



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA
ELÉTRICA E ELETRÔNICA**

Jean Alexander Gomes Braz

**Predição de velocidade de ventos para geração eólica, utilizando
aprendizado de máquina**

**Florianópolis
2023**

Jean Alexander Gomes Braz

Predição de velocidade de ventos para geração eólica, utilizando aprendizado de máquina

Trabalho de Conclusão de Curso apresentado
à Universidade Federal de Santa Catarina
como parte dos requisitos necessários para a
obtenção do Título de Engenheiro Eletricista.
Orientador: Prof. Dr. Telles Brunelli Lazzarin.
Co-orientador: Prof. Dr. Mateus Grellert da
Silva.

Florianópolis
2023

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

Braz, Jean Alexsander Gomes
Predição de velocidade de ventos para geração eólica,
utilizando aprendizado de máquina / Jean Alexsander Gomes
Braz ; orientador, Telles Brunelli Lazzarin,
coorientador, Mateus Grellert da Silva, 2023.
57 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Elétrica, Florianópolis, 2023.

Inclui referências.

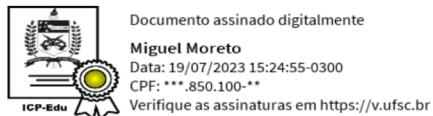
1. Engenharia Elétrica. 2. Predição de Velocidade de
vento. 3. Energia Eólica. 4. Aprendizado de máquina. 5.
Energia Renovável. I. Lazzarin, Telles Brunelli . II. da
Silva, Mateus Grellert. III. Universidade Federal de Santa
Catarina. Graduação em Engenharia Elétrica. IV. Título.

Jean Alexander Gomes Braz

Predição de velocidade de ventos para geração eólica, utilizando aprendizado de máquina

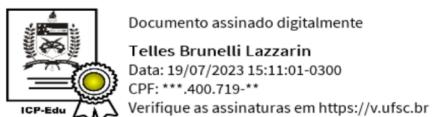
Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Elétrica” e aceito, em sua forma final, pelo Curso de Graduação em Engenharia Elétrica.

Florianópolis, 17 de julho de 2023.

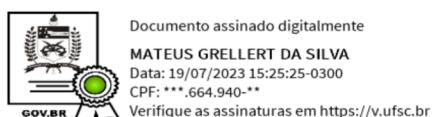


Prof. Miguel Moreto, Dr.
Coordenador do Curso de Graduação em Engenharia Elétrica

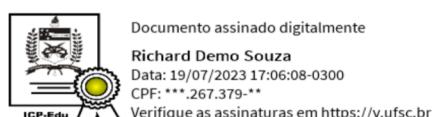
Banca Examinadora:



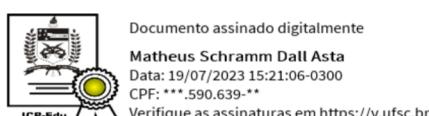
Prof. Telles Brunelli Lazzarin, Dr.
Orientador
Universidade Federal de Santa Catarina



Prof. Mateus Grellert da Silva, Dr.
Co-orientador
Universidade Federal do Rio Grande do Sul



Prof. Richard Demo Souza, Dr.
Universidade Federal de Santa Catarina



Eng. Matheus Schramm Dall Asta.
INEP – Instituto de Eletrônica de Potência - UFSC

Dedico este trabalho a todos aqueles que, de uma forma ou de outra contribuíram e me auxiliaram para a conclusão desta etapa.

AGRADECIMENTOS

Gostaria de agradecer primeiramente aos meus pais, por todo apoio, amor e dedicação que sempre me ofereceram e nesta etapa da minha vida especialmente, mesmo com tanta dificuldade e adversidades que ocorreram neste período, pelo apoio financeiro e emocional, por ter acreditado em mim, por querer o melhor para mim, eu já sou e serei eternamente grato por tudo.

A minha namorada que hoje é minha esposa Sabrina, por ser essa minha grande companheira, pela paciência neste período de estresse, pelo apoio que me deu, em todos os sentidos, por acreditar em mim, por me incentivar a crescer e participar dos meus planos e dos meus sonhos.

A minha irmã que sempre me apoiou, me incentivou e acreditou em mim, apesar de alguns afastamentos devido ao estudo, sempre me deu confiança e força, também se formou na UFSC o que me traz muito orgulho.

A meus amigos que fiz nesta jornada que em muitas horas na BU e em videochamadas unimos forças para superar cada desafio durante esta empreitada.

Ao engenheiro Felipe da secretaria do estado da saúde que também se formou na UFSC, que me deu oportunidade, e que me apoiou durante o curso, me emprestou muitos livros, me ensinou muito sobre o trabalho no dia a dia de engenharia, me aconselhou sobre estudo, trabalho e organização.

“A persistência é o menor caminho do êxito”.
(Charles Chaplin)

RESUMO

A energia eólica é uma das principais tecnologias de eletrificação com enorme potencial de crescimento. É conhecido por sua compatibilidade com o meio ambiente, economizando combustíveis fósseis e reduzindo poluentes. No entanto, embora haja um aumento significativo na capacidade eólica instalada a cada ano, essa ferramenta também acompanha seu próprio conjunto de desafios. Ao contrário das usinas convencionais, a energia eólica é altamente dependente de condições meteorológicas, que tornam a produção de energia difícil de prever. À medida que a integração de energias renováveis, como a energia eólica aumenta, algoritmos preditivos são necessários no atual sistema de energia elétrica, para fornecer previsões confiáveis, para reduzir os riscos técnicos e financeiros de todos participantes do mercado. A tecnologia de mineração de dados constitui-se em solução para extrair, de forma semi automática e inteligente, conhecimento relevante de enormes conjuntos de dados. Este trabalho apresenta uma abordagem de mineração de dados para previsão da velocidade do vento que tem baixo custo, contempla relevantes algoritmos de inteligência artificial e fornece recursos eficientes para tratamento de bancos de dados. A predição de ventos para geração eólica já é estudada há algum tempo utilizando diversas técnicas diferentes, este trabalho tem o diferencial de utilizar de como fonte de dados o MERRA-2 (Retrospectiva da Era Moderna para Pesquisa e Aplicações versão 2) é um conjunto de dados global de reanálise atmosférica desenvolvido pelo NASA's Global Modeling and Assimilation Office (GMAO). Ele fornece informações abrangentes sobre as condições atmosféricas passadas em todo o mundo, combinando observações de satélite, medições em terra e modelos de previsão do tempo estes dados são disponibilizados gratuitamente na internet e podem fornecer parâmetros pontuais em qualquer lugar no mundo mesmo que não tenha um sensor no local onde se deseja obter os dados. O objetivo deste trabalho é realizar uma predição de velocidade de vento de no mínimo uma hora em qualquer ponto utilizando esta fonte de dados, estes dados foram analisados para ver sua qualidade e realizado um tratamento de dados faltantes, foram usados modelos de aprendizado de máquina, como por exemplo o modelo Random Forest Regressor onde foi possível obter um coeficiente de determinação (R^2) de aproximadamente 0,96 e um valor NRMSE de 0,0306, além ferramentas de redes neurais recorrente como o LSTM e modelos mais tradicionais como ARIMA.

Palavras-chave: Recursos Renováveis; Energia Eólica; Velocidade do Vento; Mineração de Dados; Bancos de Dados; Inteligência Artificial.

ABSTRACT

Wind energy is one of the leading electrification technologies with enormous growth potential. It is known for its environmental compatibility, saving fossil fuels and reducing pollutants. However, despite a significant increase in installed wind capacity each year, this technology also comes with its own set of challenges. Unlike conventional power plants, wind energy is highly dependent on weather conditions, making energy production difficult to predict. As the integration of renewable energies, such as wind power, increases, predictive algorithms are needed in the current electricity system to provide reliable forecasts and reduce technical and financial risks for market participants. Data mining technology constitutes a solution to extract relevant knowledge from massive datasets in a semi-automatic and intelligent manner. This work presents a data mining approach for wind speed prediction that is cost-effective, incorporates relevant artificial intelligence algorithms, and provides efficient database handling capabilities. Wind prediction for wind power generation has been studied for some time using various techniques, and this work stands out by utilizing the MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) as a data source. MERRA-2 is a global atmospheric reanalysis dataset developed by NASA's Global Modeling and Assimilation Office (GMAO). It provides comprehensive information on past atmospheric conditions worldwide, combining satellite observations, ground measurements, and weather prediction models. These freely available data can provide point parameters anywhere in the world, even where no local sensors are present. The objective of this work is to predict wind speed at least one hour in advance at any given point using this data source. The data were analyzed for quality and missing data were treated. Machine learning models such as Random Forest Regressor were used, achieving a coefficient of determination (R^2) of approximately 0.96 and an NRMSE value of 0.0306. Recurrent neural network tools like LSTM and traditional models like ARIMA were also employed.

Keywords: Renewable Resources; Wind Energy; Wind Speed; Data Mining; Databases; Artificial Intelligence.

LISTA DE FIGURAS

Figura 1 - Atmosfera terrestre.	15
Figura 2 – Ranking mundial do potencial eólico instalado.	17
Figura 3 - Conjuntos de itens que compõem um modelo de aerogerador moderno.	18
Figura 4 – Exemplo de gráficos plotados com Matplotlib.	23
Figura 5 – Exemplo de gráficos plotados com Seaborn.	23
Figura 6 - Inteligência artificial.	24
Figura 7- Aprendizado de máquina.	25
Figura 8 - Exemplo de Algoritmo Random Forest.	27
Figura 9 - Loops Redes neurais recorrentes.	29
Figura 10 - O módulo de repetição em um LSTM contém quatro camadas de interação.	30
Figura 11 - Fluxograma da Metodologia.	32
Figura 12 - Interface do POWER Data Access Viewer.	33
Figura 13 - Gráfico de dados faltantes nas colunas.	36
Figura 14 - Distribuição de velocidade de vento no Parque Eólico de Osório.	37
Figura 15 - Boxplot da velocidade de vento pelas estações do ano.	37
Figura 16 - -Heat map de velocidade de ventos no ano de 2021 e 2022.	38
Figura 17 - Rosa dos ventos mostrando a direção da velocidade dos ventos.	39
Figura 18 - Gráficos de autocorrelação da velocidade de vento.	40
Figura 19 - Entradas de dados e variável target a ser predita pelo modelo.	41
Figura 20 - Dispersão entre Valores reais e valores preditos nos modelos.	43
Figura 21 - Velocidade de vento Reais vs Previstos para o modelo RFR nas três regiões.	43
Figura 22 - Direção dos ventos nos três pontos estudados.	47
Figura 23- Heat map de velocidade de ventos no ano de 2021 e 2022 nos outros dois pontos analisados.	47

LISTA DE TABELAS

Tabela 1- Descrição de parâmetros	35
Tabela 2 - Descrição de pontos analisados.	35
Tabela 3 - Resultados usando modelos de predição para o Parque Eólico de Osório.	45
Tabela 4 - Valor real e predição de velocidade de ventos para o Parque Eólico de Osório.	46
Tabela 5 - Resultados para as três regiões utilizando do modelo Random Forest Regressor.	48
Tabela 6 - Porcentagem de quanto os resultados estão acima ou abaixo do Baseline.	49
Tabela 7 - - Resultados para 5 passos de tempos a frente da previsão de velocidade de vento.	49

LISTA DE SIGLAS E ABREVIATURAS

UFSC Universidade Federal de Santa Catarina

PV Fotovoltaico

IA Inteligência artificial

STC Condições de teste padrão

RFR Random Forest Regressor

ML Machine Learning

MAE Erro Médio Absoluto

RMSE Erro quadrático médio

ANEEL Agência nacional de energia elétrica

PR Performance Ratio

MF Modelo Físico

LSTM Long Short-Term Memory

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 OBJETIVO.....	14
1.2 OBJETIVOS ESPECÍFICOS.....	14
2 FUNDAMENTAÇÃO TEÓRICA.....	15
2.1 METEOROLOGIA.....	15
2.2 ENERGIA EÓLICA.....	16
2.3 PREVISÃO NA GERAÇÃO ENERGIA EÓLICA.....	19
2.3.1 COEFICIENTE DE HELLMANN.....	20
2.4 MINERAÇÃO DE DADOS.....	21
2.4.1 TRATAMENTO DE DADOS.....	21
2.4.2 VISUALIZAÇÃO E ANÁLISE EXPLORATÓRIA DE DADOS.....	22
2.4.3 Inteligência Artificial.....	23
2.4.5 TÉCNICAS E FERRAMENTAS DE APRENDIZADO DE MÁQUINA.....	24
2.4.5.1 Aprendizado de Máquina.....	24
2.4.5.2 Python.....	25
2.4.5.3 Random Forest.....	26
2.5 ANÁLISE E PREDIÇÃO DE SÉRIES TEMPORAIS.....	27
2.5.1 ARIMA.....	27
2.5.2 LSTM.....	28
2.5.3 Indicadores em séries temporais.....	29
2.6 TRABALHOS RELACIONADOS.....	30
3 METODOLOGIA.....	31
3.1 Bases de Dados.....	32
3.2 Pré-processamento da Base de Dados.....	33
3.3 Visualização.....	35
3.4 Modelos de predição de velocidade de vento.....	39
3.4.1 Validação Walk-Forward.....	40
3.4.2 Random Forest Regressor.....	40
3.4.3 ARIMA Walk-Forward.....	40
3.4.4 LSTM.....	41
4 RESULTADOS.....	42
4.1 Métricas de avaliação.....	43
4.2 Baseline.....	44
4.3 Tabela de Resultados.....	44
4.8 Comparação com outras regiões.....	45
5 Conclusão.....	50
REFERÊNCIAS.....	52

1 INTRODUÇÃO

A demanda de energia vem aumentando em muito em todo o planeta devido ao aumento da população desde a revolução industrial e o consumo por tecnologia faz acelerar mais ainda este processo. A queima de combustíveis fósseis para geração de energia se mostrou nociva e insustentável, liberando dióxido de carbono na atmosfera prejudicando ecossistemas e se tornando um risco para a saúde humana. Temendo a escassez de energia e visando minimizar os danos ao meio ambiente, a maioria dos países utilizaram de políticas para expandir suas matrizes energéticas por meio de recursos renováveis.

A energia eólica vem como uma destas alternativas, que mais se desenvolve rapidamente no mundo inteiro, devido a sua vantagem em comparação a outras fontes renováveis na geração em larga escala. De fato, a produção elétrica vinte e quatro horas por dia, evolução dos equipamentos, incentivos governamentais e inovação são fatores que motivam os altos investimentos nas indústrias eólicas (PACHECO, 2006).

Sem dúvida, no Brasil a geração eólica mudou para sempre o Nordeste do país e outras regiões também, tendo promovido benefícios sociais, econômicos e ambientais.

No entanto, a inconsistência, imprevisibilidade e incerteza do fornecimento de energia são problemas persistentes que impedem o avanço desses sistemas. Produção eólica depende diretamente da força dos ventos, recurso natural e inesgotável, porém intermitente, imprevisível e incontrolável.

As estimativas da velocidade do vento são requisitos para o funcionamento eficiente dos sistemas eólicos. Porém, prever as condições de ventos é um problema complexo, considerando que o movimento do ar é originado pela diferença de pressão entre regiões e influenciado por fatores físicos e climáticos, o que demanda recursos tecnológicos apropriados.

Para propor soluções neste contexto, devemos considerar que existem grandes quantidades de dados meteorológicos que têm sido acumulados ao longo dos tempos, estes são captados a todo momento através de instrumentos como satélites, barômetros, anemômetros, radares e veículos aéreos não tripulados. Extrair informação desta coleção de dados tão ampla e diversificada supera a capacidade racional humana.

A tecnologia de mineração de dados fornece recursos poderosos para o tratamento eficiente de banco de dados, assim como algoritmos de inteligência artificial capazes de extrair conhecimento de enormes conjuntos de dados. Diante disso, essa tecnologia tem potencial para lidar com os problemas de previsão da velocidade do vento e pode fornecer contribuições significativas ao setor eólico.

1.1 OBJETIVO

Este trabalho tem como objetivo principal elaborar uma abordagem de mineração de dados em dados meteorológicos disponíveis, para construir modelos de previsão de velocidade de vento visando o avanço tecnológico na operação das indústrias e dar suporte a operação das usinas eólicas conectadas ao SIN (Sistema Integrado Nacional).

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos do trabalho são:

- Propor soluções baseadas em Inteligência artificial para os problemas de predição de geração de energia em das indústrias em sistemas eólicos;
- Análise dos métodos de predição de vento, com dados retirados do sistema NASA POWER que é disponibilizado gratuitamente na internet;
- Analisar 3 pontos no Brasil para aplicar o método desenvolvido, sendo que um destes pontos é offshore e não conta com nenhum sensor no local;
- Comparar métodos através de suas métricas de avaliação de modelos, a fim de obter um melhor resultado;
- Propor ferramentas e recursos computacionais para auxiliar os operadores de energia.

2 FUNDAMENTAÇÃO TEÓRICA

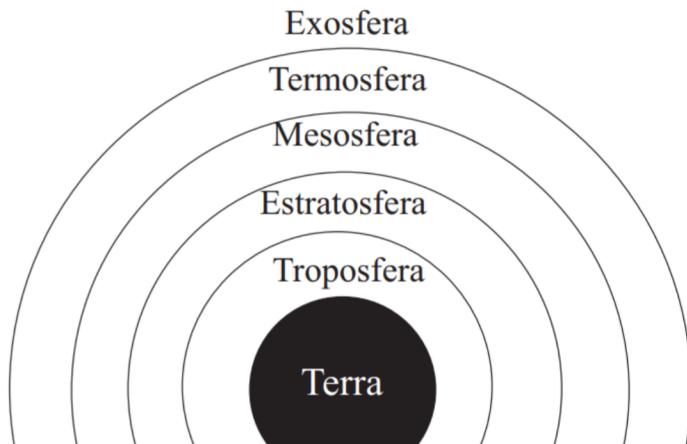
Nesta seção, conteúdos que estruturam a proposta deste trabalho são apresentados, assim como as principais bibliotecas utilizadas nos resultados numéricos, bem como os conceitos de aprendizado de máquina e de análise de dados.

2.1 METEOROLOGIA

A meteorologia (do grego meteors que significa “elevado no ar”) é a ciência que estuda os fenômenos da atmosfera terrestre. Precisamente as condições de tempo e clima. As pessoas erroneamente usam os termos tempo e clima. O tempo se refere ao estado momentâneo da atmosfera, enquanto que o clima se refere ao estado médio da atmosfera que é obtido por meio da média dos eventos de tempo durante um longo período (REBOITA et al., 2012).

A Figura 1 mostra como se divide a atmosfera terrestre, camada gasosa que envolve a Terra, é formada por cinco camadas de gases: exosfera, onde os meteoros incandescem quando passam por ela; termosfera, que absorve as radiações solares mais energéticas, mesosfera ou ionosfera, que é onde o ozônio começa a se formar e onde ocorrem as auroras boreais e austrais, estratosfera, onde os aviões trafegam e nela está a camada de ozônio e por fim troposfera, onde ocorre a vida e os fenômenos meteorológicos (BARRY; CHORLEY, 2013).

Figura 1 – Atmosfera terrestre



A diversidade de aplicações, a ampla disponibilidade de dados, as complexidades envolvidas no monitoramento da superfície terrestre e o impacto imediato na socioeconomia e na infraestrutura são fatores que impulsionam as pesquisas espaciais relacionadas à meteorologia (BARRY; CHORLEY, 2013).

2.2 ENERGIA EÓLICA

O fenômeno das mudanças climáticas ou aquecimento global é uma questão de sérias consequências para o nosso planeta. Eles dão origem a uma infinidade de problemas, incluindo, entre outros, tempestades turbulentas, inundações e ciclones.

Pesquisas indicam que as atividades humanas são responsáveis por chuvas ácidas e longos períodos de seca. A temperatura da Terra experimentou um aumento dramático, que pode ser atribuído à queima de combustíveis fósseis. A utilização de combustíveis fósseis para geração de energia contribui para a emissão de gases de efeito estufa na atmosfera, compostos predominantemente de dióxido de carbono (PACHECO, 2006).

De olho nas gerações futuras, as nações em todo o mundo estão buscando um desenvolvimento sustentável que equilibre os fatores sociais, econômicos e ambientais. Isso envolve o aproveitamento de recursos naturais como sol, vento, água e biomassa para gerar energia. Além de fornecer energia limpa de alta qualidade e econômica, diversificar e expandir as fontes de energia, pode aliviar os problemas causados pela dependência excessiva de uma única fonte de energia.

Diante do exposto, todos os recursos sustentáveis possuem relevância na geração de energia. Deve-se notar que as matrizes energéticas são independentes e, assim, uma fonte tem a capacidade de complementar a outra. Cada nação tem suas próprias características únicas, muitas vezes moldadas por seu clima, que as diferenciam das outras. Felizmente, a abundante biodiversidade do Brasil oferece amplo potencial para geração de energia renovável a partir de fontes como eólica, solar, hidrelétrica e biomassa (PACHECO, 2006).

Os bons ventos sopram forte no Nordeste do Brasil e o setor acumula recordes de geração de energia. Isso ajudou a reduzir as consequências da crise hídrica que o país enfrentou no passado devido às chuvas secas e estimulou a recuperação econômica, gerando empregos.

O Brasil está em sexto lugar no ranking global do Global Wind Energy Report 2022 (uma organização internacional especializada em energia eólica) com 21,5 GW de capacidade instalada eólica onshore (GWEC, 2022). A Figura 2 exibe o *ranking* mundial do potencial eólico instalado por países.

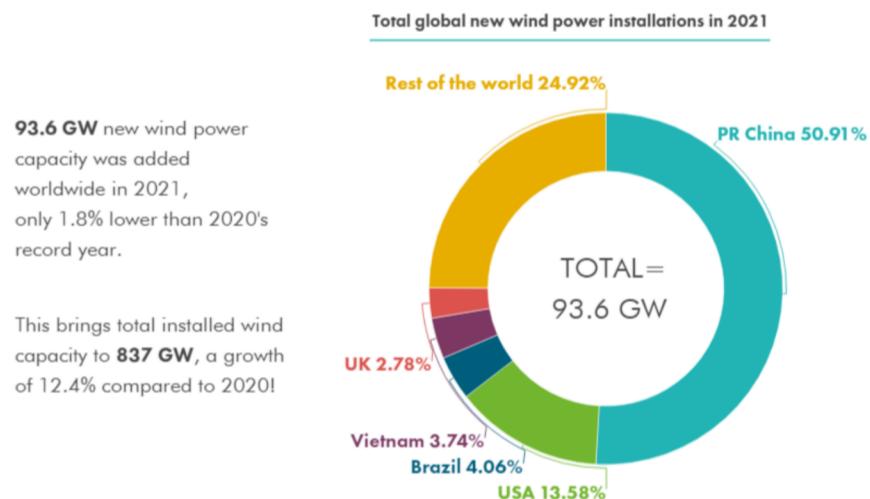


Figura 2 – Ranking mundial do potencial eólico instalado.

Fonte: Adaptada de (GWEC, 2022)

Para um breve entendimento como os sistemas eólicos funcionam, o texto a seguir introduz os conceitos básicos de elementos principais. Os aerogeradores (ou turbinas eólicas) são aparelhos geradores de energia que funcionam de forma semelhante aos moinhos de vento, ferramentas amplamente utilizadas pelos homens do campo para bombear água e moer minerais. Em resumo, a energia cinética do ar em movimento é convertida em energia mecânica pela força rotacional do rotor. Portanto, um gerador elétrico conectado ao rotor diretamente ou através de uma caixa de engrenagens transforma energia mecânica em energia elétrica (UCZAI , 2012).

Grandes turbinas modernas são montadas em uma torre acima do solo, que consiste em um rotor horizontal, uma hélice de três pás e um anemômetro (ou sensor de vento) para medir a intensidade da velocidade do vento. (normalmente a cada 10 minutos), além dos outros elementos mostrados na Figura 3.

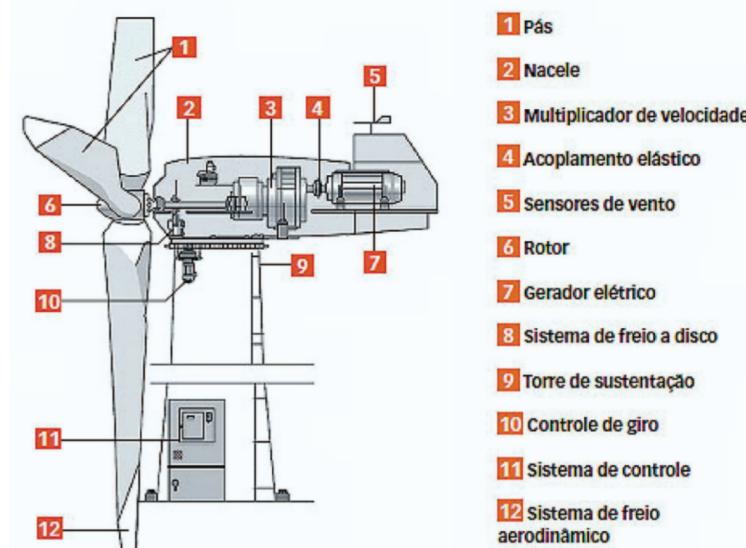


Figura 3 – Conjuntos de itens que compõem um modelo de aerogerador moderno.

Fonte: Adaptada de (STAVISSL, 2011).

Os aerogeradores possuem limitações técnicas para evitar danos às suas estruturas e garantir a viabilidade econômica da produção de energia elétrica. Essas limitações incluem um limite superior (ou cut-out) e inferior (ou cut-in) para a sua operação. Normalmente, a produção de eletricidade começa quando o vento está próximo a 3 metros por segundo (m/s), enquanto fortes rajadas de 25 m/s interrompem a produção. É importante ressaltar que os limites operacionais podem variar dependendo das características de cada aerogerador e fabricante. Embora a maioria dos parques eólicos do mundo esteja localizada em terra (ou onshore), algumas usinas na Europa estão instaladas no mar (ou offshore). Apesar dos maiores custos de instalação e manutenção, essa opção é interessante porque muitas vezes não há locais apropriados em terra e os ventos mais fortes do oceano permitem um melhor aproveitamento da eletricidade (TOLMASQUIM, 2016).

A efetividade do planejamento das operações industriais depende em grande medida das previsões de energia, pois elas proporcionam informações essenciais para a manutenção dos sistemas, a configuração das turbinas, a gestão energética e do comércio (COLAK; SAGIROGLU; YESILBUDAK, 2012). De acordo com o relatório do Laboratório Nacional de Energia Renovável dos Estados Unidos (National Renewable Energy Laboratory - NREL), em alguns países, os operadores de energia estão sujeitos a sanções se a produção ficar abaixo das suas estimativas, e as empresas podem não receber pagamento pela energia produzida acima de uma estimativa (NREL, 2010). Portanto, é fundamental que as previsões sejam precisas e confiáveis para garantir a eficiência das operações e evitar prejuízos.

2.3 PREVISÃO NA GERAÇÃO ENERGIA EÓLICA

De fato, um dos grandes desafios das indústrias eólicas é a inconsistência nas previsões de energia, o que tem dificultado a regulamentação desses sistemas. Existem quatro categorias de algoritmos preditivos para prever a potência de turbinas eólicas, que podem ser divididos de acordo com o horizonte de previsão. Essas categorias incluem previsões de muito curto prazo (alguns segundos a 30 minutos), previsões de curto prazo (de 30 minutos a 48-72 horas), previsões de médio prazo (de 48-72 horas a 1 semana) e previsões de longo prazo (de 1 semana a anos). Os métodos de previsão de energia eólica podem ser classificados de acordo com o tipo e disponibilidade de dados, bem como o objetivo da previsão (MEKA,R.;ALAEDDINI, A.;BHAGANAGAR, K., 2020).

Existem diferentes métodos de previsão de energia eólica, que podem ser divididos em três categorias: métodos físicos, métodos estatísticos e métodos híbridos, que são uma combinação de ambos. Os métodos físicos usam modelos numéricos de previsão do tempo para obter características de mesoescala do tempo (NWP), e fornecem previsões satisfatórias para horizontes de tempo mais longos. Já os modelos matemáticos baseados em estatísticas usam dados históricos para fazer previsões, sendo mais adequados para energia de curto prazo. Esses modelos são convencionalmente divididos em diferentes categorias, como autorregressivo (AR), média móvel (MA), média móvel autorregressiva (ARMA) e modelos de média móvel autoregressiva integrada (ARIMA). Os métodos híbridos, por sua vez, combinam dados de previsões meteorológicas e análise de séries temporais para fornecer previsões robustas de energia eólica (MEKA,R.;ALAEDDINI, A.;BHAGANAGAR, K., 2020).

Redes neurais recorrentes (RNN) são outro tipo de redes neurais conhecidas por funcionar bem com dados de séries temporais. (Barbounis et al,2006) e (Senju et al,2006) discutem o uso de RNNs para velocidade e potência do vento a longo prazo de previsão. Já (Felder et al, 2010) discutem o uso de RNNs para prever a energia eólica de curto prazo usando as últimas 24 horas de medições de potência média e dados NWP como entrada para o modelo. No entanto, os modelos RNN tem o problema de gradientes de fuga, em que a contribuição dos estados anteriores desaparece após alguns intervalos de tempo.

LSTM é um tipo especial de modelo RNN que é conhecido por resolver o problema de gradiente de desaparecimento de RNNs por um pequeno custo computacional extra. Desde a sua criação em 1997 já foi usado para uma grande variedade de soluções.

As previsões indiretas de energia podem ser obtidas através da equação 1, na qual a saída de energia P é dada em watts (W), C é o fator dependente do modelo da turbina (ou coeficiente de potência), S é a área de varredura em metros quadrados (m²), A é a densidade do ar em quilograma por metro cúbico (Kg/m³), e V a velocidade do vento em m/s .

Portanto, prever o comportamento do vento é primordial para as indústrias.

$$P = \frac{1}{2} \cdot C \cdot S \cdot A \cdot V^3 \quad (1)$$

O vento é um recurso natural, intermitente, incerto e de difícil controle que afeta diretamente a saída das turbinas eólicas.

2.3.1 COEFICIENTE DE HELLMANN

O coeficiente de Hellmann é um fator usado para relacionar a velocidade do vento em diferentes altitudes. Em 1916, Hellmann apresentou um modelo prático e eficiente para caracterizar a variação vertical da velocidade do vento (Okorie, 2014). Esse modelo sugere que a relação entre as velocidades do vento em duas alturas distintas pode ser determinada. O coeficiente de Hellmann é importante porque a velocidade do vento aumenta com a altura, devido à menor resistência oferecida pelo ar em altitudes mais elevadas. O coeficiente de Hellmann é utilizado para prever a velocidade do vento em uma determinada altura com base nas medições feitas em outra altura. Destaca-se que para obter o coeficiente de Hellmann é necessário ter a medida de duas alturas no ponto, após o cálculo do coeficiente conseguimos calcular a velocidade do vento para qualquer altura (Wharton, 2012).

A equação 2 que relaciona o coeficiente de Hellmann com as velocidades do vento em diferentes altitudes é:

$$\alpha = \frac{\log\left(\frac{v}{v_r}\right)}{\log\left(\frac{h}{h_r}\right)}. \quad (2)$$

Onde:

- α é o coeficiente de Hellmann.
- v_r é a velocidade do vento em uma altura de referência h_r .
- h_r é a altura de referência.
- v é a segunda medida do vento.
- h é a altura da segunda velocidade de vento.

O coeficiente de Hellmann é utilizado em muitas aplicações, incluindo a geração de energia eólica. Ele é uma ferramenta importante para entender como a velocidade do vento varia com a altura e como isso pode afetar as condições ambientais em diferentes locais. É possível então obter a velocidade para qualquer altura, o que para este trabalho se torna importante, no caso de a altura do aerogerador não ser a mesma altura presente no banco de dados. Abaixo está apresentada a equação 3 para obter a velocidade do vento para qualquer altura a partir do coeficiente de Hellmann.

$$v = v_r \left(\frac{h}{h_r} \right)^\alpha \quad (3)$$

2.4 MINERAÇÃO DE DADOS

Este capítulo aborda o tratamento de dados para obter dados consistentes para a posterior análise nos modelos propostos.

2.4.1 TRATAMENTO DE DADOS

O tratamento de dados consiste em diversas técnicas que visam a remoção dos dados faltantes ou duvidosos e a transformação destes para um melhor funcionamento no modelo de aprendizado. Precisamos de duas fontes de conhecimento, dados coletados em anos anteriores e conhecimento de domínio para auxiliar a previsão por julgamento, dados muitos antigos podem ser úteis, porém pode haver mudanças estruturais no sistema então dados recentes são confiáveis, quanto mais dados melhor dentro do que é razoável para a previsão.

É uma etapa crucial para a construção de modelos de aprendizado de máquina para previsão. O processo começa com a coleta dos dados relevantes, que devem ser limpos e pré-processados para garantir a qualidade e a consistência dos dados. Isso inclui a eliminação de dados duplicados, a correção de valores faltantes e a normalização dos dados. Em seguida, os dados são divididos em conjuntos de treinamento e teste para avaliar a precisão do modelo. É importante selecionar recursos relevantes e excluir aqueles que não têm importância, bem como aplicar técnicas de redução de dimensionalidade para evitar a complexidade desnecessária do modelo.

2.4.2 VISUALIZAÇÃO E ANÁLISE EXPLORATÓRIA DE DADOS

A visualização e a análise exploratória de dados são etapas cruciais no processo de análise de dados. A visualização de dados refere-se à representação gráfica dos dados, a fim de extrair informações e percepções relevantes. Já a análise exploratória de dados é o processo de investigação e descoberta de padrões, tendências e relações nos dados. A visualização de dados pode ser feita de várias maneiras, incluindo gráficos de barras, gráficos de dispersão, histogramas, mapas de calor, entre outros. A escolha do tipo de gráfico dependerá dos dados e das informações que se deseja extrair. A visualização de dados permite a identificação de tendências, padrões e relacionamentos, além de tornar os dados mais acessíveis e compreensíveis para pessoas que não possuem experiência em análise de dados (HAN, J.; KAMBER, M.; PEI, J. 2011).

A análise exploratória de dados, por sua vez, permite a identificação de valores discrepantes, summarização e o entendimento da distribuição dos dados, a detecção de padrões e tendências, além de fornecer informações importantes sobre os dados. A análise exploratória de dados pode incluir a verificação da qualidade dos dados, a seleção de recursos relevantes, a identificação de correlações, a detecção de outliers e a verificação da normalidade dos dados.

A ferramenta Python dispõe de bibliotecas como *matplotlib* e *Seaborn* por exemplo, ambas usadas para plotar gráficos e são ferramentas imprescindíveis na visualização dos resultados, a fim de obter informações sobre os dados analisados. Com as Figuras 4 e 5 pode-se verificar exemplos do que estas ferramentas são capazes de fornecer ajuda na análise.

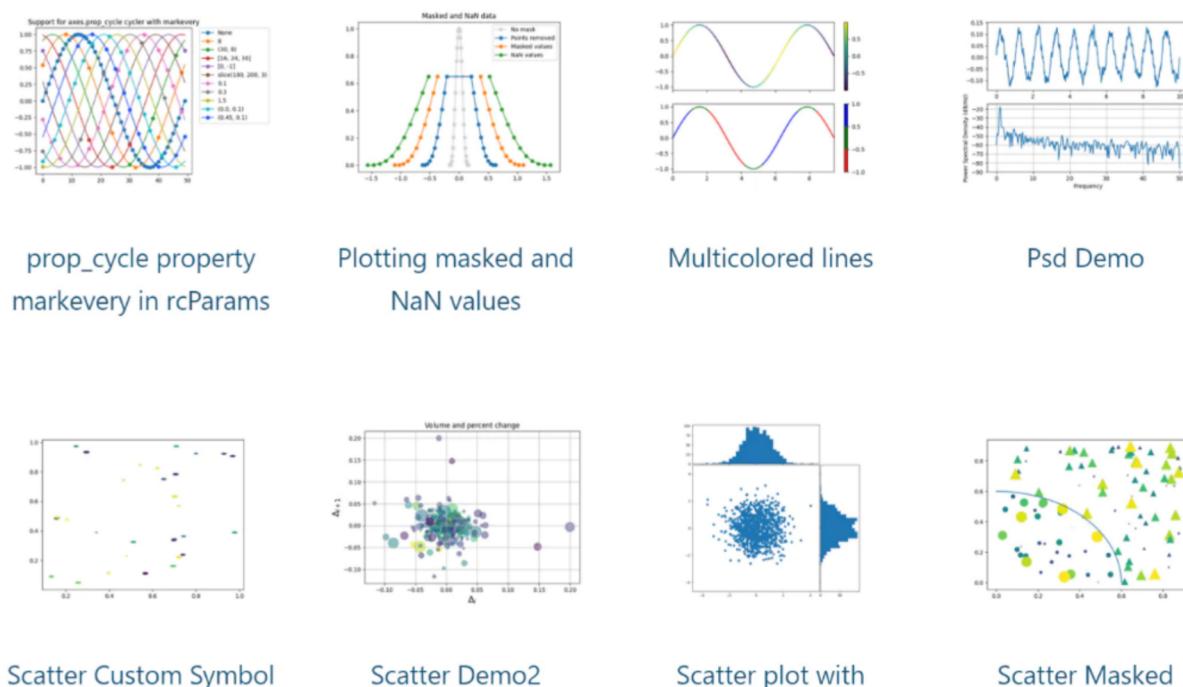
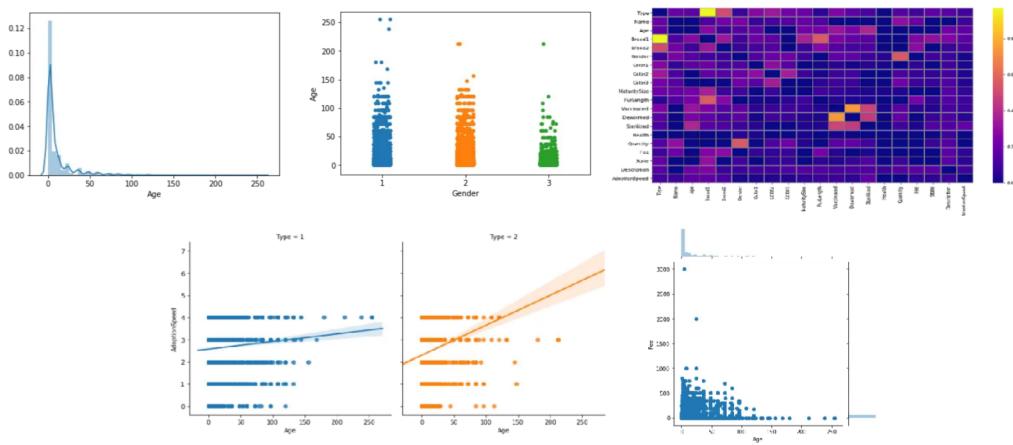


Figura 4 – Exemplo de gráficos plotados com *Matplotlib*.
Fonte: Adaptada de (NUNES, 2022).

Figura 5 – Exemplo de gráficos plotados com *Seaborn*.

Fonte: Adaptada de (ALVES, 2019).

2.4.3 Inteligência Artificial

A inteligência artificial (IA) é um campo da ciência da computação que busca desenvolver algoritmos e sistemas capazes de realizar tarefas que, até então, eram exclusivas de seres humanos, como reconhecimento de voz, análise de dados, tomada de decisões e até mesmo a criação de arte.

Os sistemas de IA são baseados em modelos matemáticos e estatísticos que permitem que os computadores aprendam a partir de dados e exemplos, o que é conhecido como aprendizado de máquina. Com o tempo, esses sistemas podem melhorar sua precisão e eficiência, tornando-se cada vez mais úteis em diversas áreas, desde a saúde e a educação até a indústria e o comércio. A Figura 6 mostra a abrangência da IA e quanto o aprendizado de máquina abrange e ainda é apenas uma parte da base de conhecimento de inteligência artificial.

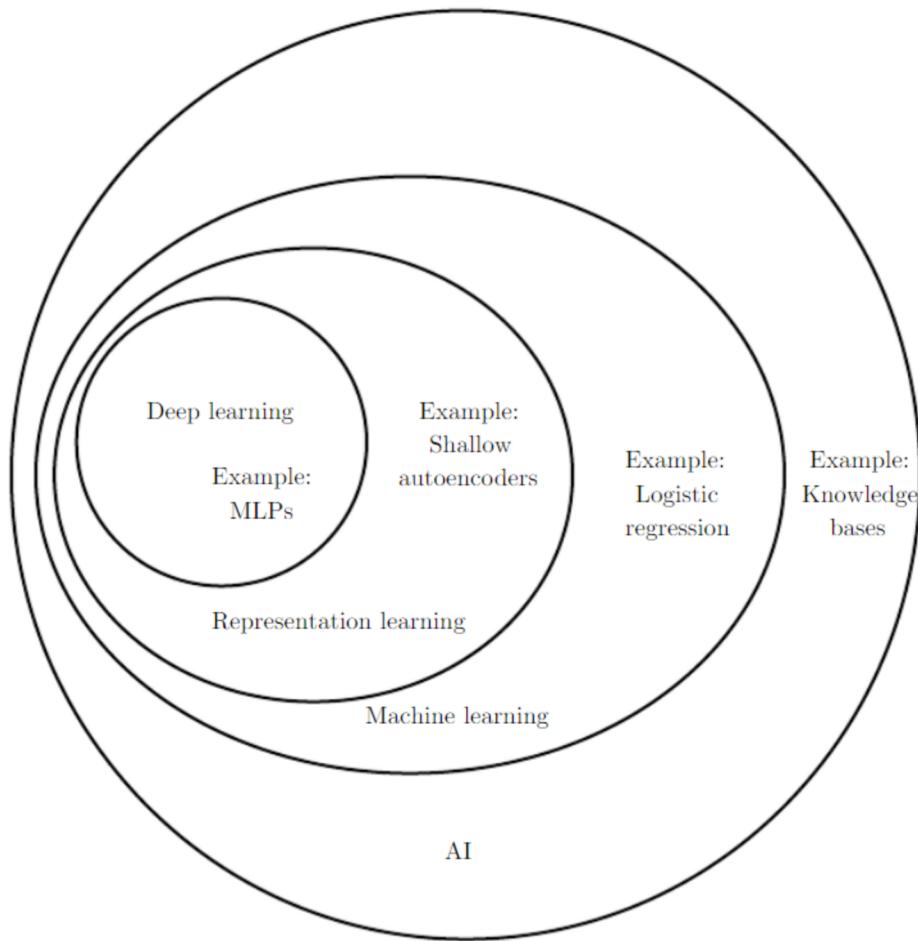


Figura 6 - Inteligência artificial.
Fonte: Adaptada de (GOODFELLOW, 2016).

2.4.4 TÉCNICAS E FERRAMENTAS DE APRENDIZADO DE MÁQUINA

2.4.4.1 Aprendizado de Máquina

A Inteligência Artificial (IA) pode ser definida como a tentativa de imitar habilidades humanas, já o Aprendizado de Máquina é uma parte específica da IA que se concentra em treinar máquinas para aprender a partir de dados. Com o avanço das tecnologias de conectividade, como a Internet das Coisas, a quantidade de dados disponíveis é imensa. Na segunda metade do século XX, o Aprendizado de Máquina evoluiu como uma sub-área da IA, com algoritmos capazes de aprender automaticamente a partir de dados e fazer diversas previsões futuras com base nesse conhecimento (Raschka, Mirjalili, 2015).

É possível fazer previsões sobre o futuro com aprendizado supervisionado. O principal objetivo do aprendizado supervisionado é aprender um modelo a partir de um treinamento rotulado. Aqui o termo "supervisionado" refere-se a um conjunto de exemplos de treinamento

(entradas de dados) onde os sinais de saída desejados (rótulos) já são conhecidos. A Figura 7 a seguir resume um típico fluxo de trabalho de aprendizado supervisionado, onde os dados de treinamento rotulados são passados a um algoritmo de aprendizado de máquina para ajustar um modelo. E este modelo de aprendizado será útil neste trabalho.

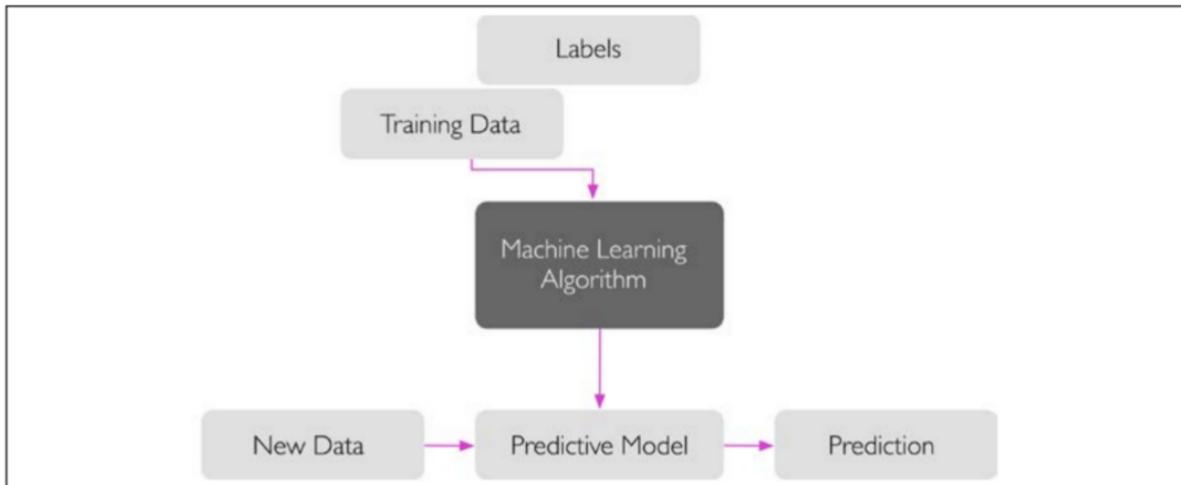


Figura 7 - Aprendizado de máquina.
Fonte: Adaptada de (Raschka, Mirjalili, 2015).

2.4.4.2 Python

Criada por Guido van Rossum em 1991 (KUHLMAN, 2009), Python é uma linguagem de programação de alto nível e propósito geral que se destaca por sua interpretação dinâmica. Devido à sua facilidade de aprendizado e sintaxe clara, Python é amplamente utilizada em áreas como desenvolvimento web, ciência de dados, automação de tarefas, jogos e inteligência artificial, consolidando-se como uma das linguagens mais populares em todo o mundo.

Python é uma das linguagens de programação mais importantes para machine learning, pois tem uma ampla gama de bibliotecas e frameworks especializados em machine learning, como o TensorFlow, Keras, PyTorch, Scikit-learn, entre outros. Essas ferramentas oferecem uma série de funções e recursos que tornam o desenvolvimento de modelos de machine learning mais fácil e rápido, por este mesmo motivo foi escolhida para este trabalho.

2.4.4.3 Random Forest

A técnica de aprendizado de máquina supervisionada, Random Forest (RF), é um algoritmo que pode ser aplicado tanto em problemas de classificação, em que a saída alvo é um rótulo, quanto em problemas de regressão, em que a saída alvo é um número(Shafi, 2023).

Durante o treinamento do algoritmo, uma árvore binária é construída, na qual cada nó interno é um teste de um atributo, cada ramificação é o resultado do teste e cada nó folha contém um rótulo de classe. O objetivo não é criar apenas um modelo de aprendizado, mas sim vários modelos, onde a saída final é uma função de todas as saídas intermediárias de cada árvore. É possível definir o número de árvores utilizado pelo algoritmo, normalmente entre 100 e 1000, enquanto a seleção dos atributos usados em cada árvore é aleatória. A **Figura 8** fornece um exemplo conciso de como a construção de uma Random Forest é realizada, onde exemplifica a quantidade de árvores que determinamos para o modelo e o resultado vindo da média de resultados de todas estas árvores.

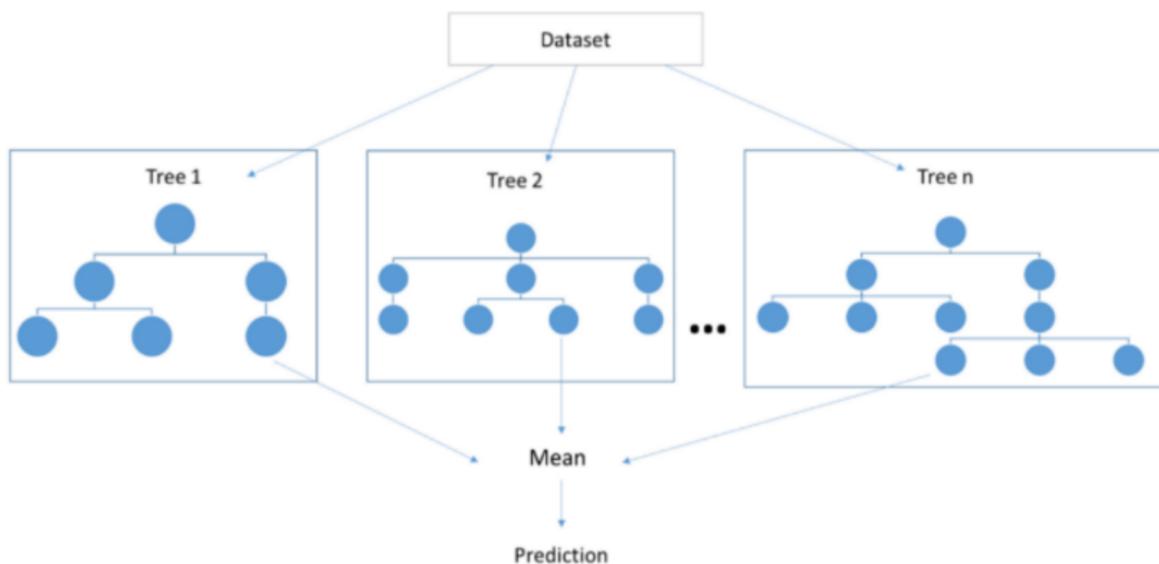


Figura 8 - Exemplo de Algoritmo Random Forest.
Fonte: Adaptada de (Deep AI, 2020).

2.5 ANÁLISE E PREDIÇÃO DE SÉRIES TEMPORAIS

Em conjuntos de dados de aprendizado de máquina convencionais, o tempo desempenha um papel importante, principalmente quando se faz previsões para novos dados em que o resultado real não é conhecido até uma data futura. Embora o futuro seja o foco, todas as observações anteriores são geralmente tratadas de maneira similar.

Já em séries temporais, há uma dependência explícita da ordem das observações, adicionando uma dimensão temporal extra. Essa dimensão adicional não só é uma restrição, mas também fornece uma estrutura que pode ser uma fonte valiosa de informações adicionais (BROWNLEE, 2016).

2.5.1 ARIMA

O ARIMA é uma técnica estatística bastante conhecida e utilizada para fazer previsões em séries temporais. Sua sigla, em inglês, significa "Autoregressive Integrated Moving Average", que é uma generalização da média móvel autoregressiva simples, incluindo a noção de integração. Esse modelo é capaz de capturar diversas estruturas temporais presentes em dados de séries temporais, o que o torna uma classe de modelo bastante versátil.

O modelo ARIMA é descrito pelo seu acrônimo, que captura de forma resumida os principais aspectos do modelo. O **AR** representa a autoregressão, que usa a relação dependente entre uma observação e um número de observações anteriores. O **I** representa a integração, que usa a diferenciação de observações brutais para tornar a série temporal estacionária. Por fim, o **MA** representa a média móvel, que usa a dependência entre uma observação e um erro residual de um modelo de média móvel aplicado a observações anteriores. Juntos, esses aspectos formam uma classe de modelo capaz de capturar diferentes estruturas temporais em dados de séries temporais. O modelo ARIMA usa uma notação padrão de ARIMA(p,d,q) para especificar cada um dos componentes do modelo. Isso permite uma rápida identificação do modelo ARIMA utilizado. Os parâmetros **p**, **d** e **q** são números inteiros que definem cada um dos componentes do modelo. O parâmetro **p** define a ordem de atraso, ou seja, o número de observações passadas que são consideradas no modelo. Já o parâmetro **d** define o grau de diferenciação, ou o número de vezes que as observações brutas são diferenciadas para tornar a série temporal estacionária. Por fim, o parâmetro **q** define o tamanho da janela da média móvel, ou seja, o número de observações passadas que são usadas para calcular a média móvel. Cada um desses parâmetros é especificado explicitamente no modelo ARIMA e pode ser ajustado para melhorar a precisão da previsão (Brownlee, 2017).

2.5.2 LSTM

O LSTM utiliza princípios de Redes Neurais Recorrentes. Ao contrário das redes neurais tradicionais, que começam a processar informações do zero a cada nova entrada, as redes neurais recorrentes permitem que informações anteriores sejam mantidas e utilizadas para processar novas entradas. Isso significa que o contexto é levado em consideração e a persistência de pensamentos é simulada, tornando-as mais eficientes. Para conseguir isso, essas redes utilizam loops, onde as informações são passadas de uma etapa para a próxima. No caso do diagrama da Figura 9, uma parte da rede neural, chamada de "A", analisa uma entrada x_t e emite um valor h_t , que é repassado para a próxima etapa por meio do loop. É importante destacar que essa capacidade é uma grande vantagem em relação às redes neurais tradicionais (GERS, 2015).

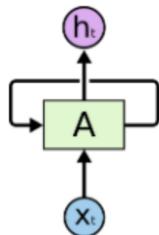


Figura 9 - Redes neurais recorrentes possuem loops.

Fonte: Adaptada de (GERS, 2015).

Porém, por mais eficientes que sejam as RNNs existe um problema das dependências de longo prazo, onde nas redes neurais recorrentes se refere à dificuldade que essas redes têm em lembrar informações relevantes em sequências de dados muito longas. Em outras palavras, as informações transmitidas em etapas anteriores da rede podem se tornar "esquecidas" ou desvanecer gradualmente, o que pode afetar negativamente a precisão da previsão ou classificação realizada pela rede (GERS, 2015).

Essa dificuldade ocorre porque, à medida que as informações passam pelas diferentes etapas da rede, elas são sujeitas a transformações não lineares, que podem fazer com que certas informações sejam amplificadas ou atenuadas. Com o tempo, informações relevantes podem ser perdidas ou se tornarem cada vez mais fracas em relação a outras informações mais recentes. Isso é especialmente problemático em tarefas em que as informações de etapas anteriores são fundamentais para a correta classificação ou previsão, como tradução automática ou reconhecimento de fala, por exemplo (GERS, 2015).

Para resolver esse problema, foram desenvolvidas arquiteturas mais sofisticadas de redes neurais recorrentes, como as chamadas redes LSTM (Long Short-Term Memory), que utilizam uma estrutura de células de memória que permite a persistência de informações relevantes ao longo do tempo, mantendo assim a capacidade de lidar com dependências de longo prazo (GERS, 2015).

Os LSTM também têm uma estrutura semelhante a uma cadeia de módulos repetitivos de

rede neural, mas o módulo de repetição tem uma estrutura diferente como podemos ver na Figura 10. Em vez de ter uma única camada de rede neural, são quatro, interagindo de maneira muito especial (GERS, 2015).

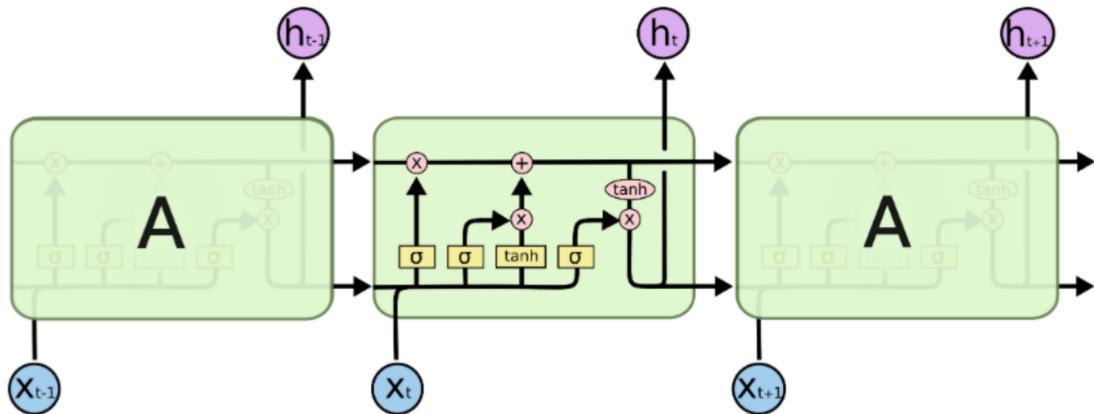


Figura 10 - O módulo de repetição em um LSTM contém quatro camadas de interação.

Fonte: Adaptada de (GERS, 2015).

2.5.3 Indicadores em séries temporais

Os indicadores de regressão são usados para medir a qualidade de um modelo de regressão e avaliar o quanto bem ele se ajusta aos dados (KUTNER, NACHTSHEIM, NETER, 2004). Aqui estão alguns dos indicadores mais comuns de regressão:

R-quadrado (R^2): uma medida de quanto bem a linha de regressão se ajusta aos dados. O R^2 varia de 0 a 1, onde 1 significa que a linha de regressão se ajusta perfeitamente aos dados e 0 significa que não há relação entre as variáveis.

$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (4)$$

Erro médio absoluto (MAE): a média das diferenças absolutas entre as previsões do modelo e os valores reais. O MAE mede a magnitude média do erro.

$$MAE(y, p) = \frac{1}{N} \sum_{i=1}^N |y_i - p_i| \quad (5)$$

Erro quadrático médio (MSE): a média dos quadrados das diferenças entre as previsões do modelo e os valores reais. O MSE penaliza mais os erros maiores do que os erros menores.

$$MSE(y, p) = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (6)$$

Raiz do erro quadrático médio (RMSE): a raiz quadrada do MSE, que fornece uma medida do erro médio na mesma unidade da variável dependente.

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

Erro percentual absoluto médio (MAPE): a média das diferenças percentuais absolutas entre as previsões do modelo e os valores reais.

$$MAPE(y, p) = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - p_i}{y_i} \right| \quad (8)$$

2.6 TRABALHOS RELACIONADOS

Um exemplo de trabalho de pesquisa no assunto é o artigo (MEKA,R.;ALAEDDINI, A.;BHAGANAGAR, K., 2020). apresenta um modelo robusto de aprendizado profundo para previsão de curto prazo da geração de energia eólica em um parque eólico em escala real. O modelo foi desenvolvido com base em redes convolucionais temporais, que capturam as dinâmicas temporais da geração de energia eólica e as relações entre as variáveis meteorológicas locais. Os autores utilizaram um método de ajuste de matriz ortogonal para otimizar os hiperparâmetros do modelo proposto e validaram o modelo utilizando dados de um parque eólico de 130 MW com 86 turbinas eólicas. Os resultados mostraram que o

modelo proposto superou alguns dos métodos existentes na literatura. O artigo destaca a importância dos modelos de aprendizado profundo para melhorar a previsão da geração de energia eólica em parques eólicos, reduzindo assim os riscos técnicos e econômicos associados à integração de energia eólica no sistema de fornecimento de eletricidade.

A previsão de curto prazo (inferior a 1 hora) da energia gerada por turbinas eólicas em um parque eólico é extremamente desafiadora devido à falta de dados confiáveis de torres meteorológicas e modelos numéricos de previsão do tempo em escalas de tempo curtas. Um modelo robusto de aprendizado profundo foi desenvolvido para previsões de curto prazo da energia gerada por turbinas eólicas em um parque eólico, usando as mais recentes redes convolucionais temporais (TCN) para capturar simultaneamente a dinâmica temporal da energia gerada pela turbina eólica e a relação entre as variáveis meteorológicas locais. Um método de ajuste de matriz ortogonal baseado no projeto de experimentos de Taguchi é utilizado para otimizar os hiperparâmetros do modelo TCN proposto. O método de ajuste de matriz ortogonal de Taguchi é uma técnica utilizada para otimizar os hiperparâmetros de modelos de aprendizado de máquina. Ele é baseado na matriz de Taguchi, que é uma matriz ortogonal que permite a identificação de todos os possíveis níveis de um conjunto de fatores de controle. O método de ajuste de matriz ortogonal de Taguchi consiste em determinar o conjunto ideal de valores de hiperparâmetros que levam a um desempenho ótimo do modelo (TAGUCHI, G. S, 1987).

3 METODOLOGIA

Este capítulo aborda a metodologia utilizada neste trabalho, incluindo a seleção da base de dados, suas características, o pré-processamento dos dados, os ajustes necessários para cada modelo, a divisão de dados, a implementação de diferentes modelos de previsão de velocidade do vento para o sistema de energia eólica A Figura 11 exemplifica a metodologia utilizada no trabalho.

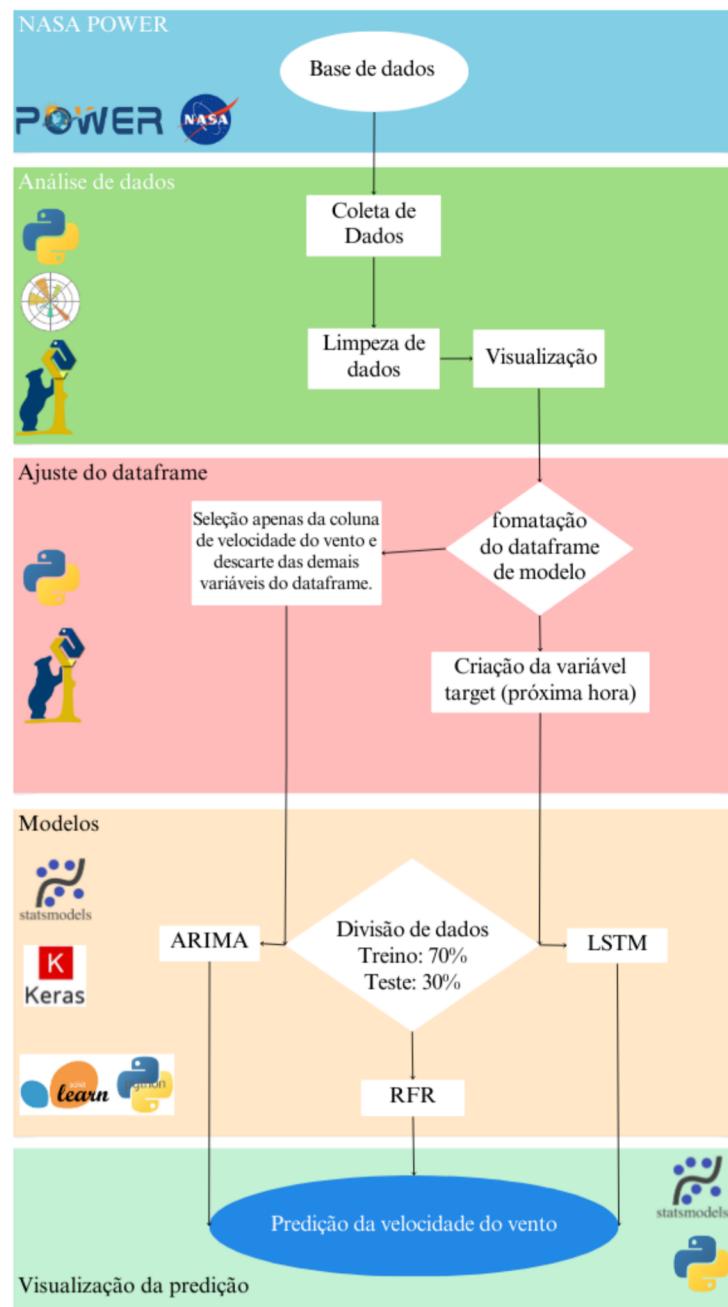


Figura 11 - Fluxograma da metodologia.

3.1 Bases de Dados

Este trabalho utilizou os dados disponibilizados pelo projeto NASA POWER, o qual é um programa de pesquisa da NASA voltado para as Ciências da Terra, sistemas de satélite e estudos que oferecem informações cruciais para a compreensão do clima e dos processos climáticos. Esses dados consistem em estimativas climatológicas de longo prazo de quantidades meteorológicas e fluxos de energia solar na superfície. Além disso, os valores diários médios de dados básicos de meteorologia e energia solar são disponibilizados em formato de séries temporais. Esses produtos são baseados em satélites e modelos que apresentam alta precisão e confiabilidade, permitindo obter informações precisas de recursos solares e meteorológicos em regiões com poucas ou nenhuma medição de superfície disponível (NASA POWER , 2020). Com esta tecnologia é possível obter dados meteorológicos pontuais em qualquer localização do planeta, o que se torna uma ferramenta chave para pesquisas de energia renováveis. Permite baixar informações meteorológicas diárias e horárias em vários formatos que podem ser usados em pesquisa, por exemplo para este trabalho foi usado o formato CSV (Comma-separated values), pelo fato de ser um formato que é fácil de se trabalhar tanto no Pandas quanto de editar no Excel. A Figura 12 mostra a interface do site, com alguns parâmetros selecionados, justamente nesta imagem está o ponto offshore que é estudado neste trabalho (STACKHOUSE, 2020).

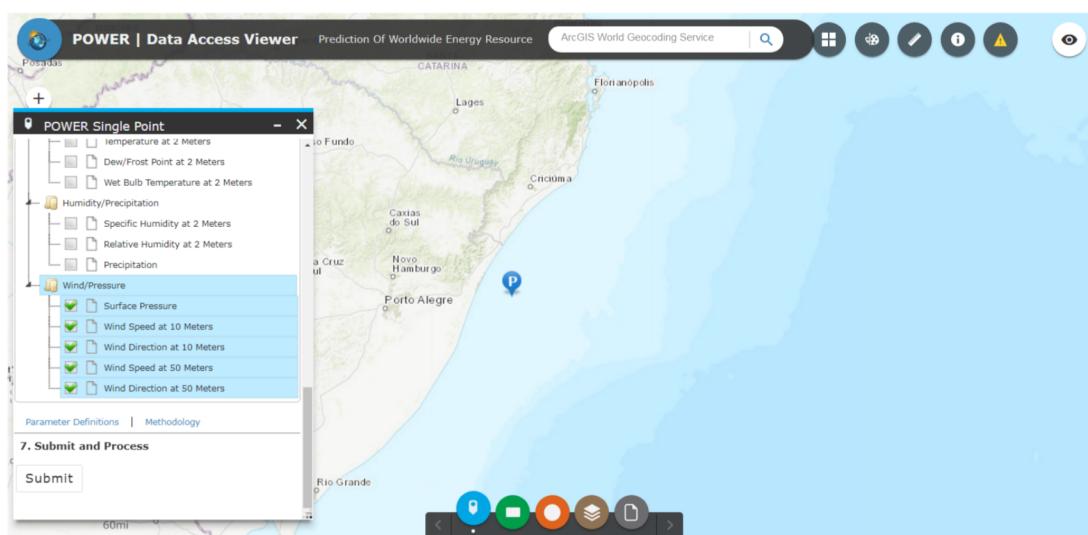


Figura 12 - Interface do POWER | Data Access Viewer.
Fonte: Adaptada de (POWER NASA, 2023).

Os dados meteorológicos são obtidos através de um processo chamado reanálise. Reanálise na meteorologia de dados é um processo que envolve o uso de modelos matemáticos para "reconstruir" o histórico do clima em uma região específica, a partir de dados observacionais e outras informações relevantes, como medições de temperatura, pressão atmosférica, umidade do ar e vento. O objetivo da reanálise é produzir conjuntos de dados meteorológicos coerentes, consistentes e precisos, que podem ser usados para estudar as variações climáticas

em longo prazo, identificar tendências e anomalias climáticas, e desenvolver previsões climáticas sazonais.

A reanálise geralmente é realizada em escala global e envolve o uso de modelos numéricos de simulação climática, que levam em conta a dinâmica atmosférica, oceânica e terrestre, bem como os processos de transferência de calor e umidade entre esses componentes. Esses modelos são alimentados com dados observacionais históricos, tais como medições de satélite, estações meteorológicas, bóias oceânicas, balões meteorológicos e outros dados, que são assimilados para criar um conjunto completo de dados climáticos para uma determinada região ou período (American Meteorological Society, 2017).

O sistema de reanálise utilizado pela NASA e que disponibiliza os dados utilizados neste trabalho é o MERRA-2 (Retrospectiva da Era Moderna para Pesquisa e Aplicações versão 2) e foi realizada para fornecer uma substituição oportuna para o MERRA e para sustentar o compromisso do GMAO (Global Modeling and Assimilation, Escritório de Modelagem e Assimilação Global da NASA) de ter uma análise climática contínua quase em tempo real. Ele utiliza dados observacionais de satélites, aviões, navios e estações meteorológicas terrestres para reconstruir a história do clima e do ambiente terrestre ao longo de várias décadas. Os modelos numéricos são utilizados para preencher as áreas onde não existem dados observacionais diretos. Esses modelos são complexos sistemas de equações matemáticas que simulam o comportamento da atmosfera e do sistema terrestre. Eles levam em conta princípios físicos, leis da termodinâmica, dinâmica dos fluidos e outros fatores para estimar os valores dos parâmetros atmosféricos em pontos específicos (American Meteorological Society, 2017). Um ponto importante para o uso desta fonte de dados para a predição de ventos, é que os dados relacionados a ventos são disponibilizados em duas alturas a 50 metros e a 10 metros de altitude (GELARO, 2017).

3.2 Pré-processamento da Base de Dados

Foi coletado dados horários meteorológicos em três pontos para esta pesquisa sendo estes a localização do parque eólico de Osório, o segundo a localização do parque eólico ventos da Bahia e um terceiro ponto offshore localizado na mesma linha de longitude do Parque Eólico de Osório porém no oceano próximo à costa.

Os dados são retirados do NASA/MERRA2 em resolução horária pelo período de 01/01/2021 a 01/01/2023. O valor para dados de origem ausentes, que não podem ser computados ou estão fora do intervalo de disponibilidade de fontes é -999, desta forma foi necessária a substituição deste valor numérico por NaN (*not a number*), que é reconhecido pelo algoritmo como a falta de dados e desta forma não atrapalha na sumarização do *dataset* podendo então eliminar de forma coerente as lacunas no banco de dados. A Tabela 1 descreve cada parâmetro usado no trabalho e a tabela 2 exibe os dados geográficos sobre o local onde foi feita a análise.

Parâmetro	Fonte	Nome	Descrição
WS10M	MERRA-2	Velocidade do vento a 10 metros	A direção do vento a 10 metros acima da superfície da terra.
WD10M	MERRA-2	Direção do vento a 10 metros	A velocidade do vento a 10 metros acima da superfície da terra.
WD50M	MERRA-2	Direção do vento a 50 metros	A direção do vento a 50 metros acima da superfície da terra.
WS50M	MERRA-2	Velocidade do vento a 50 metros	A velocidade do vento a 50 metros acima da superfície da terra.
T2M	MERRA-2	Temperatura a 2 metros de altura	A temperatura do ar (bulbo seco) a 2 metros acima da superfície da terra.
PS	MERRA-2	Pressão de superfície	A pressão superficial na superfície da Terra.
QV2M	MERRA-2	Umidade específica a 2 metros	A relação entre a massa de vapor de água e a massa total de ar a 2 metros (kg de água/kg de ar total).
RH2M	MERRA-2	Umidade relativa a 2 metros	A relação entre a pressão parcial real do vapor de água e a pressão parcial na saturação, expressa em porcentagem.

Tabela 1 - Descrição de parâmetros.

Local	Latitude	Longitude
Parque Eólico de Osório	-29,9118°	-50,2864°
Parque Eólico Ventos da Bahia	-11,9999°	-41,4455°
Ponto Offshore	-29,9161°	-49.60°

Tabela 2 - Descrição de pontos analisados.

Foi utilizada a biblioteca *missingno* para tratar e analisar a falta de dados em certos horários, algo que é esperado em qualquer *dataset* de medições. Ao usar a ferramenta *matrix()* dessa biblioteca, é possível verificar uma relação entre as colunas, sendo observada a falta de um parâmetro com a falta de todos os outros na mesma hora, exceto pelas colunas que se referem à irradiância solar. E, essa discrepância levou à decisão de não utilizar a irradiância como entrada. A visualização dos dados faltantes auxilia na tomada de decisão para excluir parâmetros relacionados à irradiância solar, e excluir as poucas linhas com dados faltantes. A Figura 13 exibe um gráfico, gerado pela ferramenta *matrix()* da biblioteca *missingno*, que mostra em branco os dados faltantes nas colunas.

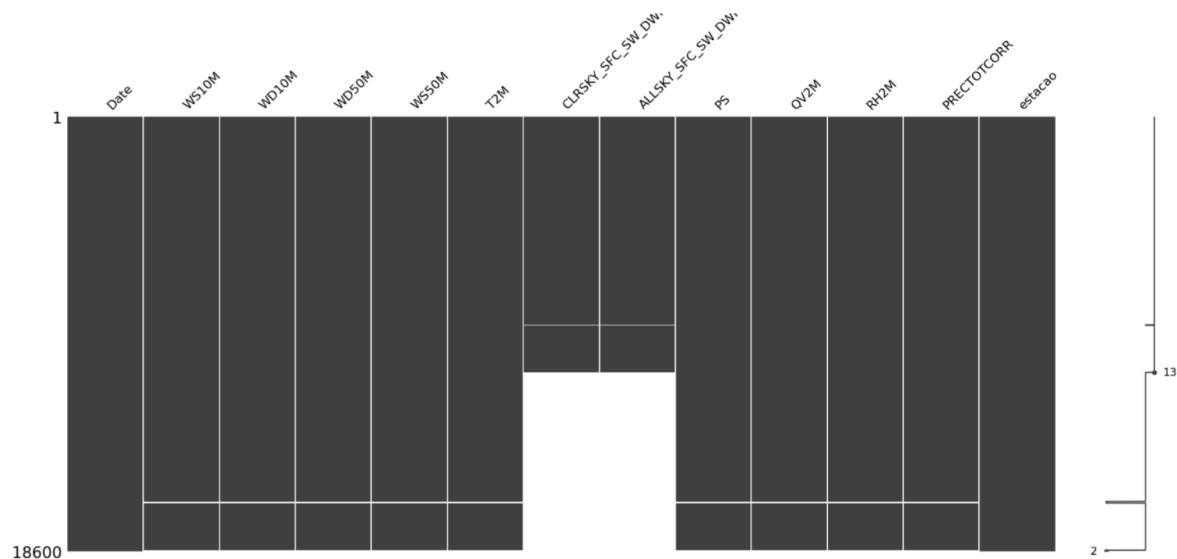


Figura 13 - Gráfico de dados faltantes nas colunas.

3.3 Visualização

Para melhor compreensão dos dados devemos obter uma visualização de como se comporta o vento naquele ponto durante o período em questão. Conforme já foi relatado, o modelo foi testado para três pontos, e abaixo é apresentado o comportamento do vento na região do parque eólico de Osório utilizando da biblioteca WindRose, Matplotlib e Seaborn.

Parque Eólico de Osório:

Com a Figura 14 é possível observar como os valores de velocidade de ventos se distribuem, é notável que para a região do parque de Osório a velocidade de vento tem uma tendência de ficar entre 4 e 7,5 metros por segundo.

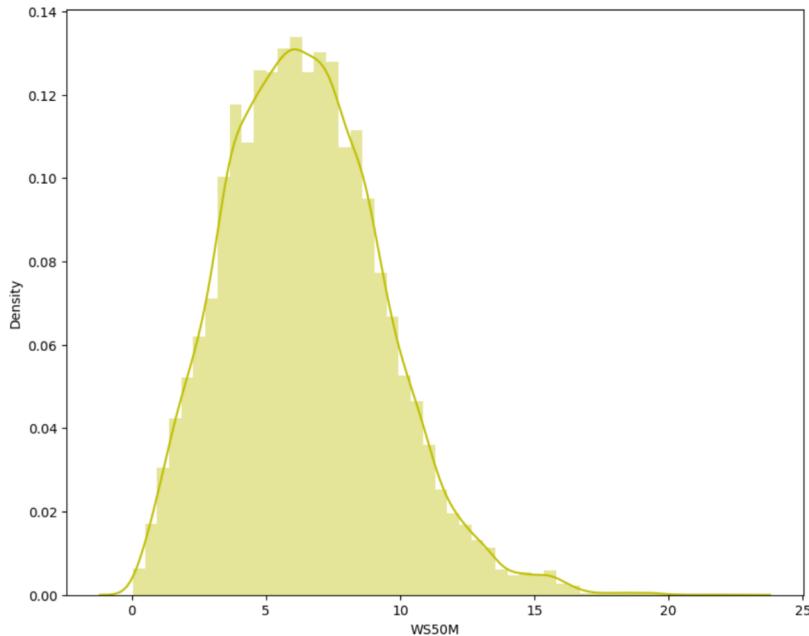


Figura 14 - Distribuição de velocidade de vento (m/s) no Parque Eólico de Osório.

A velocidade de vento pode mudar conforme as estações do ano, assim visualizar como a velocidade do vento vem mudando durante os meses, traz uma intimidade com os dados tratados, e melhora a tomada de decisão na hora de escolher a tecnologia do modelo que melhor irá se adaptar ao trabalho de predição. A Figura 15 mostra um *boxplot* da velocidade de vento nos meses como se estuda dois anos isto aparece de forma agrupada e generalizada, já a Figura 16 apresenta um *heatmap* da mediana dos valores de velocidade de vento de cada mês dos dois anos analisados.

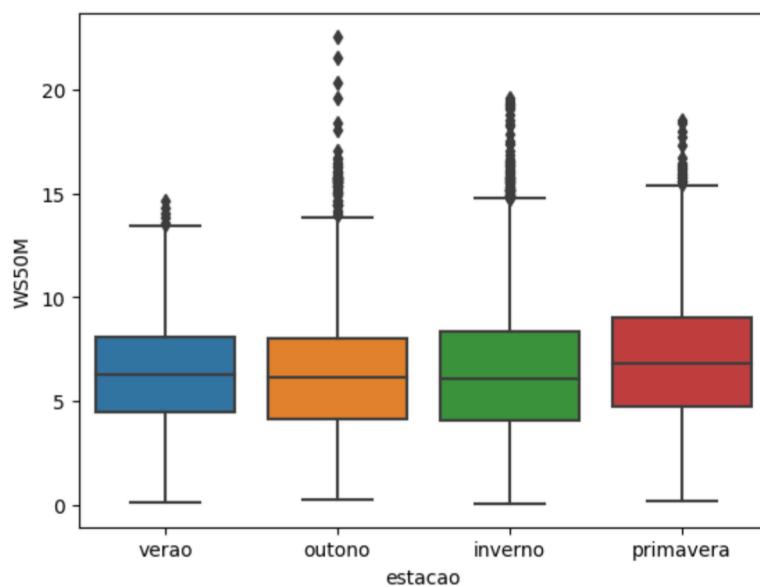


Figura 15 -Boxplot da velocidade de vento (m/s) pelas estações do ano.

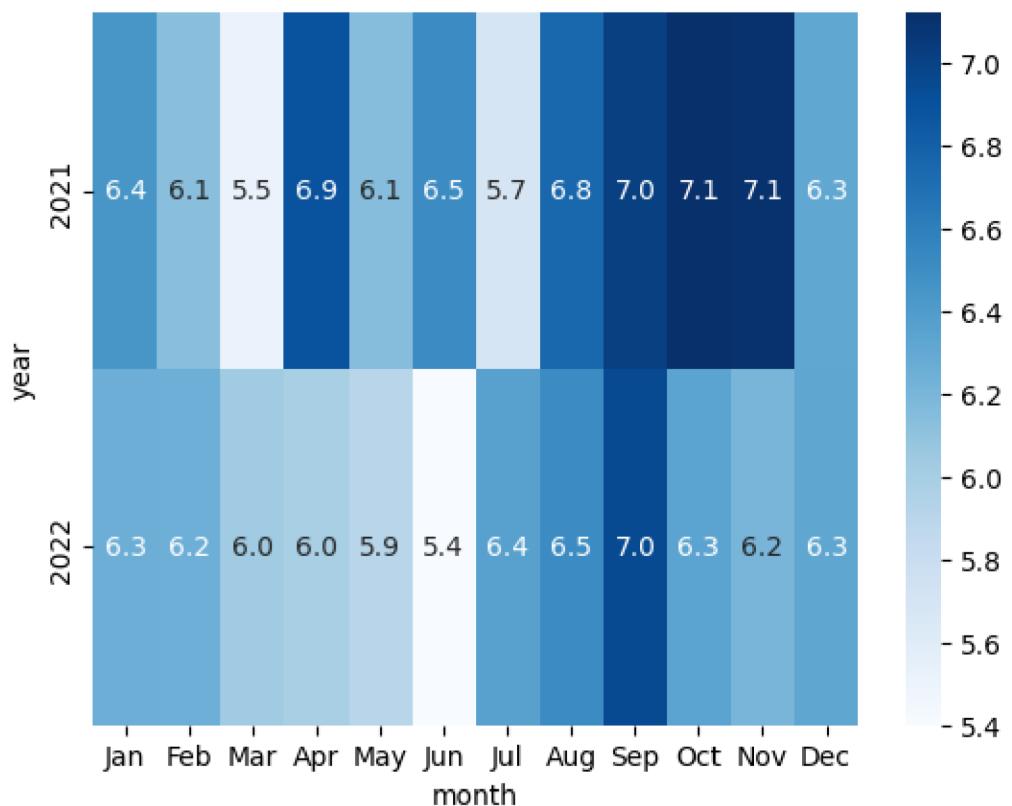


Figura 16 -Heat map de velocidade de ventos (m/s) no ano de 2021 e 2022.

Outro parâmetro de vento disponível pelo banco de dados além da velocidade, é a direção do vento. No MERRA-2, a direção do vento é referenciada em relação ao norte geográfico. O norte geográfico é definido como 0° ou 360° , o leste como 90° , o sul como 180° e o oeste como 270° . Este parâmetro pode ser visualizado por uma Rosa dos ventos, mostrando a direção da velocidade dos ventos neste período, que é mostrado na Figura 17.

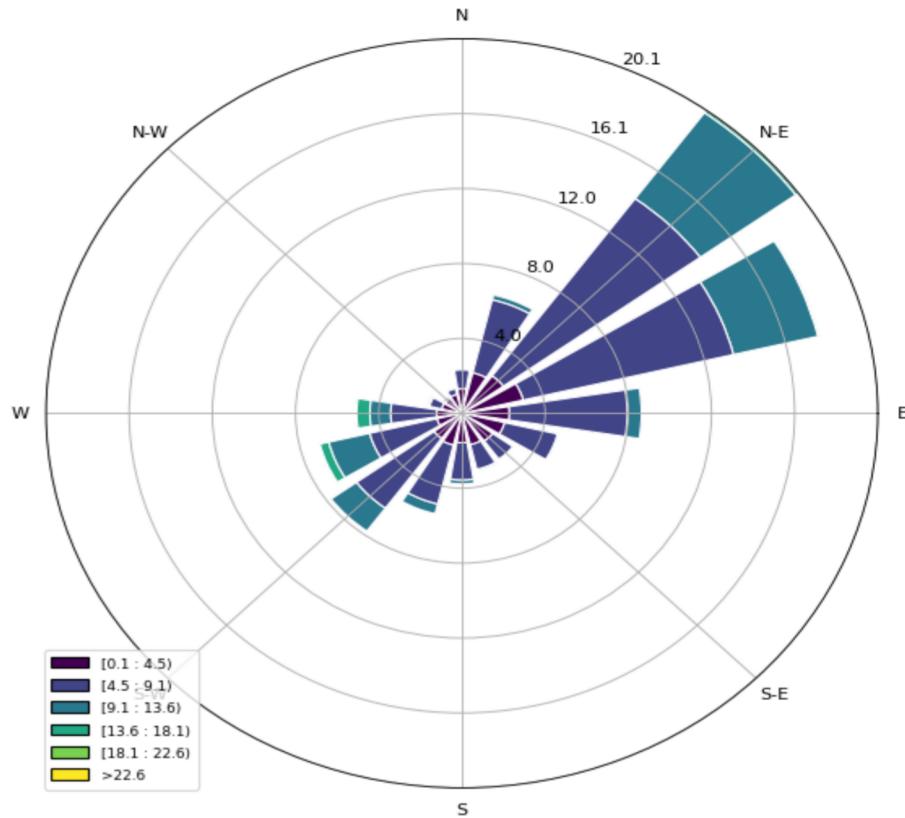


Figura 17 -Rosa dos ventos mostrando a direção da velocidade dos ventos.

A visualização da autocorrelação dos dados é de extrema importância antes de utilizar modelos para a predição. A autocorrelação permite identificar se existe algum padrão ou dependência temporal nos dados, ou seja, se os valores atuais estão correlacionados com os valores passados. Ao analisar a autocorrelação, é possível identificar se há uma tendência ou sazonalidade nos dados, o que pode afetar significativamente a capacidade de previsão dos modelos. Além disso, a visualização da autocorrelação ajuda a escolher a ordem adequada dos modelos, como o ARIMA, que levam em consideração as correlações passadas. A Figura 18 demonstra um gráfico resultante que fornece insights sobre a estrutura de autocorrelação dos dados em diferentes ordens de diferenciação.

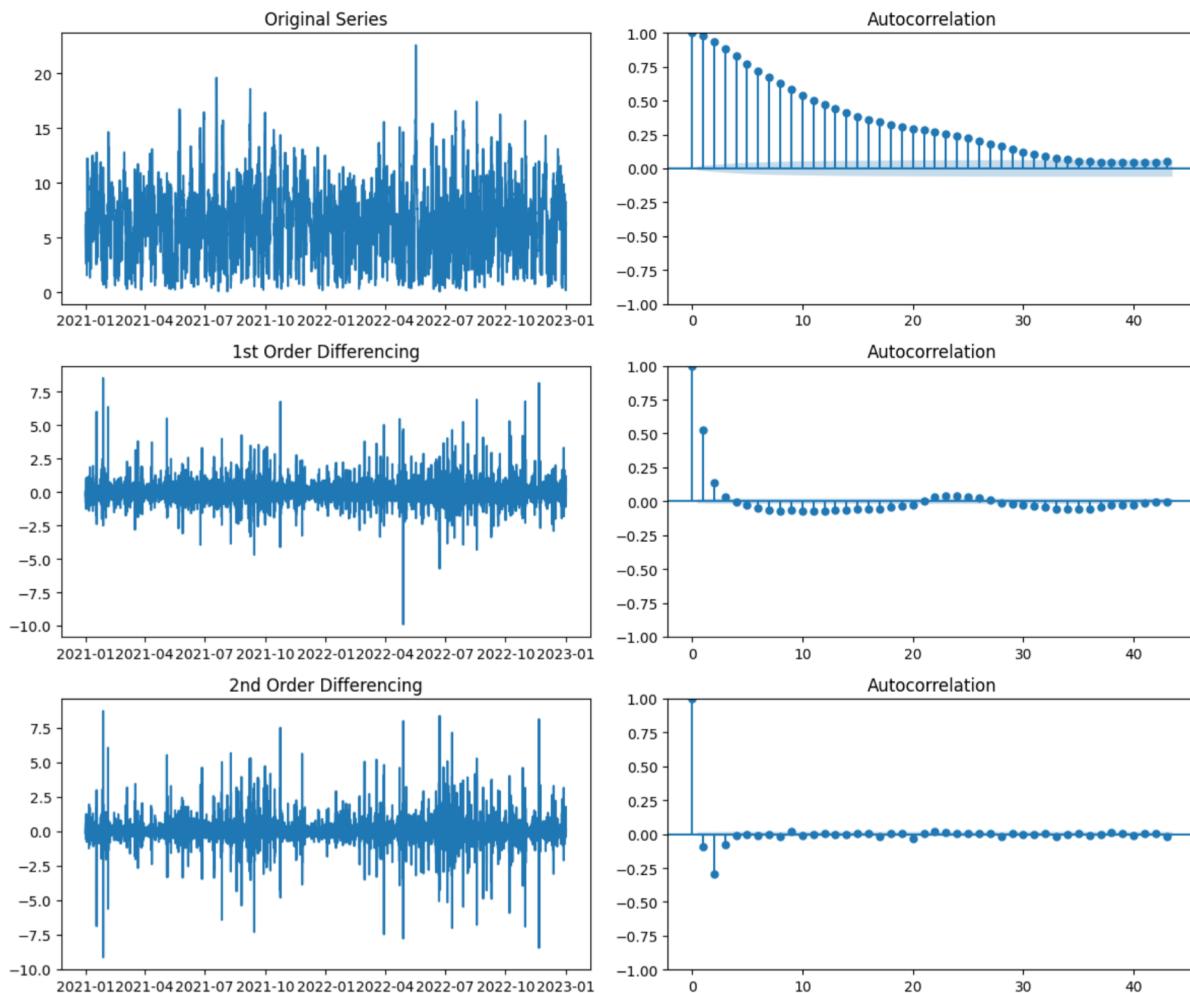


Figura 18 - Gráfico de autocorrelação da velocidade de vento.

3.4 Modelos de predição de velocidade de vento.

Esta seção trata da metodologia dos modelos de predição implementados no trabalho. Com o *dataframe* formatado para ser treinado com seu devido modelo. O primeiro sendo o modelo Random Forest Regressor de aprendizado de máquina, o segundo utilizando o algoritmo ARIMA com a validação Walk-forward e o último é o modelo que utiliza rede neurais recorrentes LSTM, todos os modelos foram treinados e testados para os dados com período de uma hora. Foi utilizado para ambos os modelos a mesma separação de dados para treinamento e para teste, onde, aproximadamente, 70% dos dados destes dois anos foi usado para treinamento e 30% para teste. Foram utilizados outros modelos porém com resultado muito abaixo do esperado. Isso se deve porque os métodos não se mostraram adequados ao tipo de dado em análise. Um dos métodos testados foi o Xgboost porém não conseguiu chegar a bons resultados para esta predição de ventos.

3.4.1 Validação Walk-Forward.

A validação walk-forward (ou validação avanço temporal) é uma técnica utilizada para avaliar a performance de modelos de previsão de séries temporais. A ideia básica é simular o processo de previsão em tempo real, onde o modelo é atualizado a cada nova observação disponível.

Essa técnica ajuda a avaliar a capacidade do modelo de se adaptar a mudanças ao longo do tempo, uma vez que é atualizado a cada nova observação, e também ajuda a evitar o problema de "vazamento de informação" (*data leakage*), onde o modelo é testado com dados futuros que ele já "viu" durante o treinamento (VALENTE, 2022).

3.4.2 Random Forest Regressor

O Random Forest Regressor foi implementado por meio da biblioteca sklearn, uma popular e amplamente utilizada biblioteca em trabalhos de machine learning na atualidade. Foi criada uma variável target que é a velocidade do vento porém deslocada no tempo, sendo que esta será a velocidade do vento na próxima hora, assim utilizamos todos os dados da hora atual, inclusive a velocidade atual, para prever esta variável como mostrado na Figura 19. Foi separado 70% dos dados para treino e 30% para teste, para essa separação foi utilizada a função `train_test_split()` da biblioteca sklearn. Foi testado e escolhido o número de 1000 árvores para a predição do modelo, oferecendo um bom resultado sem muita carga computacional. As previsões e resultados geram valores R² e RMSE que serão apresentados na seção 4. A Figura 20 mostra um gráfico de dispersão para avaliarmos os valores reais e os preditos pelo modelo.

Date	WS10M	WD10M	WD50M	WS50M	T2M	PS	QV2M	RH2M	wind_speed50	target
2021-01-01 14:00:00	3.50	161.93	164.58	3.53	23.56	101.22	10.44	57.88	3.53	3.74
2021-01-01 15:00:00	3.69	150.45	152.20	3.74	23.50	101.21	10.44	58.19	3.74	3.80
2021-01-01 16:00:00	3.71	137.91	138.67	3.80	23.34	101.21	10.50	59.12	3.80	3.79
2021-01-01 17:00:00	3.59	124.10	124.08	3.79	23.10	101.23	10.56	60.50	3.79	3.89

Figura 19 - Entradas de dados e variável target a ser predita pelo modelo.

3.4.3 ARIMA Walk-Forward.

O modelo ARIMA com a validação walk-forward foi utilizado, que é o indicado para séries temporais, o modelo ARIMA necessita da escolha de três parâmetros para sua execução, assim foi usado a ferramenta `auto_arima` da biblioteca `pmdarima` para a estimativa dos

parâmetros do modelo. Feito isso, foi executado o modelo com apenas 10% dos dados mais recentes, o que dá próximo dos últimos dois meses, foi feito isto por questões de capacidade computacional, assim como no primeiro modelo foi separado 70% para treino e 30% para teste.

3.4.4 LSTM.

Utilizamos a Biblioteca Keras que é bem popular para construção de redes neurais e aprendizado de máquinas em Python. O modelo LSTM é definido usando a classe Sequential do Keras, que permite empilhar camadas de forma sequencial. No modelo, é adicionada uma camada LSTM com 50 unidades de memória, pois problemas mais complexos, como previsão de séries temporais altamente voláteis, geralmente requer mais unidades para capturar padrões complexos. A função de ativação ReLU(Rectified Linear Unit) que é computacionalmente eficiente de calcular em comparação com algumas outras funções de ativação, como a função sigmóide e tangente hiperbólica, e uma camada de saída densa com uma única unidade, utilizando de 100 épocas para o treino do modelo.

3.4.4.2 LSTM com regularização Dropout

A regularização Dropout é uma técnica de regularização comumente usada para redes neurais que ajuda a prevenir o overfitting (sobreajuste). Ela envolve aleatoriamente "desligar" (zerar) algumas unidades (neurônios) da rede durante o treinamento, com uma probabilidade específica para cada unidade. Isso ajuda a rede a aprender padrões mais robustos e gerais, em vez de simplesmente memorizar o conjunto de treinamento (HOCHREITER, 1997).

Em uma rede LSTM com regularização Dropout, a técnica de Dropout é aplicada às conexões entre as unidades de uma camada LSTM. Isso significa que, durante o treinamento, algumas conexões são aleatoriamente "desligadas", ajudando a prevenir o overfitting. Isso pode ser particularmente útil em problemas de aprendizado de sequências, no qual as redes LSTM podem ser propensas a overfitting devido à complexidade e dependência temporal dos dados.

3.4.4.3 LSTM com múltiplas camadas e regularização L2

LSTM com múltiplas camadas é uma técnica de aprendizado profundo usada para lidar com séries temporais, nas quais várias camadas de unidades LSTM são empilhadas para melhorar o desempenho do modelo. A regularização L2 é uma técnica de regularização que adiciona uma penalidade na função de perda do modelo, a fim de evitar o overfitting. Juntas, essas técnicas permitem que o modelo aprenda padrões complexos em séries temporais e evite o overfitting, aumentando assim sua capacidade de generalização para dados não vistos (LUO,2017).

Para concluir este capítulo, com uma aplicação pragmática para a indústria de energia eólica. Pode-se fazer a predição com as duas altitudes disponíveis pela fonte de dados e com o auxílio do cálculo de coeficiente de Hellmann, pode-se encontrar a velocidade do vento para a altura de qualquer Aerogerador.

4 RESULTADOS

Esta seção apresenta resultados e análises da metodologia neste trabalho, a Figura 20 mostra um gráfico de dispersão entre os valores reais e os valores preditos pelos modelos. E para complementar o entendimento da análise de valores reais versus valores previstos, é apresentado na Figura 21 um gráfico Valor Real vs Valor Previsto do resultado do modelo RFR para as três regiões, levando em consideração os últimos 100 intervalos de tempo em relação à data final do *dataset*.

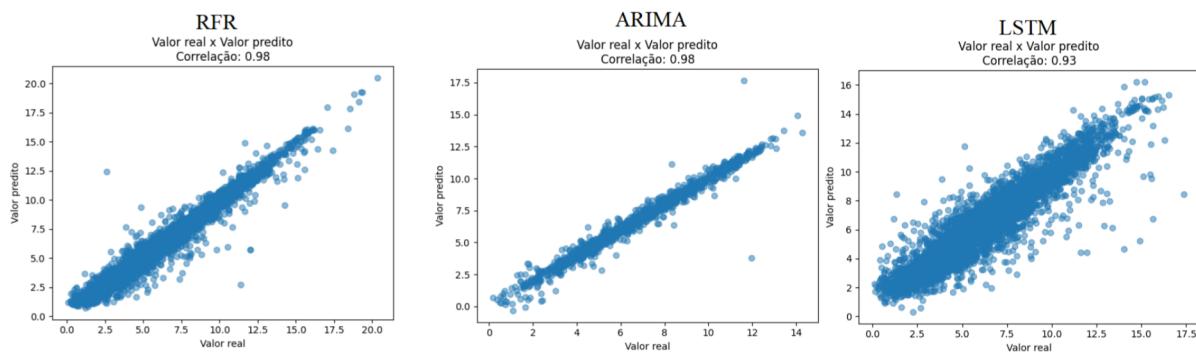


Figura 20- Dispersão entre Valores reais e valores preditos nos modelos.
Fonte: Autor.

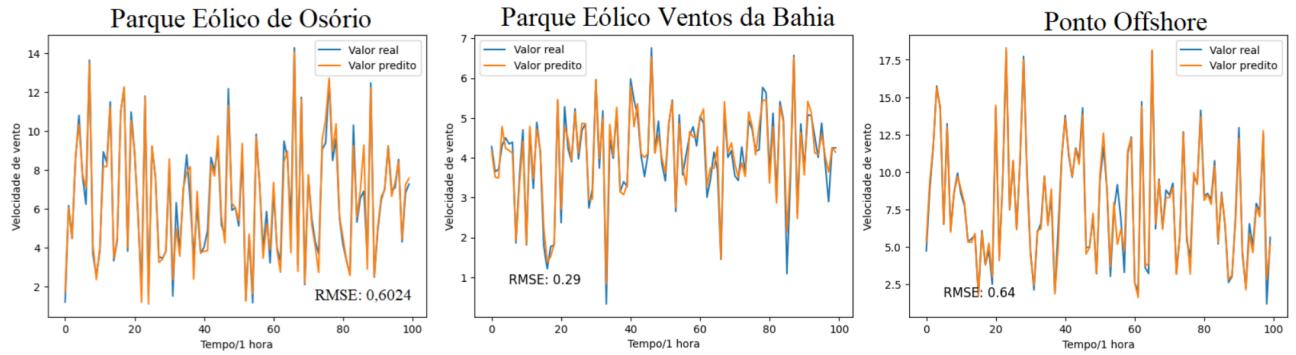


Figura 21- Velocidade de vento Reais vs Previstos para o modelo RFR nas três regiões.

Fonte: Autor.

4.1 Métricas de avaliação.

Na avaliação os modelos utilizam-se métricas que dependem do contexto e do tipo de dados. Neste trabalho é considerado duas métricas que são mais adequadas para o nosso problema que é uma série temporal.

As métricas de avaliação de séries temporais permitem medir o desempenho dos modelos de previsão e ajuste em relação aos dados reais. O coeficiente de determinação (R^2) indica a proporção da variabilidade dos dados explicada pelo modelo, sendo valores mais próximos de 1 indicativo de um melhor ajuste. O erro médio absoluto (MAE) representa a média das diferenças absolutas entre os valores previstos e os valores reais, fornecendo uma medida direta do erro médio cometido pelo modelo, independentemente da direção do erro. O erro médio quadrático (MSE) é a média dos erros ao quadrado, dando mais peso aos erros maiores, e valores mais baixos indicam um melhor ajuste. A raiz quadrada do MSE (RMSE) mede o desvio padrão dos resíduos do modelo, representando o erro típico em relação aos dados reais. O erro percentual absoluto médio (MAPE) é uma métrica de erro relativo que mede a média percentual das diferenças absolutas em relação aos valores reais, e valores menores de MAPE indicam uma maior precisão do modelo. Essas métricas são valiosas para avaliar o desempenho dos modelos de séries temporais e auxiliar na seleção e aprimoramento dos mesmos. Além disso, o RMSE é uma métrica mais adequada para avaliar modelos de série temporal, pois leva em consideração a estrutura temporal dos dados.

Portanto, é recomendado que sejam usadas ambas as métricas na avaliação de modelos de série temporal, sendo o R^2 uma métrica para avaliar a qualidade do ajuste do modelo e o RMSE para avaliar a qualidade das previsões do modelo (MAHMOOD, 2018).

4.2 Baseline.

É necessário primeiramente adotar um limiar para comparar os resultados dos modelos, portanto uma predição ingênua é útil para avaliar se os resultados dos modelos são consideráveis, e se o modelo está prevendo bem, este é o *baseline*. A ideia é usar uma técnica simples, como a média móvel dos últimos 7 dias, como uma previsão ingênua para comparar com as previsões de modelos mais avançados. Se os modelos mais complexos não conseguirem superar a performance dessa previsão simples, então é provável que eles não sejam úteis na prática.

4.3 Tabela de Resultados

Os resultados obtidos no trabalho utilizando os métodos proposto, em todos os modelos foram usados os mesmo dados, o *dataset* foi tratado e reformulado para ser treinado em todos estes modelos, particularmente o modelo ARIMA utiliza apena um parâmetro que é o que desejamos predizer, e por questões práticas de capacidade computacional, neste modelo foi usado apenas 10% dos dados do *dataset*, mesmo assim isto representa 1752 horas de medições de velocidade do vento, ainda sim com esta particularidade em questão foi possível alcançar um expressivo resultado para este modelo, junto com o modelo Random Forest Regressor que também se mostrou muito eficiente. A tabela 3 mostra estes resultados e também inclui os valores de base do *baseline* para comparação com os resultados dos modelos.

Modelo	R ²	MAE	MSE	RMSE	MAPE
Baseline	0,6224	0,5534	0,4985	0,7060	0,1679
Random Forest Regressor	0,9584	0,3962	0,3628	0,6024	0,1009
ARIMA WALK-FORWARD	0,9670	0,2806	0,2510	0,5010	0,0625
LSTM Básico	0,9515	0,4150	0,3889	0,6236	0,0955
LSTM múltiplas camadas e regularização L2	0,8831	0,6516	0,8521	0,9231	0,1499
LSTM com regularização Dropout	0,8710	0,6616	0,8770	0,9365	0,1818

Tabela 3 - Resultados usando modelos de predição para o Parque Eólico de Osório.

A Tabela 4 apresenta os resultados de diferentes modelos de previsão, incluindo *Baseline*, Random Forest Regressor, ARIMA Walk-Forward, LSTM Básico, LSTM Múltiplas Camadas com Regularização L2 e LSTM com Regularização Dropout. Ao analisar os valores de R², observamos que os modelos ARIMA Walk-Forward, Random Forest Regressor e LSTM Básico obtiveram resultados significativamente superiores ao *Baseline*. Isso indica que esses modelos foram capazes de explicar uma proporção maior da variabilidade dos dados. No que diz respeito às métricas de erro, como o MAE, MSE e RMSE, os modelos ARIMA Walk-Forward e Random Forest Regressor mostraram um desempenho melhor em

comparação com o *Baseline*, enquanto o LSTM Básico apresentou resultados mais elevados nessas métricas.

Além disso, é importante destacar que o modelo ARIMA Walk-Forward obteve valores bastante baixos para o MAPE, indicando uma menor discrepância percentual entre as previsões e os valores reais. Em termos gerais, os resultados indicam que os modelos ARIMA Walk-Forward e Random Forest Regressor se destacaram na previsão, oferecendo uma melhor explicação da variabilidade dos dados e menores erros absolutos e percentuais em comparação com o *Baseline*.

A Tabela 4 exemplifica a predição de velocidade de vento, o valor real e o valor predito pelo modelo é apresentado e pode-se acompanhar o quanto próximo a predição de cada modelo pode se aproximar na prática.

RFR		ARIMA		LSTM	
Valor real	Valor Previsto	Valor real	Valor Previsto	Valor real	Valor Previsto
5,77	6,45	6,55	6,47	5,77	6,65
9,44	9,44	6,81	6,66	9,44	9,01
6,05	5,77	6,64	6,95	6,05	6,20
12,80	12,36	6,76	6,48	12,80	12,62
1,35	1,43	7,37	6,90	1,35	1,33
...
7,07	7,35	5,86	6,08	7,07	7,05
8,53	8,41	6,28	6,07	8,53	8,01
4,29	4,43	6,00	5,98	4,29	4,31
6,82	7,24	5,94	6,07	6,82	7,03
7,26	7,55	6,14	6,32	7,26	7,42

Tabela 4 - Valor real e predição de velocidade de ventos para o Parque Eólico de Osório.

4.8 Comparação com outras regiões

Foi abordado outras duas regiões, com a mesma metodologia para avaliar seus resultados e verificar se o modelo funciona para regiões diferentes com parâmetros de valores diferentes. As regiões em questão são além do Parque eólico de Osório, o Parque eólico Ventos da Bahia e um ponto offshore na mesma linha de longitude do Parque de Osório. As Figuras a seguir mostram as características dos comportamentos do vento nas três regiões, a Figura 22 exibe uma rosa dos ventos, onde indica a direção média com maior intensidade de vento para a região. A Figura 23 permite visualizar um heatmap da mediana de velocidade de vento nas duas regiões que foram incluídas no trabalho, o heatmap de ventos do parque de Osório já foi

mostrado na Figura 15.

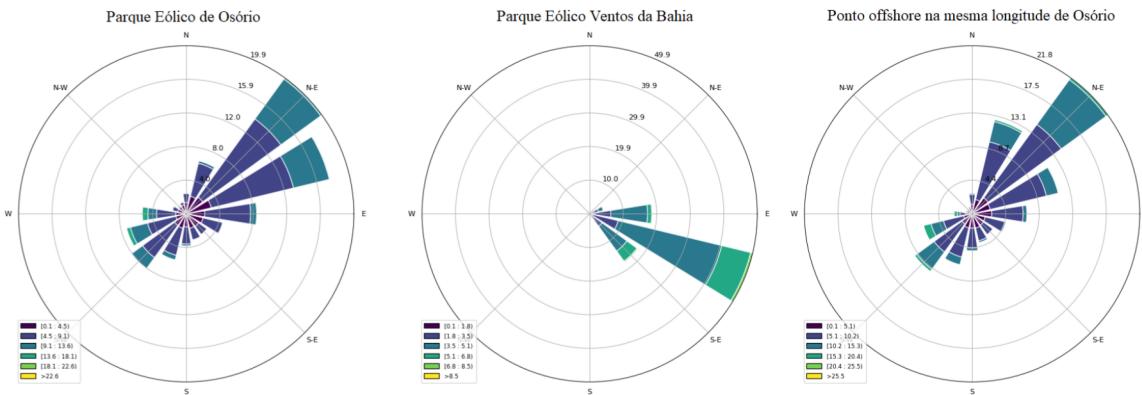


Figura 22 - Direção dos ventos nos três pontos estudados.

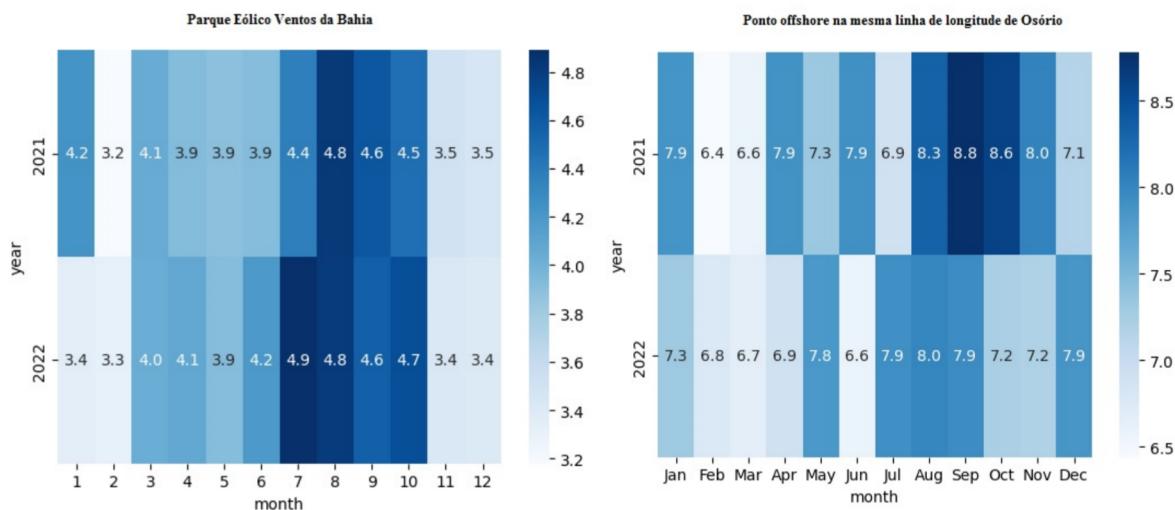


Figura 23 -Heat map de velocidade de ventos no ano de 2021 e 2022 nos outros dois pontos analisados.

Foi utilizado o mesmo método e dados da mesma fonte para esta comparação de regiões, isto é, foi coletado os mesmos tipos de parâmetros, tratado da mesma maneira os dados e colocados no mesmo formato para serem treinados, posteriormente foram usados os mesmos modelos em todas as regiões. O modelo RFR apresentou um resultado favorável, e em média teve tempo de processamento de 50 segundos, já o ARIMA por melhor que foi seus resultados, tem a grande dependência de apenas de uma variável e teve o pior tempo de processamento em média 5 minutos e por fim o modelo LSTM teve seu tempo de processamento de em média 2 minutos. Assim foi tomada a decisão de utilizar o modelo RFR para comparar os resultados das regiões em questão. Para exemplificar, tem-se a Tabela 5 a seguir os resultados do modelo Random Forest Regressor, e suas devidas *baselines* de cada região.

Local	Método	R ²	MAE	MSE	RMSE	MAPE
Parque Eólico de Osório	Baseline	0,6224	0,5534	0,4985	0,7060	0,1679
	RFR	0,9584	0,3962	0,3628	0,6024	0,1009
Parque Eólico Ventos da Bahia	Baseline	0,6224	0,5534	0,4985	0,7060	0,1679
	RFR	0,9290	0,2256	0,0936	0,3059	0,0668
Ponto Offshore	Baseline	0,7799	1,2138	2,7610	1,6616	0,2580
	RFR	0,9685	0,4284	0,4078	0,6386	0,0986

Tabela 5- Resultados para as três regiões utilizando do modelo Random Forest Regressor.

A análise dos resultados obtidos a partir do modelo Random Forest Regressor aplicado em diferentes regiões, utilizando o mesmo conjunto de dados da mesma fonte e eliminando dados faltantes e inconsistentes de maneira uniforme, revela a capacidade de generalização desse modelo para prever o comportamento do vento em contextos diversos. As conclusões extraídas dos dados mostram que o modelo Random Forest Regressor superou consistentemente as *baselines* em todas as regiões consideradas.

Ao examinar o coeficiente de determinação (R^2), constatou-se que o modelo Random Forest Regressor foi capaz de explicar uma proporção maior da variabilidade dos dados em relação às *baselines* em todas as regiões. Isso sugere que o modelo capturou adequadamente os padrões e tendências do vento, independentemente das particularidades de cada região.

Além disso, as métricas de erro, como o erro médio absoluto (MAE), o erro médio quadrático (MSE), a raiz do erro médio quadrático (RMSE) e o erro percentual médio absoluto (MAPE), demonstraram consistentemente que o modelo Random Forest Regressor obteve resultados mais precisos em todas as regiões em comparação com as *baselines*. Essa maior precisão reflete-se em valores menores de erro absoluto e percentual, indicando uma previsão mais confiável dos valores de vento para cada região considerada.

Portanto, com base nessas análises, podemos concluir que o modelo Random Forest Regressor demonstrou sua capacidade de fornecer resultados favoráveis, sendo aplicável em diferentes regiões com comportamentos de vento distintos. O fato de ter utilizado o mesmo conjunto de dados, eliminando de forma consistente os dados faltantes e inconsistentes, bem como a adoção dos mesmos parâmetros meteorológicos como entrada do modelo em todas as regiões, fortalece a validade dos resultados obtidos. Essa generalização bem-sucedida sugere que o modelo Random Forest Regressor pode ser uma ferramenta promissora para a previsão precisa da geração eólica em diferentes regiões, considerando as características específicas do

vento em cada uma delas, a Tabela 6 nos traz um percentual de quanto o modelo está acima ou abaixo do *baseline*.

Local	R ² (%)	MAE (%)	MSE (%)	RMSE (%)	MAPE (%)
Parque Eólico de Osório	53,98	-28,41	-27,22	-14,67	-39,90
Parque Eólico Ventos da Bahia	49,26	-59,23	-81,22	-56,67	-60,21
Ponto Offshore	24,18	-64,71	-85,23	-61,57	-61,78

Tabela 6 - Porcentagem de quanto os resultados estão acima ou abaixo do Baseline.

Para concluir foi feita uma predição de velocidade de vento para energia eólica de 5 passos à frente no tempo, ou seja com 1,2,3,4 e 5 horas de antecedência, utilizando o modelo RFR. A Tabela 8 contém as métricas normalizadas de cada região.

Local	Parque Ventos da Bahia		Parque de Osório		Ponto Offshore	
	Horas	NRMSE	NMAE	NRMSE	NMAE	NRMSE
1	0,0758	0,0565	0,0962	0,0621	0,08361	0,0543
2	0,1193	0,0911	0,156	0,1053	0,1447	0,0962
3	0,1434	0,1121	0,1871	0,1329	0,1857	0,1287
4	0,1612	0,1259	0,213	0,1566	0,2149	0,1538
5	0,1681	0,1311	0,2333	0,1744	0,2353	0,1714

Tabela 7- Resultados para 5 passos de tempos a frente da previsão de velocidade de vento.

5 Conclusão

A previsão de velocidade de vento com uma hora de antecedência desempenha um papel crucial na programação adequada da geração eólica. Com base nessas previsões, os operadores do sistema elétrico possuem a capacidade de ajustar a operação das turbinas eólicas para otimizar a utilização dos recursos disponíveis. Por exemplo, se a previsão indicar um aumento na velocidade do vento e a ocorrência de rajadas mais intensas, os operadores podem aumentar a operação das turbinas para aproveitar ao máximo a energia gerada por essas condições favoráveis. Isso resulta em uma maior produção de energia renovável durante esses períodos. Da mesma forma, a previsão de velocidade de vento permite que os operadores antecipem períodos de vento fraco. Se a previsão indicar uma redução na velocidade do vento nos próximos períodos, os operadores podem reduzir a operação das turbinas ou até mesmo desligá-las temporariamente. Essa abordagem evita o desperdício de recursos e garante que a geração de energia seja direcionada de forma mais eficiente para atender à demanda.

Além disso, a previsão de velocidade de vento desempenha um papel fundamental na gestão da estabilidade da rede elétrica. Se a previsão indicar uma queda abrupta na velocidade do vento, os operadores podem se preparar para compensar essa redução de geração eólica ajustando a operação de outras fontes de energia, como usinas térmicas. Dessa forma, a estabilidade da rede é mantida, garantindo um fornecimento contínuo e confiável de energia elétrica. Por fim, a previsão de velocidade de vento também desempenha um papel importante no planejamento eficiente das atividades de manutenção. Com a previsão em mãos, os operadores podem identificar momentos de redução na velocidade do vento e aproveitá-los para realizar tarefas de manutenção em turbinas eólicas ou outros componentes do sistema. Isso minimiza o impacto na produção de energia durante esses períodos em que a geração já está reduzida devido às condições de vento. Dessa forma, a previsão de velocidade de vento com uma hora de antecedência proporciona aos operadores da geração eólica a capacidade de tomar decisões estratégicas. Essas decisões incluem ajustar a operação das turbinas, gerenciar a estabilidade da rede, otimizar o planejamento de manutenção e direcionar de forma mais eficiente a geração de energia. Essas práticas resultam em uma operação mais eficiente, maximizando a utilização de energia renovável e garantindo um fornecimento confiável de energia elétrica.

Considerando os resultados obtidos e os modelos utilizados para a previsão de velocidade de vento, é possível fazer um resumo das principais conclusões. Foram coletados dados de diferentes regiões, tratados de maneira consistente e utilizados os mesmos parâmetros meteorológicos como entrada nos modelos de previsão. A comparação dos resultados entre as regiões revelou que o modelo Random Forest Regressor apresentou um bom desempenho em todas as localidades. Em relação aos valores de R^2 , que indicam o quão bem o modelo se ajusta aos dados, observou-se uma melhoria significativa em relação à *baseline* em todas as regiões. Isso indica que o modelo foi capaz de capturar padrões e realizar previsões mais precisas em comparação com a simples extração dos dados históricos.

Além disso, ao analisar as métricas de erro, como MAE, MSE, RMSE e MAPE, observou-se uma redução nos valores em comparação com a *baseline*. Isso significa que o modelo foi capaz de fazer previsões mais acuradas em relação à velocidade de vento real. Essa melhoria é essencial para a operação da geração eólica, pois permite uma melhor programação e aproveitamento da energia gerada pelo vento. Esses resultados demonstram a utilidade da previsão de velocidade de vento na operação da geração eólica. Com a capacidade de prever a velocidade do vento com antecedência, os operadores podem tomar decisões mais informadas sobre a programação da geração, aproveitando rajadas de vento mais intensas, antecipando períodos de vento fraco e ajustando a operação das turbinas de acordo com as condições previstas. Isso resulta em uma operação mais eficiente, maximizando a geração de energia renovável e contribuindo para a estabilidade da rede elétrica.

Em resumo, os resultados dos modelos de previsão de velocidade de vento evidenciam a importância e a eficácia dessa abordagem na operação da geração eólica. A capacidade de prever o comportamento do vento permite otimizar a programação da geração, melhorar a eficiência na utilização dos recursos e garantir uma operação estável e confiável dos sistemas ligados à rede elétrica.

Para trabalhos futuros pode-se explorar novas tecnologias de predição como a união de dois modelos como CNN (Redes neurais convolucionais) com LSTM, estudar métodos para otimizar hiperparâmetros incluindo o tamanho do histórico de entrada, adicionar ao método mais dois pontos, perto de onde se pretende realizar a predição de velocidade de vento, estudar uma maneira de incluir no cálculo o resultado neste dois pontos para ter uma maior abrangência pois a energia eólica é altamente dependente das condições meteorológicas. Executar ensaios em usinas de energia eólicas reais para se obter resultados pragmáticos, utilizando dados de sensores dos aerogeradores para aumentar a resolução da predição de ventos. Por fim, pesquisar uma api para a captação automática e em tempo real dos dados disponíveis na internet, explorar cada vez mais a diversidade de dados que estão sendo acumulados e procurar novas aplicações para transformar a geração de energia eólica em um sistema cada vez mais inteligente e automatizado.

REFERÊNCIAS

REBOITA, M. S. et al. Entendendo o Tempo e o Clima na América do Sul. *Terra e Didat.*, v. 8, n. 1, p. 34–50, 2012. Citado na página 15.

BARRY, R. G.; CHORLEY, R. J. *Atmosfera, Tempo e Clima*. 9 edt. ed. Porto Alegre, Brasil: Bookman, 2013. 495 p. ISBN 9788565837392. Citado na página 15.

PACHECO, F. *Energias Renováveis : Breves Conceitos. Conjunt. e Planej.*, Salvador, Brazil, SEI, n. 149, p. 4–11, oct 2006. Disponível em: . Citado na página 13 e 16.

GWEC. *Global Wind Statistics 2016*. Brussels, Belgium, 2022. 4 p.
Disponível em: <<https://gwec.net/global-wind-report-2022/>>. Citado na página 17.

UCZAI, P. *Energias Renováveis - Riqueza Sustentável ao Alcance da Sociedade*. Brasilia, Brasil: Edições Câmara, 2012. 273 p. ISBN 9788573659740. Citado na página 17.

STAVISS, B. *Usina Eólica de Alegria I. Infraestrutura urbana Proj. custos e construção*, publicado na Web, n. 6, p. 3, aug 2011. Disponível em:
<<http://infraestruturaurbana17.pini.com.br/solucoes-tecnicas/6/artigo227165-2.aspx>>. Citado na página 18.

TOLMASQUIM, M. T. *Energia Renovável: Hidráulica, Biomassa, Eólica, Solar, Oceânica*. Rio de Janeiro, Brasil: Empresa de Pesquisa Energética (EPE), 2016. 452p p. ISBN 978-85-60025-06-0. Citado na página 18.

COLAK, I.; SAGIROGLU, S.; YESILBUDAK, M. Data mining and wind power prediction: A literature review. *Renew. Energy*, Elsevier Ltd, v. 46, p. 241–247, oct 2012. ISSN 09601481.
Disponível em: <<http://dx.doi.org/10.1016/j.renene.2012.02.015>>. Citado na página 18.

NREL. *Wind Power Today 2010*. Oak Ridge, TN, EUA, 2010. 1–32 p. Disponível em:
<<https://www.nrel.gov/docs/fy10osti/47531.pdf>>. Citado na página 18.

MEKA,R.;ALAEDDINI, A.;BHAGANAGAR, K.A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables.p. 2–23,December 2020. Citado na página 19, 31 e 49.

Le QV, et al. A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. Google Brain; 2015. p. 1e20.R. Meka, A. Alaeddini and K. Bhaganagar Energy 221 (2021). Citado na página 19.

Barbounis TG, Theocharis JB, Alexiadis MC, Dokopoulos PS. Long-term wind

speed and power forecasting using local recurrent neural network models. *IEEE Trans Energy Convers* 2006;21(1):273e84.. Citado na página 19.

Senju T, Yona A, Urasaki N, Funabashi T. Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator. In: 2006 IEEE PES power systems conference and exposition. IEEE; 2006. p. 1260e5.. Citado na página 19.

Felder M, Kaifel A, Graves A. Wind power prediction using mixture density recurrent neural networks. In: Poster presentation Gehalten auf der European wind energy conference; 2010. Citado na página 19.

OKORIE, M.; INAMBAO, F.; CHIGUVARE, Z. Evaluation of Wind Shear Coefficients, Surface Roughness and Energy Yields over Inland Locations in Namibia. *Procedia Manufacturing*, v. 3, p. 89–96, 2014. Disponível em: . Citado na página 20.

WHARTON, S.; LUNDQUIST, J. K. Assessing atmospheric stability and its impacts on rotor-disk wind characteristics at an onshore windfarm. *Wind Energy*, v. 15, n. 4, p. 525–546, 2012. Disponível em: . Citado na página 20.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*, 3rd edition, Morgan Kaufmann, 2011. Citado na página 22.

NUNES, TASSIA. Visualizando gráficos usando matplotlib em processamento de sinal neural de EEG. brainlatam, 2022. Disponível em:
<https://www.brainlatam.com/blog/ciencia-de-dados-visualizando-graficos-usando-matplotlib-em-processamento-de-sinal-neural-de-eeg-4224>. Citado na página 23.

ALVES, GISELY. Visualização de dados com Seaborn. medium, 2019. Disponível em:
<https://medium.com/@gisely.alves/visualiza%C3%A7%C3%A3o-de-dados-com-seaborn-2fd0defd9adb>. Citado na página 23.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. MIT Press, 2016. Disponível em: <http://www.deeplearningbook.org>. Acesso em: 08 maio 2023.. Citado na página 24.

KUHLMAN, D. A python book: Beginning python, advanced python, and python exercises. 2009. Disponível em: <http://www.davekuhlman.org>. Citado na página 26.

RASCHKA, S.; MIRJALILI, V. *Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2*. [S.l.: s.n.], 2015. 741 p. ISBN 9781789955750. Citado na página 25.

BROWNLEE, J. What Is Time Series Forecasting?. 2016. Disponível em:
<https://machinelearningmastery.com/time-series-forecasting/>. Citado na página 27 e 28.

Deep AI. Deep Learning. 2022. Acessado em 06 out 2022. Disponível em:

<<https://deepai.org/>>. Citado na página 27.

SHAFI, A. Random Forest Classification with Scikit-Learn. 2023. Disponível em: <<https://www.datacamp.com/tutorial/random-forests-classifier-python>>. Citado na página 26.

GERS, Felix et al.Understanding LSTM Networks. colah, 2015. Disponível em:<<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 14 abr. 2023. Citado na página 28.

KUTNER, M. H.; NACHTSHEIM, C. J.; NETER, J. Applied Linear Statistical Models. 4. ed. McGraw-Hill, 2004. Citado na página 30.

STACKHOUSE Jr., Paul. POWER Data Methodology. NASA POWER, 2020. Disponível em: <<https://power.larc.nasa.gov/docs/methodology/>>. Acesso em: 26 mar. 2023. Citado na página 33.

TAGUCHI, G. System of experimental design; engineering methods to optimize quality and minimize costs. 1987. Citado na página 31.

GELARO, Ronald et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). American Meteorological Society journals, [s.l.], v. 30, n. 14, p. 5419-5454, 2017. Disponível em: <[https://journals.ametsoc.org/conFigurable/content/journals\\$002fc1im\\$002f30\\$002f14\\$002fjcli-d-16-0758.1.xml?t:ac=journals%24002fc1im%24002f30%24002f14%24002fjcli-d-16-0758.1.xml](https://journals.ametsoc.org/conFigurable/content/journals$002fc1im$002f30$002f14$002fjcli-d-16-0758.1.xml?t:ac=journals%24002fc1im%24002f30%24002f14%24002fjcli-d-16-0758.1.xml)>. Acesso em: 27 mar. 2023. Citado na página 34.

VALENTE, Kaio. O guia definitivo do Walk Forward Analysis (WFA). Medium, 2022. Disponível em: <https://medium.com/devtrader/o-guia-definitivo-do-walk-forward-analysis-wfa-e755c2c33542#:~:text=Consiste%20em%20executar%20uma%20bateria,mercado%20e%20a%20diferentes%20para%20metriza%C3%A7%C3%B5es>. Acesso em: 14 abr. 2023. Citado na página 41.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. Neural computation, v. 9, n. 8, p. 1735-1780, 1997. Citado na página 42.

LUO, Hui; WANG, Shuang; LI, Jie. Multiple layers Long Short-Term Memory (LSTM) with L2 Regularization for Time Series Prediction. In: 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2017. p. 862-867. Citado na página 43.

MAHMOOD, Tahir; HONG, Tao. Evaluating time series forecasting models: An empirical study on performance improvement. Expert Systems with Applications, v. 92, p. 1-17, 2018. DOI: <https://doi.org/10.1016/j.eswa.2017.08.037>. Citado na página 44.