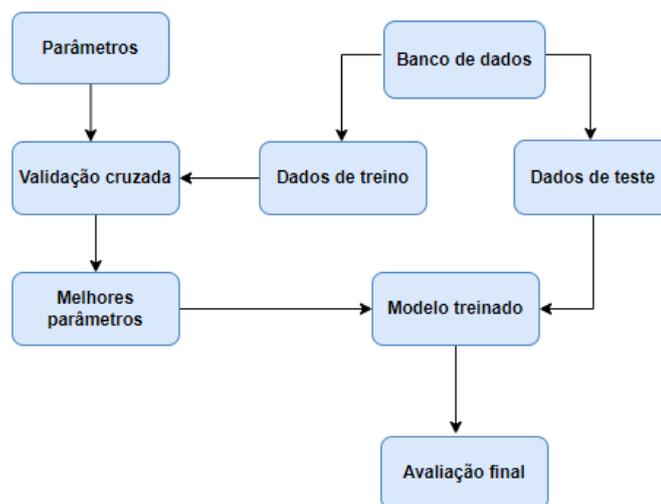
Onde u é média das variáveis de treinamento, s é o desvio-padrão, x é a variável original e z é a variável padronizada. Os dados da variável de saída (velocidade média do vento) estão na mesma ordem de grandeza. Portanto, não foram padronizados para utilização nos modelos.

3.2 TREINAMENTO, APLICAÇÃO E AVALIAÇÃO DOS MODELOS

Cada algoritmo possui variados parâmetros de funcionamento e otimização. Para automatizar a busca pelos melhores parâmetros utilizou-se do *GridSearchCV* (Pesquisa em Grade) com validação cruzada (*Cross Validation* – CV) da biblioteca *scikit-learn*, também chamada de *sklearn*, em linguagem Python. Ele tem por objetivo combinar todas as possibilidades, dentro de um conjunto de parâmetros, e retornar a melhor opção (GOODFELLOW; BENGIO; COURVILLE, 2016). Assim, a validação cruzada é utilizada para avaliar o desempenho, pois, é necessário escolher um modelo que se ajuste aos dados futuros (RUSSELL et al., 2016). Além disso, conforme os autores, se um modelo apresentar um erro pequeno no conjunto de treino, não significa que ele terá um bom desempenho na sua capacidade de generalização.

Na Figura 13 é apresentado um fluxograma de validação cruzada no treinamento de um modelo. O desempenho dos parâmetros foi avaliado com a base de teste, que não foi utilizada na etapa de seleção (SCIKIT-LEARN, 2022a). Assim, usando a técnica, o modelo consegue prever com dados não vistos e reduziria o risco de *overfitting* (SCIKIT-LEARN, 2022b).

Figura 13 - Fluxograma *Cross Validation*.



Fonte: SCIKIT-LEARN (2022b), adaptado.

O critério de escolha dos hiperparâmetros foi baseado no menor valor de RMSE encontrado durante a etapa de *CrossValidation*. Em cada iteração parte dos dados são apresentados como um conjunto de teste e o restante utilizado como treinamento do modelo (RUSSELL et al., 2016). Os modelos de previsão foram validados usando a técnica *CrossValidation* duas vezes, devido ao tempo e custo computacional quando aliado ao *GridSearch*.

Na validação cruzada de 2 vezes, o banco de dados foi embaralhado aleatoriamente em dois conjuntos B1 e C1, de forma que os conjuntos tiveram o mesmo tamanho. Em seguida, B1 foi treinado e o teste aconteceu em C1, na próxima iteração, ocorreu o treinamento em C1 e teste em B1 (SCIKIT-LEARN, 2022b). Dessa forma, os dados de treino foram divididos em 2 partes (treinamento e teste), cada uma sendo 50% do banco de dados de treino completo.

Esse método foi realizado para as Redes Neurais Artificiais (RNAs) e *Extreme Gradient Boosting* (XGBoost). Já para a Regressão Linear (RL) foi mantida uma configuração padrão por ser o algoritmo base para comparação de desempenho entre os modelos. A seguir são apresentados os modelos de *machine learning* com seus respectivos parâmetros e configurações.

3.2.1 Regressão Linear Multivariável

A biblioteca *sklearn LinearRegression* foi utilizada para o desenvolvimento. A regressão linear foi considerada o modelo a ser superado, dada a sua simplicidade de desenvolvimento em linguagem Python e sua ampla utilização. Neste estudo, são consideradas diferentes variáveis independentes, conforme Quadro 5 da seção 3.1. A velocidade média do vento é considerada como uma variável dependente (alvo).

3.2.2 Parâmetros da Rede Neural Artificial

A RNA baseou-se no algoritmo de treinamento *backpropagation*. Considerando, os seguintes parâmetros: arquitetura do tipo *Multilayer Perceptron*, uma camada de entrada, uma camada oculta com função de ativação ReLu, uma camada de saída com função de ativação linear, *batch_size = 512* e número de épocas de 1000.

Contudo, muitas épocas podem levar ao *overfitting* (sobreajustado) do conjunto de dados de treinamento, enquanto poucas podem resultar em um modelo *underfit* (subajustado). A parada antecipada (*EarlyStop*) é um método que permite especificar um grande número arbitrário de épocas de treinamento e interromper o treinamento assim que o desempenho do modelo parar de melhorar em um conjunto de dados de validação (PRECHELT, 1998; IOSIFIDIS; TEFAS, 2022). Dessa forma, utilizou-se como número de épocas de 1000. Sendo assim, a RNA pode ter seu treinamento interrompido caso não haja mais ganhos de desempenho.

O número de neurônios na camada oculta é um fator fundamental para o desempenho da RNA. De acordo com Kalogirou (2014), no algoritmo *Backpropagation* ao utilizar muitos

neurônios, a rede pode apresentar a tendência de memorizar o problema e perder a capacidade de generalizar após algum tempo. Entretanto, o autor afirma que se poucos neurônios forem usados, a rede tem maiores chances de apresentar boa capacidade de generalização, porém, pode não ter desempenho suficiente para aprender os padrões de maneira eficaz.

Por meio do *GridSearchCV* a otimização foi realizada variando a quantidade de neurônios na camada oculta de 2 a 100 em intervalos de dois. O objetivo dessa etapa foi verificar a existência de alguma tendência ou um ganho significativo em desempenho em detrimento do aumento da complexidade do modelo.

3.2.3 Parâmetros do XGBoost

O algoritmo permite uma grande possibilidade de parâmetros e estes podem ser otimizados pelas mais variadas técnicas. Contudo, manteve-se a busca por meio do *GridSearchCV*, onde os parâmetros em destaque tiveram como base o trabalho de Dosdoğru e İpek (2022), em itálico os recomendados pela documentação oficial do XGBoost e os restantes para análise experimental (Tabela 2).

Tabela 2 - Lista dos parâmetros analisados utilizando o método do *GridSearchCV*.

<i>Parâmetro</i>	<i>Intervalo de pesquisa no GridSearch</i>
<i>max_depth</i>	[4,5,6]
<i>learning_rate</i>	[0.01]
<i>n_estimators</i>	[500,1000]
<i>reg_alpha</i>	[0.067, 1.2]
<i>reg_lambda</i>	[0.507, 1.2]
<i>Subsample</i>	[0.5, 1]
<i>Gamma</i>	[0, 0.75,1]
<i>min_child_weight</i>	[0,8, 1,5]
<i>colsample_bytree</i>	[0.8, 1]

3.3 VALIDAÇÃO E AVALIAÇÃO DO MODELO SELECIONADO

O ano de 2020 foi utilizado para validação do modelo com melhor desempenho obtido na base de testes. Dessa forma, os índices de desempenho: MAE, RMSE, MBE e MAAPE, foram avaliados mensalmente. De maneira complementar, os resultados foram analisados graficamente por meio dos histogramas de resíduos.

Para fins de análise do modelo, com melhor desempenho, foram utilizados indicadores estatísticos mensais para cada variável analisada, tais como: média e desvio padrão. Foram elaborados figuras e gráficos para representação dos resultados. Essa avaliação proporciona uma análise mais detalhada dos resultados.

4 RESULTADOS E DISCUSSÃO

Na Tabela 3 é apresentada a análise da base de dados de treinamento (2018, 2019, 2021), totalizando 16071 dados tratados. Destacam-se a velocidade média do vento com valor médio de 4,93 m/s e a média máxima horária registrada de 19,31 m/s, demonstrando a presença de valores velocidade de vento intensos.

Tabela 3 - Análise exploratória das variáveis referentes ao banco de dados horários de treinamento.

Variáveis	Média	Desvio Padrão	Mínima	Máxima
Temperatura Média (° C)	19,95	3,96	6,22	36,58
Temperatura Mínima (° C)	19,61	3,97	5,93	36,08
Temperatura Máxima (° C)	20,38	3,96	6,66	37,54
Umidade (%)	87,26	9,59	29,59	100
Direção (°)	163,03	105,81	0,029	360
Pressão (hPa)	1014,59	5,14	998	1032
Razão da Velocidade (adimensional)	0,66	0,13	0,10	0,97
Diferença de velocidade (m/s)	2,15	1,18	0,148	21,05
Velocidade Máxima (m/s)	7,16	3,82	0,92	27,70
Velocidade Média (m/s)	4,93	3,17	0,13	19,31

Fonte: Elaborada pela autora (2022).

O conjunto de dados do ano de 2020 foi adotado para validação do modelo com melhor desempenho e sua análise exploratória está descrita na Tabela 4. Esse conjunto apresenta máximo de velocidade máxima horária de 28,84 m/s, média da velocidade de 5,01 m/s, ligeiramente superiores ao conjunto de dados treino. Enquanto a média da razão e da diferença de velocidade são muito semelhantes.

Tabela 4 - Análise exploratória do conjunto de dados de 2020.

Variáveis	Média	Desvio Padrão	Mínima	Máxima
Temperatura Média (° C)	19,52	3,75	5,28	32,19
Temperatura Mínima (° C)	19,17	3,77	4,67	31,89
Temperatura Máxima (° C)	19,96	3,73	5,77	33,28
Umidade (%)	86,35	10,51	26,47	100
Direção (°)	162,57	105,33	0,19	359,90
Pressão (hPa)	1014,48	5,03	993	1029
Razão da Velocidade (adimensional)	0,67	0,13	0,10	0,91
Diferença de velocidade (m/s)	2,15	1,21	0,33	21,60
Velocidade Máxima (m/s)	7,16	3,86	0,98	28,84
Velocidade Média (m/s)	5,01	3,19	0,23	19,14

Fonte: Elaborada pela autora (2022).

No Quadro 6 são apresentados o número de dias utilizados em cada mês do banco de dados de validação (2020). Percebe-se o conjunto de dados representou 97,35% da totalidade de dias, com 356 dos 366 dias para 2020 (Ano bissexto).

Quadro 6 - Quantidade de dias utilizados da base de dados de validação - mensal.

	Total dados (horários)	Dias	%
Jan	733	30,54	98,52
Fev	691	28,79	99,28
Mar	739	30,79	99,33
Abr	715	29,79	99,31
Mai	739	30,79	99,33
Jun	698	29,08	96,94
Jul	690	28,75	92,74
Ago	729	30,38	97,98
Set	688	28,67	95,56
Out	728	30,33	97,85
Nov	689	28,71	95,69
Dez	712	29,67	95,70
Anual	8551	356,29	97,35

Fonte: Elaborada pela autora (2022).

Conforme apresentado na literatura, normalmente é necessária uma combinação de diferentes métricas para avaliar o desempenho da previsão, visto que não existe uma métrica padrão que pode ser utilizada em todos os modelos de *machine learning*. Assim, conforme visto na seção 2, os valores próximos de zero são adequados para MAE e RMSE, e, valores próximos de 0% são adequados para MAAPE.

Dessa forma, de acordo com a Tabela 5, o algoritmo XGBoost foi superior em todos os modelos analisados. Conforme destaque, no Modelo 4 as métricas avaliativas foram relativamente superiores quando comparada com os outros modelos, e, há uma notável redução no Erro máximo (*Max Error*). A capacidade de acertos para o modelo 4, conforme o MBE, foi considerada levemente subestimada, com valor de - 0,01 m/s. Além disso, conforme González-Sopeña, Pakrashi e Ghosh (2021), o MAPE pode produzir valores elevados quando os valores reais estão próximos de zero, e, é indefinido quando o valor real é igual a zero. Nesse caso, ocorreu uma diferença de aproximadamente 3% no modelo 4 entre o valor do MAPE e a sua opção alternativa, o MAAPE. Esse comportamento é justificado, pois existem velocidades do vento próximas a zero na base de dados. A velocidade média mínima real do conjunto de dados de teste é de 0,13 m/s, dessa maneira, valores semelhantes podem impactar a métrica MAPE.

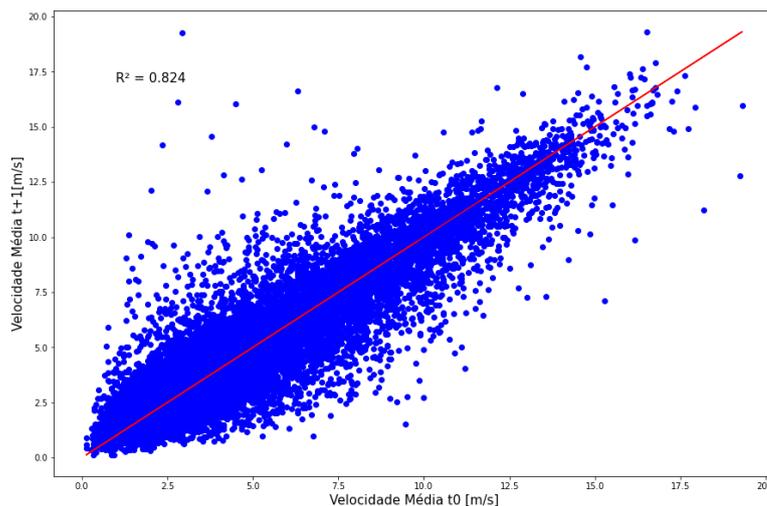
Tabela 5 - Resultados das métricas para a base de dados de teste com diferentes modelos de *machine learning* e variáveis de entrada.

MÉTRICAS	M1			M2			M3			M4		
	RL	RNA	XGB	RL	RNA	XGB	RL	RNA	XGB	RL	RNA	XGB
RMSE (m/s)	1,32	1,29	1,21	2,58	2,60	1,87	2,50	2,47	2,01	1,81	1,28	1,15
R²	0,83	0,84	0,85	0,36	0,35	0,66	0,40	0,41	0,61	0,67	0,83	0,87
MAPE (%)	27	28	26	59	60	43	54	51	43	43	28	25
MAAPE (%)	24	24	22	42	42	32	41	39	34	34	24	22
MBE (m/s)	-0,02	0,00	-0,03	0,02	-0,02	-0,07	0,07	0,11	-0,01	-0,03	-0,02	-0,01
MAE (m/s)	0,93	0,94	0,85	1,98	2,00	1,40	1,93	1,88	1,50	1,40	0,96	0,84
Max Error (m/s)	13,08	10,21	11,78	8,86	9,17	8,60	10,31	10,23	8,17	7,34	6,21	5,52

Fonte: Elaborada pela autora (2022).

O conjunto de parâmetros de entrada do Modelo 4 foi selecionado para uma análise mais detalhada utilizando o ano de 2020 (validação) com o algoritmo XGBoost. Esse procedimento de avaliação foi adotado para testar a capacidade do modelo. Além de apresentar métricas relativamente superiores, o modelo não utilizou somente a variável alvo (velocidade média) no conjunto de entrada durante as fases de treinamento, teste e validação. Essa escolha também se deve ao fato de que velocidade alvo tem uma forte correlação com a velocidade de entrada onde $R^2 = 0,824$, Figura 14, e, então, os modelos ficam enviesados com alta dependência de uma única variável.

Figura 14 - Gráfico de dispersão base de treino.



Fonte: Elaborada pela autora (2022).

Os hiperparâmetros ótimos foram selecionados por meio da aplicação da estratégia busca em grade (*Grid Search*) com validação Cruzada (*Cross Validation*), cujos valores selecionados são mostrados na Tabela 6. Vale destacar que essa abordagem testa todas as

possibilidades para descobrir o melhor conjunto de parâmetros, explorando todas as combinações possíveis e aplicada para todos os modelos de XGBoost.

Tabela 6 - Resultados dos parâmetros conforme o *GridSearchCV* aplicado para o modelo selecionado (M4) XGBoost.

<i>Parâmetro</i>	<i>Resultado</i>
<i>max_depth</i>	6
<i>learning_rate</i>	0.01
<i>n_estimators</i>	1000
<i>reg_alpha</i>	1.2
<i>reg_lambda</i>	1.2
<i>Subsample</i>	0.5
<i>Gamma</i>	1
<i>min_child_weight</i>	5
<i>colsample_bytree</i>	1

Fonte: Elaborada pela autora (2022).

Com a otimização dos parâmetros foi possível realizar previsões com a base de dados de validação. As previsões mensais utilizam dados do ano de 2020 para validar e avaliar a capacidade de previsão. A Tabela 7 apresenta os resultados estatísticos mensais (média/desvio padrão) para a base de validação. Dessa forma, ao realizar a previsão horária, a velocidade média mensal ficou próxima para todos os meses.

Tabela 7 - Resultados da análise estatística mensal com base no ano de validação (2020).

Velocidade [m/s]	Validação (2020)			
	Média Real	Média Prevista	Desvio Padrão Real	Desvio Padrão Previsto
Mês				
Janeiro	4,95	5,01	3,25	3,02
Fevereiro	5,34	5,29	3,16	2,90
Março	4,67	4,66	2,95	2,75
Abril	5,27	5,21	3,30	3,01
Mai	4,45	4,49	2,69	2,51
Junho	4,47	4,47	3,02	2,71
Julho	4,61	4,60	2,98	2,70
Agosto	4,90	4,87	3,50	3,20
Setembro	5,18	5,21	3,52	3,24
Outubro	6,20	6,14	3,48	3,26
Novembro	5,14	5,15	3,15	2,93
Dezembro	5,17	5,16	3,13	2,88

Fonte: Elaborada pela autora (2022).

Os resultados das métricas mensais estão dispostos na Tabela 8. Os tons mais frios (verde) representam as melhores métricas: menores valores para RMSE, MAE, MAPE, MAAPE e Erro Máximo, e, maiores valores para R². As cores que apresentam variação entre

frio e quente, são os valores intermediários. Em tons mais quentes (vermelho) referem-se aos valores com desempenhos relativamente inferiores, com valores maiores para RMSE, MAE, MAAPE e Erro Máximo, e, menores valores para R^2 . Em cinza são apresentados os resultados de MBE, visto que podem ser positivos ou negativos.

Destaca-se que, quanto mais próximo de zero melhor os resultados, exceto para o R^2 . Entretanto, dada a intermitência dos ventos, em especial para a região de estudo, não se espera que as métricas de desempenho sejam exatamente zero. Nos meses em que o MBE apresentou valores negativos o algoritmo XGBoost obteve resultados de maneira subestimada, e, em caso positivo sobrestimada. Conforme os resultados, MBE ficou próximo de zero em ambos os meses.

Os valores de MAE foram menores que 1,00 m/s em onze meses de 2020, e, em setembro o valor ficou em 1,02 m/s. O que implica que o XGBoost apresenta uma boa capacidade de previsão. Os resultados são semelhantes aos apresentados por Hiruma (2016), que utilizando RNA com dados da Usina Eólico-Elétrica de Palmas (PR), obteve os valores das métricas R^2 e RMSE de 0,845 e 1,055 m/s, respectivamente, para Janeiro e Fevereiro. Para Julho e Agosto o valor de R^2 foi de 0,829, enquanto que a métrica RMSE ficou em 1,088 m/s. Cabe destacar que os resultados não podem ser comparados diretamente, visto que cada região de estudo tem suas características físicas e climáticas.

Tabela 8 - Resultados de métricas mensais – validação (2020).

Mês	Métricas - Validação						
	MAE (m/s)	RMSE (m/s)	R^2	MBE (m/s)	MAPE (%)	MAAPE (%)	Erro Máx. (m/s)
Jan	0,75	1,10	0,89	-0,06	26,00	21,00	7,39
Fev	0,84	1,17	0,86	0,05	23,00	20,00	5,22
Mar	0,70	0,98	0,89	0,01	24,00	20,00	6,35
Abr	0,92	1,30	0,85	0,07	24,00	21,00	6,82
Mai	0,89	1,25	0,79	-0,04	27,00	23,00	6,55
Jun	0,87	1,26	0,83	0,00	27,00	23,00	6,30
Jul	0,89	1,25	0,82	0,01	28,00	23,00	6,32
Ago	0,87	1,31	0,86	0,03	28,00	23,00	14,76
Set	1,02	1,50	0,82	-0,03	29,00	24,00	7,97
Out	0,94	1,27	0,87	0,06	24,00	20,00	5,74
Nov	0,76	1,05	0,89	-0,02	22,00	19,00	5,10
Dez	0,81	1,11	0,87	0,01	24,00	21,00	5,27

Fonte: Elaborada pela autora (2022).

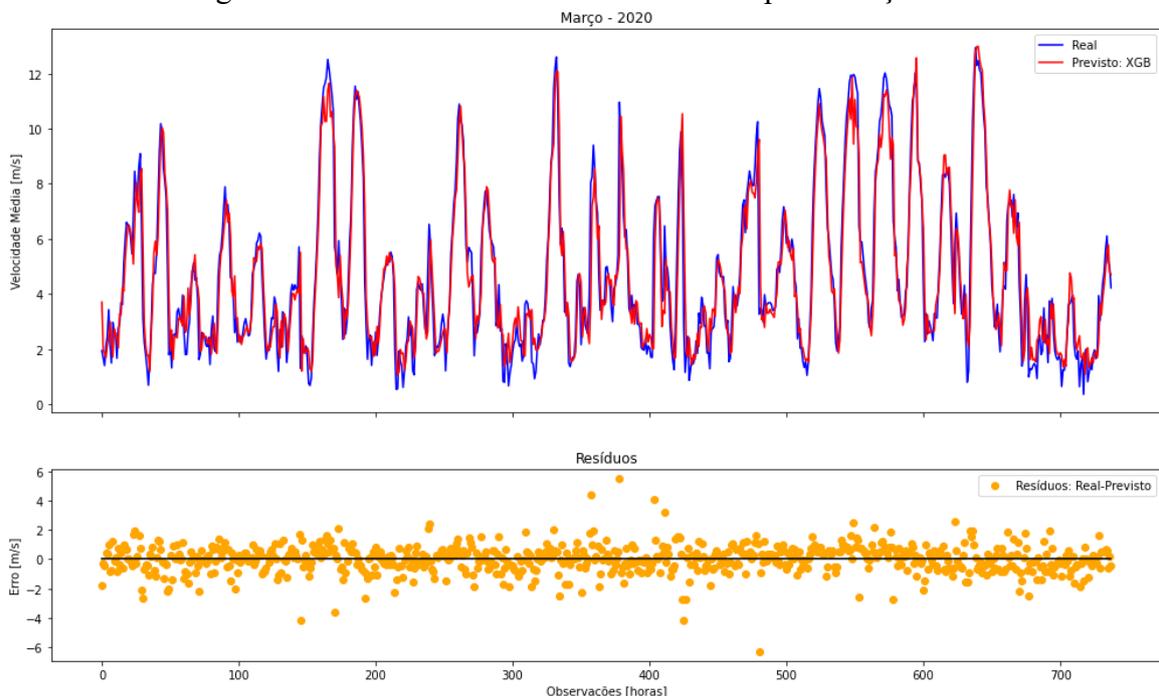
O cálculo de RMSE penaliza grandes erros de previsão, Equação 3, seção 2.4. Assim, a métrica expressa possíveis custos financeiros ao utilizar o modelo proposto. Destaca-se que os

valores mensais encontrados demonstram que foram obtidas previsões da velocidade do vento com pequena diferença para o valor real.

Assim, por meio das métricas de avaliação foi possível determinar a capacidade de previsão do modelo utilizando XGBoost. Contudo, conforme os resultados, o mês de março foi o que apresentou grande parte das melhores métricas de desempenho, a saber: MAE, RMSE e R^2 . O mês de setembro apresentou uma maior discrepância na previsão, visto que suas métricas apresentaram erros relativamente maiores, a citar: MAE, RMSE e MAAPE. E, ainda, o mês de agosto apresentou o maior erro máximo: 14,76 m/s.

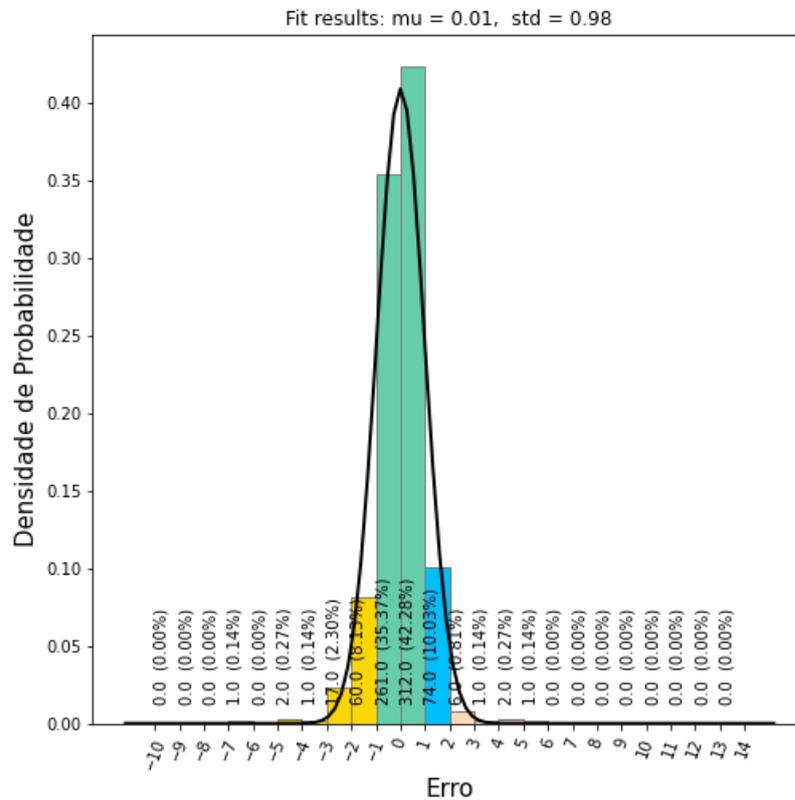
Para analisar esses meses com mais detalhes, os resultados para a velocidade real e prevista são representados graficamente. Observa-se na Figura 15, mês de Março, que o modelo apresentou valores próximos aos dados reais. Na Figura 16, identifica-se uma maior frequência de erros localizados entre -1 a 0 e entre 0 a 1, onde combinados, representam 77,65% do total. Isso evidencia os acertos das previsões. Entretanto, ainda existem limitações no modelo ao analisar cerca de 18,16% entre -2 a -1 e 1 a 2.

Figura 15 - Previsão da velocidade do vento para Março de 2020.



Fonte: Elaborada pela autora (2022).

Figura 16 - Resíduos XGBoost: Março de 2020.

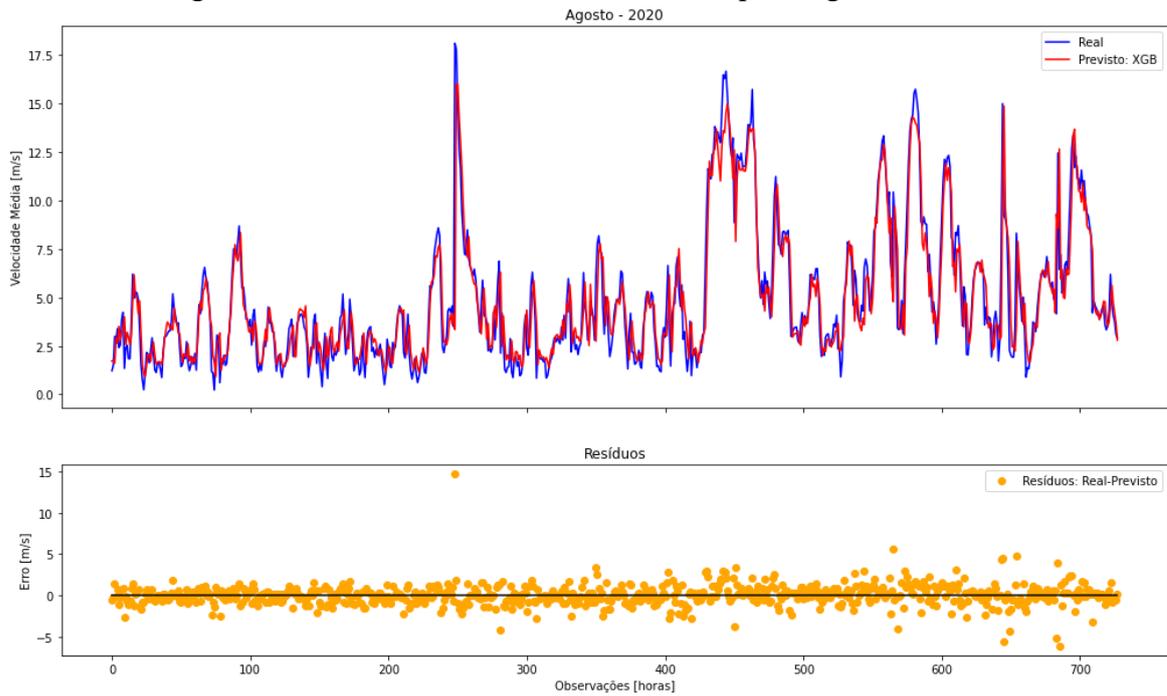


Fonte: Elaborada pela autora (2022).

Conforme a Figura 17, a previsão para Agosto foi levemente subestimada em alguns pontos. O modelo apresenta uma boa capacidade de previsão, com uma maior frequência de erros localizados entre -1 a 0 e entre 0 a 1 (Figura 18), os quais, representam 70,00% do total.

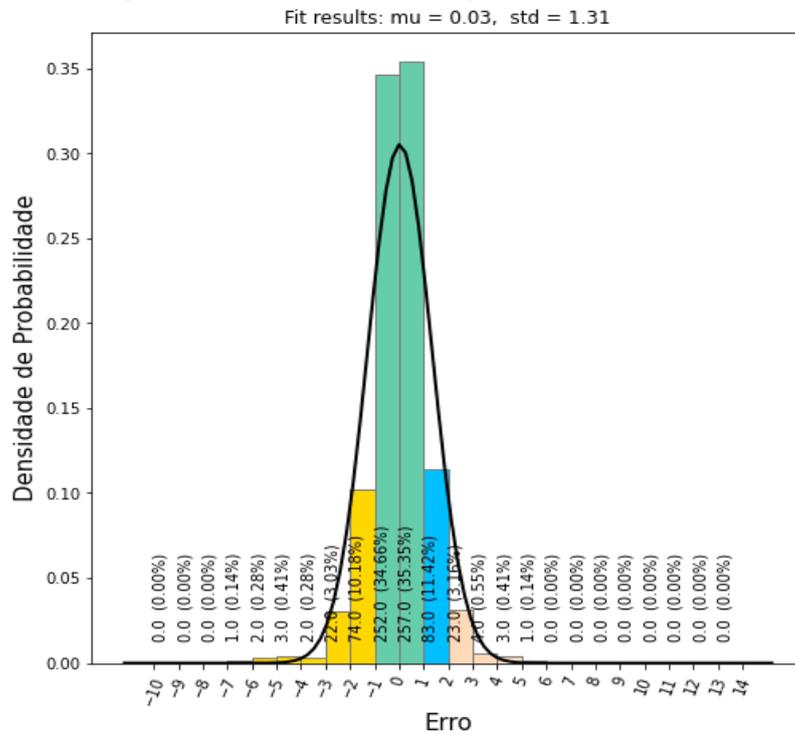
Por outro lado, nesse mês ocorreu o maior erro máximo na base de dados de validação (14,76 m/s). Essa limitação pode ser devida ao fato de a velocidade do vento ter apresentado um salto na hora seguinte, saindo de 3,50 m/s às 8 h:00 min UTC para 18,11 m/s às 9 h:00 min UTC no dia 11 de agosto. Nesse dia, a Defesa Civil de Santa Catarina emitiu um alerta para evento de rajadas de vento, noticiada às 07h:23 min da seguinte forma: “Rajadas de vento com intensidade moderada a forte (entre 50 e 65 km/h) com direção noroeste passando para sudoeste/sul para a região Litoral Sul nas próximas 2 horas.” (SANTA CATARINA, 2020). O modelo foi capaz de prever a alta velocidade, porém não atingiu os valores máximos reais influenciados pelas rajadas.

Figura 17 - Previsão da velocidade do vento para Agosto de 2020.



Fonte: Elaborada pela autora (2022).

Figura 18 - Resíduos XGBoost: Agosto de 2020.

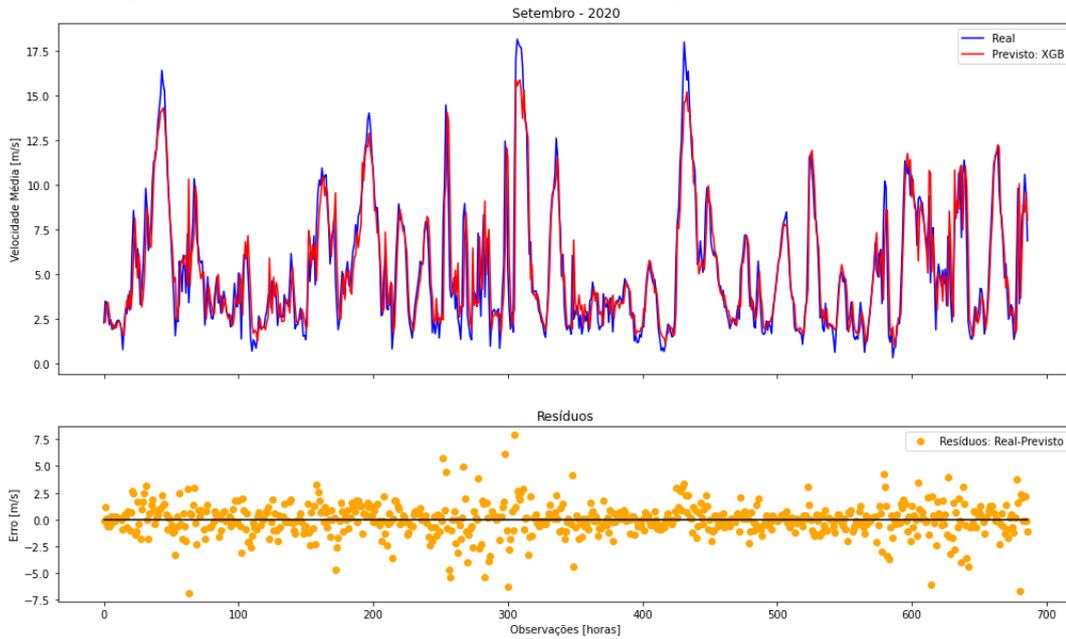


Fonte: Elaborada pela autora (2022).

Para o mês de Setembro (Figura 19), o modelo apresenta uma maior frequência de erros localizados entre -1 a 0 e entre 0 a 1, que representam 64,19% do total, conforme a Figura 20. Comparado com os outros meses apresentados anteriormente, essa foi a menor frequência de

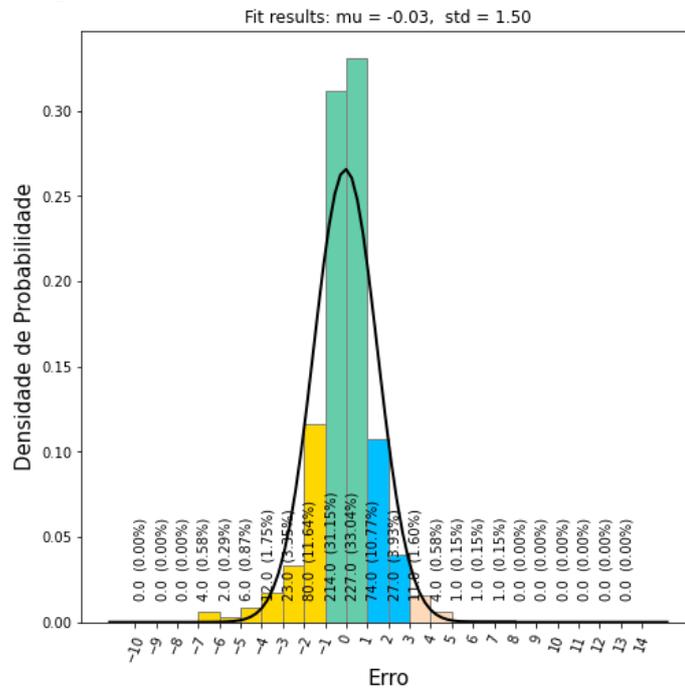
acertos. O modelo XGBoost teve a capacidade subestimada na previsão de valores próximos ou superiores à 15,00 m/s, Figura 19. Cabe destacar que não houveram grandes diferenças nas métricas em relação aos outros meses avaliados.

Figura 19 - Previsão da velocidade do vento para Setembro de 2020.



Fonte: Elaborada pela autora (2022).

Figura 20 - Resíduos XGBoost: Setembro de 2020.



Fonte: Elaborada pela autora (2022).

Nota-se que a maior frequência de erros nos meses de Março, Agosto e Setembro está localizada entre -1 a 0 e entre 0 a 1, sendo que esse intervalo fica entre 64,14% e 77,65%. Isso evidencia os acertos das previsões. Para os demais meses do ano, os gráficos mensais e os histogramas de resíduos estão nos Apêndices A até I.

Após a avaliação mensal, a análise exploratória anual foi realizada para comparar os dados reais e previstos, Tabela 9. Quando comparado com a média real, a média prevista apresentou um erro percentual de apenas 0,20%.

Tabela 9 - Resultados da análise estatística para a base de dado de validação (2020), comparando os dados reais para o modelo de previsão selecionado.

Validação (2020)		
Velocidade [m/s]	Média	Desvio Padrão
Real	5,02	3,20
XGBoost	5,01	2,95

Fonte: Elaborada pela autora (2022).

Li e Shi (2010) utilizaram dados de dois locais na Dakota do Norte, para Hannaford os melhores valores de MAE e RMSE foram de 0,951 m/s e 1,254 m/s, respectivamente. Para a cidade de Kulm, a métrica MAE foi de 1,112 m/s e o RMSE foi de 1,444 m/s. Observando a Tabela 10, os resultados para as métricas MAE e RMSE ficaram relativamente abaixo desses valores.

Tabela 10 - Resultados das métricas anuais para 2020 calculadas para o modelo selecionado.

Modelo	Métricas - Validação						
	MAE (m/s)	RMSE (m/s)	R ²	MBE (m/s)	MAPE (%)	MAAPE (%)	Máx. Erro (m/s)
XGB	0,84	1,22	0,86	0,01	26	22	14,76

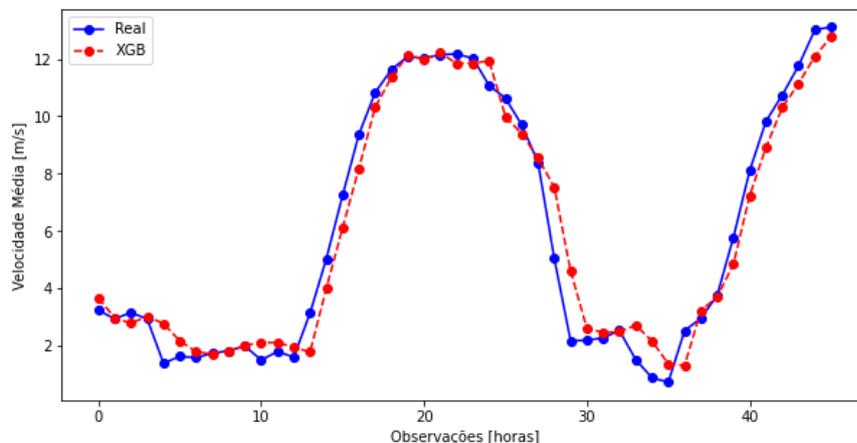
Fonte: Elaborada pela autora (2022).

Os valores anuais do MAE, RMSE e R² foram de 0,84 m/s, 1,22 m/s e 0,86 respectivamente. A métrica R² foi semelhante ao valor obtido por Zucatelli et al. (2019) no Uruguai, que para altura de 10 m obteve 0,82, e a 25,7 m alcançou 0,81. Além disso, o resultado do MAPE vai ao encontro do que foi afirmado por González-Sopeña, Pakrashi e Ghosh (2021). Assim, o MAAPE, apresentou cerca de 4% de diferença percentual quando comparado com a MAPE.

De acordo com o valor anual do MBE, a capacidade de acertos anual foi considerada levemente superestimada (0,01 m/s), Tabela 10. Cabe destacar que, o algoritmo XGBoost é relativamente novo e existem lacunas quanto as pesquisas científicas relacionadas a velocidade do vento costeiro, principalmente, na região Sul do Brasil e na previsão com uma hora de antecedência.

A Figura 21 apresenta uma previsão para um conjunto de 50 dados da base de validação (2020), representando uma série temporal onde não houveram falhas de medição. O comportamento da previsão da velocidade do vento (vermelho) segue a tendência dos dados observados (azul), apresentando pequenas diferenças dentro dos percentuais de erros descritos. Assim, pode-se concluir que o modelo utilizando XGBoost teve um bom desempenho para a previsão dessa série temporal.

Figura 21 - Previsão com 50 dados horários considerando a base de dados de validação para o modelo selecionado (2020).



Fonte: Elaborada pela autora (2022).

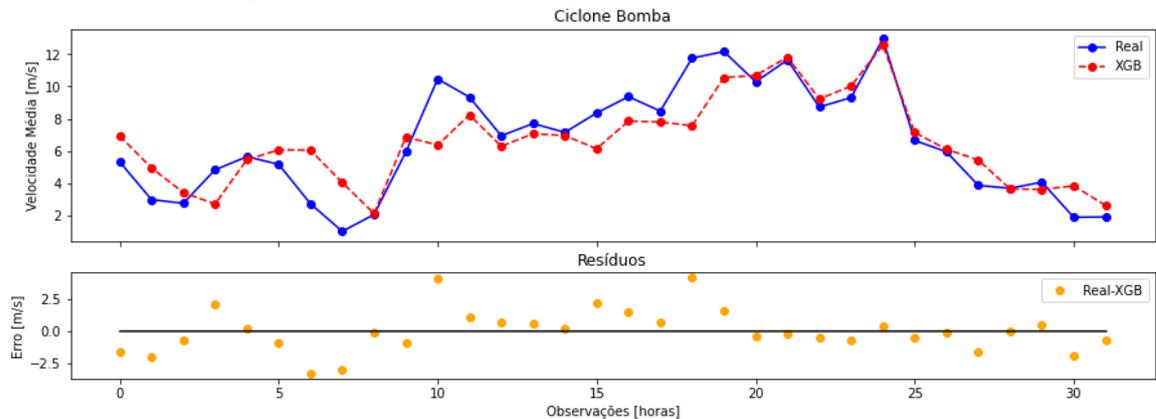
Considerando a presença de parques eólicos no sul do Brasil e a ocorrência de eventos extremos na região, investigou-se a capacidade de previsão do modelo em um dos fenômenos mais noticiados pela imprensa após o furacão Catarina. Em 2020 verificou-se a ocorrência de um intenso fenômeno climático no Estado catarinense, que ficou conhecido e divulgado na mídia nacional, como Ciclone Bomba.

O evento extremo (Ciclone Bomba) teve início, em Balneário Arroio do Silva - SC (local da estação meteorológica BOOA que forneceu os dados para o presente estudo), no dia 30 de junho durante o período da noite. Os equipamentos da estação meteorológica, coletaram dados durante sua ocorrência, totalizando 10 horas de medições antes de apresentarem falhas/lacunas.

A Figura 22 apresenta dados observados (azul) e previstos (vermelho) com o modelo XGBoost a partir do dia 30/06 às 9:00 UTC. A primeira falha ocorrida durante o período de investigação do evento, foi observada às 19:00 UTC com reinício do registro das medições às 20:00 UTC. A atuação do ciclone no local de medições começou a ser identificada a partir da 8ª observação (dia 30/06 19:00 UTC), e, o pico de velocidade média observado ficou em torno de 14,00 m/s que ocorreu às 12:00 UTC do dia seguinte (01/07) com máxima registrada de 23,43 m/s, antes do desligamento total da estação meteorológica. De maneira geral, o modelo

conseguiu reproduzir o padrão do fenômeno climático extremo (Figura 22 superior). Destaca-se que no pico da velocidade média, por volta de 14,00 m/s, o modelo se aproximou do valor real.

Figura 22 - Observações e previsões do Ciclone Bomba.

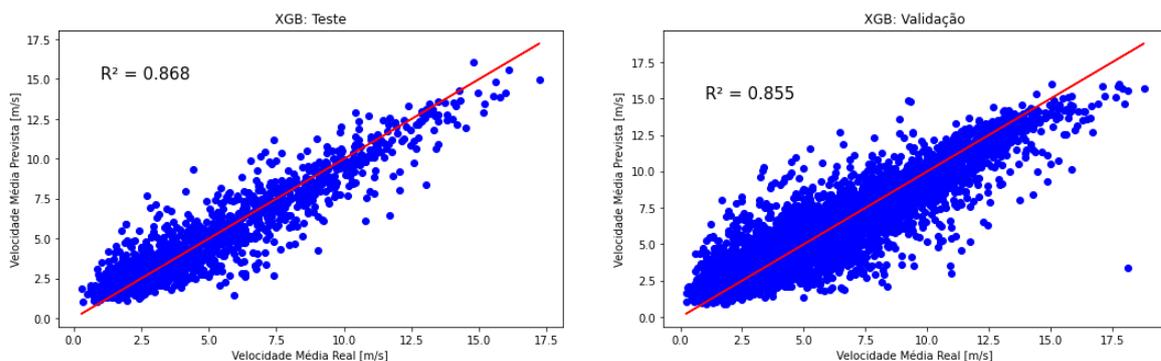


Fonte: Elaborada pela autora (2022).

É importante salientar que, mesmo sem nenhum evento deste tipo e magnitude presente nos dados de treinamento, observa-se que os valores previstos com o modelo XGBoost foram relativamente próximos dos valores observados. Esse resultado é promissor, observando os resíduos (Figura 22 inferior), existe margem para que o desempenho seja aprimorado, e que a capacidade do modelo possa ser testada em fenômenos climáticos extremos que apresentem menos falhas na coleta de dados.

Na Figura 23, é apresentado o gráfico de dispersão para o melhor modelo. Os valores para os bancos de teste e validação foram de 0,868 e 0,855, respectivamente, apresentando uma boa proximidade dos valores previstos com os reais. De acordo com os resultados de validação, o modelo apresentou valores promissores de R^2 , o que representa que o modelo se ajusta bem aos dados.

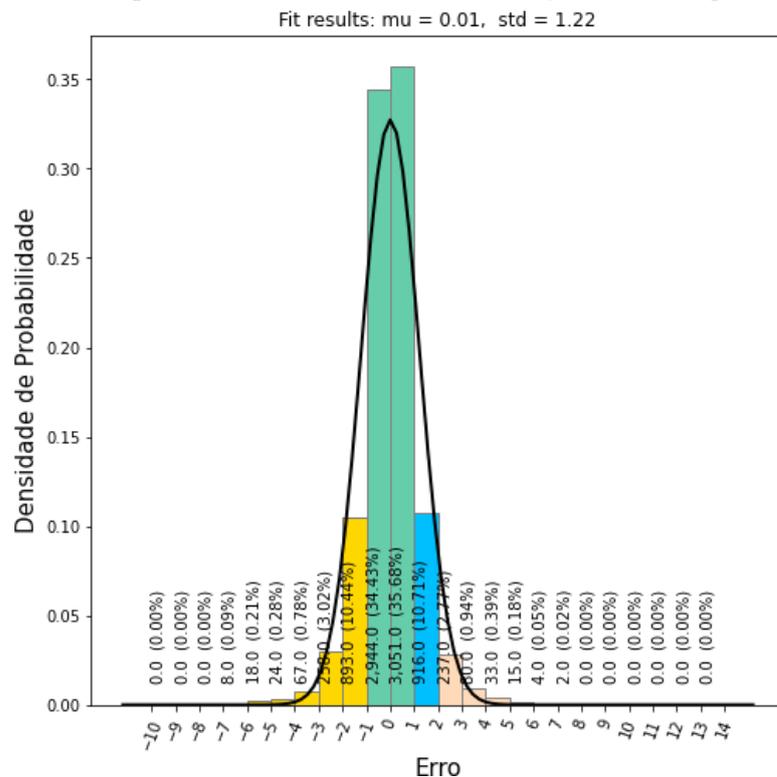
Figura 23 - Gráfico de dispersão (R^2) para o melhor modelo (M4) utilizando XGBoost.



Fonte: Elaborada pela autora (2022).

Conforme a Figura 24, percebe-se que a maior frequência de erros é entre -1 a 0 e entre 0 a 1 onde combinados representam 70,11% do total. Entretanto, ainda existem limitações no modelo ao analisar cerca de 21,15% dos erros que se encontram na faixa entre -2 a -1 e 1 a 2. Algumas hipóteses são levantadas: as relações entre as variáveis meteorológicas utilizadas no treino, quantidade de dados disponíveis para o processo de aprendizagem e, as falhas nas medições, são fatores que influenciam diretamente os resultados obtidos. Cabe destacar que a análise foi realizada com todas as 8551 medições de 2020.

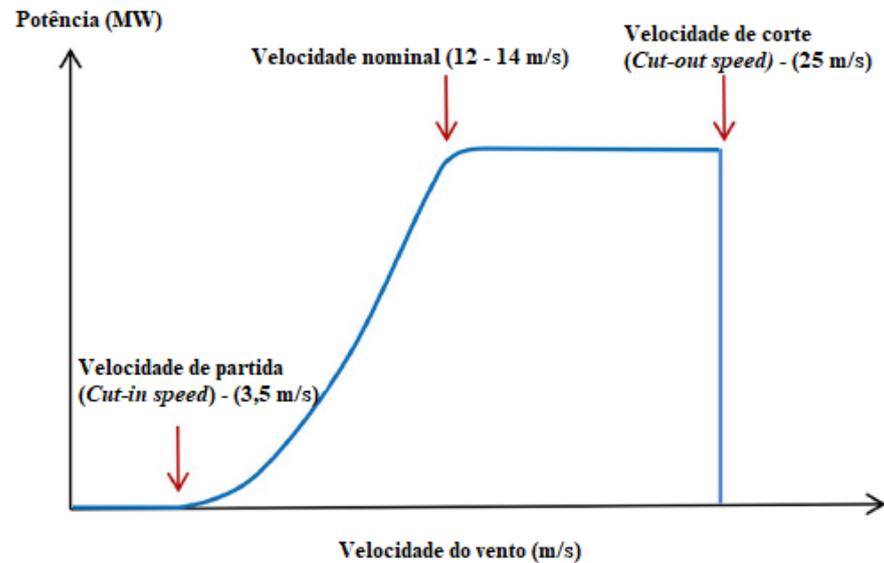
Figura 24 - Resíduos para o banco de dados de validação com o algoritmo XGBoost.



Fonte: Elaborada pela autora (2022).

Visando a utilização futura do modelo no setor de energia eólica, realizou-se uma avaliação de acordo com a curva de potência típica de um aerogerador, conforme Figura 25. São apresentadas as velocidades de partida (cerca de 3,5 m/s) e de velocidade de corte (cerca de 25 m/s).

Figura 25 - Curva de potência típica de um aerogerador.



Fonte: Neill e Hashemi (2018), adaptado

A faixa de velocidades da curva de potência e as métricas utilizadas para sua avaliação são apresentadas na Tabela 11. Conforme os resultados, para velocidades médias $\leq 3,5$ m/s, os valores anuais de RMSE e MAAPE foram de 1,03 m/s e 32%, respectivamente. Para a faixa de operação da turbina, o valor de RMSE foi de 1,33 m/s e MAAPE de 14%. Nesse intervalo de velocidade o RMSE apresentou um valor relativamente maior devido aos maiores valores de velocidade. Contudo, o MAAPE reduziu, pois, as velocidades próximas de zero impactam no cálculo dessa métrica. No banco de dados não foram observadas medições acima de 25 m/s.

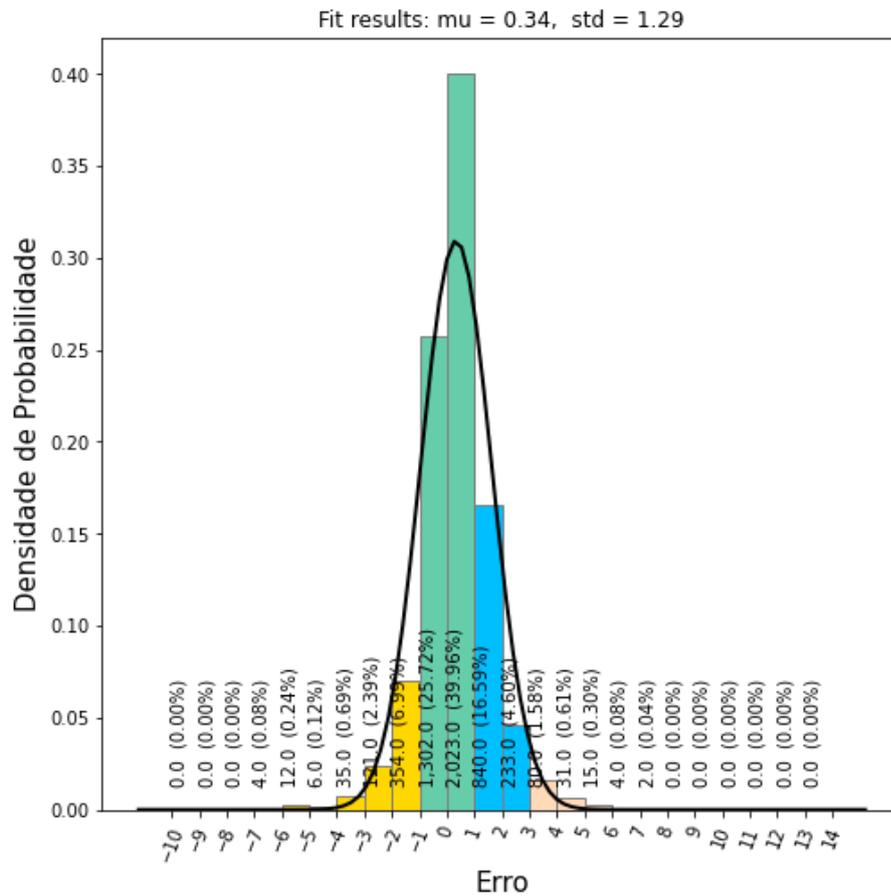
Tabela 11 - Faixa de velocidades baseada na curva de potência - ano de 2020.

Intervalo de velocidade	RMSE (m/s)	MAAPE (%)
Velocidade $\leq 3,5$ m/s	1,03	32,00
$3,5$ m/s < Velocidade < 25 m/s (Faixa de operação de uma turbina típica)	1,33	14,00
Velocidade ≥ 25 m/s	-	-

Fonte: Elaborada pela autora (2022).

Os resíduos entre os valores reais e previstos, utilizando a faixa de operação de uma turbina, são apresentados na Figura 26. Conforme os resultados, percebe-se que a maior frequência de erros está entre -1 a 0 e entre 0 a 1 onde combinados representam 65,68% do total. Além disso, dos 8551 dados pertencentes ao banco de validação, 5063 estão na faixa de operação de uma turbina, correspondendo à cerca de 59,20% do total.

Figura 26 - Resíduos para o banco de dados de validação utilizando o algoritmo XGBoost para a faixa de operação de uma turbina.



Fonte: Elaborada pela autora (2022).

A região de estudo apresenta grande variação de velocidades, com médias entre 5,00 e 7,00 m/s e máximas entre 19,00 e 28,00 m/s. Dessa forma, o desempenho do modelo com o algoritmo XGBoost demonstrou-se promissor. Assim, espera-se que esse trabalho contribua para o desenvolvimento e evolução de soluções voltadas para a previsão da velocidade do vento utilizando modelos de *machine learning* com foco no recurso eólico costeiro.

5 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi analisar o desempenho de algoritmos de *machine learning* na previsão da velocidade do vento costeiro com uma hora de antecedência. Dessa forma, com base nessa análise foi possível selecionar o melhor modelo e, assim, testar seu desempenho para o conjunto de dados destinados para validação.

Os três métodos de aprendizado de máquina (Redes Neurais Artificiais, XGBoost e Regressão Linear Multivariável) foram avaliados em termos de acurácia utilizando as métricas: MAE, RMSE, MBE, MAAPE, Máximo Erro e R^2 . O XGBoost foi selecionado como o método de melhor desempenho.

O conjunto de parâmetros do Modelo 4 com XGBoost foi selecionado e validado utilizando o ano de 2020. Cabe destacar que, além de apresentar métricas relativamente superiores, o modelo não utilizou a variável alvo (velocidade média) no conjunto de entrada durante as fases de treinamento, teste e validação. Dessa forma, não ocorreu enviesamento pela dependência de uma única variável.

Conforme as métricas anuais destacam-se os valores de MAE, RMSE e R^2 que apresentaram valores de 0,84 m/s, 1,22 m/s e 0,86, demonstrando a capacidade de previsões promissoras com antecedência de 1 hora. De acordo com os resultados, a maior frequência de resíduos ficou entre -1 a 0 e entre 0 a 1 m/s, onde combinados representaram 70,11% do total de dados.

Destacam-se que os valores de MAE abaixo de 1,00 em 11 meses do ano, exceto para setembro (1,02 m/s) e MBE próximo de zero para todos os meses, também os valores de RMSE próximos de 1,00 m/s. Os resultados obtidos demonstram que as previsões com o horizonte de previsão de 1 hora ficaram próximas ao valor real da velocidade do vento. O XGBoost conseguiu reproduzir um fenômeno meteorológico extremo registrado no ano de 2020, o ciclone bomba. Para o evento, o modelo conseguiu alcançar o máximo de velocidade do vento registrado pela base de dados da estação meteorológica.

A previsão da velocidade do vento desempenha um papel importante na mitigação da intermitência e incerteza do recurso. Um modelo de previsão preciso é relevante para sistemas de energia quanto ao planejamento e despacho antecipado de eletricidade. O crescimento da geração eólica e a demanda energética aumentam a necessidade de soluções de previsão de velocidade do vento em diferentes horizontes. Diante do crescimento da participação da energia eólica na matriz elétrica brasileira nos últimos anos, destaca-se que um aspecto importante do estudo é que a previsão do recurso eólico pode ser aplicada em qualquer tecnologia de turbina.

Ao analisar os resultados, para a utilização futura do modelo no setor de energia eólica, o desempenho foi promissor. Considerando a faixa de operação de uma turbina, o valor de MAAPE ficou em 14%, e, a maior frequência de erros localizada entre -1 e 1m/s, com 65,68% do total. Dessa forma, os resultados apresentados e discutidos nesse trabalho são de grande importância para o setor de energia.

Os resultados apontam informações relevantes para estudos posteriores, existem questões que precisam ser consideradas nos modelos além de, somente, otimizações nos algoritmos de ML. Portanto, sugere-se para trabalhos futuros combinações de variáveis de entrada, adição de recursos espaciais, como diferentes recursos ambientais e geográficos, um período maior de coleta de dados, incremento da altura e horizonte de previsão e, também, a adoção de técnicas de otimização, tais como: *Isolation Forest* e algoritmo de Redução de Dimensionalidade PCA (*Principal Component Analysis*). Espera-se que isso permita uma flexibilização e uma maior disponibilidade na obtenção de relações potenciais e importantes.

Assim, recomenda-se a continuação e aperfeiçoamento dos algoritmos e métodos de *Machine Learning* focando na previsão de ventos costeiros, como por exemplo: Redes Neurais Artificiais Profundas/ *Deep Learning*, *Long Short-Term Memory* (LSTM), *Light Gradient Boosting Machine*, *CatBoost*, combinações de técnicas, entre outros. Recomenda-se também que o modelo seja testado em diferentes regiões geográficas, com padrão climatológico diferente, para comparar os resultados e obter uma nova perspectiva.

Os resultados apresentados são específicos para a região de estudo e isso ocorre pela particularidade das características e relações climáticas e físicas. Porém, vale destacar que a metodologia adotada nesse trabalho pode ser aplicada e reproduzida em outros locais. Existem lacunas quanto as pesquisas científicas, principalmente, na região Sul do Brasil, mais investigações estão em andamento para o desenvolvimento de modelos híbridos combinando diferentes métodos. Salienta-se que esses resultados são promissores e contribuem com a comunidade científica, até o momento, não foram identificados na literatura estudos com diferentes técnicas computacionais para previsão da velocidade do vento costeiro no Extremo Sul Catarinense.

REFERÊNCIAS

ABEEÓLICA - Associação Brasileira de Energia Eólica. **INFOVENTO Brasil nº. 25 – 2022**. Disponível em: https://abeeolica.org.br/wp-content/uploads/2022/04/2022_03_InfoVento25.pdf. Acesso em: 10 mai. 2022.

ABEEÓLICA - Associação Brasileira de Energia Eólica. **Boletim Anual: dados de 2020**. https://abeeolica.org.br/wp-content/uploads/2022/04/PT_Boletim-Anual-de-Geracao_2020.pdf. Acesso em: 10 mai. 2022.

AGUILAR MADRID, Ernesto; ANTONIO, Nuno. Short-Term Electricity Load Forecasting with Machine Learning. **Information**, v. 12, n. 2, p. 50, 22 jan. 2021. DOI 10.3390/info12020050. Disponível em: <https://www.mdpi.com/2078-2489/12/2/50>. Acesso em: 4 mai. 2022.

ALENCAR, David Barbosa de. **Modelo híbrido baseado em séries temporais e redes neurais para previsão da geração de energia eólica**. 2018. 120 f. Tese (Engenharia Elétrica) - Universidade Federal do Pará, Belém, 2018. Disponível em: <http://repositorio.ufpa.br/jspui/handle/2011/10416>. Acesso em: 1 mai. 2022.

ALKHAYAT, Ghadah; MEHMOOD, Rashid. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. **Energy and AI**, v. 4, p. 100060, jun. 2021. DOI 10.1016/j.egyai.2021.100060. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666546821000148>. Acesso em: 9 maio 2022.

ANEEL. Agência Nacional de Energia Elétrica. **Sistema de Informações de Geração da ANEEL - SIGA**. Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiNjc4OGYyYjQtYWM2ZC00YjllLWJlYmEtYzdkNTQ1MTc1NjM2IiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBIMSIsImMiOiR9>. Acesso em: 02 mai. 2022.

AZAD, Hanieh Borhan; MEKHILEF, Saad; GANAPATHY, Vellapa Gounder. Long-Term Wind Speed Forecasting and General Pattern Recognition Using Neural Networks. **IEEE Transactions On Sustainable Energy**, [S.L.], v. 5, n. 2, p. 546-553, abr. 2014. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tste.2014.2300150>.

AZADI, Sama; KARIMI-JASHNI, Ayoub. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: a case study of fars province, Iran. **Waste Management**, [S.L.], v. 48, p. 14-23, fev. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.wasman.2015.09.034>.

BARHMI, Soukaina; ELFATNI, Omkaltoume; BELHAJ, Ismail. Forecasting of wind speed using multiple linear regression and artificial neural networks. **Energy Systems**, [S.L.], v. 11, n. 4, p. 935-946, 11 maio 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s12667-019-00338-y>.

BEGAM, K. Maruliya; DEEPA, S.N.. Optimized nonlinear neural network architectural models for multistep wind speed forecasting. **Computers & Electrical Engineering**, [S.L.], v. 78, p. 32-49, set. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.compeleceng.2019.06.018>.

BILAL, B.; NDONGO, M.; ADJALLAH, K. H.; SAVA, A.; KEBE, C. M. F.; NDIAYE, P. A.; SAMBOU, V.. Wind turbine power output prediction model design based on artificial neural networks and climatic spatiotemporal data. **2018 IEEE International Conference On Industrial Technology (Icit)**, [S.L.], fev. 2018. IEEE.
<http://dx.doi.org/10.1109/icit.2018.8352329>

BILGILI, Mehmet; SAHIN, Besir. Comparative analysis of regression and artificial neural network models for wind speed prediction. **Meteorology and Atmospheric Physics**, v. 109, n. 1–2, p. 61–72, nov. 2010. <https://doi.org/10.1007/s00703-010-0093-9>.

BRAGA, Antônio de Pádua; CARVALHO, André Carlos Ponce de Leon Ferreira; LUDERMIR, Teresa Bernarda. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC Editora, 2007.

BRANNSTROM, Christian; GORAYEB, Adryane; SOUZA, Wallason Farias de; LEITE, Nicolly Santos; CHAVES, Leilane Oliveira; GUIMARÃES, Rodrigo; GÊ, Dweynny Rodrigues Filgueira. Perspectivas geográficas nas transformações do litoral brasileiro pela energia eólica. **Revista Brasileira de Geografia**, [S.L.], v. 63, n. 1, p. 3-28, 31 ago. 2018. IBGE. http://dx.doi.org/10.21579/issn.2526-0375_2018_n1.

BROWNLEE, Jason. **Train Faster, Reduce Overfitting, and Make Better Predictions**. Si: Machine Learning Mastery, 2018. 575 p.

BULHÕES, Diego Bitencourt. **ESTRATÉGIAS DE PREVISÃO DE GERAÇÃO DE ENERGIA EÓLICA UTILIZANDO REDES NEURAI ARTIFICIAIS**. 2015. 86 f. Dissertação (Engenharia Elétrica) - Universidade Federal de Sergipe, São Cristóvão, 2015. Disponível em:
https://ri.ufs.br/bitstream/riufs/5015/1/DIEGO_BITENCOURT_BULHOES.pdf. Acesso em: 4 mai. 2022.

CAI, Ren; XIE, Sen; WANG, Bozhong; YANG, Ruijiang; XU, Daosen; HE, Yang. Wind Speed Forecasting Based on Extreme Gradient Boosting. **IEEE Access**, v. 8, p. 175063–175069, 2020. DOI 10.1109/ACCESS.2020.3025967. Disponível em:
<https://ieeexplore.ieee.org/document/9203893/>. Acesso em: 4 mai. 2022.

CAMELO, Henrique do Nascimento. **MODELOS HÍBRIDOS ESTOCÁSTICO - MATEMÁTICO PARA PREVISÃO DE VELOCIDADE DO VENTO**. 2018. 138 f. Tese (Doutorado) - Curso de Ciências Climáticas, Universidade Federal do Rio Grande do Norte, Natal, 2018. Disponível em: <https://repositorio.ufrn.br/handle/123456789/26651>. Acesso em: 5 mai. 2022.

CARTA, José A.; VELÁZQUEZ, Sergio; CABRERA, Pedro. A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. **Renewable And Sustainable Energy Reviews**, [S.L.], v. 27, p. 362-400, nov. 2013. Elsevier BV. <http://dx.doi.org/10.1016/j.rser.2013.07.004>.

CHAI, T.; DRAXLER, R. R.. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, [S.L.], v. 7, n. 3, p. 1247-1250, 30 jun. 2014. Copernicus GmbH.
<http://dx.doi.org/10.5194/gmd-7-1247-2014>.

CHANG, Wen-Yeau. A Literature Review of Wind Forecasting Methods. **Journal Of Power And Energy Engineering**, [S.L.], v. 02, n. 04, p. 161-168, 2014. Scientific Research Publishing, Inc., <http://dx.doi.org/10.4236/jpee.2014.24023>.

CHEN, Tianqi. **Introduction to Boosted Trees**. 2014. Disponível em: https://web.njit.edu/~usman/courses/cs675_spring20/BoostedTree.pdf. Acesso em: 1 mai. 2022.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost. **Proceedings Of The 22Nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining**, [S.L.], p. 785-794, 13 ago. 2016. ACM. <http://dx.doi.org/10.1145/2939672.2939785>.

CHEN, Yong; ZHANG, Shuai; ZHANG, Wenyu; PENG, Juanjuan; CAI, Yishuai. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. **Energy Conversion and Management**, v. 185, p. 783–799, abr. 2019. <https://doi.org/10.1016/j.enconman.2019.02.018>.

CHOUDHURY, Manabendra D.; ATTEYA, Walid A.; DAHAL, Keshav; CHETIA, Pankaj; CHOUDHURY, Karabi D.; PARADKAR, Anant. Virtual Screening of Phytochemicals. **Computational Phytochemistry**, [S.L.], p. 301-334, 2018. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-812364-5.00011-0>.

CHU, Kyung-Su; OH, Cheong-Hyeon; CHOI, Jung-Ryel; KIM, Byung-Sik. Estimation of Threshold Rainfall in Ungauged Areas Using Machine Learning. **Water**, v. 14, n. 6, p. 859, 10 mar. 2022. DOI 10.3390/w14060859. Disponível em: <https://www.mdpi.com/2073-4441/14/6/859>. Acesso em: 27 abr. 2022.

COIMBRA, Carlos F.M.; PEDRO, Hugo T.C. Stochastic-Learning Methods. **Solar Energy Forecasting and Resource Assessment**. [S. l.]: Elsevier, 2013. p. 383–406. DOI 10.1016/B978-0-12-397177-7.00015-2.

CORRÊA, Ana Garcia. **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. Disponível em: https://movlidar.paginas.ufsc.br/files/2018/03/Ana_Correa2018-1.pdf. Acesso em: 18 fev. 2022.

CRESESB – Centro de Referência para as Energias Solar e Eólica de S. Brito. **O recurso eólico**. Disponível em: http://www.cresesb.cepel.br/index.php?section=com_content&lang=pt&cid=211. Acesso em: 20 mar. 2022.

CRUZ, Karla Dayane Bezerra. **Os Impactos da Instalação De Parques Eólicos nas Comunidades Urbanas e Rurais da Serra De Santana/RN**. 2016. 130 p. Dissertação (Engenharia de Produção) - Universidade Federal do Rio Grande do Norte, Natal, 2016. Disponível em: https://repositorio.ufrn.br/jspui/bitstream/123456789/22877/1/KarlaDayaneBezerraCruz_DISERT.pdf. Acesso em: 21 set. 2021.

DEMOLLI, Halil; DOKUZ, Ahmet Sakir; ECEMIS, Alper; GOKCEK, Murat. Wind power forecasting based on daily wind speed data using machine learning algorithms. **Energy Conversion And Management**, [S.L.], v. 198, p. 111823, out. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.enconman.2019.111823>.

DHIMAN, Harsh S.; DEB, Dipankar; BALAS, Valentina Emilia. **Supervised machine learning in wind forecasting and ramp event prediction**. London, United Kingdom: Academic Press, 2020a (Wind energy engineering series).

DHIMAN, Harsh S.; DEB, Dipankar. A Review of Wind Speed and Wind Power Forecasting Techniques. **arXiv:2009.02279 [cs, eess]**, 2 set. 2020b. Disponível em: <http://arxiv.org/abs/2009.02279>. Acesso em: 10 abr. 2022.

DONG, Wei; HUANG, Yimiao; LEHANE, Barry; MA, Guowei. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. **Automation in Construction**, v. 114, p. 103155, jun. 2020. DOI 10.1016/j.autcon.2020.103155. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0926580519311148>. Acesso em: 5 maio 2022.

DOSDOĞRU, Ayşe Tuğba; İPEK, Aslı Boru. Hybrid boosting algorithms and artificial neural network for wind speed prediction. **International Journal Of Hydrogen Energy**, [S.L.], v. 47, n. 3, p. 1449-1460, jan. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.ijhydene.2021.10.154>.

EPE - Empresa de Pesquisa Energética. **Balanco Energético Nacional 2022**. Rio de Janeiro. Disponível em: https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-675/topico-631/BEN_S%C3%ADntese_2022_PT.pdf Acesso em: 10 jun. 2022.

EPE. Empresa de Pesquisa Energética. **Roadmap Eólica Offshore Brasil: Perspectivas e caminhos para a energia eólica marítima**. 2020. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/roadmap-eolica-offshore-brasil>. Acesso em: 10 jun. 2022.

ERMIS, Kemal. Experimental and Numerical Investigations of Condensation Heat Transfer in Multiport Tubes. **Exergetic, Energetic And Environmental Dimensions**, [S.L.], p. 657-676, 2018. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-813734-5.00038-x>.

FADARE, D.A. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. **Applied Energy**, v. 87, n. 3, p. 934–942, mar. 2010. <https://doi.org/10.1016/j.apenergy.2009.09.005>.

FENG, Cong; CUI, Mingjian; HODGE, Bri-Mathias; ZHANG, Jie. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. **Applied Energy**, [S.L.], v. 190, p. 1245-1257, mar. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.apenergy.2017.01.043>.

FILIK, Ümmühan Başaran; FILIK, Tansu. Wind Speed Prediction Using Artificial Neural Networks Based on Multiple Local Measurements in Eskisehir. **Energy Procedia**, [S.L.], v. 107, p. 264-269, fev. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.egypro.2016.12.147>.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, v. 38, n. 4, p. 367–378, fev. 2002. DOI 10.1016/S0167-9473(01)00065-2. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167947301000652>. Acesso em: 5 mai. 2022.

GANDHI, Hir. Comparative analysis between Wind Speed Prediction architectures using a two-hidden layer Neural Network. **2020 Fourth International Conference On I-Smac (Iot In Social, Mobile, Analytics And Cloud) (I-Smac)**, [S.L.], p. 928-931, 7 out. 2020. IEEE. <http://dx.doi.org/10.1109/i-smac49090.2020.9243425>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge, Massachusetts: The MIT Press, 2016(Adaptive computation and machine learning).

GONZÁLEZ-SOPEÑA, J.M.; PAKRASHI, V.; GHOSH, B.. An overview of performance evaluation metrics for short-term statistical wind power forecasting. **Renewable And Sustainable Energy Reviews**, [S.L.], v. 138, p. 110515, mar. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.rser.2020.110515>.

GRAF, Renata; KOLERSKI, Tomasz; ZHU, Senlin. Predicting Ice Phenomena in a River Using the Artificial Neural Network and Extreme Gradient Boosting. **Resources**, [S.L.], v. 11, n. 2, p. 12, 26 jan. 2022. MDPI AG. <http://dx.doi.org/10.3390/resources11020012>.

GUPTA, Aditya; GUSAIN, Kunal; POPLI, Bhavya. Verifying the value and veracity of extreme gradient boosted decision trees on a variety of datasets. In: 2016 11TH INTERNATIONAL CONFERENCE ON INDUSTRIAL AND INFORMATION SYSTEMS (ICIIS), dez. 2016. **2016 11th International Conference on Industrial and Information Systems (ICIIS)** [...]. Roorkee: IEEE, dez. 2016. p. 457–462. DOI 10.1109/ICIINFS.2016.8262984. Disponível em: <http://ieeexplore.ieee.org/document/8262984/>. Acesso em: 5 mai. 2022.

GWEC - Global Wind Energy Council. **Global Wind Report 2022**. Disponível em: <https://gwec.net/wp-content/uploads/2022/03/GWEC-GLOBAL-WIND-REPORT-2022.pdf>. Acesso em: 21 fev. 2022.

HANIFI, Shahram; LIU, Xiaolei; LIN, Zi; LOTFIAN, Saeid. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. **Energies**, [S.L.], v. 13, n. 15, p. 3764, 22 jul. 2020. MDPI AG. <http://dx.doi.org/10.3390/en13153764>.

HAYKIN, Simon. **Redes neurais: princípios e prática**. 2ed. - Porto Alegre: Bookman, 2001.

HIRUMA, Felipe Shogo. **Previsão da geração de energia eólica com redes neurais artificiais**. 2016. 168 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, - Universidade Federal do Paraná, Curitiba, 2016. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/43520/R%20-%20D%20-%20FELIPE%20SHOGO%20HIRUMA.pdf?sequence=1&isAllowed=y>. Acesso em: 12 mar. 2022.

HUANG, Yuwei; YANG, XI; ZHUANG, Fuzhen; ZHANG, Lishan; YU, Shengquan. Automatic Chinese Reading Comprehension Grading by LSTM with Knowledge Adaptation. **Advances In Knowledge Discovery And Data Mining**, [S.L.], p. 118-129, 2018. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-93034-3_10.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Indicadores de Desenvolvimento Sustentável. Brasil. 2015.** Rio de Janeiro: IBGE, 2015.

IEA - International Energy Agency. **Wind Power.** 2021. Disponível em: <https://www.iea.org/reports/wind-power>. Acesso em: 20 mar. 2022.

IMRAN, Mohammed; ALSUHAIBANI, Sarah A.. A Neuro-Fuzzy Inference Model for Diabetic Retinopathy Classification. **Intelligent Data Analysis For Biomedical Applications**, [S.L.], p. 147-172, 2019. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-815553-0.00007-0>.

IOSIFIDIS, Alexandros; TEFAS, Anastasios. **Deep Learning for Robot Perception and Cognition.** Si: Elsevier Academic Press, 2022.

JIADING, Jiang; FENG, Wang; RUI, Tang; LINGLING, Zhang; XIN, Xu. TS_XGB: ultra-short-term wind power forecasting method based on fusion of time-spatial data and xgboost algorithm. **Procedia Computer Science**, [S.L.], v. 199, p. 1103-1111, 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.procs.2022.01.140>.

JIANG, Ping; LIU, Zhenkun; NIU, Xinsong; ZHANG, Lifang. A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. **Energy**, v. 217, p. 119361, fev. 2021. <https://doi.org/10.1016/j.energy.2020.119361>.

JIAO, Wenjiang; HAO, Xingwei; QIN, Chao. The Image Classification Method with CNN-XGBoost Model Based on Adaptive Particle Swarm Optimization. **Information**, v. 12, n. 4, p. 156, 9 abr. 2021. DOI 10.3390/info12040156. Disponível em: <https://www.mdpi.com/2078-2489/12/4/156>. Acesso em: 4 mai. 2022.

JIN, Qingwen; FAN, Xiangtao; LIU, Jian; XUE, Zhuxin; JIAN, Hongdeng. Estimating Tropical Cyclone Intensity in the South China Sea Using the XGBoost Model and FengYun Satellite Images. **Atmosphere**, v. 11, n. 4, p. 423, 22 abr. 2020. DOI 10.3390/atmos11040423. Disponível em: <https://www.mdpi.com/2073-4433/11/4/423>. Acesso em: 4 mai. 2022.

JUNG, Jaesung; BROADWATER, Robert P.. Current status and future advances for wind speed and power forecasting. **Renewable And Sustainable Energy Reviews**, [S.L.], v. 31, p. 762-777, mar. 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.rser.2013.12.054>.

KALOGIROU, Soteris. **Solar energy engineering: processes and systems.** Second edition. Amsterdam; Boston: Elsevier, AP, Academic Press is an imprint of Elsevier, 2014.

KATO, Takeyoshi. 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>.

KHOSRAVI, A.; MACHADO, L.; NUNES, R.O.. Time-series prediction of wind speed using machine learning algorithms: a case study Osorio wind farm, brazil. **Applied Energy**, [S.L.], v. 224, p. 550-566, ago. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.apenergy.2018.05.043>.

KIM, Sungil; KIM, Heeyoung. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal Of Forecasting**, [S.L.], v. 32, n. 3, p. 669-679, jul. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.ijforecast.2015.12.003>.

KUMAR, Vijay; PAL, Yash; TRIPATHI, M. M. PSO-Tuned ANN-Based Prediction Technique for Penetration of Wind Power in Grid. In: SINGH, Pradeep Kumar; KAR, Arpan Kumar; SINGH, Yashwant; KOLEKAR, Maheshkumar H.; TANWAR, Sudeep (orgs.). **Proceedings of ICRIC 2019**. Cham: Springer International Publishing, 2020. v. 597, p. 485–494. DOI 10.1007/978-3-030-29407-6_35. Disponível em: http://link.springer.com/10.1007/978-3-030-29407-6_35. Acesso em: 12 mar. 2022.

LEDERER, Johannes. Activation Functions in Artificial Neural Networks: A Systematic Overview. **arXiv:2101.09957** [cs, stat], 25 jan. 2021. Disponível em: <http://arxiv.org/abs/2101.09957>. Acesso em: 22 mai. 2022.

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

LETCHER, Trevor M. (Org.). **Wind energy engineering: a handbook for onshore and offshore wind turbines**. London ; San Diego: Academic Press, an imprint of Elsevier, 2017.

LI, Gong; SHI, Jing. On comparing three artificial neural networks for wind speed forecasting. **Applied Energy**, [S.L.], v. 87, n. 7, p. 2313-2320, jul. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.apenergy.2009.12.013>.

LI, Pin; ZHANG, Jin-Suo. A New Hybrid Method for China's Energy Supply Security Forecasting Based on ARIMA and XGBoost. **Energies**, v. 11, n. 7, p. 1687, 27 jun. 2018. DOI 10.3390/en11071687. Disponível em: <http://www.mdpi.com/1996-1073/11/7/1687>. Acesso em: 28 abr. 2022.

LIU, Hui; MI, XI-Wei; LI, Yan-Fei. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. **Energy Conversion And Management**, [S.L.], v. 156, p. 498-514, jan. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.enconman.2017.11.053>.

LIU, Xiangjie; ZHANG, Hao; KONG, Xiaobing; LEE, Kwang Y.. Wind speed forecasting using deep neural network with feature selection. **Neurocomputing**, [S.L.], v. 397, p. 393-403, jul. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.neucom.2019.08.108>.

LIU, Ling; WANG, Jujie. Super multi-step wind speed forecasting system with training set extension and horizontal-vertical integration neural network. **Applied Energy**, [s. l.], v. 292, jun. 2021. <https://doi.org/10.1016/j.apenergy.2021.116908>.

LONDRES, Guilherme Lopes. **Desenvolvimento de Redes Neurais Artificiais para Classificação de Ensaios Não Destrutivos de Ultrassom em PVDF**. 2018. 100 f. Dissertação (Mestrado) - Curso de Programa de Engenharia Metalúrgica e de Materiais,

UFRJ/Coppe, Rio de Janeiro, 2018. Disponível em:
<https://pantheon.ufrj.br/handle/11422/12897>. Acesso em: 20 abr. 2022.

LYNN, Paul A. **Onshore and offshore wind energy: an introduction**. Chichester: Wiley, 2012.

MA, Yanhong; FENG, Shuanglei. 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>.

MABEL, M. Carolin; FERNANDEZ, E.. Analysis of wind power generation and prediction using ANN: a case study. **Renewable Energy**, [S.L.], v. 33, n. 5, p. 986-992, maio 2008. Elsevier BV. <http://dx.doi.org/10.1016/j.renene.2007.06.013>.

MANERO, Jaume; BÉJAR, Javier; CORTÉS, Ulises. “Dust in the Wind...”, Deep Learning Application to Wind Energy Time Series Forecasting. **Energies**, [S.L.], v. 12, n. 12, 21 jun. 2019. MDPI AG. <http://dx.doi.org/10.3390/en12122385>.

MARUGÁN, Alberto Pliego; MÁRQUEZ, Fausto Pedro García; PEREZ, Jesus María Pinar; RUIZ-HERNÁNDEZ, Diego. A survey of artificial neural network in wind energy systems. **Applied Energy**, [S.L.], v. 228, p. 1822-1836, out. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.apenergy.2018.07.084>.

MASRUR, Hasan; NIMOL, Meas; FAISAL, Mohammad; MOSTAFA, Sk. Md. Golam. Short term wind speed forecasting using Artificial Neural Network: a case study. **2016 International Conference On Innovations In Science, Engineering And Technology (ICISSET)**, [S.L.], out. 2016. IEEE. <http://dx.doi.org/10.1109/iciset.2016.7856485>.

MAYRINK, Victor Teixeira de Melo. **Avaliação do algoritmo Gradient Boosting em aplicações de previsão de carga elétrica a curto prazo**. 2016. 91 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora, Juiz de Fora, 2016. Disponível em: https://www2.ufjf.br/pgmc/wp-content/uploads/sites/236/2009/08/Dissertacao_Victor.pdf. Acesso em: 7 abr. 2022.

MEDEIROS, Amira Rose Costa; ARAÚJO, Yana Balduino de; VIANNA, Rodrigo Pinheiro de Toledo; MORAES, Ronei Marcos de. Decision support model applied to the recognition of non-adherent individuals to antihypertensive therapy. **Saúde em Debate**, v. 38, n. 100, 2014. DOI 10.5935/0103-104.20140016. Disponível em:
<https://www.scielo.br/j/sdeb/a/rPQMWDgncmxjZ9spRygfPMj/abstract/?lang=en>. Acesso em: 5 mai. 2022.

MENG, Xiangping; PIAN, Zhaoyu. **Intelligent coordinated control of complex uncertain systems for power distribution network reliability**. Amsterdam, Netherlands ; Waltham, MA: Elsevier, 2016.

MOSAVI, Amir; SALIMI, Mohsen; FAIZOLLAHZADEH ARDABILI, Sina; RABCZUK, Timon; SHAMSHIRBAND, Shahaboddin; VARKONYI-KOCZY, Annamaria. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. **Energies**, v. 12, n. 7, p. 1301, 4 abr. 2019. DOI 10.3390/en12071301. Disponível em:
<https://www.mdpi.com/1996-1073/12/7/1301>. Acesso em: 27 abr. 2022.

MOVLIDAR-UFSC. **Base de Observação do Oceano e Atmosfera**. 2022. Disponível em: <https://movlidar.paginas.ufsc.br/estruturas/booa>. Acesso em: 20 fev. 2022.

MULTAZAM, Teuku; PUTRI, Ratna Ika; PUJANTARA, Margo; LYSTIANINGRUM, Vita; PRIYADI, Ardyono; HERYP, Mauridhi. Short-Term Wind Speed Prediction Base on Backpropagation Levenberg-Marquardt Algorithm; Case Study Area Nganjuk. **2017 5Th International Conference On Instrumentation, Communications, Information Technology, And Biomedical Engineering (Icici-Bme)**, [S.L.], p. 163-166, nov. 2017. IEEE. <http://dx.doi.org/10.1109/icici-bme.2017.8537769>.

NANA, Huai; LEI, Dong; LIJIE, Wang; YING, Hao; ZHONGJIAN, Dai; BO, Wang. Short-term Wind Speed Prediction Based on CNN_GRU Model. **2019 Chinese Control And Decision Conference (Ccdc)**, [S.L.], p. 2243-2247, jun. 2019. IEEE. <http://dx.doi.org/10.1109/ccdc.2019.8833472>.

NASSIF, Felipe de Barros. **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, Universidade Federal de Santa Catarina, Florianópolis, 2017. Disponível em: <https://repositorio.ufsc.br/handle/123456789/186188>. Acesso em: 12 abr. 2022.

NAVAS, R Kaja Bantha; 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**, [S.L.], v. 542, mar. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.physa.2019.123383>.

NEILL, Simon P.; HASHEMI, M. Reza. **Fundamentals of ocean renewable energy: generating electricity from the sea**. London [England] ; San Diego, CA: Academic Press, 2018.

NOOROLLAHI, Younes; JOKAR, Mohammad Ali; KALHOR, Ahmad. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. **Energy Conversion And Management**, [S.L.], v. 115, p. 17-25, maio 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.enconman.2016.02.041>.

ONS - Operador Nacional do Sistema Elétrico. **O Sistema Interligado Nacional**. 2022 a. Disponível em: <http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>. Acesso em: 28 abr. 2022.

ONS - Operador Nacional do Sistema Elétrico. **Balanco de Energia**. 2022b. Disponível em: <http://www.ons.org.br/paginas/energia-agora/balanco-de-energia>. Acesso em: 28 abr. 2022.

OSMAN, Ahmedbahaaldin Ibrahim Ahmed; AHMED, Ali Najah; CHOW, Ming Fai; HUANG, Yuk Feng; EL-SHAFIE, Ahmed. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. **Ain Shams Engineering Journal**, [S.L.], v. 12, n. 2, p. 1545-1556, jun. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.asej.2020.11.011>.

OZSAHIN, Ilker; OZSAHIN, Dilber Uzun. Neural network applications in medicine. **Biomedical Signal Processing And Artificial Intelligence In Healthcare**, [S.L.], p. 183-206, 2020. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-818946-7.00007-x>.

PAJANKAR, Ashwin; JOSHI, Aditya. **Hands-on Machine Learning with Python: Implement Neural Network Solutions with Scikit-learn and PyTorch**. Berkeley, CA: Apress, 2022.

PANDEY, Daya Shankar; DAS, Saptarshi; PAN, Indranil; LEAHY, James J.; KWAPINSKI, Witold. Artificial neural network based modelling approach for municipal solid waste gasification in a fluidized bed reactor. **Waste Management**, [S.L.], v. 58, p. 202-213, dez. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.wasman.2016.08.023>.

PELLETIER, Francis; MASSON, Christian; TAHAN, Antoine. Wind turbine power curve modelling using artificial neural network. **Renewable Energy**, v. 89, p. 207–214, abr. 2016. <https://doi.org/10.1016/j.renene.2015.11.065>.

PHAN, Quoc-Thang; WU, Yuan-Kang; PHAN, Quoc-Dung. A Comparative Analysis of XGBoost and Temporal Convolutional Network Models for Wind Power Forecasting. In: **2020 INTERNATIONAL SYMPOSIUM ON COMPUTER, CONSUMER AND CONTROL (IS3C)**, nov. 2020. 2020 International Symposium on Computer, Consumer and Control (IS3C) [...]. Taichung City, Taiwan: IEEE, nov. 2020. p. 416–419. DOI 10.1109/IS3C50286.2020.00113. Disponível em: <https://ieeexplore.ieee.org/document/9394194/>. Acesso em: 4 mai. 2022.

PHILIPPOPOULOS, Kostas; DELIGIORGI, Despina. Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography. **Renewable Energy**, v. 38, n. 1, p. 75–82, fev. 2012. <https://doi.org/10.1016/j.renene.2011.07.007>.

PICOLO, Ana Paula; BÜHLER, Alexandre J.; RAMPINELLI, Giuliano Arns. Uma abordagem sobre a energia eólica como alternativa de ensino de tópicos de física clássica. **Revista Brasileira de Ensino de Física**, [S.L.], v. 36, n. 4, p. 01-13, dez. 2014. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1806-11172014000400007>.

PIRES, César Henrique Mattos; PIMENTA, Felipe M.; D'AQUINO, Carla A.; SAAVEDRA, Osvaldo R.; MAO, Xuerui; ASSIREU, Arcilan T.. Coastal Wind Power in Southern Santa Catarina, Brazil. **Energies**, [S.L.], v. 13, n. 19, p. 5197, 5 out. 2020. MDPI AG. <http://dx.doi.org/10.3390/en13195197>.

PIRES, César Henrique Mattos. **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. Disponível em: <https://repositorio.ufsc.br/handle/123456789/215111>. Acesso em: 5 mai. 2022

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

PUROHIT, Shantanu; NG, E.y.K.; KABIR, Ijaz Fazil Syed Ahmed. Evaluation of three potential machine learning algorithms for predicting the velocity and turbulence intensity of a wind turbine wake. **Renewable Energy**, [S.L.], v. 184, p. 405-420, jan. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.renene.2021.11.097>.

RANGANAYAKI, V.; DEEPA, S. N.. An Intelligent Ensemble Neural Network Model for Wind Speed Prediction in Renewable Energy Systems. **The Scientific World Journal**, [S.L.], v. 2016, p. 1-14, 2016. Hindawi Limited. <http://dx.doi.org/10.1155/2016/9293529>.

REN, Xudie; GUO, Haonan; LI, Shenghong; WANG, Shilin; LI, Jianhua. A Novel Image Classification Method with CNN-XGBoost Model. In: KRAETZER, Christian; SHI, Yun-Qing; DITTMANN, Jana; KIM, Hyoung Joong (orgs.). **Digital Forensics and Watermarking**. Cham: Springer International Publishing, 2017. v. 10431, p. 378–390. DOI 10.1007/978-3-319-64185-0_28. Disponível em: http://link.springer.com/10.1007/978-3-319-64185-0_28. Acesso em: 4 mai. 2022.

REN, He; CHEN, Xi; CHEN, Yong. **Reliability based aircraft maintenance optimization and applications**. London: Academic Press, 2017 (Shanghai Jiao Tong University Press aerospace series).

RIOS, Jorge D.; ALANIS, Alma Y.; ARANA-DANIEL, Nancy; LOPEZ-FRANCO, Carlos; SANCHEZ, Edgar N. **Neural networks modeling and control: applications for unknown nonlinear delayed systems in discrete time**. 1. ed. Waltham: Elsevier, 2019.

RODRÍGUEZ, Fermín; FLOREZ-TAPIA, Ane M.; FONTÁN, Luis; GALARZA, Ainhoa. Very short-term wind power density forecasting through artificial neural networks for microgrid control. **Renewable Energy**, [S.L.], v. 145, p. 1517-1527, jan. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.renene.2019.07.067>.

RUSSELL, Stuart J.; NORVIG, Peter; DAVIS, Ernest; EDWARDS, Douglas. **Artificial intelligence: a modern approach**. Third edition, Global edition: Pearson, 2016.

SAIRAMYA, N.J.; SUSMITHA, L.; GEORGE, S. Thomas; SUBATHRA, M.s.P.. Hybrid Approach for Classification of Electroencephalographic Signals Using Time–Frequency Images With Wavelets and Texture Features. **Intelligent Data Analysis For Biomedical Applications**, [S.L.], p. 253-273, 2019. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-815553-0.00013-6>.

SAMET, Haidar; REISI, Mohammad; MARZBANI, Fatemeh. Evaluation of neural network-based methodologies for wind speed forecasting. **Computers & Electrical Engineering**, [S.L.], v. 78, p. 356-372, set. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.compeleceng.2019.07.024>.

SANTA CATARINA. Defesa Civil. **Monitoramento 24 horas - 11/08/2020**. Disponível em: <https://www.defesacivil.sc.gov.br/2020/08/11/?arcf=cat:245>. Acesso em: 01 jun. 2022.

SCIKIT-LEARN. **Parameter estimation using grid search with cross-validation**. 2022a. Disponível em: <https://scikit->

learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html. Acesso em: 05 abr. 2022.

SCIKIT-LEARN. **Cross-validation: evaluating estimator performance**. 2022b. Disponível em: https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation. Acesso em: 05 abr. 2022.

SEWDIEN, V.N.; PREECE, R.; TORRES, J.L. Rueda; RAKHSHANI, E.; MEIJDEN, M. van Der. Assessment of critical parameters for artificial neural networks based short-term wind generation forecasting. **Renewable Energy**, [S.L.], v. 161, p. 878-892, dez. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.renene.2020.07.117>.

SIEFERT, Cesar Augusto Crovador; DOMBROWSKI NETTO, Nicolas; MARANGON, Fernando Helmuth Syring; SCHULTZ, Gilson Bauer; SILVA, Luan Marcus dos Reis; FONTENELLE, Thiago Henriques; SANTOS, Irani dos. Avaliação de Séries de Velocidade do Vento de Produtos de Reanálises Climáticas para o Brasil. **Revista Brasileira de Meteorologia**, v. 36, n. 4, p. 689–701, dez. 2021. DOI 10.1590/0102-7786360026. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862021000500689&tlng=pt. Acesso em: 14 mai. 2022.

SILVA, Pollyanna Kelly de Oliveira. **Análise e previsão de curto prazo do vento através de modelagem estatística em áreas de potencial eólico no Nordeste do Brasil**. 2017. 166 f. Tese (Doutorado) - Curso de Meteorologia, Universidade Federal de Campina Grande, Campina Grande, 2017. Disponível em: http://www.dca.ufcg.edu.br/posgrad_met/teses/PollyannaKellydeOliveiraSilva_2017.pdf. Acesso em: 10 mar. 2022.

SILVA, Ramon Gomes da; RIBEIRO, Matheus Henrique dal Molin; MORENO, Sinvaldo Rodrigues; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting. **Energy**, [S.L.], v. 216, p. 119174, fev. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.energy.2020.119174>.

SILVEIRA, Edvana da Silva da; CAPO, Políbio Ícaro Moro; D'AQUINO, Carla de Abreu. Previsão de Curtíssimo prazo da velocidade do vento costeiro utilizando Redes Neurais Artificiais. In: X SIMPÓSIO DE INTEGRAÇÃO CIENTÍFICA E TECNOLÓGICA DO SUL CATARINENSE, SITC SUL: A TRANSVERSALIDADE DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O PLANETA, 10., 2021, Tubarão. **Anais [...]**. Tubarão: IFSC, 2021. p. 694-701. Disponível em: <http://criciuma.ifsc.edu.br/anais-sict-sul/>. Acesso em: 7 mar. 2022.

SILVERA, Oscar Churio; CHAMORRO, Marley Vanegas; OCHOA, Guillermo Valencia. Wind and solar resource assessment and prediction using Artificial Neural Network and semi-empirical model: case study of the colombian caribbean region. **Heliyon**, [S.L.], v. 7, n. 9, p. e07959, set. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.heliyon.2021.e07959>.

SINGH, Upma; RIZWAN, Mohammad; ALARAJ, Muhannad; ALSAIDAN, Ibrahim. A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments. **Energies**, v. 14, n. 16, p. 5196, 23 ago. 2021. DOI 10.3390/en14165196. Disponível em: <https://www.mdpi.com/1996-1073/14/16/5196>. Acesso em: 27 abr. 2022.

SUN, Wei; WANG, Yuwei. 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**, [S.L.], v. 157, p. 1-12, fev. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.enconman.2017.11.067>.

TAHMASSEBI, Amirhessam; WENGERT, Georg J.; HELBICH, Thomas H.; BAGO-HORVATH, Zsuzsanna; ALAEI, Sousan; BARTSCH, Rupert; DUBSKY, Peter; BALTZER, Pascal; CLAUSER, Paola; KAPETAS, Panagiotis; MORRIS, Elizabeth A.; MEYER-BAESE, Anke; PINKER, Katja. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients: **Investigative Radiology**, v. 54, n. 2, p. 110–117, fev. 2019. DOI 10.1097/RLI.0000000000000518. Disponível em: <http://journals.lww.com/00004424-201902000-00008>. Acesso em: 4 mai. 2022.

TORRES, J.L.; GARCÍA, A.; DE BLAS, M.; DE FRANCISCO, A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). **Solar Energy**, v. 79, n. 1, p. 65–77, jul. 2005. <https://doi.org/10.1016/j.solener.2004.09.013>.

TRENTINI, Caroline. **Viabilidade técnico-econômica de implantação de aerogeradores em Regiões Brasileiras**. 2016. 71 p. Trabalho de Conclusão de Curso (Engenharia Elétrica) - Universidade Federal de Juiz de Fora, Juiz de Fora, 2016. Disponível em: https://www.ufjf.br/eletrica_energia/files/2016/11/TCC_CAROLINE-TRENTINI.pdf. Acesso em: 21 nov. 2021.

VALERIO, Lorenzo; NARDINI, Franco Maria; PASSARELLA, Andrea; PEREGO, Raffaele. Dynamic hard pruning of Neural Networks at the edge of the internet. **Journal Of Network And Computer Applications**, [S.L.], v. 200, p. 103330, abr. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.jnca.2021.103330>.

VAFAEIPOUR, Majid; RAHBARI, Omid; ROSEN, Marc A.; FAZELPOUR, Farivar; ANSARIRAD, Pooyandeh. Application of sliding window technique for prediction of wind velocity time series. **International Journal Of Energy And Environmental Engineering**, [S.L.], v. 5, n. 2-3, 18 maio 2014. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s40095-014-0105-5>.

VEIT, Dieter (Org.). **Simulation in textile technology: theory and applications**. Oxford ; Philadelphia: Woodhead Pub, 2012(Woodhead Publishing series in textiles, no. 136).

VELO, Ramón; LÓPEZ, Paz; MASEDA, Francisco. Wind speed estimation using multilayer perceptron. **Energy Conversion And Management**, [S.L.], v. 81, p. 1-9, maio 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.enconman.2014.02.017>.

VERMA, Manoj; GHRITLAHRE, Harish Kumar; CHANDRAKAR, Ghritanchi. Wind Speed Prediction of Central Region of Chhattisgarh (India) Using Artificial Neural Network and Multiple Linear Regression Technique: a comparative study. **Annals Of Data Science**, [S.L.], 2 maio 2021. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s40745-021-00332-1>.

WANG, Jian; YANG, Zhongshan. Ultra-short-term wind speed forecasting using an optimized artificial intelligence algorithm. **Renewable Energy**, [S.L.], v. 171, p. 1418-1435, jun. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.renene.2021.03.020>.

WANG, Jianzhou; LI, Qiwei; ZENG, Bo. Multi-layer cooperative combined forecasting system for short-term wind speed forecasting. **Sustainable Energy Technologies And Assessments**, [S.L.], v. 43, p. 100946, fev. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.seta.2020.100946>.

WONG, W.K.; ZENG, X.H.; AU, K.F.. Selecting the location of apparel manufacturing plants using neural networks. **Optimizing Decision Making In The Apparel Supply Chain Using Artificial Intelligence (Ai)**, [S.L.], p. 41-54, 2013. Elsevier. <http://dx.doi.org/10.1533/9780857097842.41>.

XGBOOST. **XGBoost Parameters**. 2021. Disponível em: <https://xgboost.readthedocs.io/en/stable/parameter.html>. Acesso em: 05 abr. 2022.

XU, Lei; MAO, Jiandong. Short-term wind power forecasting based on Elman neural network with particle swarm optimization. **2016 Chinese Control And Decision Conference (CCDC)**, [S.L.], maio 2016. IEEE. <http://dx.doi.org/10.1109/ccdc.2016.7531436>

ZHA, Wenting; LIU, Jie; LI, Yalong; LIANG, Yingyu. Ultra-short-term power forecast method for the wind farm based on feature selection and temporal convolution network. **ISA Transactions**, p. S0019057822000398, jan. 2022. DOI 10.1016/j.isatra.2022.01.024. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0019057822000398>. Acesso em: 28 abr. 2022.

ZHANG, Wenyu; ZHANG, Lifang; WANG, Jianzhou; NIU, Xinsong. Hybrid system based on a multi-objective optimization and kernel approximation for multi-scale wind speed forecasting. **Applied Energy**, [S.L.], v. 277, p. 115561-55555, nov. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.apenergy.2020.115561>.

ZHENG, Huiting; YUAN, Jiabin; CHEN, Long. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. **Energies**, v. 10, n. 8, p. 1168, 8 ago. 2017. DOI 10.3390/en10081168. Disponível em: <http://www.mdpi.com/1996-1073/10/8/1168>. Acesso em: 4 mai. 2022.

ZHENG, Huan; WU, Yanghui. A XGBoost Model with Weather Similarity Analysis and Feature Engineering for Short-Term Wind Power Forecasting. **Applied Sciences**, [S.L.], v. 9, n. 15, p. 3019, 26 jul. 2019. MDPI AG. <http://dx.doi.org/10.3390/app9153019>.

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, maio 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.heliyon.2019.e01664>.

ANEXO A - Descrição matemática da função objetivo - XGBoost com base nos trabalhos de Chen (2014) e Chen e Guestrin (2016)

Existem duas partes características das funções objetivo do algoritmo XGBoost: perda de treinamento e termo de regularização, dadas pela equação:

$$obj(\varphi) = L(\varphi) + \Omega(\varphi)$$

Onde L é a função de perda do treinamento, Ω é o termo de regularização e φ significa que os parâmetros que serão descobertos durante a etapa de treinamento. A perda mede o quão preditivo é o modelo, ou seja, avalia se o modelo se ajusta aos dados de treinamento, e, a regularização controla a complexidade do modelo e ajuda a evitar o *overfitting* (CHEN, 2014; CHEN; GUESTRIN, 2016).

O XGBoost utiliza um modelo baseado em conjuntos de árvore de decisão que utiliza K funções de adição para previsão é dado por:

$$\hat{y}_i = \sum_k^K f_k(x_i), \quad f_k \in F$$

Onde \hat{y}_i representa a previsão, K é o número de árvores, f_k é uma função no espaço funcional F , cada f_k corresponde a uma estrutura de árvore independente q com peso ω em suas folhas, x_i é a entrada, F é o conjunto de árvores de regressão (CHEN, GUESTRIN, 2016). A nova função objetivo regularizada é dada por:

$$obj(\varphi) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Onde $\Omega(f_k)$ é a complexidade das árvores/termo de regularização da árvore f_k , sendo que este selecionará um modelo adotando funções simples, $l(y_i, \hat{y}_i)$ é a função de perda, que representa a diferença entre o valor real (y_i) e previsto (\hat{y}_i), n é o tamanho da amostra (CHEN, 2014; CHEN; GUESTRIN, 2016). A regularização $\Omega(f_k)$ é definida como:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \|\omega_j\|^2$$

Onde λ é um valor constante e pequeno, T é o número de folhas na árvore, ω são os pesos das folhas, γ é uma constante. O primeiro termo avalia o número de folhas e o segundo calcula a pontuação das folhas (CHEN, 2014).

Durante cada iteração, uma subamostra dos dados é extraída de forma aleatória do banco de dados de treino para ajustar a árvore e calcular a atualização do modelo (FRIEDMAN, 2002). Dessa maneira, o treinamento aditivo é aplicado durante a etapa de aprendizagem, assim, uma nova árvore que otimiza o sistema é adicionada de cada vez no modelo, \hat{y}_i^t é descrito como o modelo na etapa de treinamento t , \hat{y}_i^{t-1} é a função adicionada na rodada anterior e $f_t(x_i)$ é a nova função, assim, escrevendo o valor de previsão na etapa como \hat{y}_i^t , (CHEN, 2014):

$$\begin{aligned}\hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)\end{aligned}$$

Assim, a função objetivo do modelo acima pode ser definida como:

$$\begin{aligned}\text{obj}^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^K \Omega(f_t) \\ \text{obj}^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{k=1}^K \Omega(f_t)\end{aligned}$$

Considerando o erro quadrático médio $(y_i - \hat{y}_i)^2$ como o termo de perda, é feita a expansão de Taylor para aproximação de segunda ordem para otimizar a função objetivo, que demonstra como seria feita a divisão dos nós. G e H são definidas nas equações abaixo, onde g e h são as derivadas parciais de primeira e segunda ordem na função de perda, $I_j = \{i | q(x_i) = j\}$ é o grupo de índices de pontos de dados atribuídos à j-ésima folha e $q(x)$ é a estrutura da árvore (CHEN, 2014).

$$\text{Obj}^{(t)} = \sum_{j=1}^T [G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2] + \gamma T$$

$$G_j = \sum_{i \in I_j} g_i$$

$$H_j = \sum_{i \in I_j} h_i$$

Onde g_i e h_i são definidos como

$$g_i = \partial_{y_i} \hat{l}(y_i, \hat{y}_i^{(t-1)})$$

$$h_i = \partial_{y_i}^2 \hat{l}(y_i, \hat{y}_i^{(t-1)})$$

Nesta equação, w_j são independentes entre si, o valor ótimo w_j de uma folha j com uma estrutura $q(x)$, e, a função objetivo ótima correspondente são apresentadas abaixo (CHEN; GUESTRIN, 2016).

$$w_j^* = -\frac{G_j}{H_j + \lambda}$$

$$\text{obj}^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T$$

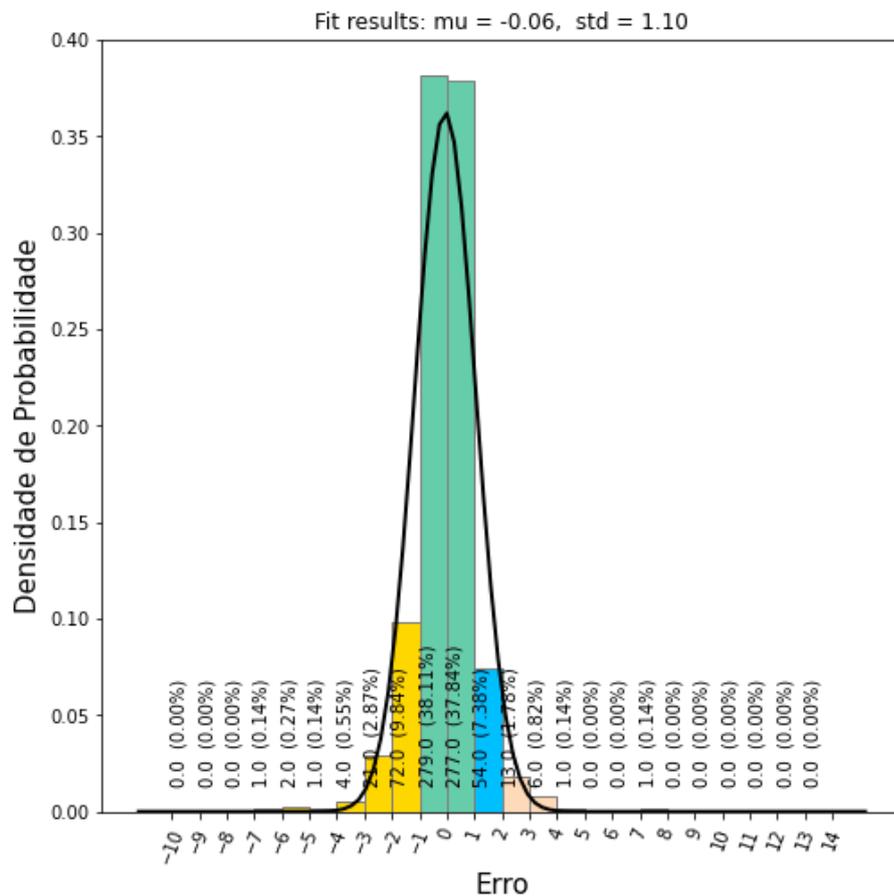
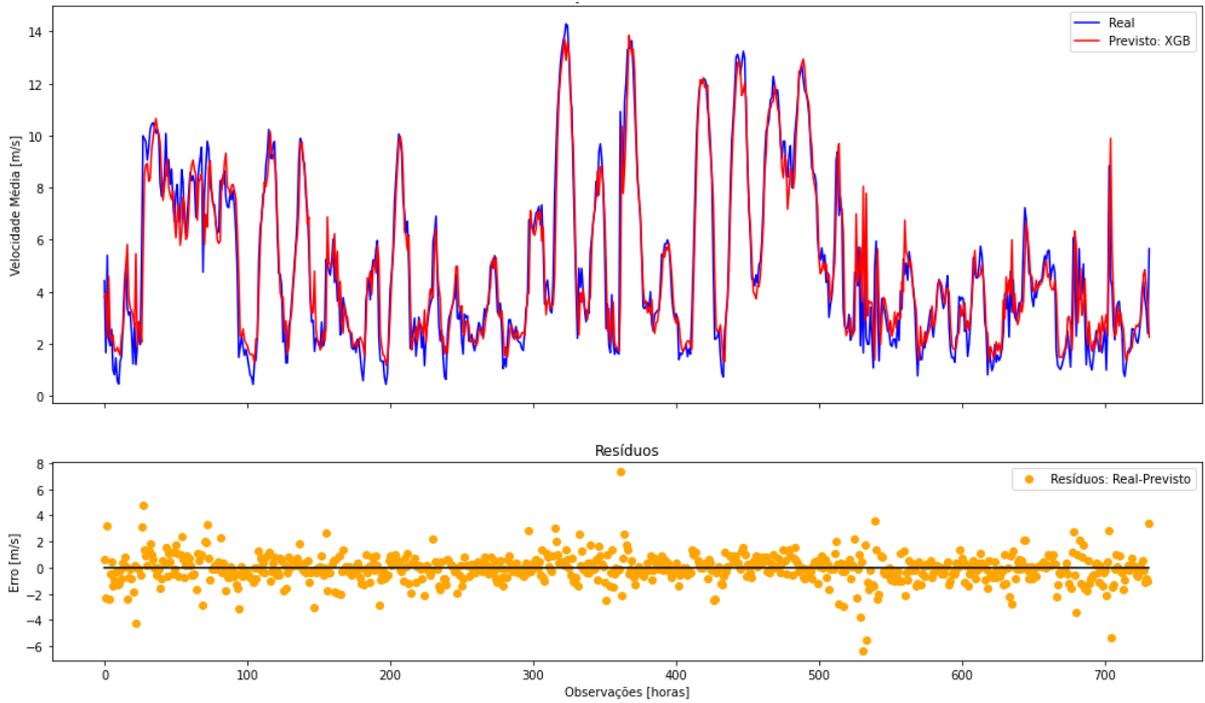
λ controla o peso w_j do nó da folha j , e, γ é um parâmetro de regularização, utilizado para controlar o número de nós na folha T (ZHA et al., 2022). λ e γ são adotados para garantir o equilíbrio. Assim, a última equação mede o desempenho da estrutura de árvore $q(x)$, assim, se a pontuação for menor, a estrutura será melhor (CHEN, 2014).

Visto que pode haver infinitas estruturas de árvores possíveis, as árvores são cultivadas com o algoritmo “guloso” (*Greedy algorithm*) (CHEN; GUESTRIN, 2016). Assim, no XGBoost, um nível da árvore é otimizado por vez, partindo de uma árvore com profundidade zero. Para cada nó folha da árvore, o algoritmo tenta adicionar uma divisão/ramificação. Assim, o ganho é expresso conforme a equação abaixo (CHEN, 2014)

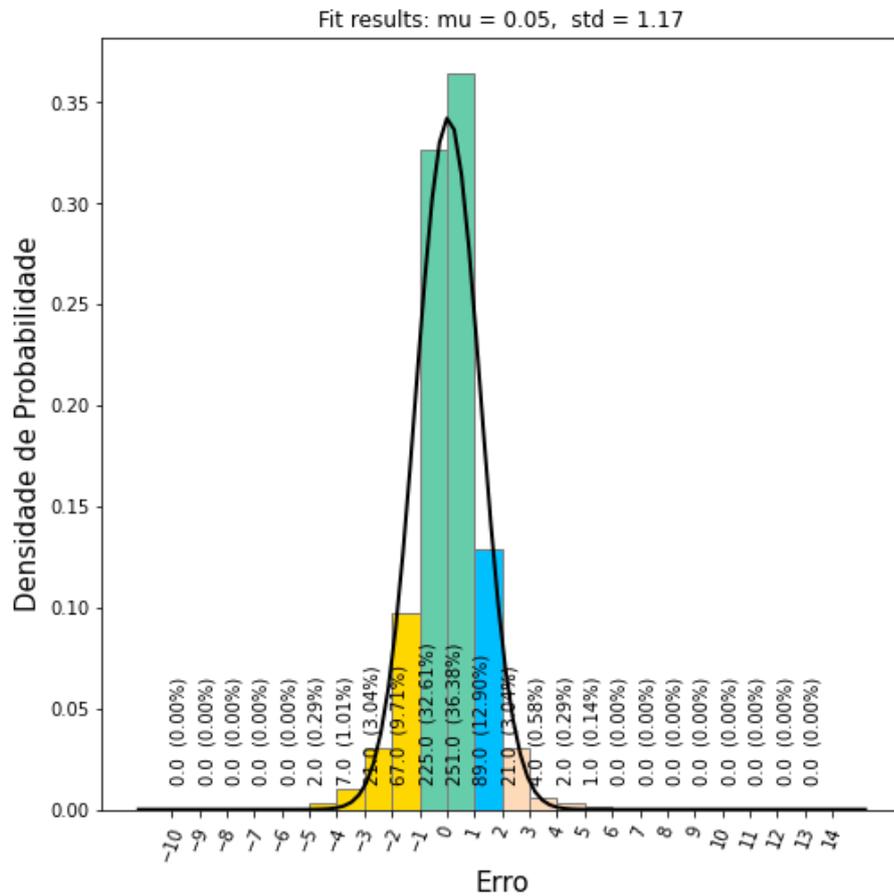
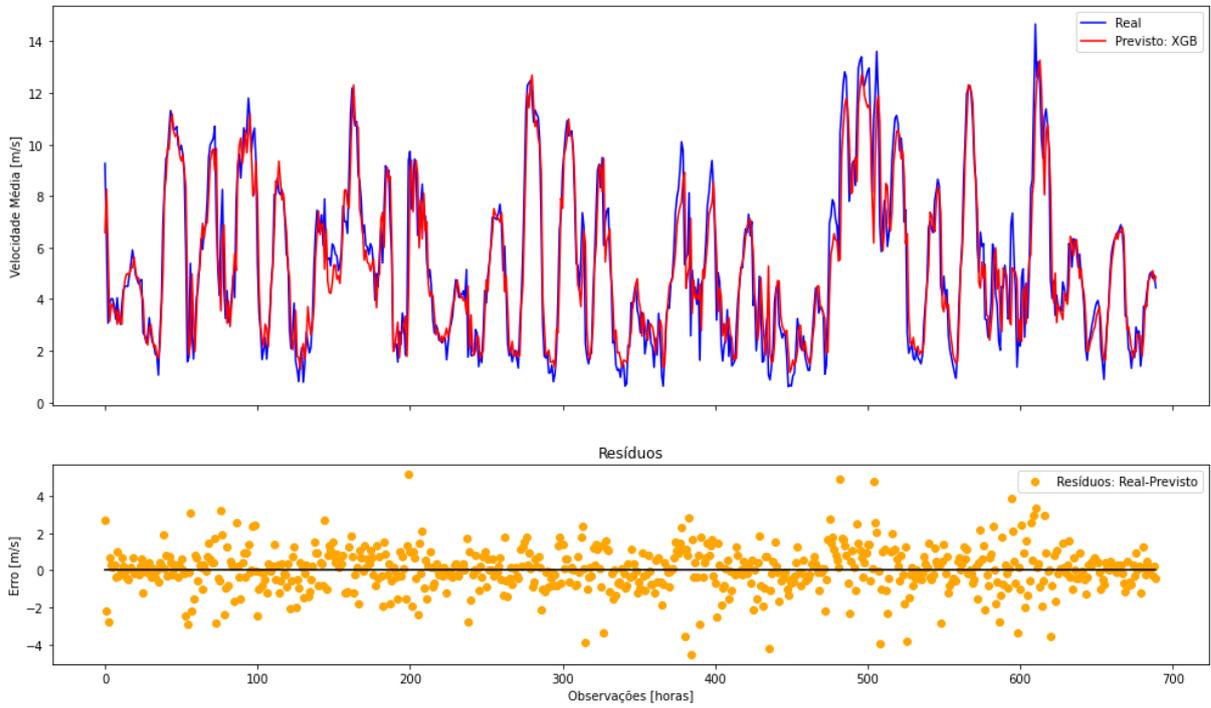
$$\text{Ganho} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$

Na equação anterior pode ser decomposta como: 1) pontuação do filho esquerdo; 2) pontuação do filho direito; 3) a pontuação sem a divisão, e o último termo considera um custo de complexidade devido a introdução de folhas adicionais (CHEN, 2014). O autor destaca um aspecto importante: é melhor não adicionar uma divisão (ramificação) se o ganho for menor que γ . Ou seja, parar o processo de divisão se a equação acima apresentar valor negativo. Na técnica denominada de poda, a árvore cresce até uma profundidade máxima, e as divisões de folhas com valores negativos são podadas (CHEN, 2014).

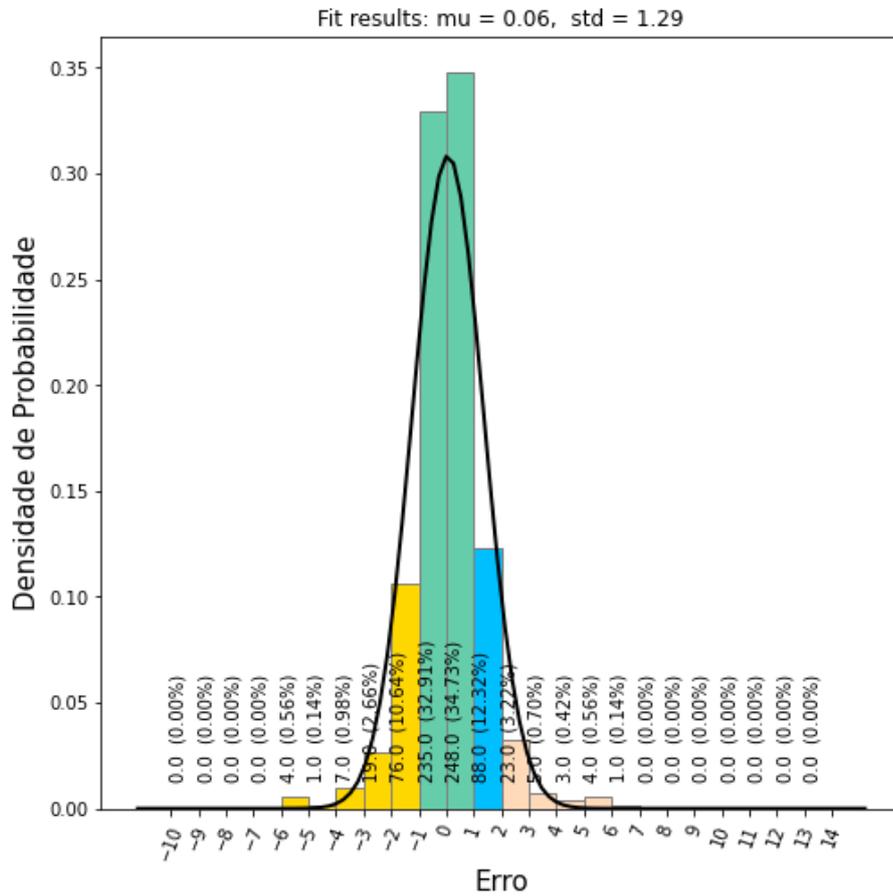
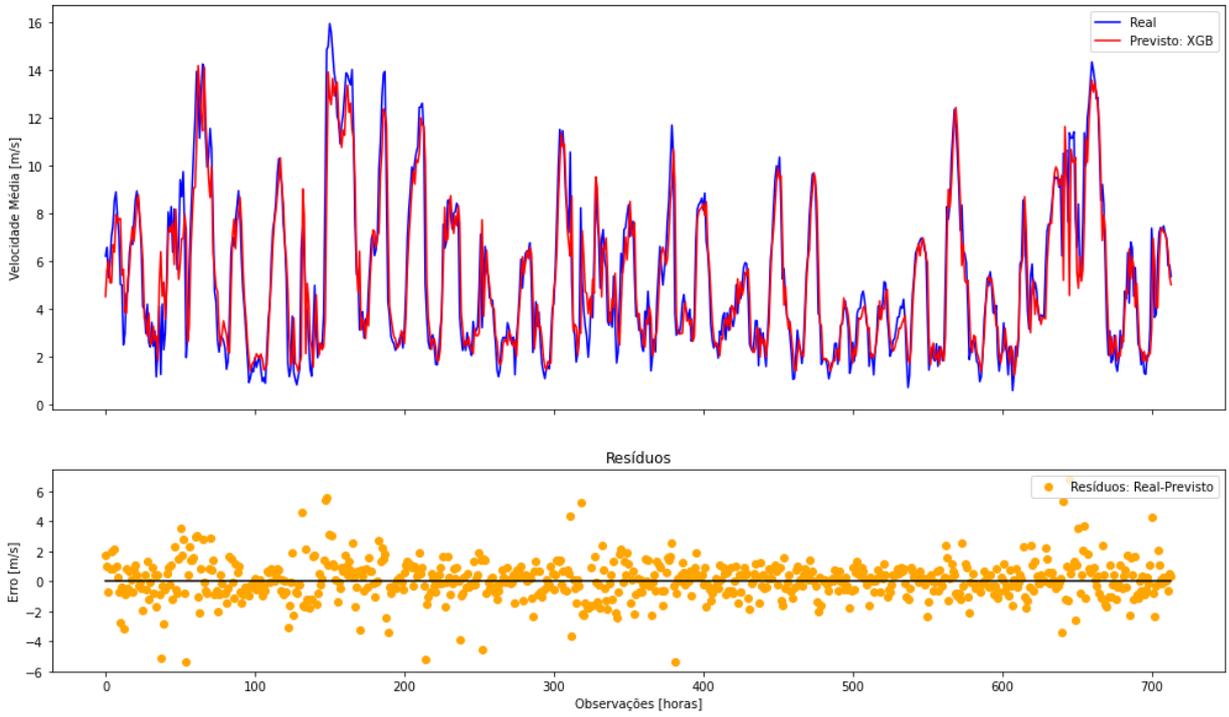
APÊNDICE A - Previsão da velocidade do vento e resíduos para Janeiro de 2020



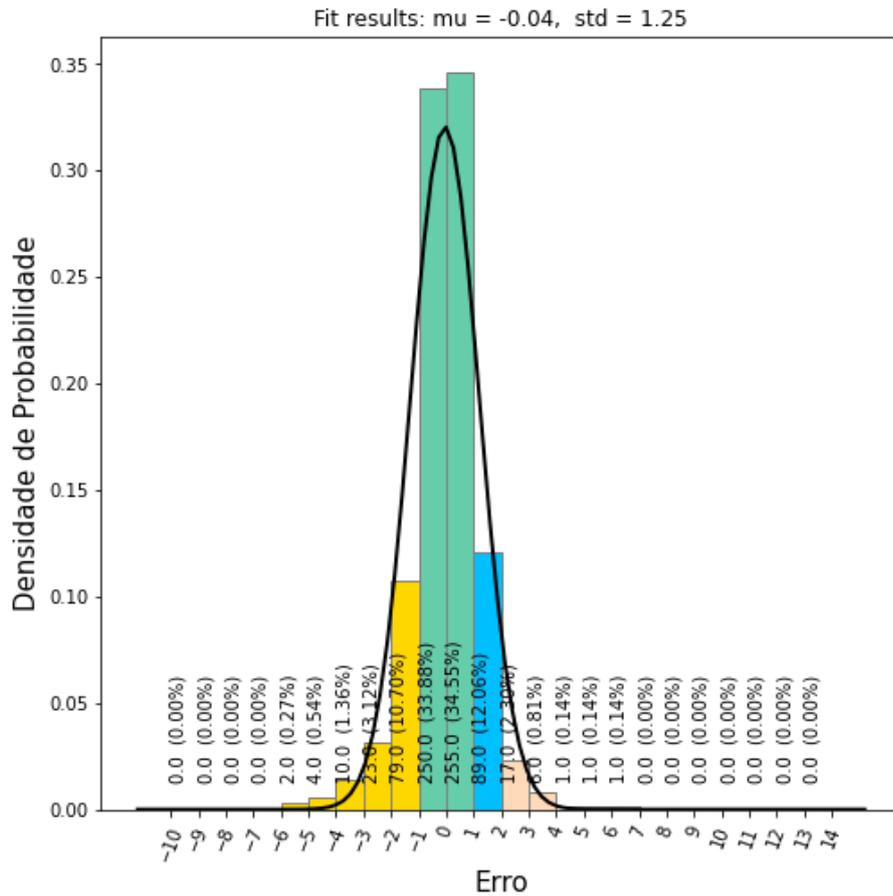
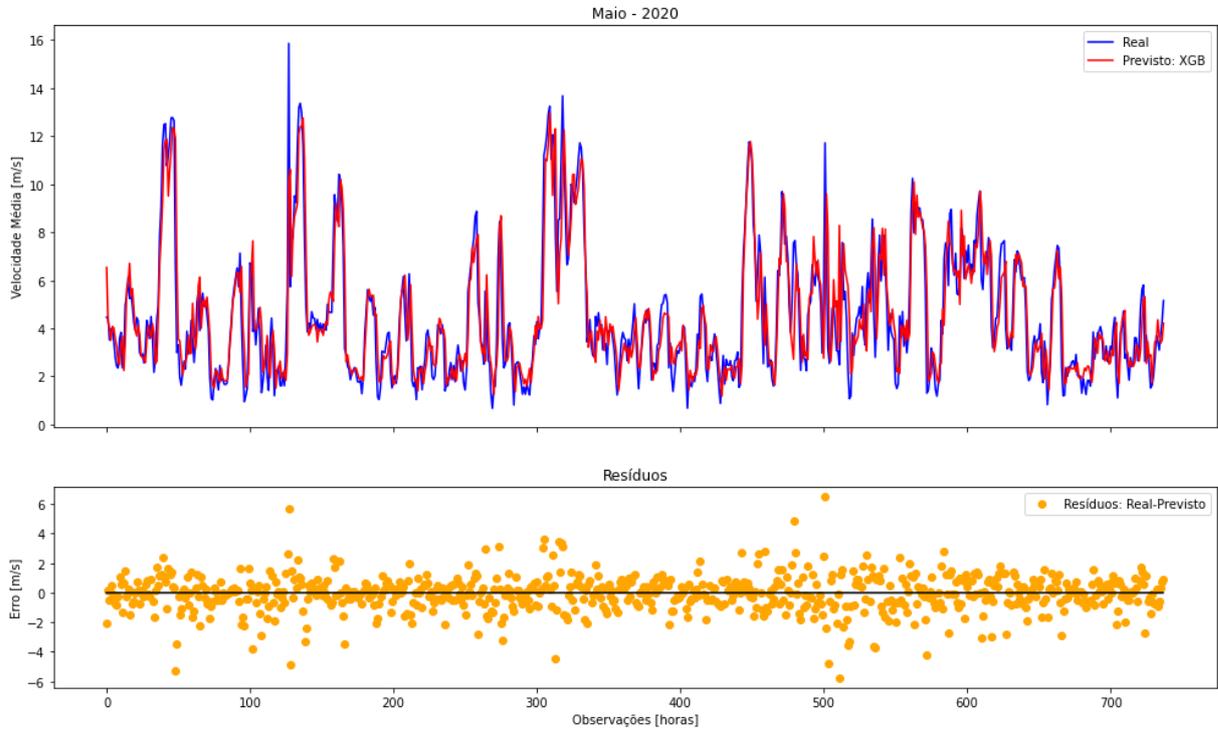
APÊNDICE B - Previsão da velocidade do vento e resíduos para Fevereiro de 2020



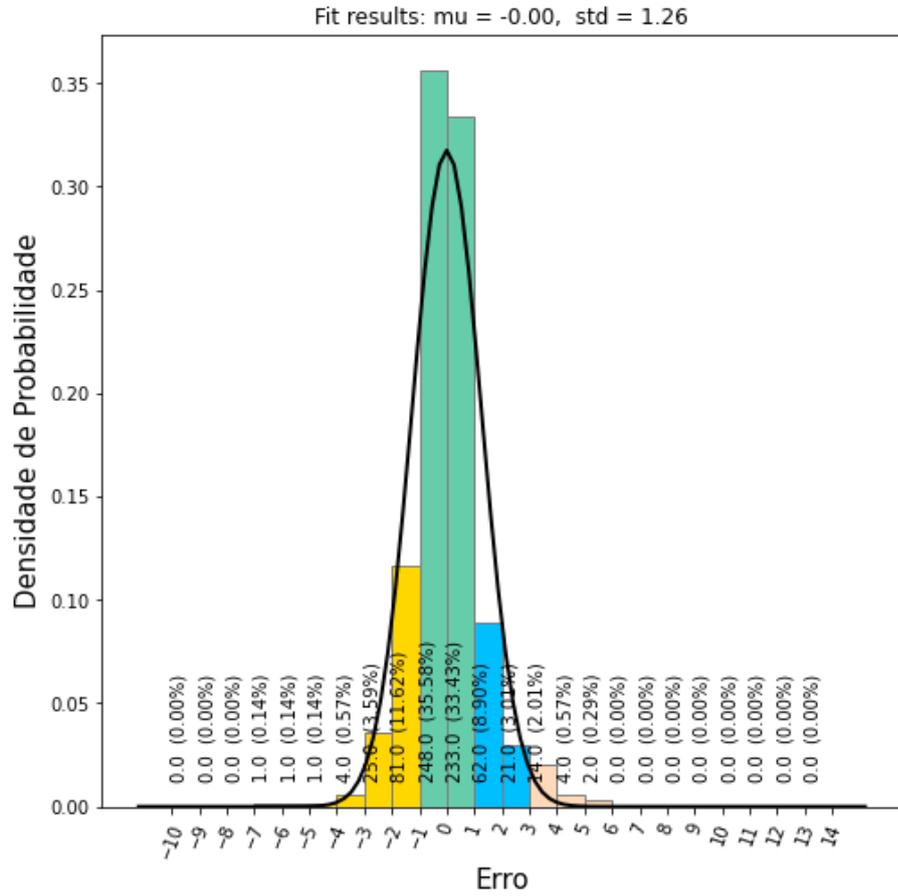
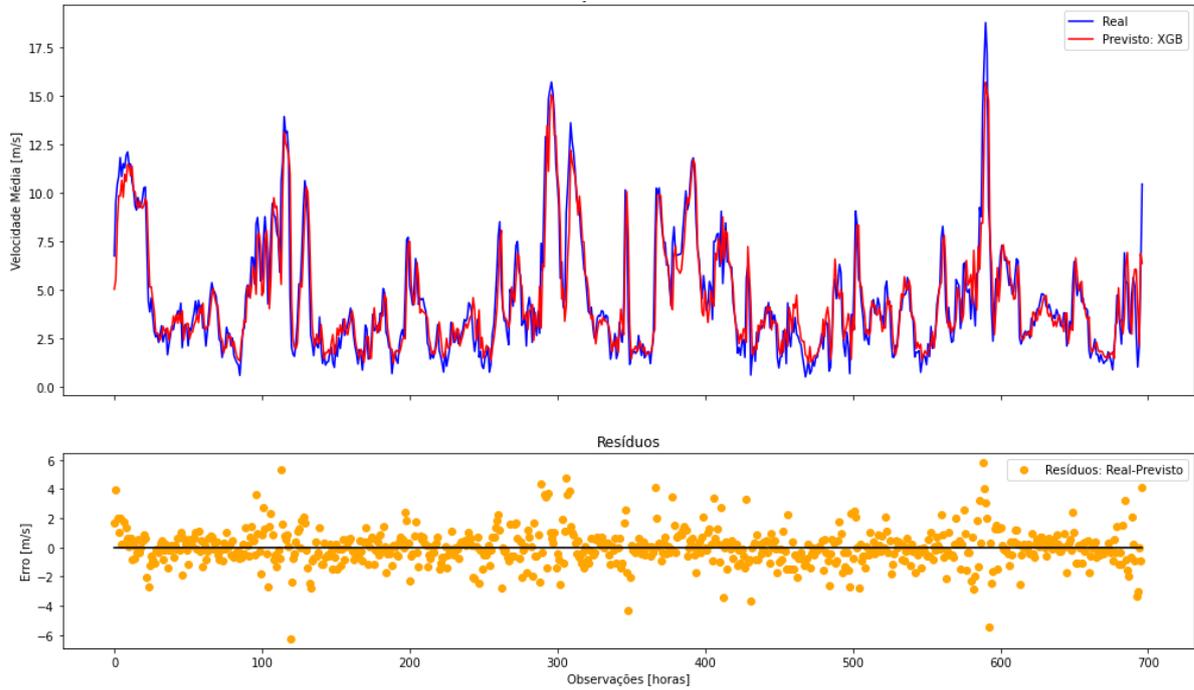
APÊNDICE C - Previsão da velocidade do vento e resíduos para Abril de 2020



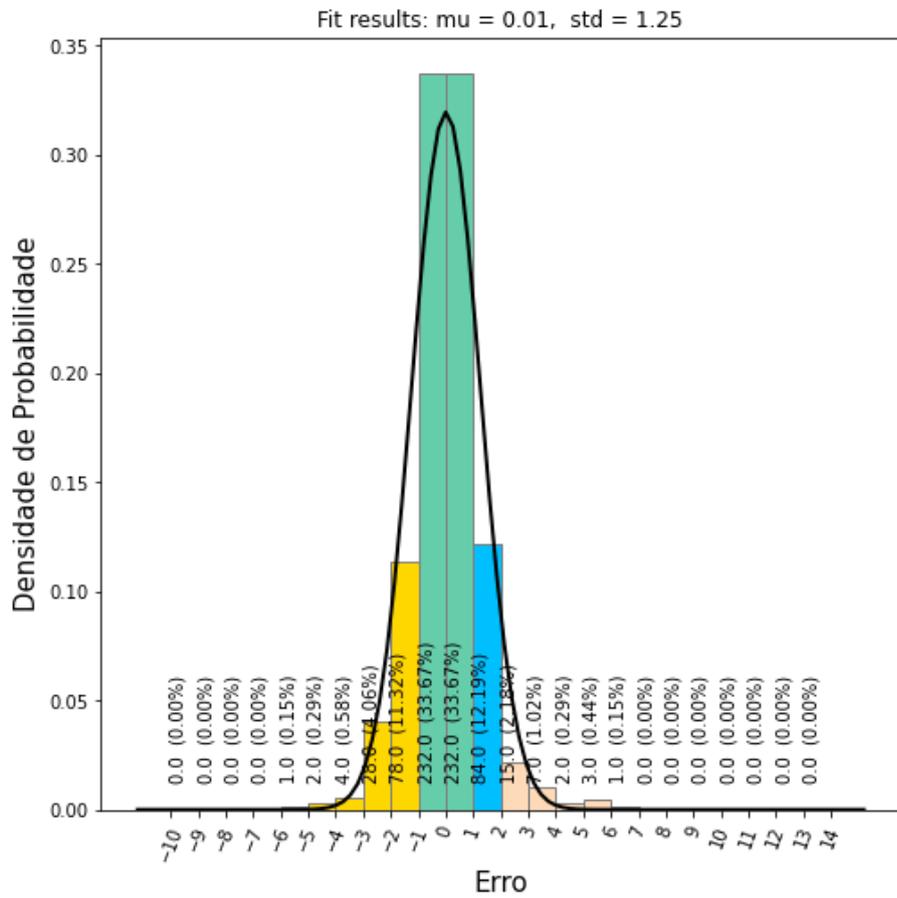
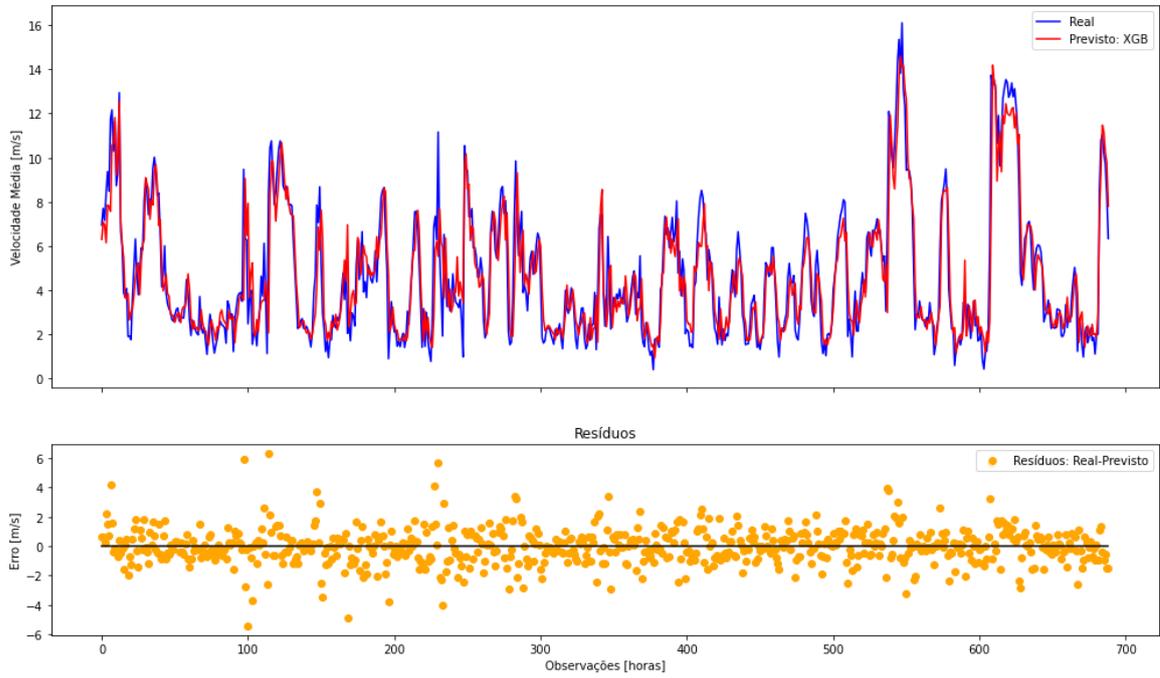
APÊNDICE D - Previsão da velocidade do vento e resíduos para Maio de 2020



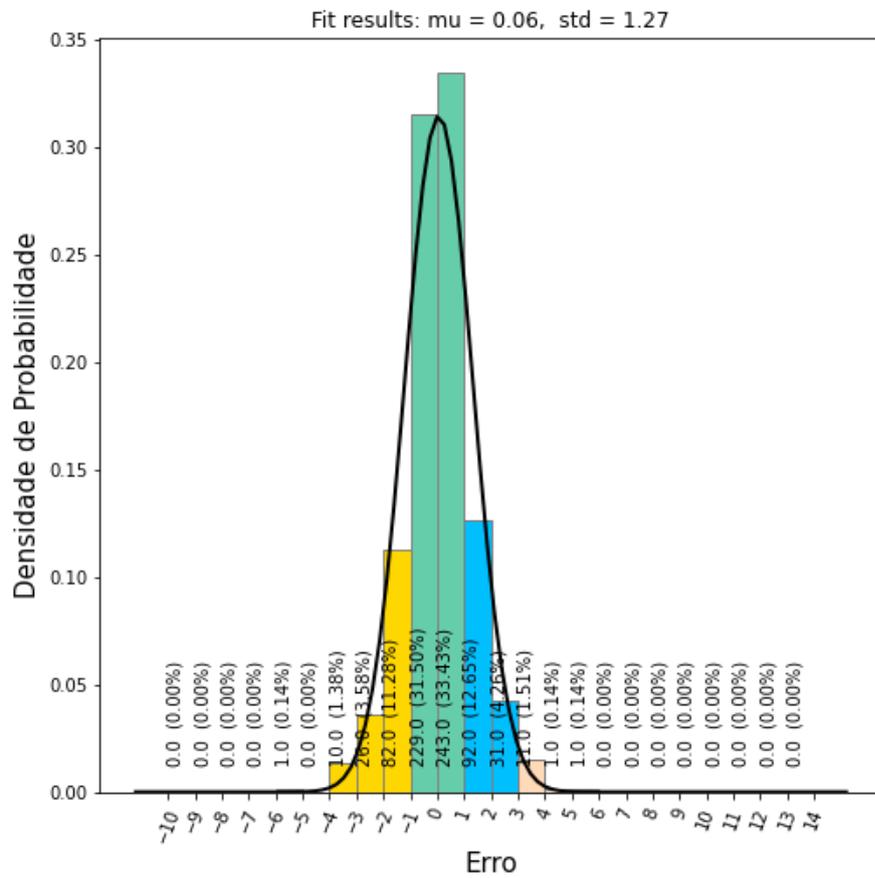
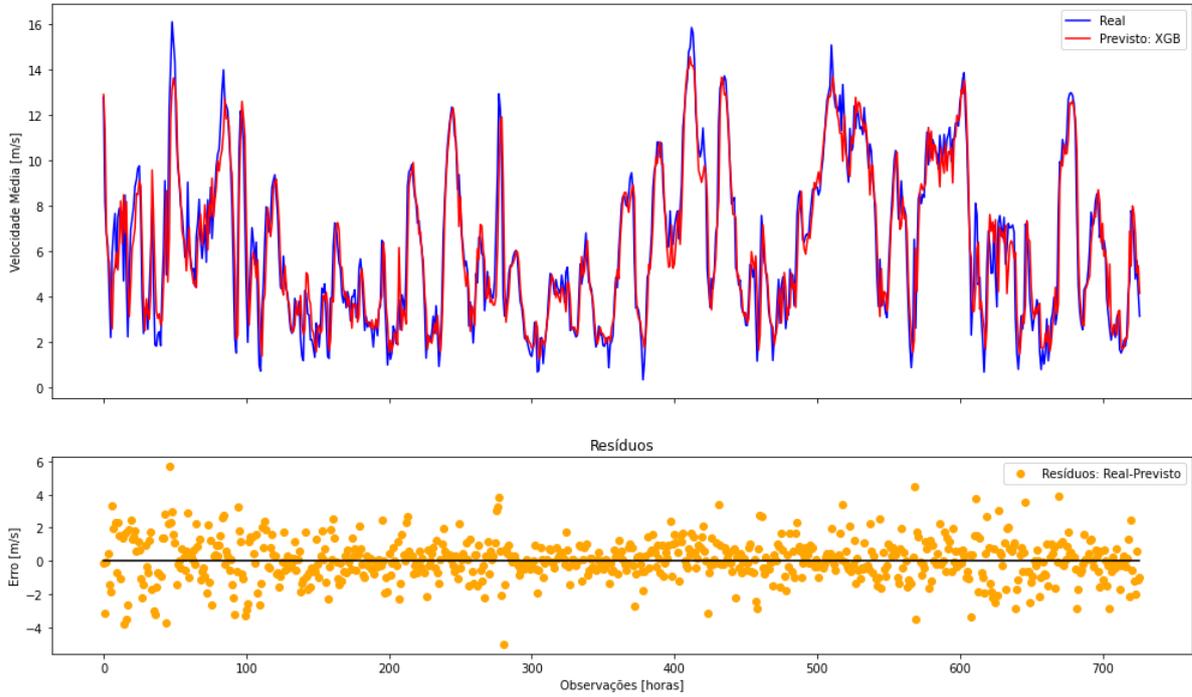
APÊNDICE E - Previsão da velocidade do vento e resíduos para Junho de 2020



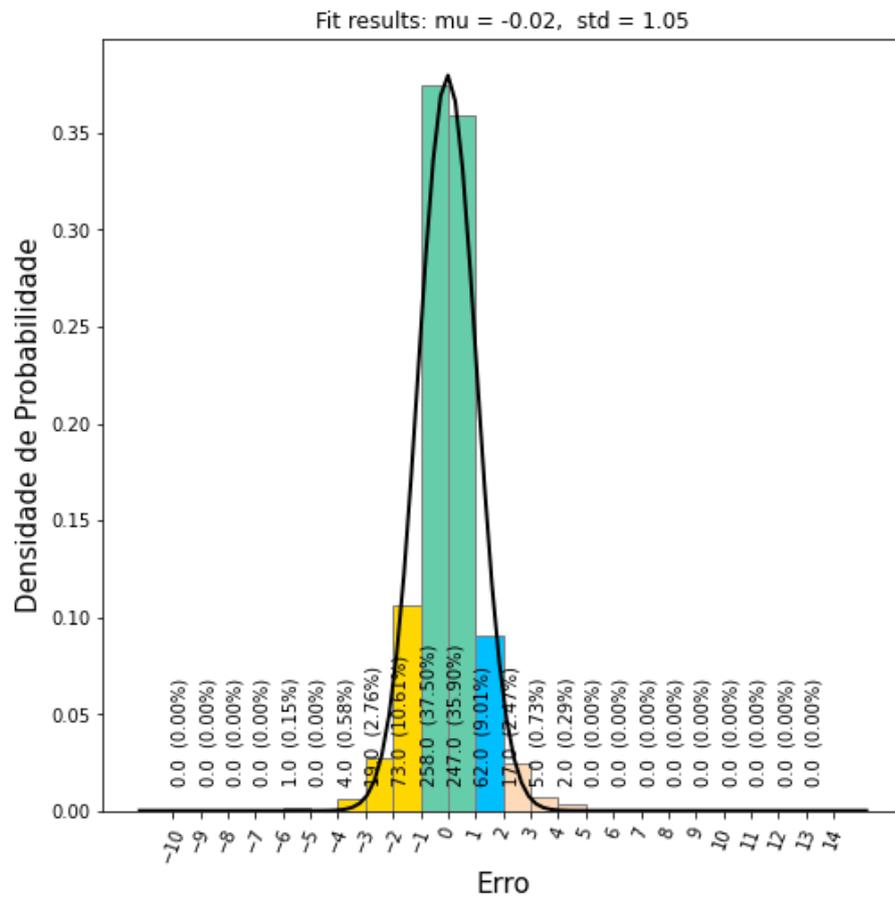
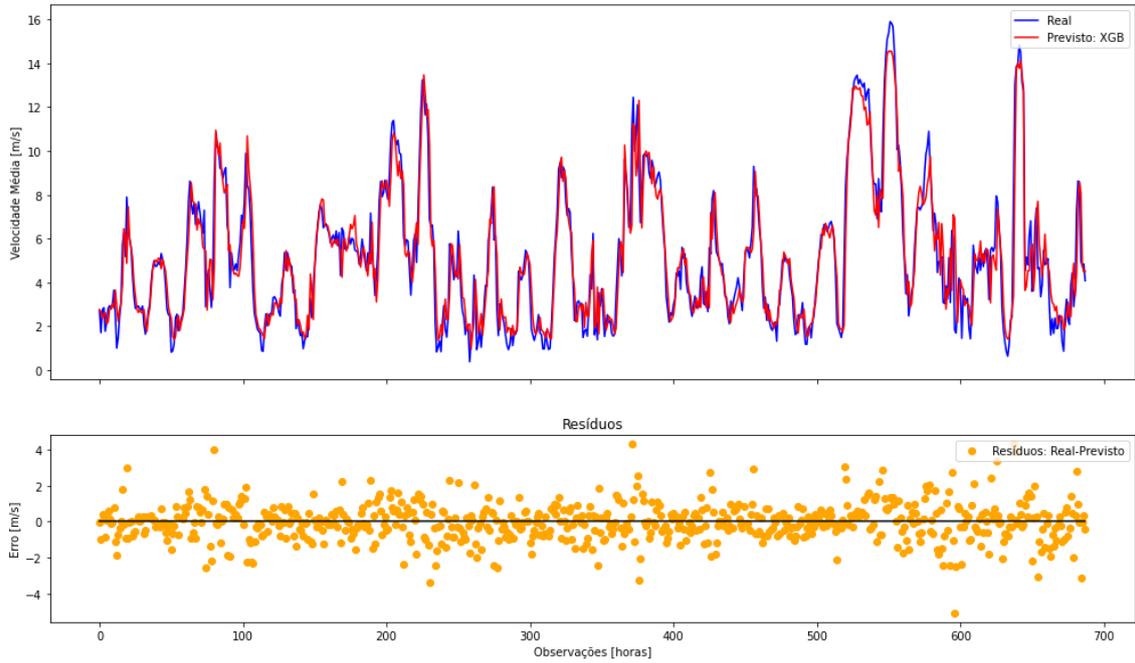
APÊNDICE F - Previsão da velocidade do vento e resíduos para Julho de 2020



APÊNDICE G - Previsão da velocidade do vento e resíduos para Outubro de 2020



APÊNDICE H - Previsão da velocidade do vento e resíduos para Novembro de 2020



APÊNDICE I - Previsão da velocidade do vento e resíduos para Dezembro de 2020

