

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
FACULDADE DE ENGENHARIA & INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM ENGENHARIA COMPUTACIONAL

# Modelagem Preditiva de Geração de Energia Eólica Utilizando Redes Neurais e Dados Climáticos

Nícolas Estanislau Moreira

JUIZ DE FORA  
DEZEMBRO, 2023

# Modelagem Preditiva de Geração de Energia Eólica Utilizando Redes Neurais e Dados Climáticos

NÍCOLAS ESTANISLAU MOREIRA

Universidade Federal de Juiz de Fora  
Faculdade de Engenharia & Instituto de Ciências Exatas  
Bacharelado em Engenharia Computacional

Orientador: Heder Soares Bernardino  
Coorientadora: Luciana Conceição Dias Campos

JUIZ DE FORA  
DEZEMBRO, 2023

# MODELAGEM PREDITIVA DE GERAÇÃO DE ENERGIA EÓLICA UTILIZANDO REDES NEURAIS E DADOS CLIMÁTICOS

Nícolas Estanislau Moreira

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DA FACULDADE DE ENGENHARIA & INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM ENGENHARIA COMPUTACIONAL.

Aprovada por:

Heder Soares Bernardino  
Doutor em Modelagem Computacional

Luciana Conceição Dias Campos  
Doutora em Engenharia Elétrica

Leonardo Goliatt da Fonseca  
Doutor em Modelagem Computacional

Stênio Sã Rosário Furtado Soares  
Doutor em Computação

JUIZ DE FORA  
15 DE DEZEMBRO, 2023

## Agradecimentos

Aos meus pais, Eduardo e Adriana, expresso minha profunda gratidão por todo o apoio ao longo desta jornada. Obrigado por acreditarem em mim, mesmo quando eu tinha dúvidas. Sou imensamente grato por tudo que fizeram por mim.

À minha irmã, Eduarda, pelo amor, parceria e por estar sempre presente, mesmo distante, nos momentos mais difíceis. Não consigo imaginar uma melhor irmã mais velha, que sempre cuidou de mim.

Aos meus amigos, Bruno, Kevin, Matheus, Leonardo e Luiz, por fazerem desta jornada mais divertida. Seu apoio e companhia fizeram toda a diferença. Crescer durante esse período ao lado de pessoas tão especiais como vocês foi um prazer inestimável.

À minha tia, Meire, expresso minha sincera gratidão por seu constante incentivo à leitura e ao estudo. Sua preocupação e interesse contínuo em minha carreira são fontes de inspiração que valorizo profundamente. Obrigado por ser uma presença tão positiva e apoiadora em minha vida acadêmica e profissional.

Ao meu orientador, Heder, e à minha coorientadora Luciana, expresso profundo agradecimento por todo o direcionamento fornecido durante este trabalho. Suas orientações foram fundamentais para o desenvolvimento desde projeto, e sou grato pela paciência e sabedoria que compartilharam comigo durante esse processo.

Agradeço também à Equipe Microraptor por ter marcado significativamente meu período na UFJF. Durante esse tempo, aprendi que o empenho e a disciplina, quando aplicados a um propósito, são fundamentais para atingir nossos objetivos.

A jornada até aqui não foi fácil, seriam páginas e mais páginas de agradecimentos a todos que de alguma forma marcaram esses 8 anos da minha vida. Obrigado a todos que fizeram de mim quem sou hoje.

*“Perder-se também é caminho”.*

*Clarice Lispector*

# Conteúdo

<b>Lista de Figuras</b>	<b>4</b>
<b>Lista de Tabelas</b>	<b>6</b>
<b>Lista de Abreviações</b>	<b>7</b>
<b>1 Introdução</b>	<b>10</b>
1.1 Objetivos . . . . .	12
1.2 Organização . . . . .	13
<b>2 Fundamentação Teórica</b>	<b>14</b>
2.1 Energia Eólica . . . . .	14
2.2 Séries Temporais . . . . .	16
2.3 Redes Neurais Artificiais . . . . .	18
2.3.1 LSTM . . . . .	20
2.4 Tratamento de <i>Outliers</i> . . . . .	22
2.4.1 Intervalo Interquartil . . . . .	22
2.5 Tratamento de Dados Ausentes . . . . .	23
2.6 Correlação . . . . .	23
2.7 Métricas de Avaliação . . . . .	24
<b>3 Trabalhos Relacionados</b>	<b>27</b>
<b>4 Metodologia</b>	<b>32</b>
4.1 Coleta de Dados . . . . .	33
4.2 Pré-processamento de Dados . . . . .	34
4.2.1 Análise de dados das estações . . . . .	34
4.2.2 Análise de dados das usinas . . . . .	39
4.3 Arquitetura do Modelo . . . . .	46
<b>5 Resultados</b>	<b>48</b>
<b>6 Considerações Finais e Trabalhos Futuros</b>	<b>58</b>
<b>Bibliografia</b>	<b>60</b>

## Lista de Figuras

1.1	Aumento da temperatura média global (CENTRAL, 2023). . . . .	10
1.2	Número de usinas eólicas no Brasil ao longo dos anos (ONS, 2023). . . . .	12
2.1	Formação dos ventos em decorrência do deslocamento das massas de ar (Fonte: (CRESESB, 2008)). . . . .	15
2.2	Esquema de funcionamento de um aerogerador (Fonte: Picolo, Buhler e Rampinelli (2014)). . . . .	16
2.3	Decomposição de uma série temporal (Fonte: Brownlee (2021)). . . . .	17
2.4	Neurônio biológico (Fonte: Ciaburro e Venkaterwaran (2017)). . . . .	18
2.5	Neurônio artificial (Fonte: Haykin (2017)). . . . .	19
2.6	RNA multi-camadas sem bias. (Fonte: Aggarwal (2018)). . . . .	19
2.7	Rede Neural Recorrente. (Fonte: Olah (2015)). . . . .	20
2.8	Arquitetura LSTM (Fonte: Olah (2015)). . . . .	21
2.9	Representação de boxplot contendo outliers (Fonte: Yukio (2018)). . . . .	22
2.10	Correlação Linear Positiva. (Fonte: USP (2017)). . . . .	24
2.11	Correlação Linear Negativa. (Fonte: USP (2017)). . . . .	24
4.1	Diagrama de metodologia (Fonte: elaborado pelo autor). . . . .	32
4.2	Correlação das variáveis climáticas pela distância (Fonte: Elaborado pelo autor). . . . .	34
4.3	Dados faltantes da Estação de Acarau (Fonte: Elaborado pelo autor). . . . .	36
4.4	Dados faltantes da Estação de Guanambi (Fonte: Elaborado pelo autor). . . . .	37
4.5	Dados faltantes da Estação de Macau (Fonte: Elaborado pelo autor). . . . .	37
4.6	Dados faltantes da Estação de Rio Grande (Fonte: Elaborado pelo autor). . . . .	38
4.7	Dados faltantes da Estação de Santana (Fonte: Elaborado pelo autor). . . . .	38
4.8	Dados faltantes da Estação de Tramandaí (Fonte: Elaborado pelo autor). . . . .	39
4.9	Dados faltantes das usinas próximas à estação de Tramandaí (Fonte: Elaborado pelo autor). . . . .	40
4.10	Dados faltantes das usinas próximas à estação de Rio Grande (Fonte: Elaborado pelo autor). . . . .	41
4.11	Dados faltantes das usinas próximas à estação de Guanambi (Fonte: Elaborado pelo autor). . . . .	42
4.12	Dados faltantes das usinas próximas à estação de Macau (Fonte: Elaborado pelo autor). . . . .	43
4.13	Decomposição da série temporal de geração da usina Alegria 2 (Fonte: Elaborado pelo autor). . . . .	44
4.14	Mapa de Isoplelas contendo as estações e usinas utilizadas no estudo (Fonte: Adaptado de Solar (2023)). . . . .	45
4.15	Arquitetura LSTM adotada aqui (Fonte: Elaborado pelo autor). . . . .	46
5.1	Resultados de MSE para os dados de teste (Fonte: elaborado pelo autor) . . . . .	49
5.2	Resultados de $R^2$ para os dados de teste (Fonte: elaborado pelo autor) . . . . .	49
5.3	Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Morrão (Fonte: elaborado pelo autor) . . . . .	50

5.4	Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Alegria 2 (Fonte: elaborado pelo autor) . . . . .	51
5.5	Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Quinta 69kV (Fonte: elaborado pelo autor) . . . . .	51
5.6	Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Cidreira 1 (Fonte: elaborado pelo autor) . . . . .	52
5.7	Resultado do <i>MSE</i> para a usina de Quinta 69kV (Fonte: elaborado pelo autor) . . . . .	52
5.8	Resultado do <i>RMSE</i> para a usina de Quinta 69kV (Fonte: elaborado pelo autor) . . . . .	53
5.9	Resultado do <i>MAE</i> para a usina de Quinta 69kV (Fonte: elaborado pelo autor) . . . . .	53
5.10	Resultado do $R^2$ para a usina de Quinta 69kV (Fonte: elaborado pelo autor)	53
5.11	Gráfico de Dispersão para os dados de teste da usina de Alegria 2 (Fonte: elaborado pelo autor) . . . . .	54
5.12	Gráfico de Dispersão para os dados de teste da usina de Quinta 138kV (Fonte: elaborado pelo autor) . . . . .	54
5.13	Comparação de métricas para os modelos sem e com as variáveis climáticas, respectivamente representados por SC e VC (Fonte: elaborado pelo autor)	55
5.14	Relação das métricas de desempenho com as distâncias entre estações e usinas (Fonte: elaborado pelo autor). . . . .	56
5.15	Relação das médias das métricas de desempenho por conjunto de usinas próximas a uma mesma estação e a porcentagem de dados faltantes (Fonte: elaborado pelo autor). . . . .	57

## Lista de Tabelas

4.1	Pares de estações e usinas dentro do raio de 33 km. . . . .	35
4.2	Estações e usinas utilizadas no estudo. . . . .	45
4.3	Camadas e respectiva quantidade de neurônios empregados no modelo. . .	47
5.1	Resultados dos dados de teste . . . . .	48

## Lista de Abreviações

DCC	Departamento de Ciência da Computação
UFJF	Universidade Federal de Juiz de Fora
INPE	Instituto Nacional de Pesquisas Espaciais
INMET	Instituto Nacional de Meteorologia
LSTM	<i>Long Short-Term Memory</i>
MSE	<i>Mean Squared Error</i>
MAE	<i>Mean Absolute Error</i>
RMSE	<i>Root Mean Squared Error</i>

## Resumo

Nos últimos anos, a energia eólica tem adquirido uma significativa importância na matriz energética brasileira, impulsionada pelas preocupações globais com as mudanças climáticas e a considerável contribuição das emissões de gases provenientes da queima de combustíveis fósseis. Diante desse cenário, a energia eólica emerge como uma alternativa essencial para mitigar e controlar o efeito estufa, despertando um interesse crescente em estudos dedicados a esta área. Este estudo investiga dados públicos das usinas de energia eólica no Brasil, com o propósito de aplicá-los em um modelo de aprendizado de máquina capaz de prever a geração elétrica dessas instalações. Para assegurar a robustez dos resultados, integram-se dados climáticos públicos, explorando a relação entre esses dados meteorológicos e a produção de energia eólica. Um modelo de rede neural LSTM foi desenvolvido, e uma análise aprofundada dos dados foi conduzida, empregando técnicas de pré-processamento como remoção de *outliers* e tratamento de dados ausentes. Além disso, foi avaliada a viabilidade do uso de dados climáticos, considerando a distância geográfica entre as usinas e as estações meteorológicas. Esse enfoque abrangente busca não apenas antecipar a geração de energia eólica, mas também compreender a influência direta das condições climáticas nesse processo. A análise dos dados resultou em um conjunto de 13 usinas que foram aplicadas no modelo de LSTM. Os resultados demonstram que o uso de variáveis climáticas influencia positivamente nas previsões de geração de energia.

**Palavras-chave:** energia eólica, LSTM, dados públicos, séries temporais.

# Abstract

In recent years, wind energy has gained significant importance in the Brazilian energy matrix, driven by global concerns about climate change and the substantial contribution of emissions from the burning of fossil fuels. In this scenario, wind energy emerges as a crucial alternative to mitigate and control the greenhouse effect, sparking a growing interest in dedicated studies in this field. This study investigates public data from wind power plants in Brazil with the purpose of applying them in a machine learning model capable of predicting the electrical generation of these facilities. To ensure the robustness of the results, public climatic data is integrated, exploring the relationship between meteorological data and wind energy production. An LSTM neural network model was developed, and a thorough analysis of the data was conducted, employing pre-processing techniques such as outlier removal and handling of missing data. Additionally, the feasibility of using climatic data was evaluated, considering the geographical distance between the wind farms and meteorological stations. This comprehensive approach seeks not only to anticipate wind energy generation but also to understand the direct influence of weather conditions on this process. The data analysis resulted in a set of 13 wind farms applied to the LSTM model. The results demonstrate that the use of climatic variables positively influences energy generation predictions.

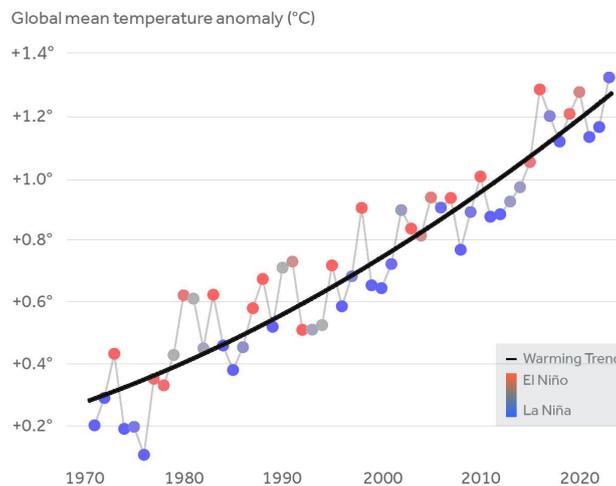
**Keywords:** wind energy, LSTM, public data, time series.

# 1 Introdução

As mudanças climáticas têm se intensificado em virtude do fenômeno do efeito estufa, uma ocorrência natural na qual determinados gases atmosféricos, tais como dióxido de carbono, metano e óxidos de nitrogênio, retêm porção do calor irradiado pela superfície terrestre. Esses gases desempenham um papel crucial na regulação térmica da Terra, proporcionando condições propícias à vida. Contudo, em razão das atividades antropogênicas, principalmente a queima de combustíveis fósseis e o desmatamento, observa-se um incremento substancial na concentração desses gases, resultando na amplificação do efeito estufa e contribuindo para o fenômeno de aquecimento global.

Um relatório divulgado pela Climate Central (CENTRAL, 2023) em novembro de 2023 analisa os “12 meses mais quentes já registrados na história”. Conforme indicado naquela pesquisa, a temperatura média global apresentou um aumento recorde de 1,3°C nesse período, como pode ser visto na Figura 1.1, com a projeção de que o fenômeno El Niño, com base em padrões históricos, apenas começou durante esse intervalo, prometendo efeitos ainda mais expressivos no ano subsequente. O documento ressalta a importância de uma redução anual ágil das emissões de carbono, declarando que isso se mostra imprescindível para interromper a tendência de aquecimento.

Figura 1.1: Aumento da temperatura média global (CENTRAL, 2023).



A emissão de gases provenientes da queima de diversos combustíveis fósseis contribui significativamente para a intensificação do efeito estufa, desencadeando repercussões alarmantes relacionadas ao aquecimento global e eventos trágicos associados a ele. Diante desse cenário, a energia eólica emerge como uma solução prática e eficiente para atenuar as consequências do aquecimento global. O crescente apelo às preocupações ambientais e a urgência em reduzir a dependência de fontes não renováveis ressaltam a importância de transições para práticas mais sustentáveis.

A transição energética emerge como um conceito cada vez mais proeminente em meio às preocupações globais com as mudanças climáticas. Este conceito refere-se à substituição de fontes de energia convencionais e poluentes por alternativas mais sustentáveis, a exemplo das energias eólica e solar.

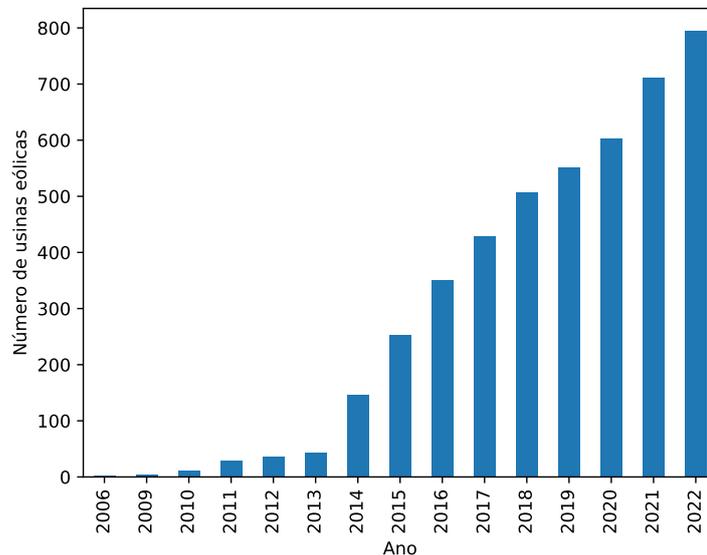
Historicamente, a humanidade transitou por diversas fases de transição energética, desde formas primitivas de obtenção de energia até a dependência de combustíveis fósseis, como petróleo e carvão. Contudo, o atual enfoque na transição energética destaca a mudança de fontes de energia menos sustentáveis para matrizes mais ecologicamente viáveis, visando a diminuição das emissões de gases de efeito estufa. Além de ser vital para a sustentabilidade ambiental, a transição energética é reconhecida como um dos pilares essenciais para o desenvolvimento econômico e social, visando conciliar geração de emprego, renda, inclusão social, combate às desigualdades, melhora na qualidade de vida, reindustrialização, preservação da biodiversidade e da qualidade ambiental, entre outros (ENERGIA, 2023).

No âmbito global e brasileiro, a energia eólica tem assumido um papel cada vez mais relevante, desempenhando um papel crucial como alternativa eficiente e sustentável para a geração de energia elétrica. No Brasil, com sua vasta extensão territorial, as condições favoráveis para a geração de energia eólica têm motivado o desenvolvimento desse setor. Segundo dados do Operador Nacional do Sistema Elétrico (ONS, 2023), estima-se que a geração eólica, que correspondia a 11,7% da matriz elétrica brasileira em 2022, chegue a 14,6% ao fim de 2027.

No início de 2022, o Brasil tomou uma nova iniciativa ao publicar o decreto nº 10.946/2022 (REPÚBLICA, 2022), emitido pelo governo federal, com o intuito de estabe-

lecer diretrizes que incentivem o progresso de projetos de geração de energia no ambiente *offshore*. Essa medida representa um marco significativo no país, demonstrando o compromisso em impulsionar o desenvolvimento desse setor. Ainda em 2022, o Senado aprovou o projeto (PL 3.386/2021) (VALÉRIO, 2021) intitulado Programa de Incentivo ao Desenvolvimento da Energia Eólica e da Solar Fotovoltaica (Pides), visando o financiamento pela União de até 500 milhões de reais por ano em projetos de desenvolvimento de energia eólica e solar. Fica evidente a importância da energia eólica no cenário atual e o interesse na expansão do setor de energia econômica e sustentável. Na Figura 1.2 é possível notar o crescimento no número de usinas eólicas no país.

Figura 1.2: Número de usinas eólicas no Brasil ao longo dos anos (ONS, 2023).



As turbinas eólicas são projetadas para aproveitar a energia cinética do vento e convertê-la em energia elétrica. No entanto, devido à natureza variável da velocidade e direção do vento, a quantidade de energia gerada também varia. Por esse motivo, é essencial estimar a quantidade de energia eólica a ser gerada para adequar a distribuição à demanda de energia elétrica.

## 1.1 Objetivos

Destacada a relevância da energia eólica na matriz energética nacional, este trabalho tem como objetivo principal criar modelos de previsão utilizando séries temporais para prever a geração de energia de usinas eólicas utilizando dados públicos. Assim, propõe-

se a utilização de resultados obtidos na literatura relacionados às tomadas de decisão e parametrização do modelo, visando obter resultados mais precisos. Como objetivos secundários:

1. Analisar a viabilidade de uso dos dados disponíveis ao público;
2. Avaliar conjuntos de estações e usinas em diferentes situações geográficas, verificando se o modelo pode ser aplicado da mesma maneira;
3. Analisar os efeitos da distância entre usinas e estações na capacidade de previsão;
4. Analisar a influência da imputação de dados desconhecidos nas previsões resultantes.

## 1.2 Organização

Este trabalho está estruturado da seguinte forma:

- No Capítulo 2 são estudados os conceitos por trás dos métodos utilizados no trabalho, abordando séries temporais e redes neurais;
- No Capítulo 3 é realizada uma revisão da literatura de temas relacionados à modelos de previsão semelhantes ao deste trabalho;
- No Capítulo 4 são abordadas as técnicas de pré-processamento utilizadas nos dados coletados das estações e usinas, a arquitetura do modelo e as métricas utilizadas na avaliação do modelo;
- No Capítulo 5 é apresentada uma análise dos resultados obtidos;
- No Capítulo 6 são apresentadas as considerações finais e trabalhos futuros.

## 2 Fundamentação Teórica

Neste capítulo, serão apresentados os fundamentos essenciais para compreensão do trabalho, assim como das técnicas utilizadas em sua proposta. O início é marcado pela introdução ao conceito de Energia Eólica, Series Temporais e Redes Neurais Artificiais (RNAs), explorando suas arquiteturas e métodos de aprendizado. Posteriormente, aborda-se o conceito de *Long Short-Term Memory* (LSTM) e sua forma de lidar com dependências a longo prazo.

Segundo Reis (2011), a energia eólica é obtida a partir da energia cinética contida nos movimentos do ar na atmosfera, o vento. Essa movimentação é gerada principalmente pelo aquecimento diferenciado das camadas de ar causado pela radiação solar e pelo movimento de rotação da Terra. O vento é gerado principalmente pela diferença de temperatura da superfície da Terra entre o Equador e os polos. Os ventos das superfícies frias circulam dos polos ao Equador de forma a substituir o ar quente que sobe nos trópicos e se move pela atmosfera superior até os polos.

### 2.1 Energia Eólica

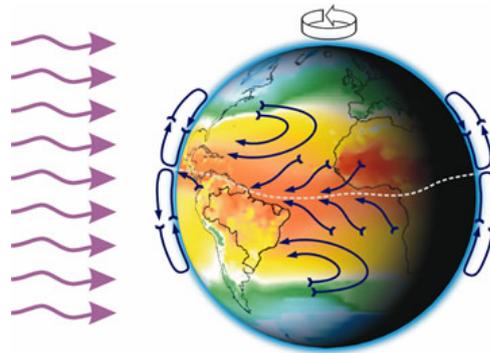
A energia eólica tem sua origem na energia solar, uma vez que os ventos são resultantes do aquecimento desigual da superfície terrestre. Uma estimativa da quantidade total de energia disponível nos ventos globais pode ser calculada considerando que cerca de 2% da energia solar absorvida pela Terra é transformada em energia cinética dos ventos, o que representa centenas de vezes a potência anual instalada nas usinas elétricas ao redor do mundo (DUTRA, 2008).

Nas zonas tropicais, onde os raios solares incidem quase perpendicularmente, ocorre um aquecimento mais intenso em comparação com as regiões polares. Como resultado, o ar quente nas baixas altitudes das áreas tropicais tende a se elevar, sendo substituído por massas de ar mais frio provenientes das regiões polares. Esse deslocamento de massas de ar é o responsável pela geração dos ventos, estabelecendo um ciclo de circulação atmosférica.

Esse processo, ilustrado na Figura 2.1 é essencial para compreender os padrões de vento globais e a dinâmica climática em escala planetária.

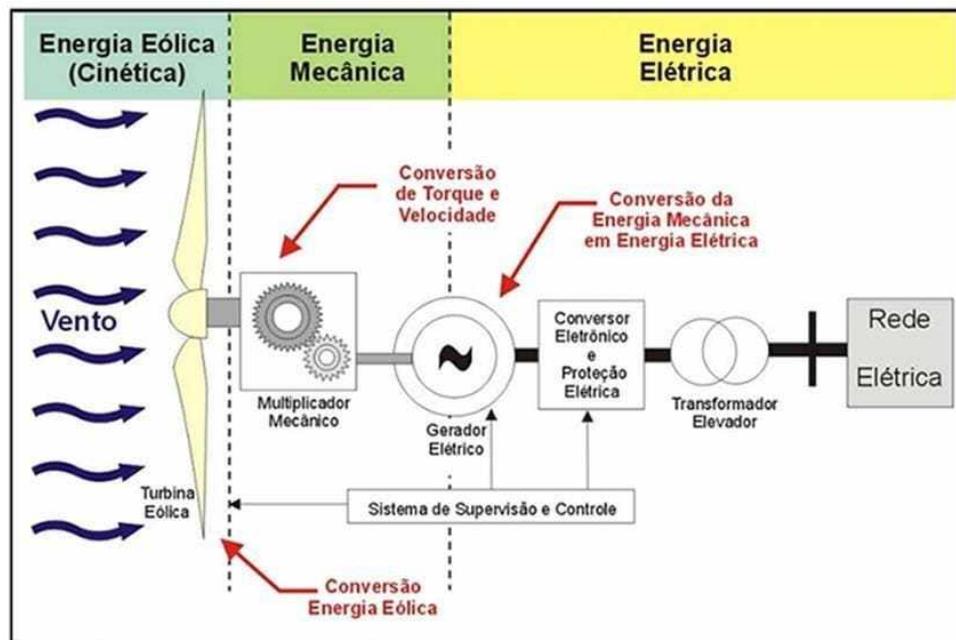
Existem regiões no planeta onde os ventos nunca cessam, pois os mecanismos que os originam, como o aquecimento no equador e o resfriamento nos polos, são constantes na natureza. Esses ventos perenes são denominados ventos planetários ou constantes.

Figura 2.1: Formação dos ventos em decorrência do deslocamento das massas de ar (Fonte: (CRESESB, 2008)).



As estruturas encarregadas de transformar a energia cinética do vento em eletricidade são os aerogeradores ou turbinas eólicas. Conforme o princípio da conservação de energia, levando em consideração algumas perdas entre os processos, o aerogerador realiza essa conversão por meio de dois processos principais: a transformação da velocidade e torque, e a conversão da energia mecânica em eletricidade. A Figura 2.2 proporciona uma visão geral desse processo.

Figura 2.2: Esquema de funcionamento de um aerogerador (Fonte: Pico, Buhler e Rampinelli (2014)).



O vento incide sobre as pás do rotor, as quais começam a girar, gerando movimento transferido para um sistema de engrenagens que amplifica a frequência do rotor. O gerador elétrico recebe a energia mecânica do rotor e a converte em eletricidade por meio do fenômeno de indução eletromagnética. Posteriormente, um transformador distribui a energia gerada para a rede elétrica.

## 2.2 Séries Temporais

Ao utilizar dados históricos de geração de energia eólica, juntamente com dados climáticos, para atingir as previsões, é evidente a natureza temporal do problema. Assim, caracterizando-o como um problema de séries temporais.

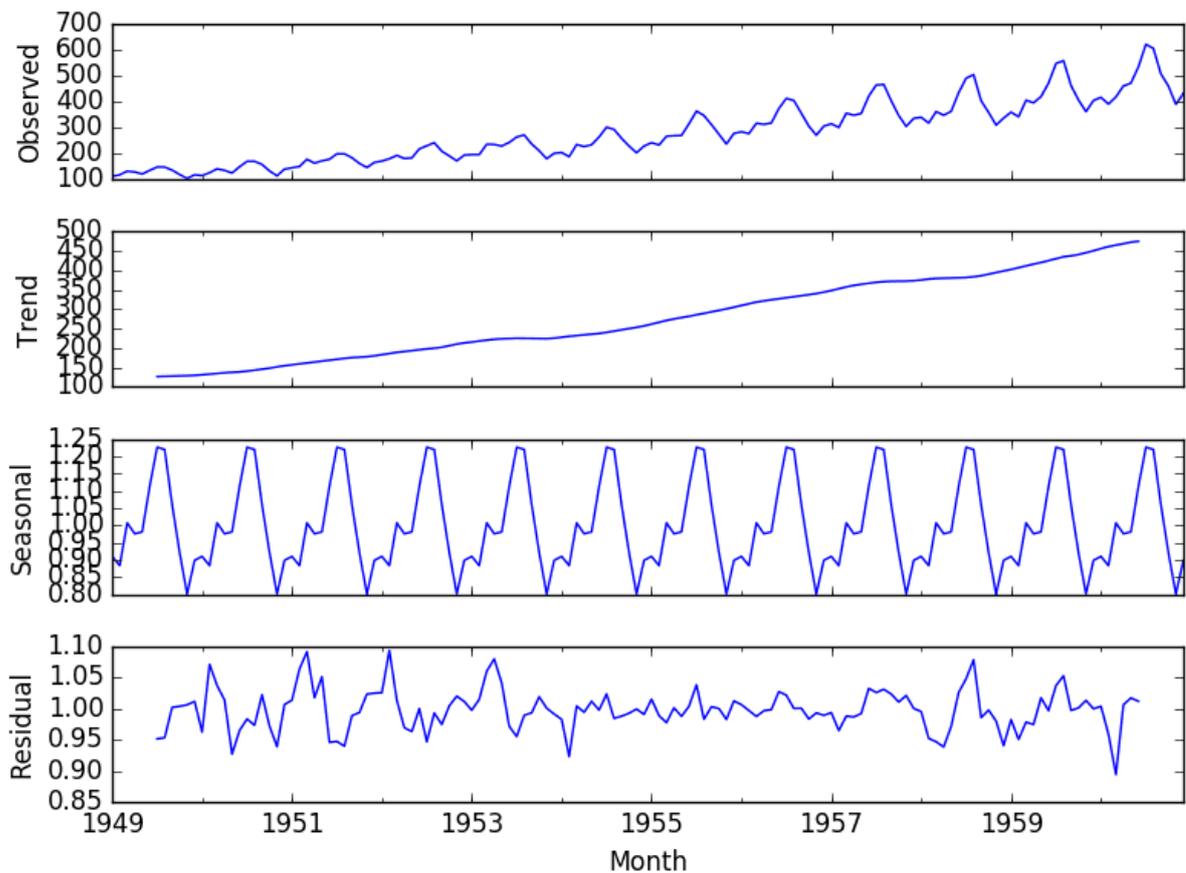
Uma série temporal refere-se a uma sequência de observações coletadas ao longo de um período de tempo. Nas análises de séries temporais, o primeiro passo se dá pela visualização de dados e identificação de padrões, como tendências e variações sazonais. Modelos simples podem ser suficientes para séries com características mais evidentes, mas séries mais complexas exigem abordagens avançadas, incluindo a consideração de *outliers* (CHATFIELD, 1996).

Os componentes principais de um série temporal podem ser classificados em:

- **Nível:** Refere-se ao valor médio da série temporal. É a linha de base em torno da qual a série flutua;
- **Tendência:** Indica a direção geral da série ao longo do tempo, mostrando se ela está aumentando, diminuindo ou permanecendo constante;
- **Sazonalidade:** Representa padrões que se repetem em períodos curtos e regulares, como sazonalidades anuais, mensais ou semanais;
- **Ruído:** Refere-se à variabilidade não sistemática na série, representando flutuações irregulares ou aleatórias que não podem ser atribuídas aos outros componentes.

Na Figura 2.3 é possível ver um exemplo de decomposição de uma série temporal em tendência, sazonalidade e ruído, respectivamente.

Figura 2.3: Decomposição de uma série temporal (Fonte: Brownlee (2021)).



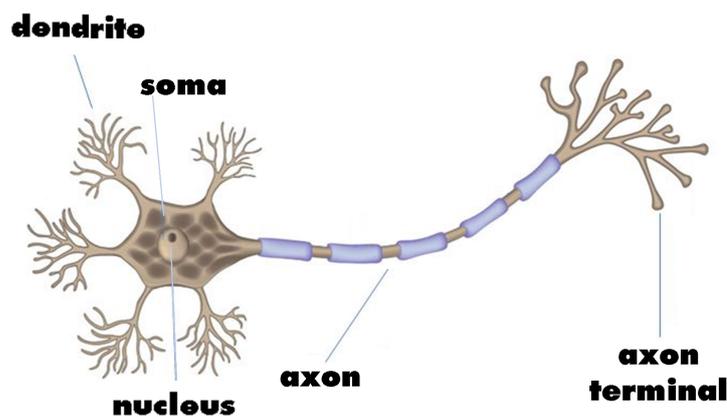
A análise de séries temporais, conforme destacada por Box et al. (2016), abrange diversas áreas de aplicação. A previsão de valores futuros é uma delas, envolvendo a

estimativa de dados com base em informações atuais e históricas. Além disso, a determinação da função de transferência destaca-se, consistindo na identificação de modelos de entrada e saída para avaliar como diferentes entradas impactam o resultado de um modelo. A utilização de variáveis indicadoras é uma prática empregada para avaliar eventos de intervenção em séries temporais, enquanto a análise de inter-relações se concentra em examinar a relação entre diferentes variáveis em modelos multivariados. Por fim, o desenvolvimento de esquemas de controle busca criar estratégias para ajustar os valores das séries de entrada, compensando desvios em relação ao resultado desejado.

## 2.3 Redes Neurais Artificiais

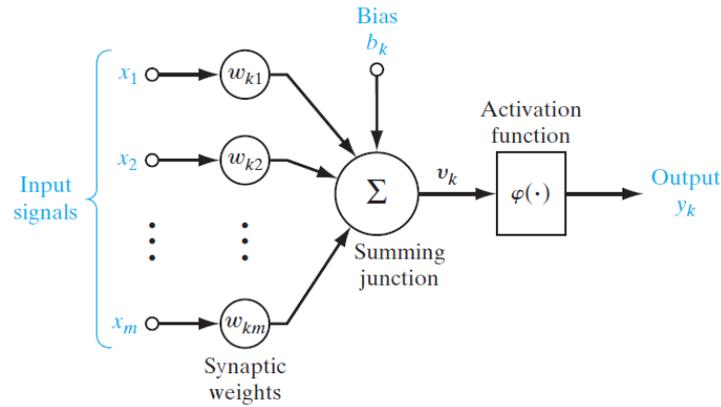
Redes neurais artificiais (ou RNA) são técnicas de aprendizado de máquina que simulam o mecanismo de aprendizado de organismos biológicos. O sistema nervoso humano é composto por células conhecidas como neurônios que se interconectam por meio de axônios e dendritos, sendo a ligação entre esses denominadas sinapses (AGGARWAL, 2018). Na Figura 2.4 é representada uma simplificação de um neurônio cerebral.

Figura 2.4: Neurônio biológico (Fonte: Ciaburro e Venkaterwaran (2017)).



De forma a simular o funcionamento do neurônio, são utilizados neurônios artificiais compostos por entradas que simulam os dendritos e saídas que simulam o axônio. Na Figura 2.5 é possível ver a representação de um neurônio artificial  $k$ . As entradas  $x_j$  possuem pesos  $w_{kj}$  aplicados, o *bias* da camada é representado por  $b_k$  e a saída do somador por  $v_k$ , descrito na equação 2.1, que define a saída do neurônio de acordo com o campo local induzido  $v$ .

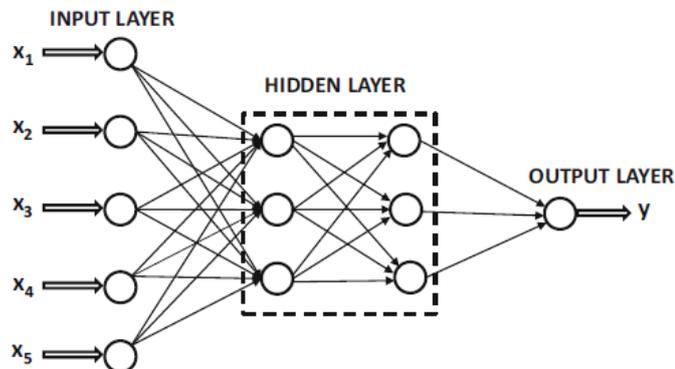
Figura 2.5: Neurônio artificial (Fonte: Haykin (2017)).



$$v_k = \left( \sum_{j=1}^n x_j w_{kj} \right) + b_k \quad (2.1)$$

Os neurônios da RNA são dispostos em camadas, sendo a primeira denominada camada de entrada e a última, a camada de saída. As redes multicamadas existem também as camadas intermediárias, geralmente chamadas de camadas ocultas pois as computações realizadas por essa parte da rede não são diretamente observáveis pelo usuário (AGGARWAL, 2018). O número de neurônios em uma camada é determinado de acordo com o problema. A Figura 2.6 exemplifica uma camada com 5 neurônios na camada de entrada, 3 em cada uma das 2 camadas oculta e 1 na camada de saída.

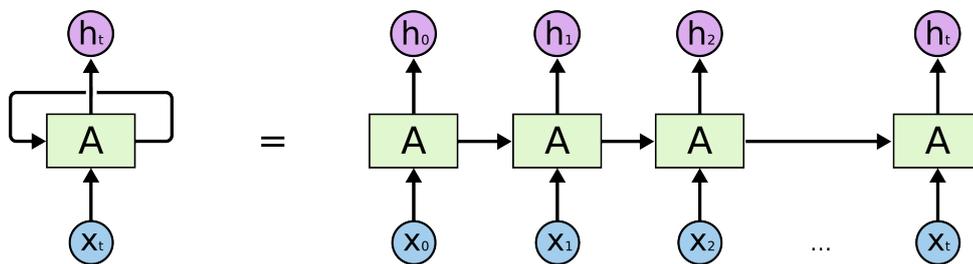
Figura 2.6: RNA multi-camadas sem bias. (Fonte: Aggarwal (2018)).



### 2.3.1 LSTM

A Rede Neural Recorrente (RNN) é um tipo de RNA projetada para lidar com dados sequenciais ou séries temporais, aplicada em tarefas como tradução de idiomas, processamento de linguagem natural (NLP), reconhecimento de fala, entre outros (IBM, 2021). Ao contrário das redes neurais *feedforward* e das convolucionais (CNNs), as RNNs possuem uma espécie de “memória”, na qual, por meio de um estado oculto atualizado a cada passo de tempo, é possível capturar dependências temporais.

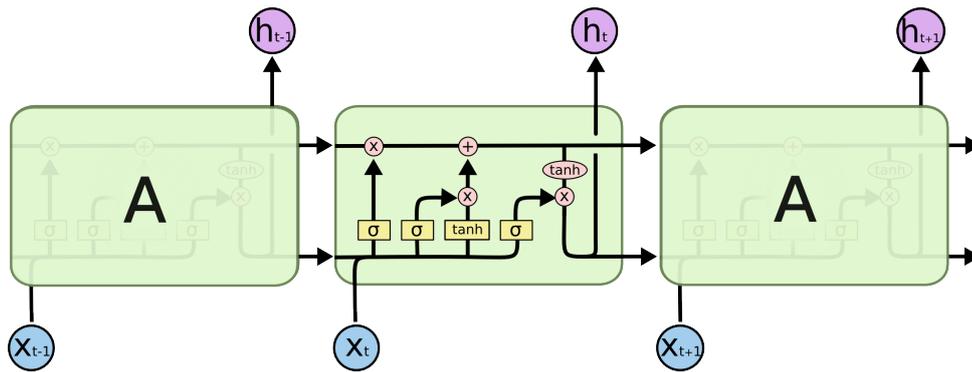
Figura 2.7: Rede Neural Recorrente. (Fonte: Olah (2015)).



A Figura 2.7 esboça o funcionamento de uma Rede Neural Recorrente (RNN), onde múltiplas instâncias da mesma rede são dispostas em sequência, cada uma transmitindo uma mensagem para a subsequente. Embora as RNNs possuam a habilidade de incorporar dados passados à tarefa presente, essa capacidade pode ser comprometida pelo intervalo entre a informação relevante e o ponto de interesse. Em situações onde essa lacuna é mínima, as RNNs demonstram eficácia. No entanto, em cenários que demandam informações mais distantes, surge a dificuldade em estabelecer conexões eficientes entre os dados, resultando em limitações para lidar com dependências de longo prazo (HOCHREITER; SCHMIDHUBER, 1997).

A rede que eficientemente incorpora informações de entrada de longo alcance no tempo e “memórias” é denominada *Long Short-Term Memory* (LSTM). Criada por Hochreiter e Schmidhuber (1997), os autores apresentaram esse modelo em resposta ao desafio das dependências de longo prazo. Em situações em que o estado anterior que impacta a previsão não está recentemente disponível, modelos convencionais podem ter dificuldades em realizar previsões precisas do estado atual.

Figura 2.8: Arquitetura LSTM (Fonte: Olah (2015)).



Como solução, foi sugerido que nas camadas ocultas da rede neural sejam incorporadas três portas distintas: uma de entrada, uma de saída e uma de esquecimento. A função da porta de esquecimento é eliminar informações consideradas irrelevantes para a resolução do problema. Na Figura 2.8 é representada a arquitetura de uma LSTM, que funciona, de forma resumida, da seguinte maneira:

- O primeiro passo é a decisão de incluir ou excluir informações no resultado destinado à próxima célula. A função sigmoide, representada por  $\sigma$ , na porta de esquecimento, opera de maneira binária: atribui valor 1 para encaminhar a informação adiante e 0 para descartá-la;
- Na segunda etapa, a célula decide, na porta de entrada, quais novas informações serão adicionadas ao estado atual;
- Em seguida, ocorre a transformação do antigo estado da célula no novo estado, descartando as informações de acordo com a primeira etapa e incorporando novos valores candidatos definidos na segunda etapa e;
- Por fim, através de uma função sigmoide, decide-se quais partes do estado da célula devem ser emitidas na saída. Nesta última etapa a memória de curto prazo é utilizada para determinar a saída.

Este processo é recorrente, é gerado um novo estado interno da célula e um novo registro na série temporal. Cada neurônio recebe um valor de entrada, o qual é processado para gerar um valor de saída. Esse resultado é então comparado com a previsão esperada

utilizando uma função de perda, que informa o nível de precisão da rede ao fazer previsões para uma determinada entrada (LA, 2019).

Por meio do algoritmo de retropropagação, também conhecido como *Backpropagation*, o valor da função de perda calculado em cada neurônio é analisado para ajustar o peso do estado. Essa adaptação visa melhorar o desempenho do neurônio no próximo ciclo, ou época, melhorando o resultado obtido (LA, 2019). O foco nesse processo é otimizar a função de perda que, de acordo com a métrica utilizada, pode ser minimizar ou maximizar seu valor.

## 2.4 Tratamento de *Outliers*

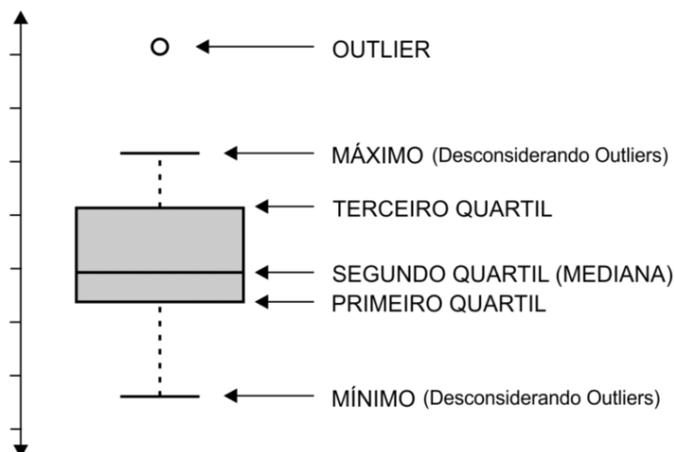
### 2.4.1 Intervalo Interquartil

O intervalo interquartil ou IQR é uma medida de dispersão utilizada em estatística descritiva. Sua fórmula é simples, sendo a subtração do terceiro quartil (Q3) pelo primeiro quartil (Q1), como apresentado na Equação 2.2.

$$IQR = Q3 - Q1 \quad (2.2)$$

Comumente utilizada em boxplots, como apresentado na Figura 2.9, o IQR é utilizado para detecção de *outliers* de forma semelhante ao *Z-score*, definindo um limite inferior e um superior.

Figura 2.9: Representação de boxplot contendo outliers (Fonte: Yukio (2018)).



O limite inferior é dado por  $Q1 - 1,5 \times IQR$  e o superior por  $Q3 + 1,5 \times IQR$ , que também estão representados na Figura 2.9. Valores considerados fora desse intervalo são então considerados *outliers* e removidos da série.

## 2.5 Tratamento de Dados Ausentes

Idealmente, uma série temporal teria todos os seus pontos alinhados conforme a discretização, mas ao lidar com dados coletados, é provável que ocorram lacunas sem informações. Fatores como erros de medição, falhas nos instrumentos e quedas de energia podem resultar na ausência de dados. A abordagem para tratar as lacunas, seja por meio de interpolação, suavização ou métodos mais avançados, impacta a precisão das análises e previsões derivadas da série temporal.

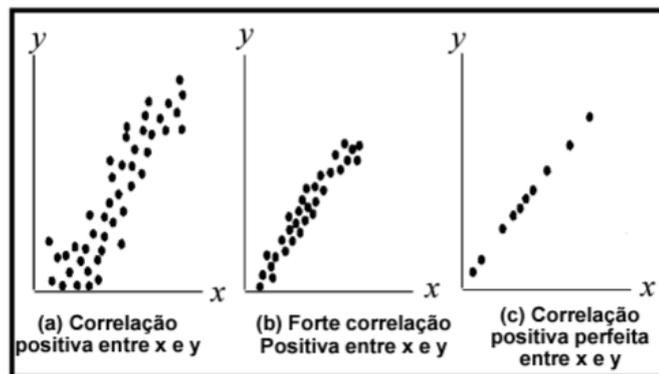
Uma das formas de lidar com a ausência de dados é a imputação de dados, método aplicado para inserir um valor para cada item ausente. Um modelo simples de imputação de dados é a regressão linear. Davidson (2003) representa o modelo de regressão linear simples como indicado na equação 2.3, na qual  $y_i$  representa a variável dependente (ou resposta) do  $i$ -ésimo elemento da amostra,  $x_i$  representa a variável independente e  $u_i$  representa o erro aleatório. Os parâmetros  $\beta$  e  $\beta_1$  são desconhecidos e exigem inferências para definir a relação entre a variável dependente e a variável independente.

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (2.3)$$

## 2.6 Correlação

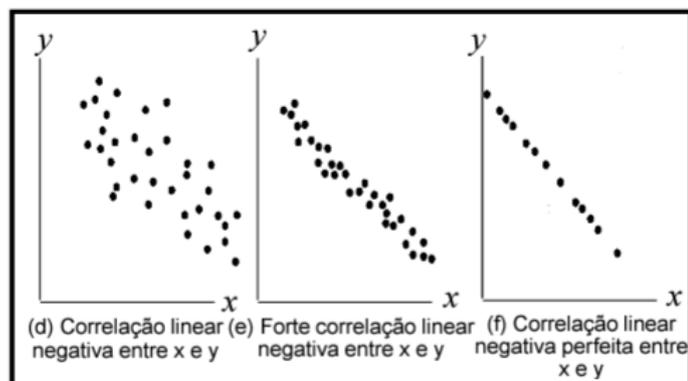
A correlação é um tipo de medida utilizada para saber se duas variáveis possuem alguma relação, de forma que as duas variam de maneira semelhante. Com esta medida, pode-se ter ideia se com os valores de uma das variáveis é possível prever valores de outra variável (ROQUE, 2017).

Figura 2.10: Correlação Linear Positiva. (Fonte: USP (2017)).



A correlação é considerada positiva se uma variável tende a aumentar quando a outra variável considerada também aumenta; a Figura 2.10 ilustra alguns casos de correlação positiva. Quando uma variável tende a diminuir com o aumento de outra, a correlação é considerada negativa, como pode ser visto na Figura 2.11.

Figura 2.11: Correlação Linear Negativa. (Fonte: USP (2017)).



Uma observação importante é a de que um índice de correlação de Pearson, que avalia a relação linear entre duas variáveis, igual a 0 não significa que não haja relação entre duas variáveis, mas de que a relação linear entre elas é baixa ou nula. Ou seja, pode existir relação não linear entre as variáveis.

## 2.7 Métricas de Avaliação

Previsões geradas por modelos de regressão, como as redes neurais, não são completamente precisas. Cada previsão contém erros que precisam ser quantificados de forma

a avaliar os resultados obtidos (PLEVRIS et al., 2022). Abaixo são exploradas algumas das métricas que podem ser utilizadas para este fim. Nas equações  $n$  representa o número total de observações,  $y_i$  representa o valor real na  $i$ -ésima observação, e  $\hat{y}_i$  é o valor predito na  $i$ -ésima observação.

O Erro Médio Quadrático (*Mean Squared Error - MSE*) é um dos indicadores de desempenho mais utilizados em modelos de regressão, e pode ser representado da seguinte forma:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2.4)$$

ou seja, é calculada a média dos quadrados das diferenças entre os valores previstos e reais. Uma das vantagens do MSE é sua sensibilidade a *outliers* no modelo, dado que há um peso maior devido à natureza quadrática da função. Se o modelo fizer uma previsão muito destoante, esta gera uma grande influência no erro calculado.

O Erro Médio Absoluto (*MAE - Mean Absolute Error*) é definido como a média da diferença absoluta entre as previsões de um modelo e os valores reais observados. Ele é calculado usando a equação seguinte:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.5)$$

O *MAE* é sempre não negativo devido ao uso do valor absoluto nas diferenças. Diferentemente do *MSE*, o *MAE* é menos sensível a *outliers* e a função de perda em um modelo de regressão fornece uma medida mais genérica de avaliação do modelo.

A Raiz do Erro Quadrático Médio (*Root Mean Squared Error - RMSE*) é outra métrica utilizada para avaliar o desempenho de modelos de regressão e pode ser representada da seguinte forma:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.6)$$

A principal diferença entre o *RMSE* e *MSE* é a raiz quadrada da média dos erros. O primeiro fornece uma medida mais intuitiva, uma vez que os resultados são apresentados na unidade de medida original dos dados de entrada.

O coeficiente de determinação  $R^2$  (*R2 Score*) varia de 0 a 1. Quanto maior o valor melhor é considerada a previsão. O  $R^2$  pode ser representado conforme a seguinte equação:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.7)$$

onde  $n$  é o número total de observações,  $y_i$  representa o valor real na  $i$ -ésima observação,  $\hat{y}_i$  é o valor predito na  $i$ -ésima observação, e  $\bar{y}$  é a média dos valores reais. O  $R^2$  indica em porcentagem o quanto o modelo consegue explicar os valores observados (GALVÃO, 2019).

A métrica a ser utilizada em métodos de aprendizado de máquina depende da solução que o modelo busca resolver. Uma abordagem comum é utilizar diferentes métricas, já que, como explicitado acima, cada uma fornece uma visão diferente com relação às previsões.

### 3 Trabalhos Relacionados

Foram considerados como trabalhos relacionados aqueles que abordam modelagens preditivas, com um foco naquelas dentro do contexto de energia eólica, incluindo ou não a integração dos dados climáticos. Além da previsão de geração, foram consideradas também modelagens preditivas das variáveis que afetam a geração, como a velocidade do vento. Com o intuito de ajustar a expectativa quanto à influência dos métodos de pré-processamento nos resultados, foram considerados também trabalhos com foco no tratamento de dados desconhecidos, comparando diferentes métodos de imputação e sua influência nos resultados obtidos.

Na literatura são reportados diversos métodos para a previsão de geração de energia eólica. Um dos principais desafios encontrados se dá pela imprevisibilidade do vento, podendo ter sua velocidade facilmente influenciada por obstáculos ou terreno próximo à usina eólica. Para lidar com a previsão da velocidade do vento, diversos métodos foram desenvolvidos (LEI et al., 2009), podendo ser divididos em duas categorias: físicos e estatísticos.

Conforme apontado por Lei et al. (2009), os métodos físicos demonstram vantagens em previsões de longo prazo, enquanto os métodos estatísticos destacam-se em previsões a curto prazo, embora essa distinção não seja uma regra inflexível. É comum encontrar modelos que adotam uma abordagem híbrida, incorporando tanto métodos físicos quanto estatísticos. Recentemente, observa-se um crescente interesse em novas abordagens, como o emprego de Inteligência Artificial em diversas formas.

É ressaltada ainda a complexidade envolvida na comparação direta entre os métodos, uma vez que a eficácia de cada um é mais pronunciada em circunstâncias específicas, além de ser sensível à metodologia de coleta de dados. Contudo, é relevante destacar que modelos fundamentados em inteligência artificial, como aquele abordado neste trabalho, sobressaem-se ao exibir uma eficiência notavelmente superior em previsões de curto prazo.

Um exemplo é o trabalho de Fujise (2018), no qual são comparados três modelos: físico, de persistência e baseado em RNA. Com testes realizados utilizando uma usina

eólica localizada no Rio Grande do Norte, denominada Parque Modelo, são utilizadas duas métricas de desempenho na comparação da aplicação da série de tal usina para os três modelos.

O modelo baseado em RNA teve resultados muito superiores ao modelo físico, que por sua vez teve resultados superiores ao modelo de persistência. Além disso, o autor destaca que, independente do modelo utilizado, é necessária a análise e pré-processamento criteriosos das séries de entrada. É notado também que os modelos de RNA e físico conseguem prever melhor as variações bruscas que ocorrem por fatores aleatórios, como, por exemplo, quedas de energia, resultando em lacunas de dados faltantes.

No artigo de Cali e Sharma (2019), uma abordagem semelhante à empregada no presente trabalho foi aplicada para prever a geração de energia na usina de Sotavento, localizada na Espanha. O estudo propõe uma análise de sensibilidade em relação a diversas variáveis climáticas, como temperatura, pressão atmosférica, velocidade e direção do vento. Foram conduzidos diferentes testes diversificando combinações de variáveis e também explorando diferentes altitudes para a velocidade e direção do vento.

Os resultados mais promissores foram alcançados ao utilizar as variáveis de temperatura, velocidade do vento e direção do vento, sendo a combinação de variáveis a ser empregada no presente trabalho. A outra variável analisada foi a pressão atmosférica, que teve impacto negativo nos resultados. Esta abordagem reforça a busca por resultados otimizados ao considerar variáveis específicas, ressaltando a importância da seleção criteriosa de parâmetros para a modelagem preditiva da geração de energia eólica.

No estudo de Bilgili, Sahin e Yasar (2007), foi proposto o emprego de uma RNA modelada especificamente para a previsão da velocidade do vento, utilizando estações meteorológicas localizadas na região leste do Mediterrâneo, na Turquia. Além da previsão da velocidade do vento, o trabalho concluiu que, para que modelos de redes neurais alcancem resultados mais precisos, é crucial que o fator de correlação entre as estações seja superior a 0,59.

No presente trabalho, não se realizou a previsão da velocidade do vento conforme o artigo mencionado. No entanto, foi conduzida uma análise de correlação entre estações meteorológicas do INMET abrangendo todas aquelas consideradas automáticas, que pos-

suem maior quantidade de dados, para determinar quais podem ser utilizadas para integrar as variáveis climáticas em conjunto com os dados históricos de determinadas usinas próximas, visando a obtenção de resultados mais robustos.

Wand et al. (2021) menciona que, embora as Redes Neurais Recorrentes (RNNs), especialmente o tipo Long Short-Term Memory (LSTM), tenham sido eficazes em lidar com sequências temporais e superar o problema do desaparecimento de gradientes, ainda existem desafios em termos de longas dependências temporais. As LSTMs, por sua natureza, são projetadas para capturar dependências de longo prazo, mas em algumas situações, modelos híbridos podem ser mais eficientes para melhorar a precisão das previsões. Para lidar com essas limitações específicas das LSTMs, pesquisadores têm explorado abordagens híbridas que combinam, por exemplo, LSTMs com Redes Neurais Convolucionais (CNNs), Redes Neurais de Memória Associativa (ELMs) ou outros tipos de redes neurais.

Tal trabalho ainda conclui por meio de experimentos que, apesar da LSTM ter um bom desempenho por si só, todos os modelos híbridos que fizeram a combinação de LSTM com algum outro modelo de rede neural teve resultado superior. Apesar disso, destaca que a combinação de diferentes modelos é muito complexa, principalmente ao lidar com previsões de velocidade do vento ou geração eólica, sendo que mesmo com um modelo extremamente robusto, se não forem aplicadas técnicas extensivas de pré-processamento, os resultados podem ser insatisfatórios devido à natureza do problema.

O texto de Karaman (2023) aborda a aplicação de métodos de aprendizado de máquina para prever a produção de energia eólica. São utilizados métodos de RNA, RNN, CNN, e LSTM. Com as métricas de desempenho MAE, RMSE, MSE e  $R^2$ , as mesmas aplicadas no presente trabalho, o desempenho dos algoritmos foi avaliado com a LSTM apresentando previsões mais próximas dos dados reais. Assim como em Wand et al. (2021), é proposto que abordagens híbridas combinando diferentes modelos de aprendizado de máquina sejam utilizados devido à complexidade do problema, mas que resultados satisfatórios puderam ser encontrados utilizando o modelo LSTM.

O trabalho de Batista (2003) aborda especificamente métodos de imputação de dados desconhecidos, incluindo a imputação pela média ou moda como método mais

simples. De maneira geral, o autor destaca a importância da fase de pré-processamento, descrevendo-a como “uma atividade que requer muito tempo e atenção por parte dos envolvidos no processo”. Entretanto, ressalta a necessidade de preencher tais dados de maneira menos trabalhosa e mais dinâmica pela comunidade de aprendizado de máquina. Caso não seja feita de maneira cuidadosa, pode haver a introdução de novos problemas mais difíceis de serem investigados.

Batista (2003), ao citar as limitações dos métodos de imputação de dados, afirma que a imputação pela média ou moda pode distorcer os dados, introduzindo falsos padrões, sobretudo quando os valores desconhecidos são introduzidos de forma não aleatória. Valores imputados são, na maioria, mais bem comportados do que os valores reais, desconhecidos, seriam. Há um grande risco de simplificação excessiva do problema em estudo, principalmente quando há grandes quantidades de valores desconhecidos. Portanto, a avaliação dos métodos de imputação se mostra uma tarefa complexa, uma vez que, uma quantidade de dados faltantes muito grande pode gerar taxas de erros menores que os valores reais.

Outro trabalho que trata especificamente sobre o impacto de dados faltantes no desempenho de métodos de aprendizado de máquina é o de Costa (2018). Neste, foram realizados experimentos com mais de 20 conjuntos de dados utilizando 5 algoritmos de imputação de dados. Foram analisadas as faixas de 5%, 35% e 60% de dados faltantes. Em 4 dos 5 métodos analisados, foram observados resultados melhores para a faixa de 35% que na faixa de 5%, com uma redução no desempenho na faixa de 60%. Isso corrobora com a afirmação de Batista (2003) de que uma quantidade maior de dados faltantes pode acabar simplificando o problema, como é o caso entre as faixas de 5% e 35%. Já na última faixa, de 60%, Costa (2018) comenta que para valores tão altos de dados faltantes há não somente a simplificação, mas a descaracterização do problema, já que mais da metade da série de entrada é composta por dados fictícios.

Os trabalhos de Batista (2003) e Costa (2018) destacam o fato de que há grande possibilidade de que séries temporais com grandes quantidades de dados ausentes podem refletir resultados “melhores” quando introduzidos em modelos de aprendizado de máquina. Entretanto, apesar de possuírem taxas de erros menores, isso pode ocorrer de-

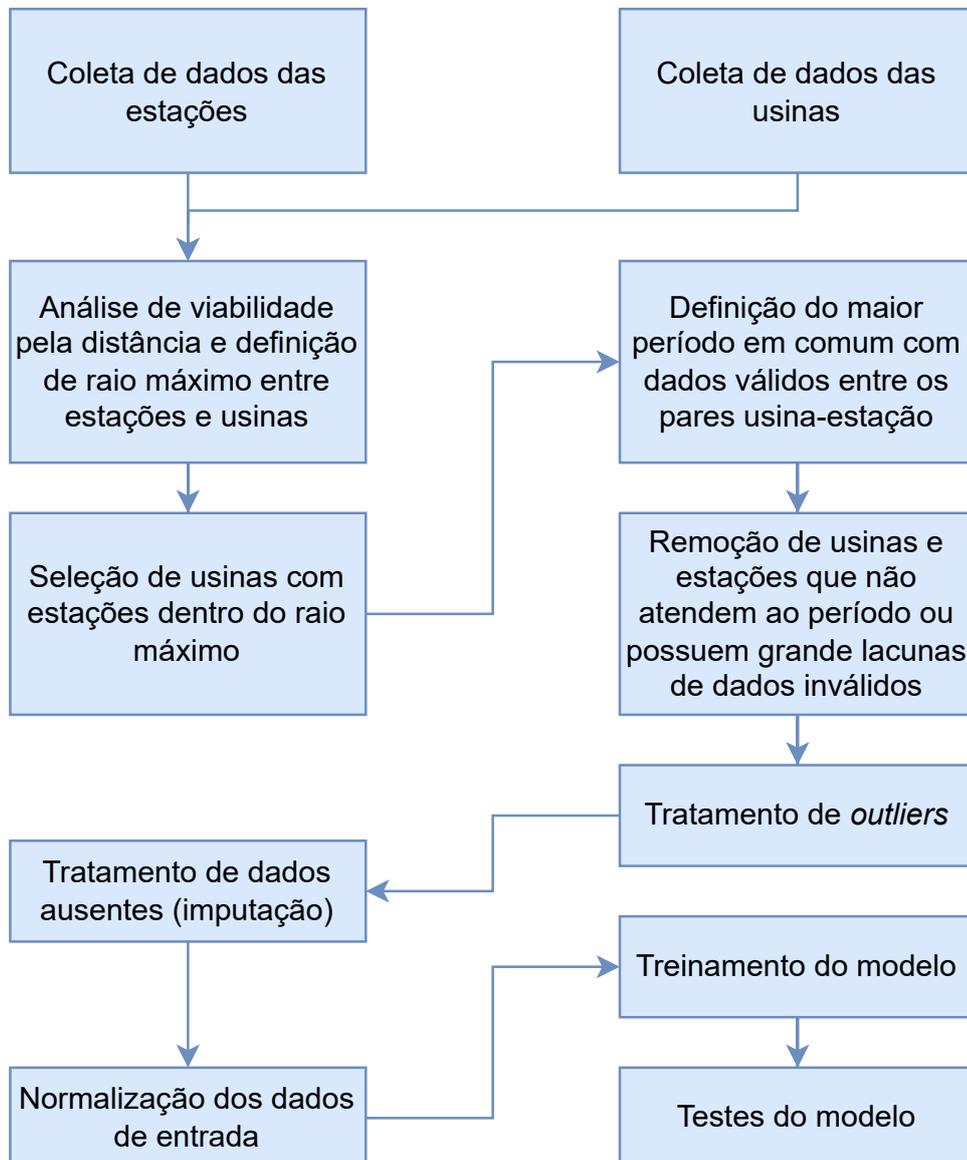
---

vido à simplificação da série, dado que a maioria dos métodos de imputação utilizados podem não possuir o mesmo caráter aleatório de dados reais.

## 4 Metodologia

Neste capítulo são abordadas todas as etapas da metodologia utilizada neste estudo. Na Seção 4.1, a origem dos dados coletados, assim como o método utilizado, são descritos. Já na Seção 4.2 o pré-processamentos dos dados é descrito, tanto das estações quanto das usinas, nas Seções 4.2.1 e 4.2.2, respectivamente. Por fim, na seção 4.3, é apresentada a arquitetura do modelo utilizado aqui. A Figura 4.1 apresenta um diagrama da metodologia seguida neste trabalho.

Figura 4.1: Diagrama de metodologia (Fonte: elaborado pelo autor).



## 4.1 Coleta de Dados

Os dados utilizados neste trabalho foram obtidos a partir de conjuntos de dados abertos disponibilizados pelo ONS e pelo INMET.

Os dados disponíveis publicamente pelo ONS contém uma discretização horária da geração de energia em GWh e MWmed (média de MW durante um período de referência) para as usinas eólicas, além de dados de localização e sua classificação dentro do SIN - Sistema Interligado Nacional.

Já os dados do INMET obtidos através das estações meteorológicas automáticas possuem dados, também com discretização horária, de precipitação, pressão, radiação, temperatura, umidade, velocidade do vento e direção do vento. Neste estudo serão utilizados somente dados de temperatura, velocidade do vento e direção do vento.

As estações meteorológicas automáticas coletam, de minuto em minuto, informações meteorológicas (temperatura, umidade, pressão atmosférica, precipitação, direção e velocidade dos ventos, radiação solar) da área em que está localizada. Os dados são integrados e disponibilizados gratuitamente por meio do site do INMET (INMET, 2011).

Em ambos os casos, a coleta foi feita de forma manual. Os arquivos relacionados às estações puderam ser obtidos em conjunto por ano no site do INMET (<https://portal.inmet.gov.br/>), sendo que foram coletados arquivos em formato *csv* para todas as estações automáticas disponíveis. Os dados das usinas foram coletados em mesmo formato, mas de forma individual através do Histórico da Operação do ONS (<https://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>). Neste, são apresentados gráficos utilizando a interface do *software* Tableau, na qual os pontos podem ser selecionados e exportados. Na Seção 4.2.1 é detalhado como as usinas foram definidas, dessa forma não foi necessário a coleta de todos os dados, somente das usinas utilizadas no trabalho.

Dado que a proposta inicial consiste em integrar dados de previsão e dados climáticos em um mesmo modelo, foram coletados dados de estações meteorológicas automáticas em todo o Brasil. Essa escolha foi motivada pela capacidade das estações automáticas de fornecerem medições precisas e contínuas. Para cada usina analisada, buscaram-se as estações meteorológicas em um raio delimitado na Seção 4.2.1, garantindo a representatividade dos dados climáticos na região de cada usina.

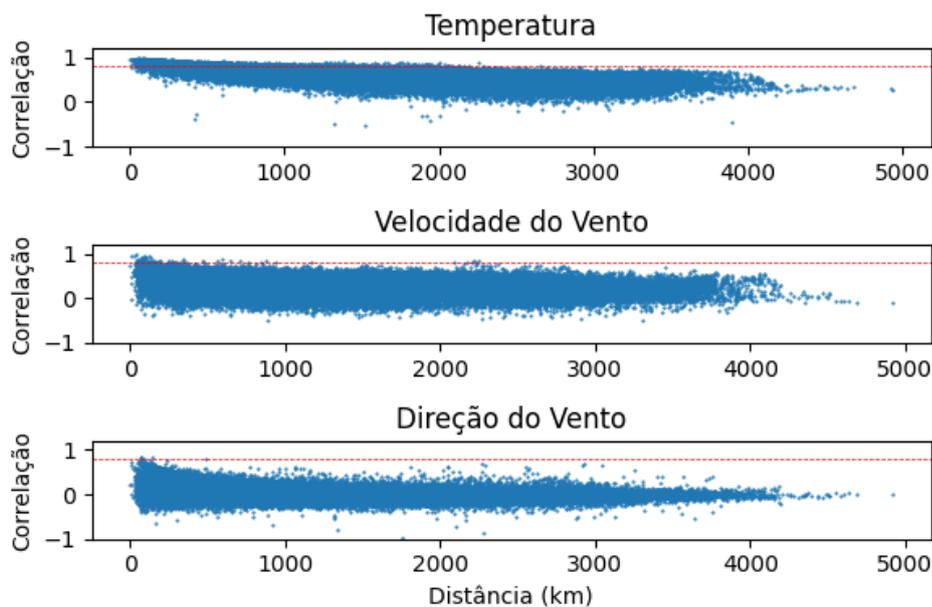
## 4.2 Pré-processamento de Dados

Nesta seção é realizada uma análise das séries temporais das estações e usinas, sendo descritas todas as etapas de tratamento dos dados aplicada neste trabalho.

### 4.2.1 Análise de dados das estações

Para definir quais estações seriam utilizadas no estudo, foi realizada um estudo da correlação entre as variáveis climáticas entre todas as estações. Dessa forma, ao traçar a correlação entre as variáveis climáticas pela distância entre as estações, é possível definir um raio máximo no qual é válido utilizar os dados destas estações também para as usinas.

Figura 4.2: Correlação das variáveis climáticas pela distância (Fonte: Elaborado pelo autor).



O fator de correlação é examinado não apenas para a velocidade do vento, mas também para a temperatura e a direção do vento, visando atingir valores próximos ou maiores de correlação que os citados em Bilgili, Sahin e Yasar (2007), o que assegura uma maior confiabilidade no uso dos dados provenientes das estações meteorológicas próximas às usinas dadas as conclusões do trabalho.

Zou, Tuncali e Silverman (2003) sugere que um coeficiente de correção superior a 0.8 representa uma forte associação. A análise das três variáveis climáticas revelou que a direção do vento é a que possui maior restrição. A correlação atinge valores superiores a

Tabela 4.1: Pares de estações e usinas dentro do raio de 33 km.

	Estação	Usina	Distância (km)
0	SANTANA DO LIVRAMENTO	UEE Cerro Chato II	6,979
1	TRAMANDAI	UEE Cidreira 1	8,484
2	SANTANA DO LIVRAMENTO	Conj. Livramento 2 (UEE)	9,371
3	SANTANA DO LIVRAMENTO	UEE Cerro Chato III	10,692
4	SANTANA DO LIVRAMENTO	UEE Cerro Chato I	10,726
5	GUANAMBI	Conj. Morrão (UEE)	11,47
6	TRAMANDAI	UEE Osório 3	11,944
7	GUANAMBI	Conj. Licínio de Almeida (UEE)	12,085
8	GUANAMBI	Conj. Alvorada (UEE)	12,142
9	TRAMANDAI	UEE Osório 2	13,344
10	MACAU	UEE Miassaba 3	13,87
11	RIO GRANDE	Conj. Quinta 138 kV (UEE)	16,435
12	RIO GRANDE	Conj. Quinta 69 kV (UEE)	16,435
13	TRAMANDAI	UEE Índios	16,443
14	GUANAMBI	Conj. Guirapá (UEE)	16,764
15	TRAMANDAI	UEE Sangradouro 2	17,851
16	TRAMANDAI	UEE Sangradouro 3	17,857
17	TRAMANDAI	UEE Sangradouro	17,912
18	MACAU	UEE Mangue Seco 5	18,505
19	GUANAMBI	Conj. Araçás (UEE)	19,107
20	TRAMANDAI	UEE Índios 2	19,156
21	MACAU	UEE Alegria II	19,815
22	TRAMANDAI	UEE Índios 3	20,859
23	MACAU	UEE Alegria I	21,27
24	TRAMANDAI	Conj. Atlântica (UEE)	21,421
25	MACAU	UEE Mangue Seco 3	23,875
26	GUANAMBI	Conj. Planaltina (UEE)	26,287
27	ACARAU	UEE Cajucoco	27,126
28	GUANAMBI	Conj. Caetité 123 (UEE)	27,712
29	MACAU	UEE Mangue Seco 2	28,155
30	ACARAU	Conj. Papagaios (UEE)	28,218
31	MACAU	UEE Mangue Seco 1	30,084
32	ACARAU	UEE Volta do Rio	32,841

0.8, representada na Figura 4.2 pelas linhas vermelhas tracejadas, apenas para distâncias superiores a 33 km. Portanto, foi estabelecido esse valor como o raio limite para a utilização dos dados climáticos. Ou seja, somente as usinas que possuem pelo menos uma estação meteorológica em um raio de 33 km serão consideradas. As estações e usinas com as distâncias entre elas podem ser vistas na Tabela 4.1.

A análise dos pares de estações e usinas revelou a presença de estações duplicadas, enquanto nenhuma repetição foi observada nas usinas. Esta discrepância ocorre devido à inclusão de algumas usinas em parques eólicos próximos a uma mesma estação.

Após a identificação dos pares de estações e usinas, conduziu-se uma análise minuciosa dos dados ausentes para determinar quais séries apresentavam dados suficientes para utilização no modelo.

Esta avaliação considerou as seguintes variáveis:

- Temperatura: uma variável crucial para a geração de energia eólica, uma vez que exerce influência sobre a velocidade do vento.
- Velocidade do vento: a variável mais significativa para a geração de energia eólica, sendo a força primordial que impulsiona as pás das turbinas.
- Direção do vento: uma variável que também pode afetar a geração de energia eólica, pois tem potencial para impactar a eficiência das turbinas.

A análise inicial foi conduzida considerando as séries temporais de cada estação, de forma independente dos pares de estação e usina. Em azul são os dados geração e os espaços em rosa indicam lacunas com ausência de dados.

Figura 4.3: Dados faltantes da Estação de Acarau (Fonte: Elaborado pelo autor).

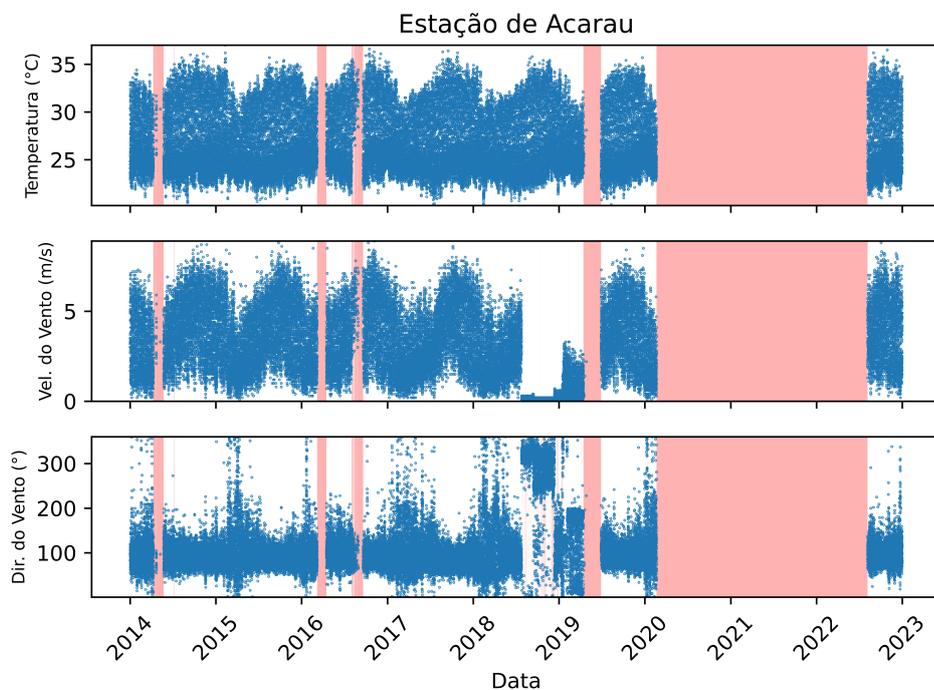


Figura 4.4: Dados faltantes da Estação de Guanambi (Fonte: Elaborado pelo autor).

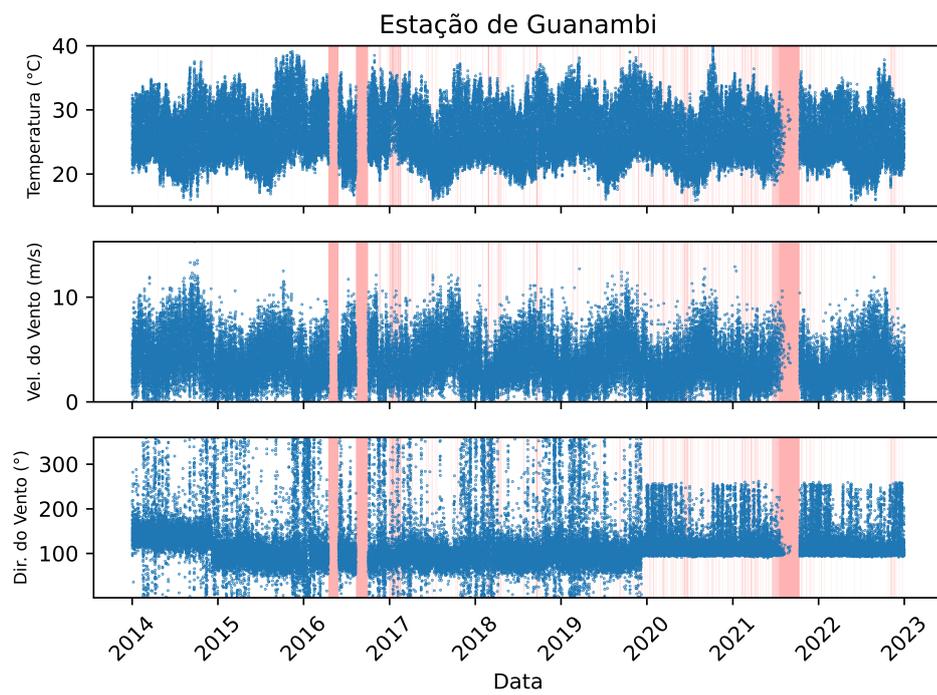


Figura 4.5: Dados faltantes da Estação de Macau (Fonte: Elaborado pelo autor).

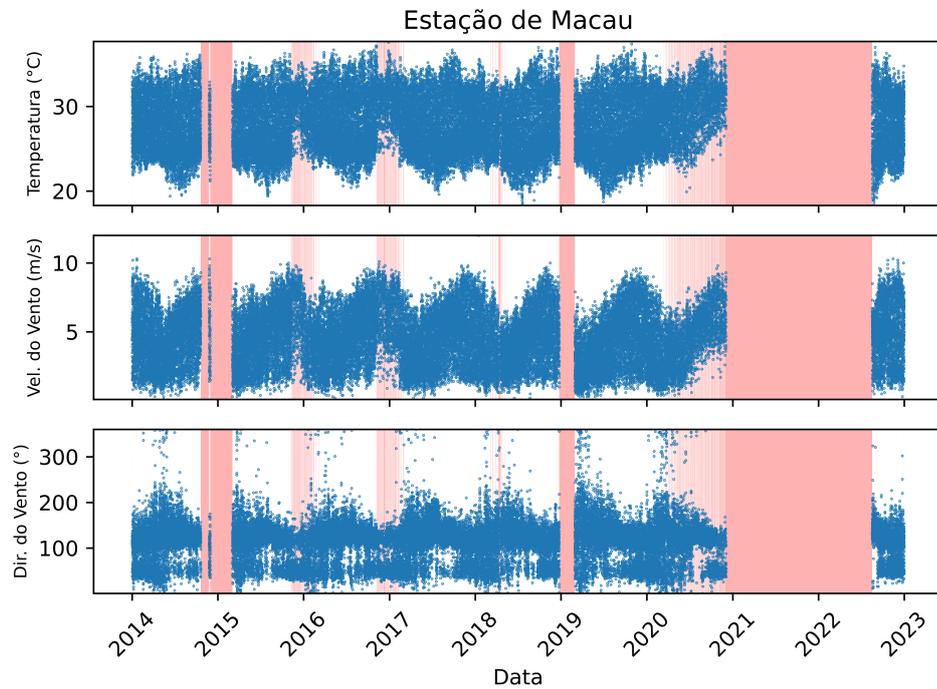


Figura 4.6: Dados faltantes da Estação de Rio Grande (Fonte: Elaborado pelo autor).

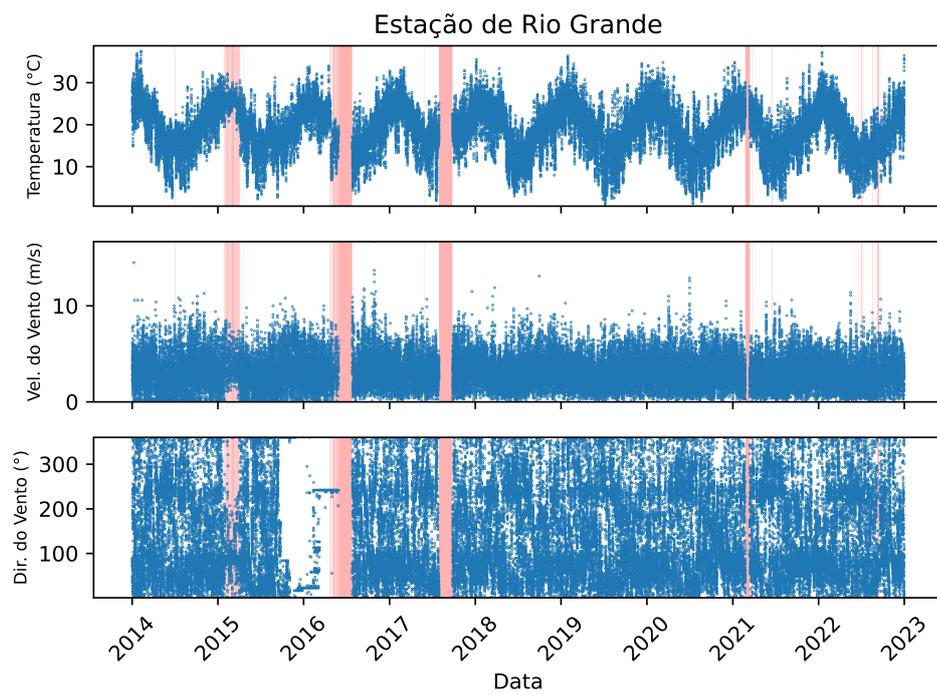


Figura 4.7: Dados faltantes da Estação de Santana (Fonte: Elaborado pelo autor).

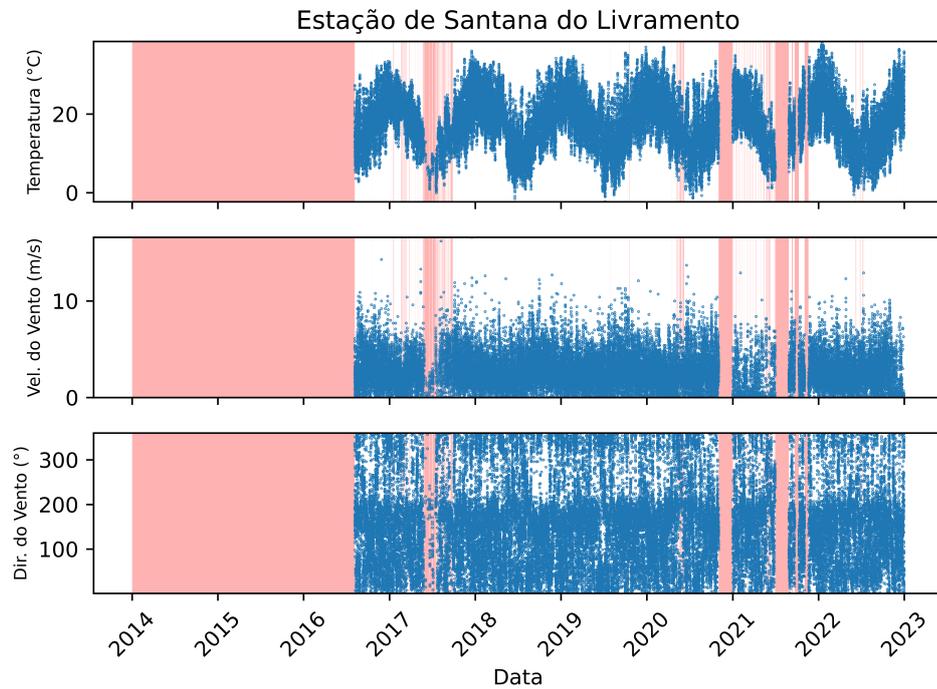
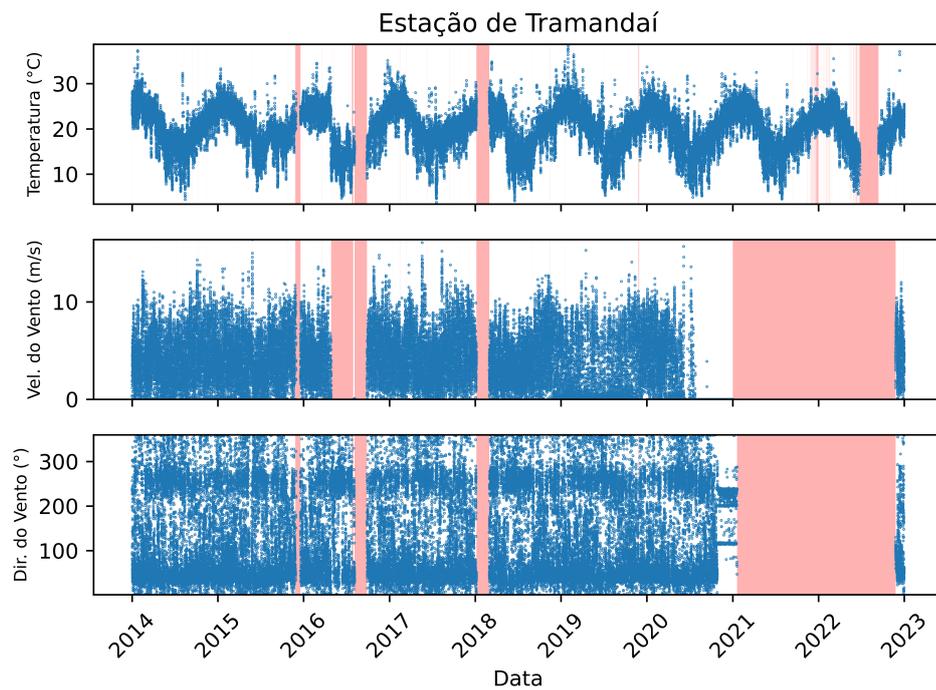


Figura 4.8: Dados faltantes da Estação de Tramandaí (Fonte: Elaborado pelo autor).

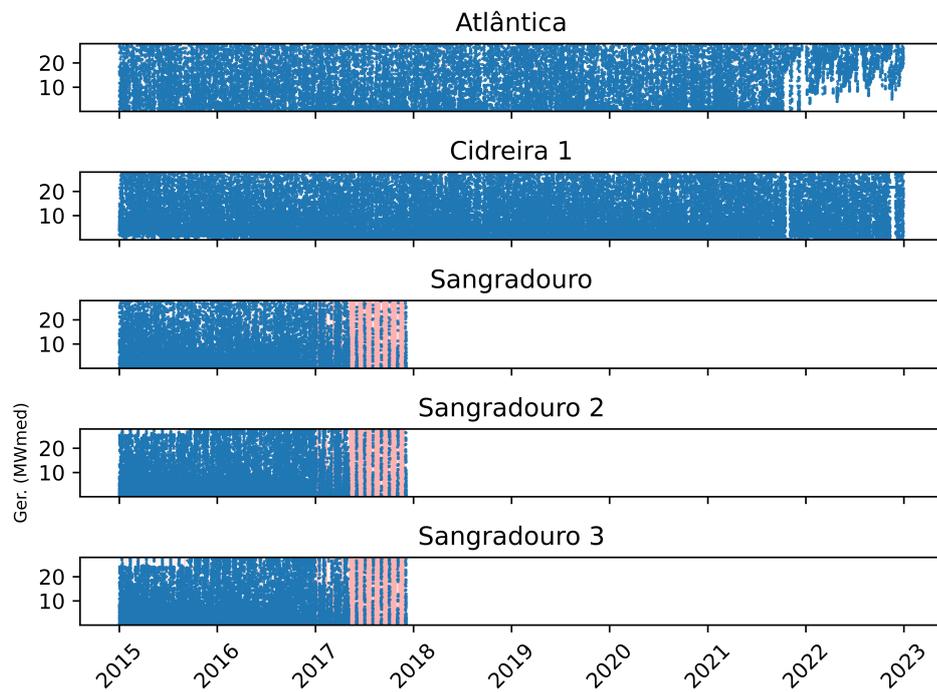


Examinando os períodos de dados ausentes nas estações, observa-se que o intervalo entre o início de 2014 e o início de 2021 exibe padrões mais favoráveis, com lacunas que podem ser preenchidas sem impactos significativos. As exceções são a estação de Santana do Livramento, cujos dados estão disponíveis apenas a partir da segunda metade de 2016, e a estação de Acaraú, que não possui dados a partir de 2020. Portanto, visando manter um intervalo consistente entre as estações, abrangendo a maioria dos dados válidos, optou-se por não considerar estas estações. Com os dados das estações analisados, procedeu-se, em seguida, a um estudo dos dados das usinas.

### 4.2.2 Análise de dados das usinas

Para a estação de Tramandaí, no Rio Grande do Sul, foram analisadas as 5 usinas eólicas dentro do raio definido.

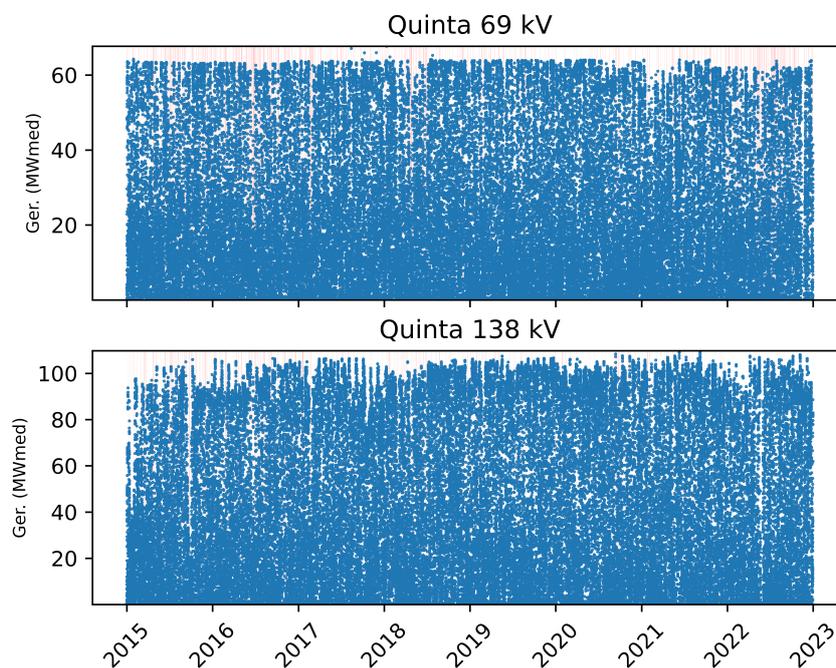
Figura 4.9: Dados faltantes das usinas próximas à estação de Tramandaí (Fonte: Elaborado pelo autor).



As usinas de Sangradouro, Sangradouro 2 e Sangradouro 3 apresentam dados somente até 2018, enquanto as usinas de Atlântica e Cidreira 1 apresentam dados válidos durante todo o período proposto.

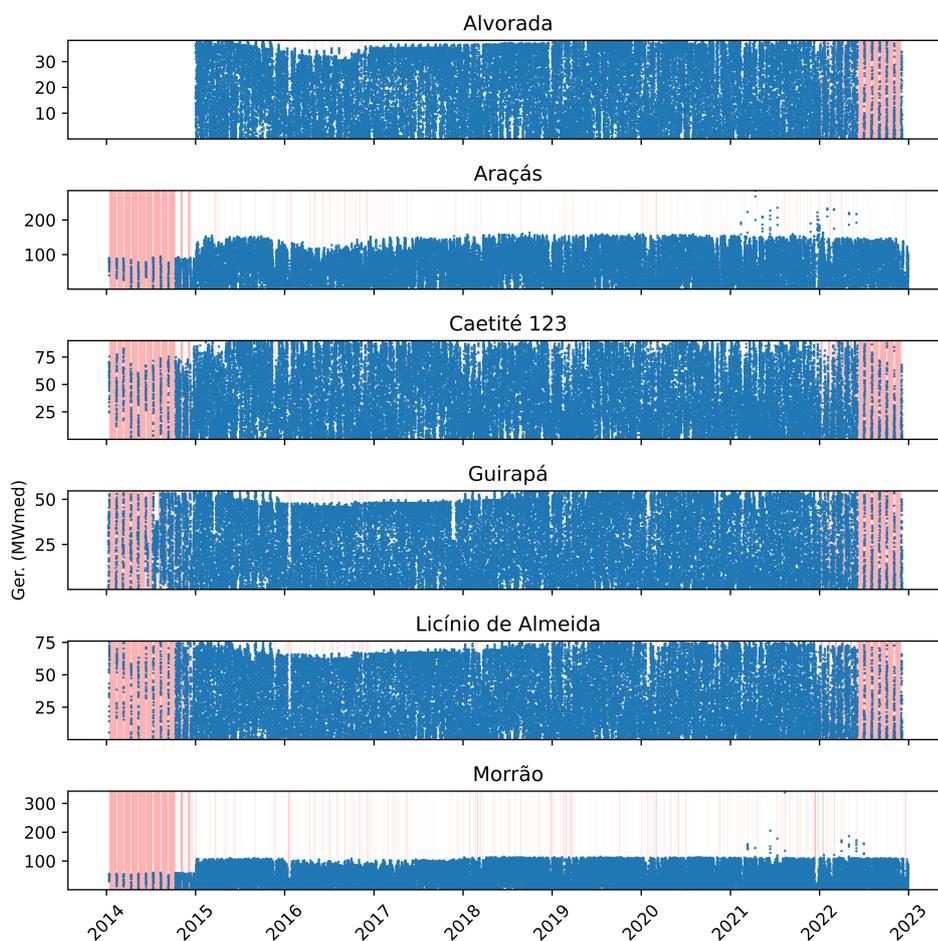
Em seguida foram analisadas as usinas Quinta 69 kV e Quinta 138 kV, próximas à estação de Rio Grande, no Rio Grande do Sul.

Figura 4.10: Dados faltantes das usinas próximas à estação de Rio Grande (Fonte: Elaborado pelo autor).



Como as duas usinas apresentaram dados válidos no período, ambas foram utilizadas no estudo. Na sequência foram analisadas as usinas próximas à estação de Guanambi, na Bahia.

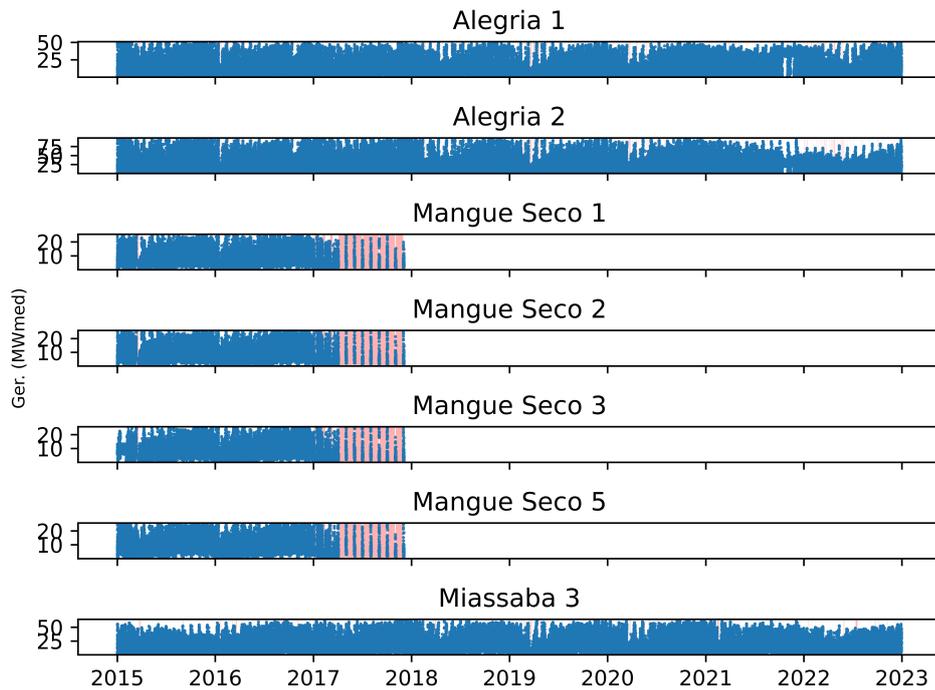
Figura 4.11: Dados faltantes das usinas próximas à estação de Guanambi (Fonte: Elaborado pelo autor).



Na Figura 4.11, é evidente o desempenho satisfatório das seis usinas ao longo do período de 2015 a 2021. Embora o intervalo inicialmente estabelecido tenha abrangido o período de 2014 a 2021, a estação de Guanambi se destacou ao possuir o maior número de usinas nas proximidades com dados válidos. Por esse motivo, optou-se por reduzir o intervalo em um ano, evitando assim a exclusão dessas usinas do escopo do estudo.

Para a estação de Macau, no Rio Grande do Norte, foram analisadas 7 usinas, como pode ser visto na Figura 4.12.

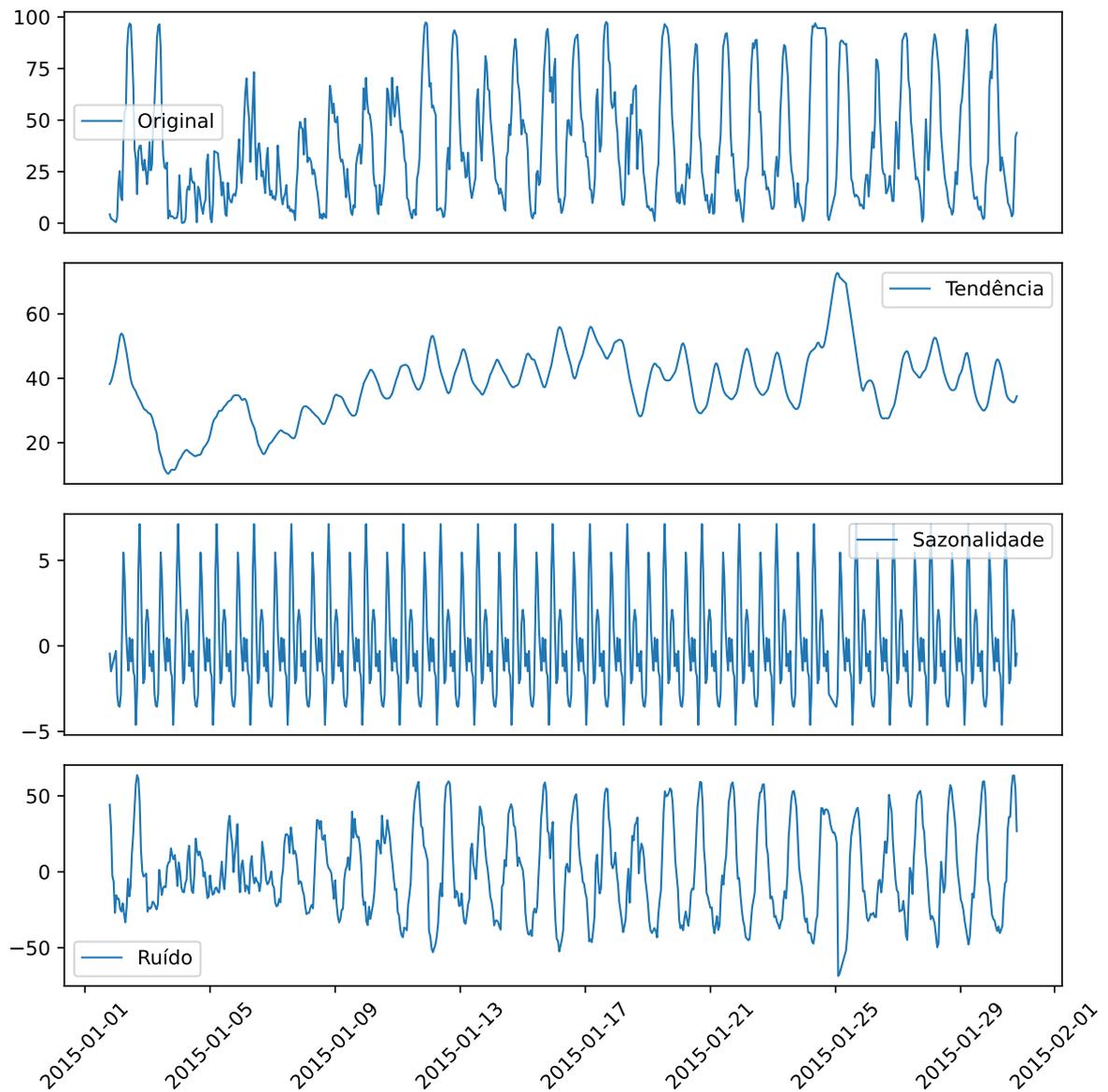
Figura 4.12: Dados faltantes das usinas próximas à estação de Macau (Fonte: Elaborado pelo autor).



As usinas vinculadas ao grupo Mangue Seco exibem padrões de comportamento bastante similares, até o final do ano de 2017. Contudo, nesse mesmo ano, observa-se a presença de lacunas extensas nos dados, levando à decisão de excluir as quatro usinas do escopo do estudo devido à falta de informações confiáveis. Portanto, para esta estação específica, os dados provenientes das usinas Alegria 1, Alegria 2 e Miassaba 3 serão utilizados como base para as próximas análises.

Com o intuito de compreender o padrão das séries de geração de energia das usinas, procedeu-se com a decomposição em tendência, sazonalidade e ruído. A análise da decomposição da série de geração da usina de Alegria 2 está ilustrada na Figura 4.13, destacando-se a clara presença de uma sazonalidade diária. Para facilitar a visualização, a série foi plotada para um período de 1 mês. Esse comportamento sazonal é consistente ao longo de todas as séries durante o período analisado, exceto nos momentos em que há lacunas nos dados. Nestas instâncias, torna-se evidente a natureza não estacionária das séries.

Figura 4.13: Decomposição da série temporal de geração da usina Alegria 2 (Fonte: Elaborado pelo autor).

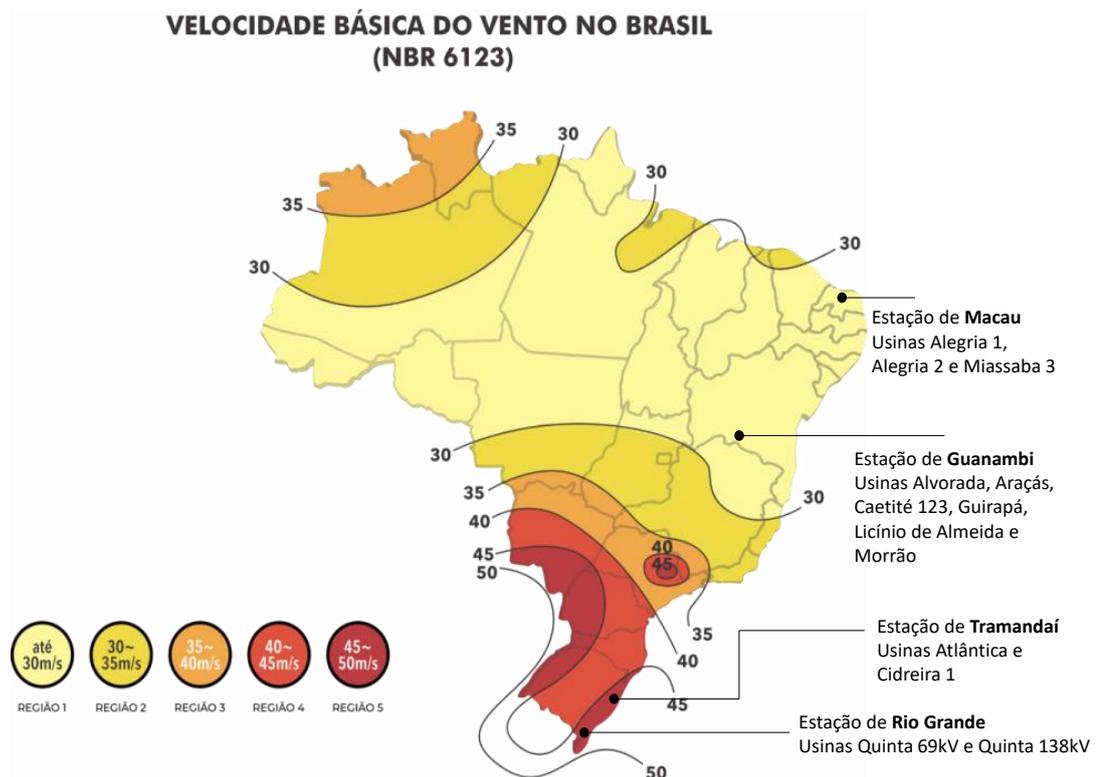


Finalmente, foram definidas as 13 usinas a serem utilizadas no estudo, conforme a Tabela 4.2. A localização das usinas e estações estão destacadas no mapa de isopletas da Figura 4.14, que apresenta as velocidades básicas do vento em diferentes regiões do Brasil.

Tabela 4.2: Estações e usinas utilizadas no estudo.

Estação	Usina	Distância (km)
Tramandaí	Atlântica	21,421
	Cidreira 1	8,484
Rio Grande	Quinta 69 kV	16,435
	Quinta 138 kV	16,435
Guanambi	Alvorada	12,142
	Araçás	19,107
	Caetité 123	27,712
	Guirapá	16,764
	Licínio de Almeida	12,085
	Morrão	11,47
Macau	Alegria 1	21,27
	Alegria 2	19,815
	Miassaba 3	13,87

Figura 4.14: Mapa de Isopletas contendo as estações e usinas utilizadas no estudo (Fonte: Adaptado de Solar (2023)).



Após a seleção das usinas para o estudo, a abordagem para preencher lacunas nos dados meteorológicos das estações foi explorada. Bier e Ferraz (2017) realizaram uma

comparação de métodos específicos para esse contexto em estações localizadas no Sul do Brasil. Destacaram-se os métodos de Regressão Linear, Ponderação Regional (PR) e o Método Tradicional do Reino Unido (UK). Os métodos PR e UK possuem dependência geográfica ao utilizar dados de outras estações próximas à estação em análise. No entanto, neste trabalho, foi estabelecido um raio de 33 km entre as estações selecionadas, e nenhuma estação selecionada possui outra estação próxima, optou-se então por utilizar o método de regressão linear para preencher lacunas nas séries temporais.

### 4.3 Arquitetura do Modelo

Foi utilizado um modelo sequencial com uma camada LSTM com 1024 neurônios seguida por uma camada densa de 512 neurônios utilizando a função de ativação ReLU, que introduz não-linearidade na rede. A vantagem de usar a função ReLU é que ela não ativa todos os neurônios da camada, tornando a rede mais eficiente e facilitando a computação devido à sua característica esparsa (DEEPLARNINGBOOK, 2022). Uma camada de *dropout* foi aplicada para regularizar a rede para prevenir problemas de *overfitting*. A última camada é densa com um número de camadas igual ao número de previsões que se deseja gerar. O treinamento foi realizado por um total de 200 épocas utilizando dados de treino e validação e um tamanho de lote de 32 amostras.

A função de perda utilizada foi a *MSE*. Como discutido no Capítulo 2 na Seção 2.7, a desvantagem do uso do *MSE* se dá devido à presença de *outliers*. Para prevenir isso, foi utilizado o Intervalo Interquartil (IQR) como medida para identificar e lidar com valores discrepantes.

Na Figura 4.15 é possível ver um esquema do modelo adotado aqui, contendo as camadas mencionadas acima, assim como as camadas de *input* e *output*.

Figura 4.15: Arquitetura LSTM adotada aqui (Fonte: Elaborado pelo autor).

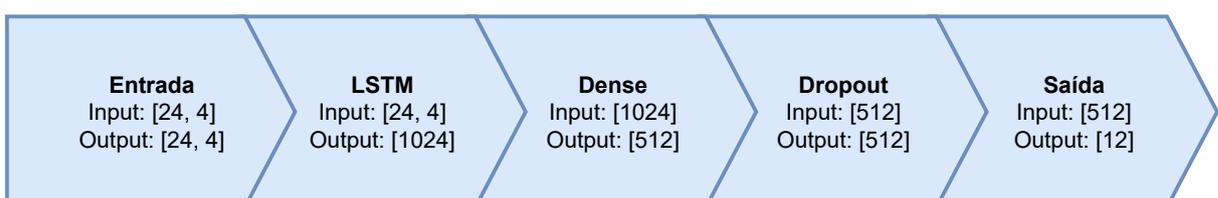


Tabela 4.3: Camadas e respectiva quantidade de neurônios empregados no modelo.

Camada	Unidades
Input (look_back)	24
LSTM	1024
Densa (Função ReLU)	512
Dropout	512
Output (look_forward)	12

Para evitar problemas de *overfitting* foi utilizado o recurso *Early Stopping*, no qual é monitorada a função de perda e, caso não haja uma melhoria  $\delta$  dentro de  $x$  épocas, o treinamento é interrompido. Neste modelo foram utilizados os valores  $\delta = 0,001$  e  $x = 30$ , ou seja, se dentro de 30 épocas não houver melhora de 0,001 na função de perda, nesse caso o *MSE*, o treinamento é concluído.

Para o funcionamento da LSTM são definidos dois padrões específicos de forma que o modelo consiga identificar padrões temporais nos dados. O *look\_back*, estabelecido em 24 horas, representa a quantidade de períodos temporais anteriores considerados para antecipar o próximo valor. Por outro lado, o *look\_forward*, com um intervalo de 12 horas, aponta para o horizonte de previsão, ou seja, o número de períodos futuros que o modelo se esforça para prever. A Tabela 4.3 apresenta as camadas do modelo e o número de neurônios em cada uma delas.

Os dados de treino, validação e teste foram separados da seguinte forma (considerando o intervalo de 2015 ao início de 2021):

- Treino: 2015, 2016, 2017;
- Validação: 2018;
- Teste: 2019, 2020.

Os códigos nos quais os modelos foram implementados podem ser encontrados em [https://github.com/nico-est/previsao\\_energiaolica](https://github.com/nico-est/previsao_energiaolica).

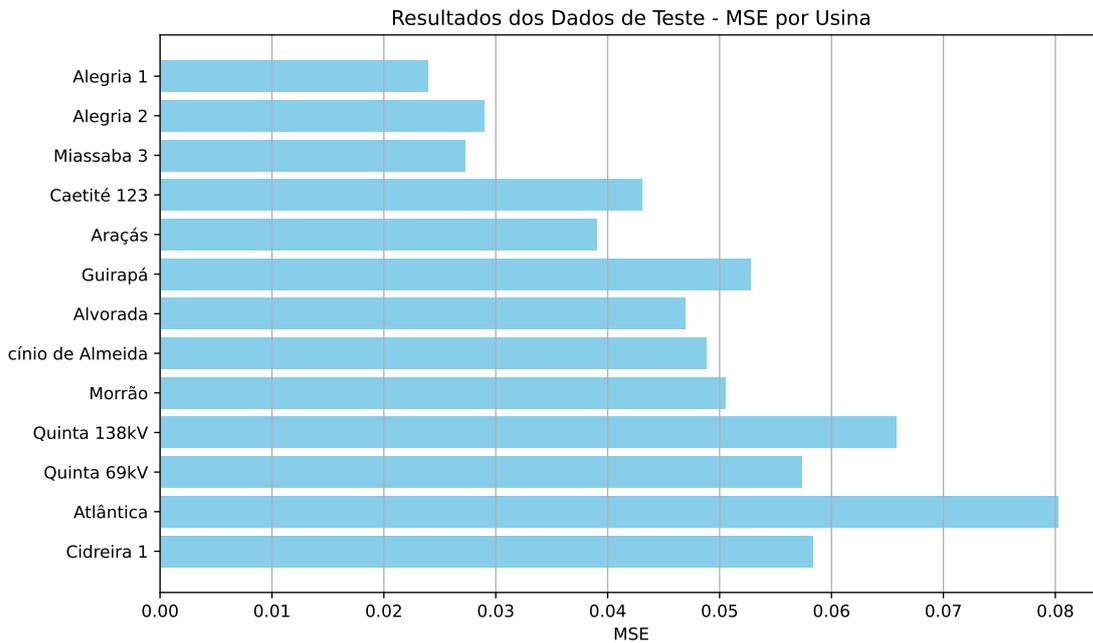
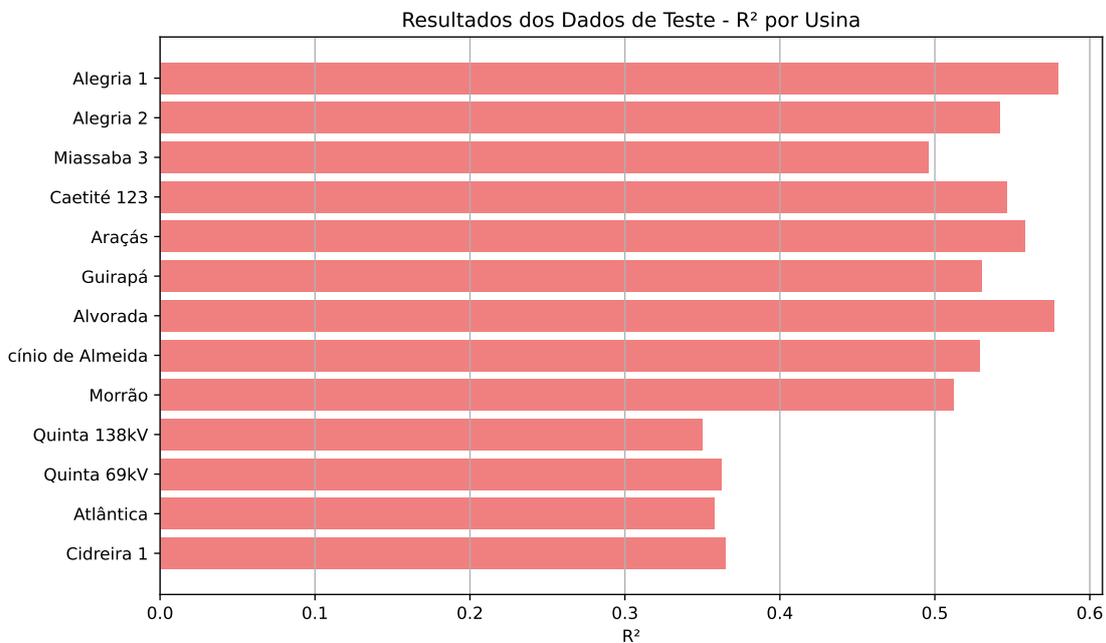
## 5 Resultados

Foram analisadas as 13 usinas em conjunto com suas referentes estações, apontadas na Tabela 4.2. Optou-se por uma comparação das métricas de  $MSE$ ,  $MAE$ ,  $R^2$  e  $RMSE$ . Os resultados obtidos para o conjunto de teste estão representados na Tabela 5.1. As Figuras 5.1 e 5.2 ilustram graficamente os dados para as métricas  $MSE$  e  $R^2$ .

Tabela 5.1: Resultados dos dados de teste

Estação	Usina	Distância (km)	MSE	MAE	R2	RMSE
Tramandaí	Cidreira 1	8,484	0,0583	0,1919	0,3647	0,2415
	Atlântica	21,421	0,0802	0,2295	0,3578	0,2833
Rio Grande	Quinta 69kV	16,435	0,0573	0,1892	0,3620	0,2395
	Quinta 138kV	16,435	0,0658	0,2074	0,3498	0,2564
Guanambi	Morrão	11,470	0,0505	0,1809	0,5121	0,2247
	Licínio de Almeida	12,085	0,0488	0,1746	0,5288	0,2209
	Alvorada	12,142	0,0469	0,1695	0,5771	0,2165
	Guirapá	16,764	0,0528	0,1822	0,5304	0,2298
	Araçás	19,107	0,0390	0,1558	0,5583	0,1976
	Caetité 123	27,712	0,0431	0,1622	0,5467	0,2075
Macau	Miassaba 3	13,870	0,0273	0,1293	0,4960	0,1653
	Alegria 2	19,185	0,0290	0,1330	0,5422	0,1703
	Alegria 1	21,270	0,0239	0,1216	0,5792	0,1547

Figura 5.1: Resultados de MSE para os dados de teste (Fonte: elaborado pelo autor)

Figura 5.2: Resultados de  $R^2$  para os dados de teste (Fonte: elaborado pelo autor)

Verifica-se que as métricas obtidas para as usinas próximas a uma mesma estação possuem valores relativamente próximos, um comportamento esperado uma vez que que

os dados das variáveis temperatura, velocidade do vento e direção do vento são os mesmos para estas usinas. Foram selecionadas 4 usinas representativas de cada conjunto, sendo a Figura 5.3 da usina de Morrão, próxima à estação de Guanambi, a Figura 5.4 da usina de Alegria 2, próxima à estação de Macau, a Figura 5.5 da usina de Quinta 69kV, próxima à estação de Rio Grande e, por fim, a Figura 5.6 da usina de Cidreira 1, próxima à estação de Tramandaí.

Figura 5.3: Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Morrão (Fonte: elaborado pelo autor)

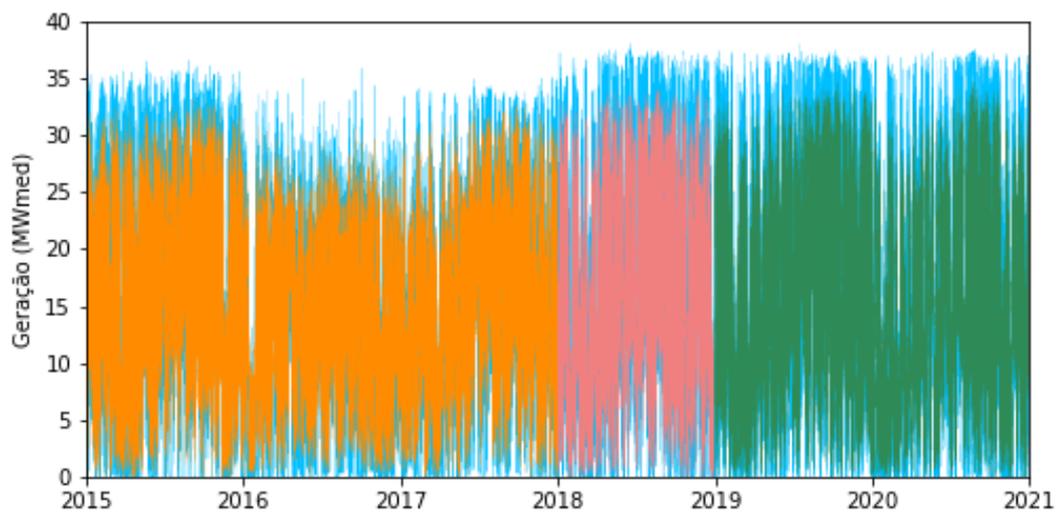


Figura 5.4: Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Alegria 2 (Fonte: elaborado pelo autor)

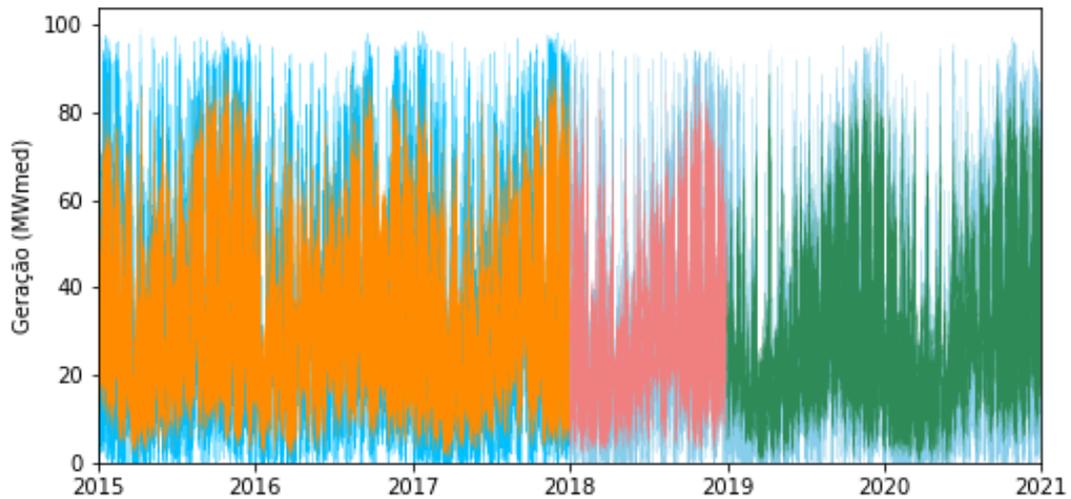


Figura 5.5: Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Quinta 69kV (Fonte: elaborado pelo autor)

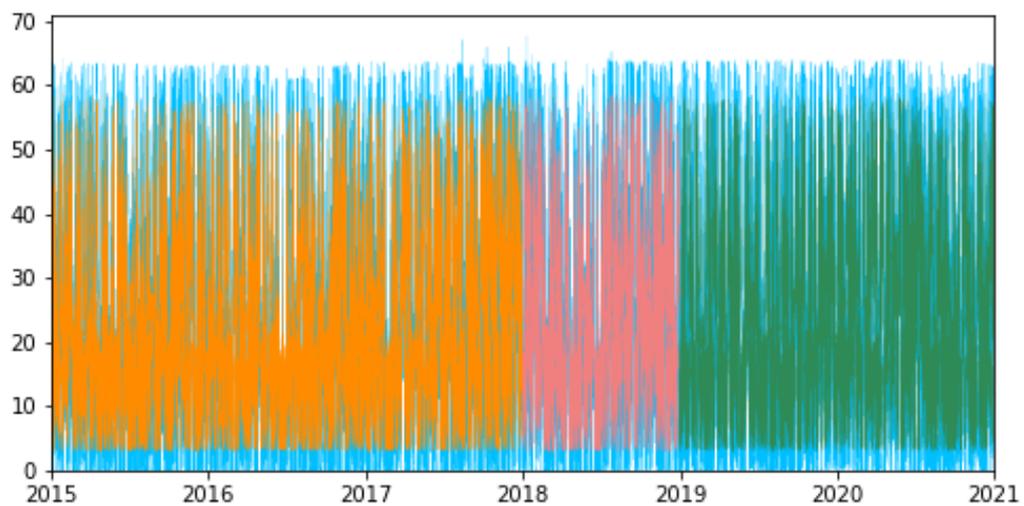
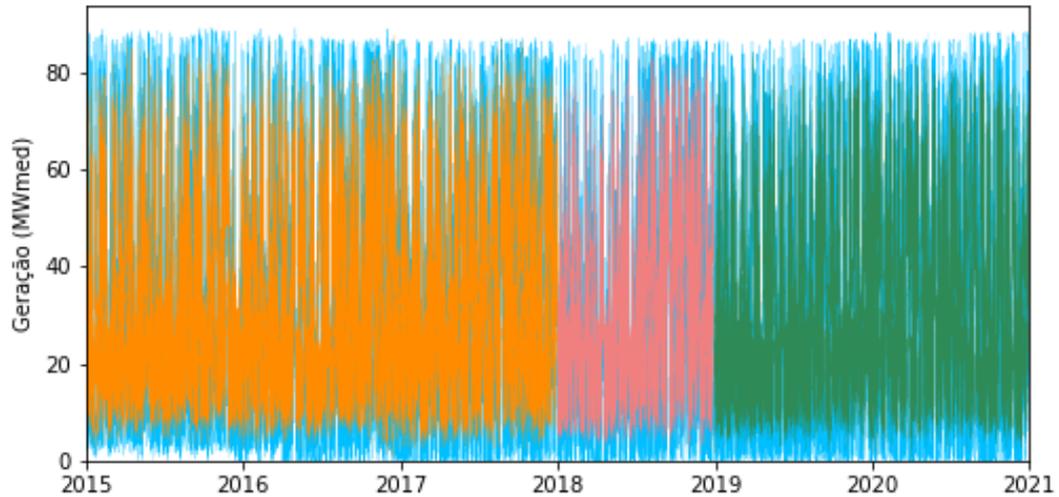


Figura 5.6: Comparação entre dados reais (azul) com previsões dos conjuntos de treino (laranja), validação (rosa) e teste (verde) para a usina de Cidreira 1 (Fonte: elaborado pelo autor)



Nas Figuras 5.7, 5.8, 5.9 e 5.10 estão representados os gráficos das métricas de desempenho para os dados de treino e validação da usina de Quinta 69kV.

Figura 5.7: Resultado do  $MSE$  para a usina de Quinta 69kV (Fonte: elaborado pelo autor)

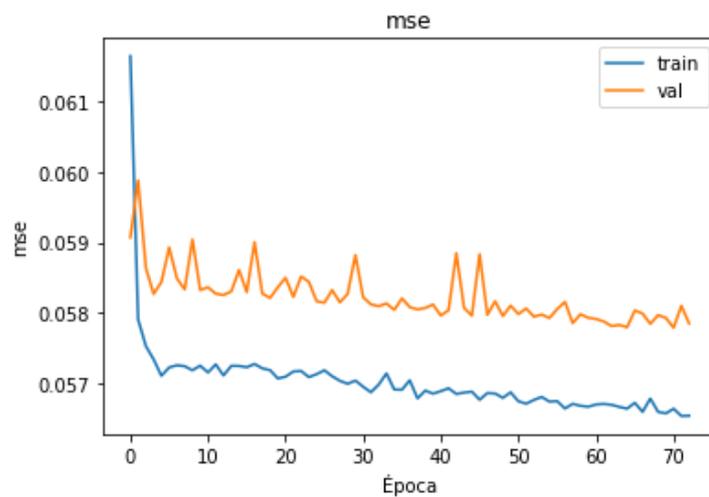


Figura 5.8: Resultado do  $RMSE$  para a usina de Quinta 69kV (Fonte: elaborado pelo autor)

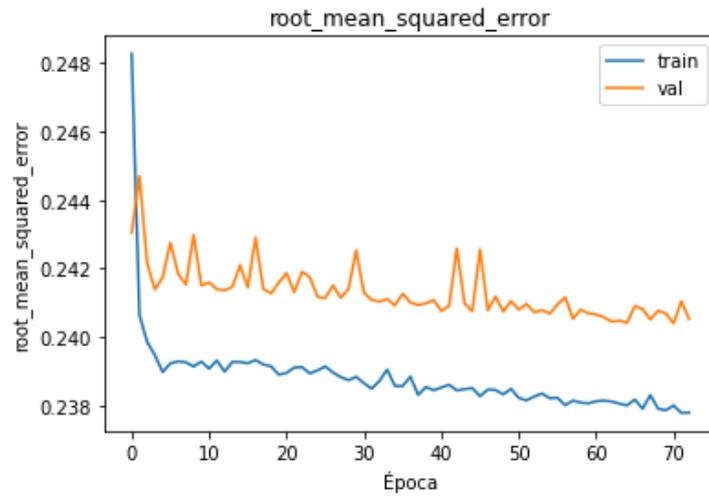


Figura 5.9: Resultado do  $MAE$  para a usina de Quinta 69kV (Fonte: elaborado pelo autor)

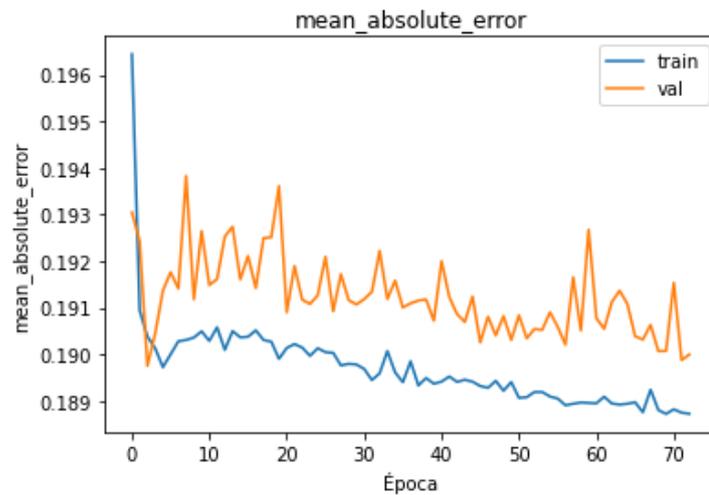
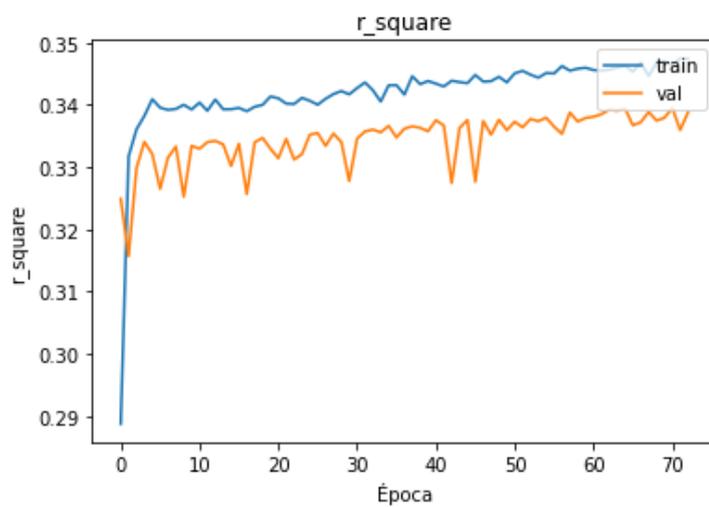


Figura 5.10: Resultado do  $R^2$  para a usina de Quinta 69kV (Fonte: elaborado pelo autor)



O mesmo comportamento é observado para todas as usinas. As métricas de desempenho  $MSE$ ,  $MAE$  e  $RMSE$  ao longo das épocas do conjunto de validação apresentam valores inferiores quando comparadas ao conjunto de treino. Este comportamento é esperado uma vez que os dados de treino são já conhecidos pelo modelo, enquanto os dados de validação são considerados novos. O coeficiente  $R^2$  apresenta comportamento contrário, tendo valor menor para o conjunto de validação, o que, dada a natureza inversa da métrica, corrobora para o comportamento esperado do modelo.

Para comparar os valores previstos com os reais foram gerados os gráficos de dispersão, ilustrados nas Figuras 5.11 e 5.12, para as usinas de Alegria 2 e Quinta 138kV, respectivamente. Os gráficos mostram que as variações entre as previsões e os dados reais são bem grandes e o modelo possui dificuldade em lidar com valores próximos aos máximos e mínimos das séries de geração.

Figura 5.11: Gráfico de Dispersão para os dados de teste da usina de Alegria 2 (Fonte: elaborado pelo autor)

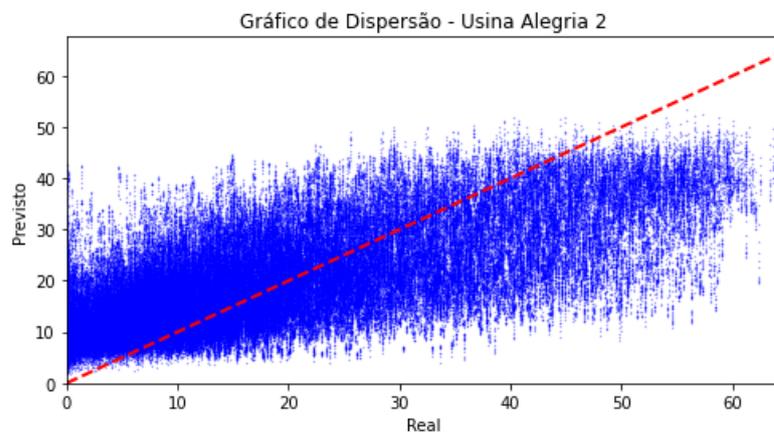
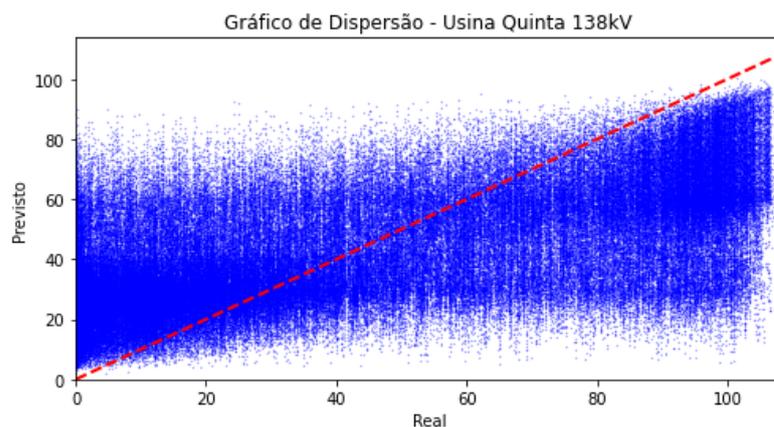
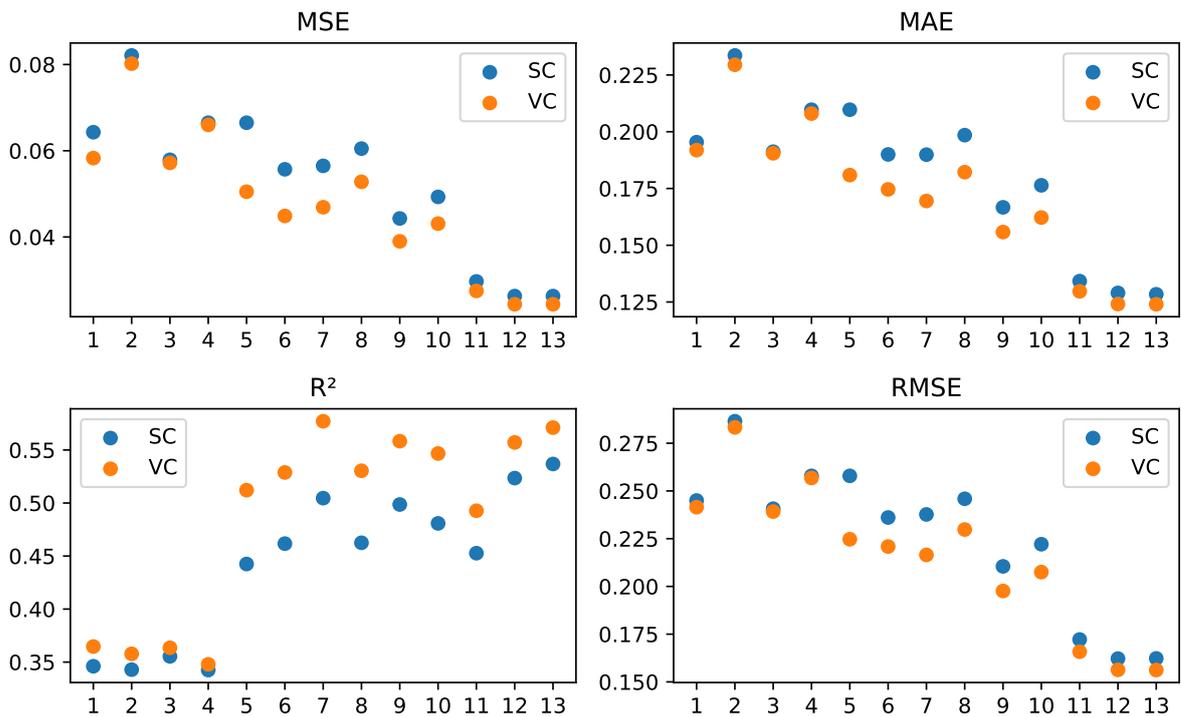


Figura 5.12: Gráfico de Dispersão para os dados de teste da usina de Quinta 138kV (Fonte: elaborado pelo autor)



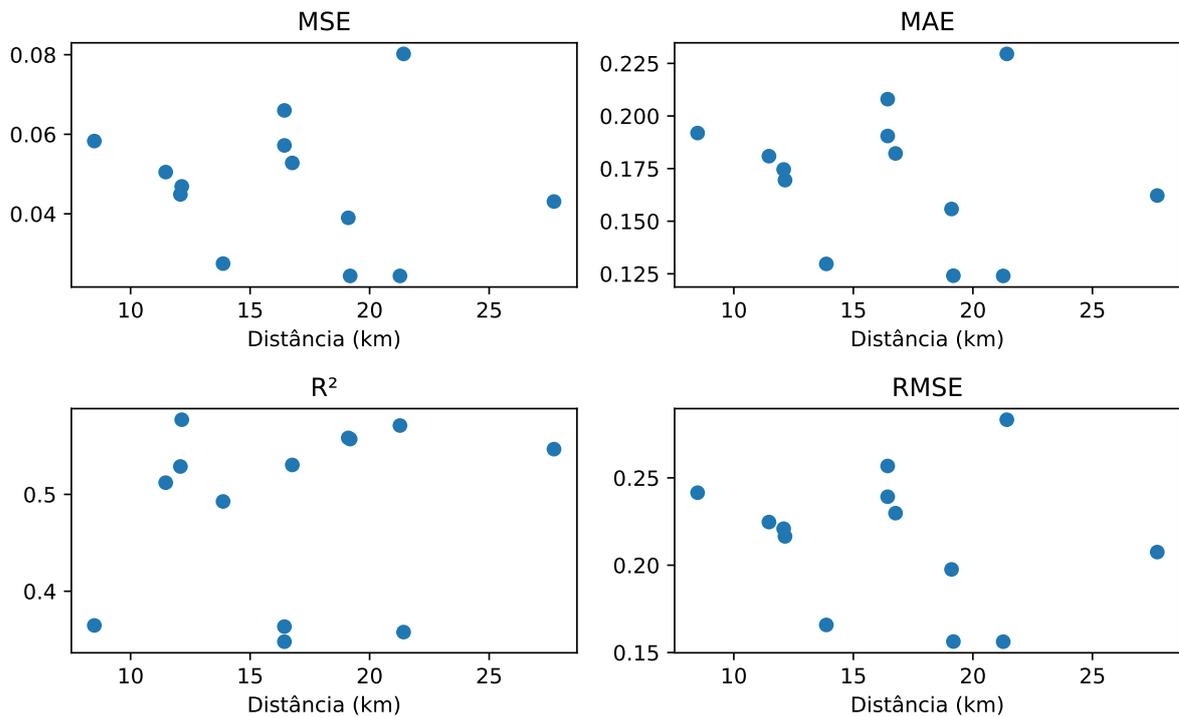
Foram também realizados testes para todas as usinas sem a inclusão das variáveis climáticas, de forma a avaliar a influência destas variáveis nos resultados. Na Figura 5.13 é possível visualizar através das métricas o desempenho superior do modelo com as variáveis climáticas, tendo valores inferiores de  $MSE$ ,  $MAE$  e  $RMSE$  e valores superiores para  $R^2$ .

Figura 5.13: Comparação de métricas para os modelos sem e com as variáveis climáticas, respectivamente representados por SC e VC (Fonte: elaborado pelo autor)



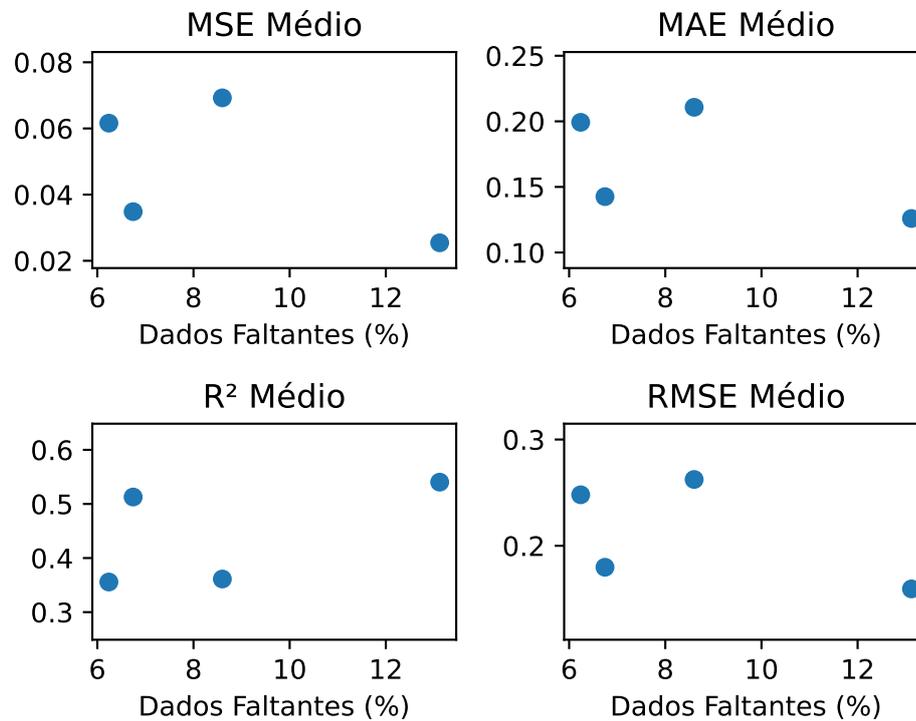
Para investigar a possível influência da distância entre as usinas e as estações na qualidade dos resultados, foram gerados os gráficos apresentados na Figura 5.14. Os resultados indicam que não há uma relação direta entre a distância e a qualidade das previsões, podendo ser considerados consistentes dado que, levando em conta o raio máximo de 33km, era pouco provável uma relação significativa.

Figura 5.14: Relação das métricas de desempenho com as distâncias entre estações e usinas (Fonte: elaborado pelo autor).



Como forma de avaliar os dados faltantes das séries em relação aos resultados, os dados faltantes foram quantificados em porcentagem com relação ao período total analisado. Foi verificado que, entre as usinas próximas a uma mesma estação, a quantidade de dados faltantes é muito similar ou igual, dessa forma, foram definidos quatro conjuntos por estação e definido a média das métricas de desempenho, como apresentado na Figura 5.15.

Figura 5.15: Relação das médias das métricas de desempenho por conjunto de usinas próximas a uma mesma estação e a porcentagem de dados faltantes (Fonte: elaborado pelo autor).



Conforme destacado na Seção 3 pelos trabalhos de Batista (2003) e Costa (2018), a análise da influência da quantidade de dados faltantes pode resultar em alta variância e, uma maior quantidade de dados faltantes não necessariamente possui taxas de erros menores. Os melhores resultados obtidos, conforme a Tabela 5.1, foram para a estação de Macau, com a maior quantidade de dados faltantes (13,12%), entretanto, não se pode considerar que esse conjunto possui as melhores previsão, dado que os dados imputados não possuem o mesmo comportamento imprevisível dos dados reais.

Outra observação importante é a de que as estações de Rio Grande e Tramandaí, que tiveram resultados inferiores, possuem uma quantidade maior de valores próximos a zero para as variáveis de velocidade do vento e direção do vento. Isso pode ter um impacto na capacidade de previsão do modelo, já que as estações de Guanambi e Macau possuem um padrão mais evidente com valores diferentes de zero e resultaram em taxas de erro menores.

## 6 Considerações Finais e Trabalhos Futuros

O objetivo principal desse trabalho era prever a geração de energia eólica utilizando séries temporais obtidas via dados públicos. Além disso, era de interesse conseguir avaliar a influência das etapas de pré-processamento nos resultados do modelo.

Embora o modelo proposto tenha demonstrado capacidade para prever a geração de energia eólica, a comparação com dados reais revela resultados insatisfatórios em situações próximas a zero ou próximas ao valor máximo das séries de geração. Isso sugere a necessidade de ajustes no modelo, visando melhorar a qualidade dos resultados. A robustez da arquitetura LSTM, aliada aos dados climáticos — comprovadamente correlacionados à geração de energia eólica, conforme indicado pela literatura — revelou-se eficaz na previsão da produção elétrica das usinas.

A ênfase na utilização de dados públicos provenientes de estações e usinas ressalta a viabilidade dessas informações para pesquisas na área. No entanto, destaca-se a importância de empregar diversas técnicas de pré-processamento para alcançar resultados otimizados. Além do estudo sobre a viabilidade do uso de dados das estações em relação à distância, a significativa quantidade de dados faltantes nas séries de entrada representou um desafio na busca pelos resultados desejados.

O estudo abrangeu estações e usinas em diversas regiões do Brasil, revelando comportamento semelhante do modelo, com variações nos resultados devido a características específicas das séries de cada usina e às condições climáticas das respectivas regiões.

Não foi possível encontrar uma relação direta entre a distância entre as usinas e estações e a capacidade de previsão do modelo. Como o raio máximo definido foi de 33km, essa distância é relativamente pequena para que essa avaliação pudesse ser feita, sendo essa considerada uma limitação do estudo.

Quanto à imputação de dados, esperava-se que uma maior porcentagem de dados faltantes nas séries resultasse em previsões menos precisas. Entretanto, a utilização de um método simples de imputação, principalmente um método linear, pode ter simplificado as séries. Uma sugestão para trabalhos futuros é explorar a comparação de resultados

sem a inclusão dos pontos imputados, mesmo que isso implique em uma abordagem não convencional na avaliação.

Como possíveis trabalhos futuros, recomenda-se a exploração de métodos alternativos para a imputação de dados faltantes, a investigação e comparação de diferentes métricas de desempenho como função de perda no modelo, o aumento do raio máximo, aliviando as restrições de correlação, e a utilização de algoritmos de otimização para ajuste de parâmetros na rede, como número de neurônios, características do Early Stopping, ou mesmo a arquitetura da rede com inclusão ou remoção de camadas.

## Bibliografia

- AGGARWAL, C. C. *Neural Networks and Deep Learning*. 1. ed. [S.l.]: Springer, 2018.
- BATISTA, G. E. de A. P. A. *Pré-processamento de Dados em Aprendizado de Máquina Supervisionado*. Dissertação (Mestrado) — ICMC - USP, São Carlos, 2003.
- BIER, A. A.; FERRAZ, S. E. T. Comparação de metodologias de preenchimento de falhas em dados meteorológicos para estações no sul do Brasil. *Revista Brasileira de Meteorologia*, v. 32, n. 2, p. 215–226, 2017.
- BILGILI, M.; SAHIN, B.; YASAR, A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renewable Energy*, v. 32, p. 2350 – 2360, 2007.
- BOX, G. E. P. et al. *Time Series Analysis: Forecasting and Control*. 5. ed. [S.l.]: WILEY, 2016.
- BROWNLEE, J. *How to Decompose Time Series Data into Trend and Seasonality*. 2021. Acesso em 24/11/2023. Disponível em: <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>.
- CALI, U.; SHARMA, V. Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. *International Journal of Smart Grid and Clean Energy*, v. 8, p. 103 – 110, 2019.
- CENTRAL, C. *The hottest 12-month stretch in recorded history*. 2023. Disponível em: <https://www.climatecentral.org/press-releases>.
- CHATFIELD, C. *The Analysis of Time Series: An Introduction*. 5. ed. [S.l.]: CHAPMAN HALL/CRC, 1996.
- CIABURRO, G.; VENKATERWARAN, B. *Neural Networks with R: Smart models using cnn, rnn, deep learning, and artificial intelligence principles*. 1. ed. [S.l.]: Packt Publishing, 2017.
- COSTA, J. C. J. *Uma Análise sobre o Impacto de Dados Faltantes no Desempenho de Métodos de Aprendizado de Máquina*. Dissertação (Mestrado) — Universidade Federal do Ceará, Fortaleza, 2018.
- CRESESB, C. de Referência para Energias Solar e Eólica Sérgio de S. B. *O Recurso Eólico*. 2008. Acesso em 18/12/2023. Disponível em: [http://www.cresesb.cepel.br/index.php?section=com\\_content&lang=pt&cid=211](http://www.cresesb.cepel.br/index.php?section=com_content&lang=pt&cid=211).
- DAVIDSON, J. G. M. R. *Econometric Theory and Methods*. illustrated edition. [S.l.]: Oxford University Press, USA, 2003. ISBN 9780195123722; 0195123727.
- DEEPLARNINGBOOK. *As 10 Principais Arquiteturas de Redes Neurais*. 2022. Acesso em 27/11/2023. Disponível em: <https://www.deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais/>.

DUTRA, R. *Energia Eólica - Princípios e Tecnologias*. [S.l.]: Centro de Referência para Energia Solar e Eólica Sérgio de Salvo Brito, 2008.

ENERGIA, M. de Minas e. *Transição energética: a mudança de energia que o planeta precisa*. 2023. Acesso em 17/12/2023. Disponível em: <https://www.gov.br/mme/pt-br/assuntos/noticias/transicao-energetica-a-mudanca-de-energia-que-o-planeta-precisa>.

FUJISE, J.

*Desenvolvimento de Modelo de Previsão de Geração Eólica para Planejamento Energético Utilizando Redes Neurais Artificiais* — Escola Politécnica da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

GALVÃO, L. *Regressão linear: um dos tipos mais importantes de análise de dados*. 2019. Acesso em 30/11/2023. Disponível em: [https://ligiagalvao.medium.com/ regresso-linear-um-dos-tipos-mais-importantes-de-anlise-de-dados-827cfdbd2229](https://ligiagalvao.medium.com/ regresso-linear-um-dos-tipos-mais-importantes-de-analise-de-dados-827cfdbd2229).

HAYKIN, S. *Neural Networks and Learning Machines*. 3. ed. [S.l.]: Pearson, 2017.

HOCHREITER, S.; SCHMIDHUBER, J. Long short term memory. *Neural Computation*, v. 9, p. 1735 – 1780, 1997.

IBM. *O que são redes neurais recorrentes?* 2021. Acesso em 27/11/2023. Disponível em: <https://www.ibm.com/br-pt/topics/recurrent-neural-networks#O+que+s%C3%A3o+redes+neurais+recorrentes%3F>.

INMET, I. N. de M. *Rede de Estações Meteorológicas Automáticas do INMET*. 2011.

KARAMAN, O. A. Prediction of wind power with machine learning models. *Applied Sciences*, v. 13, 2023.

LA, F. L. *Como as redes neurais aprendem?* 2019. Acesso em 03/12/2023. Disponível em: <https://learn.microsoft.com/pt-br/archive/msdn-magazine/2019/april/artificially-intelligent-how-do-neural-networks-learn>.

LEI, M. et al. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, v. 13, p. 915–920, 2009.

OLAH, C. *Understanding LSTM Networks*. 2015. Acesso em 27/11/2023. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.

ONS, O. N. do S. E. *EVOLUÇÃO DA CAPACIDADE INSTALADA NO SIN - NOVEMBRO 2023/ DEZEMBRO 2027*. 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>.

PICOLO, A. P.; BUHLER, A. J.; RAMPINELLI, G. A. Uma abordagem sobre a energia eólica como alternativa de ensino de tópicos de física clássica. *Caderno Brasileiro de Ensino de Física*, v. 36, n. 4, 2014.

PLEVRIS, V. et al. Investigation of performance metrics in regression analysis and machine learning-based prediction models. *The 8th European Congress on Computational Methods in Applied Sciences and Engineering - ECCOMAS*, 2022. Disponível em: [https://www.scipedia.com/public/Plevris\\_et\\_al.2022a](https://www.scipedia.com/public/Plevris_et_al.2022a).

REIS, L. B. d. *GERACAO DE ENERGIA ELÉTRICA*. [S.l.]: Manole, 2011.

- REPÚBLICA, P. da. *Decreto nº 10.946*. 2022. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2022/Decreto/D10946.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/Decreto/D10946.htm).
- ROQUE, A. *Estatística Aplicada à Educação*. 2017. Acesso em 30/11/2023. Disponível em: <https://edisciplinas.usp.br/mod/resource/view.php?id=2831726>.
- SOLAR, M. *Velocidade Básica do Vento no Brasil*. 2023. Acesso em 20/12/2023. Disponível em: <https://www.metallightsolar.com/informacoes-gerais/>.
- USP. *Correlação e Regressão*. 2017. Acesso em 30/11/2023. Disponível em: <https://edisciplinas.usp.br/mod/resource/view.php?id=1699883>.
- VALÉRIO, P. *Projeto de Lei nº 3386*. 2021. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/150062>.
- WAND, Y. et al. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, v. 304, 2021.
- YUKIO. *Localizando Outliers Através do Intervalo Interquartil*. 2018. Acesso em 03/12/2023. Disponível em: <https://estatsite.com.br/2018/12/01/localizando-outliers-atraves-do-intervalo-interquartil-boxplot-codigo-sas/>.
- ZOU, K. H.; TUNCALI, K.; SILVERMAN, S. G. Correlation and simple linear regression. *Radiology Society of North America*, v. 227, 2003.