

PREVISÃO DE VELOCIDADE DO VENTO NO NORDESTE BRASILEIRO USANDO APRENDIZADO PROFUNDO E DADOS LIDAR

Lucas Batista Pereira

Trabalho de Conclusão de Curso
Bacharelado em Ciência da Computação

Pereira, Lucas Batista

Previsão de Velocidade do Vento no Nordeste Brasileiro usando Aprendizado Profundo e dados LiDAR / Lucas Batista Pereira. - Itajubá: UNIFEI, 2024.

21 p.

Trabalho de Conclusão de Curso (Ciência da Computação) - Universidade Federal de Itajubá, Itajubá, 2024

Orientação: Vanessa Cristina Oliveira de Sousa

Coorientação: Flávio Belizário da Silva Mota

1. Previsão de vento. 2. Seq2Seq. 3. Transformada de Wavelet. 4. Aprendizado profundo. 5. Energia eólica. I. de Sousa, Vanessa Cristina Oliveira, orient. II. Mota, Flávio Belizário da Silva, coorient. III. Título.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE ITAJUBÁ
Criada pela Lei nº 10.435, de 24 de abril de 2002
Instituto de Matemática e Computação

FOLHA DE APROVAÇÃO

O Trabalho Final de Graduação em Ciência da Computação de **Lucas Batista Pereira** intitulado *Previsão de velocidade do vento no nordeste brasileiro usando aprendizado profundo e dados LiDAR* foi **APROVADO**, em sessão pública realizada no dia **09/12/2024**.

BANCA EXAMINADORA

Documento assinado digitalmente
VANESSA CRISTINA OLIVEIRA DE SOUZA
(Presidente / Orientador)
Universidade Federal de Itajubá – IMC
VANESSA CRISTINA OLIVEIRA DE SOUZA
Data: 10/12/2024 20:55:53-0300
Verifique em <https://validar.itii.gov.br>

Documento assinado digitalmente
FLÁVIO BELIZÁRIO DA SILVA MOTA
Coorientador
Descubra Soluções em Decisões Estratégicas

Documento assinado digitalmente
 FLÁVIO BELIZARIO DA SILVA MOTA
Data: 09/12/2024 15:38:14-0300
Verifique em <https://validar.itii.gov.br>

Documento assinado digitalmente
 ISABELA NEVES DRUMMOND
Data: 11/12/2024 11:06:50-0300
Verifique em <https://validar.itii.gov.br>

ISABELA NEVES DRUMMOND
Universidade Federal de Itajubá - IMC

Documento assinado digitalmente
 ARCILAN TREVENZOLI ASSIREU
Data: 09/12/2024 17:46:15-0300
Verifique em <https://validar.itii.gov.br>

ARCILAN TREVENZOLI ASSIREU
Universidade Federal de Itajubá - IRN



Previsão de Velocidade do Vento no Nordeste Brasileiro usando Aprendizado Profundo e dados LiDAR

Lucas Batista Pereira

lucas.bpereira1999@unifei.edu.br

Orientador: Vanessa Cristina Oliveira de Sousa

Coorientador: Flávio Belizário da Silva Mota

RESUMO

Este trabalho apresenta uma análise detalhada do fenômeno meteorológico "Vento-Sul" na costa nordeste do Brasil, com foco em sua previsão e impacto na operação de parques eólicos. Utilizando dados coletados pelo sensor LiDAR Windcube V2, foram investigadas variações na velocidade e direção do vento, além de parâmetros ambientais, como temperatura e cisalhamento do vento. O estudo incluiu técnicas avançadas de análise de séries temporais para identificar padrões, comprovar a estacionalidade dos dados e selecionar as variáveis mais relevantes. Foi desenvolvido um modelo seq2seq capaz de prever a velocidade do vento com até 6 horas de antecedência, alcançando um RMSE de 0.2220 (m/s), superando a meta inicial de precisão. Este trabalho contribui para o avanço no uso de sensores LiDAR na previsão de ventos, fornecendo subsídios para a gestão eficiente e segura de recursos energéticos renováveis na região, com potencial de aplicação em contextos semelhantes.

Palavras-chave: Previsão de vento; Seq2Seq; Transformada de Wavelet; Aprendizado profundo ; Energia eólica

1 Introdução

A matriz energética brasileira se destaca globalmente pelo maior uso de fontes renováveis, que são inesgotáveis e impactam menos o meio ambiente. A energia eólica é uma dessas fontes e vem desempenhando um papel chave no desenvolvimento socioeconômico e na preservação ambiental do país ([Santos & Araújo, 2023](#)). Segundo a Associação Brasileira de Energia Eólica e Novas Tecnologias (ABEEólica), a energia eólica responde atualmente por 16,1% da matriz energética brasileira, contando com 1085 parques eólicos, em 12 Estados ([ABEEólica, 2024a](#)). A região Nordeste é uma das precursoras na instalação de energia eólica no Brasil, favorecida por suas condições climáticas. Ainda segundo a ABEEólica, em Agosto de 2024, a geração das Usinas Eólicas atendeu a 100% do consumo do Nordeste e exportou 29,41% para as demais regiões. Ademais, em Julho de 2024, o Rio Grande do Norte atingiu o marco de 10 GW de capacidade instalada ([ABEEólica, 2024c](#)).

No entanto, [Ibrahim & Altun \(2024\)](#) destacam que a natureza variável e intermitente do vento impõe desafios significativos aos gestores do sistema de energia. Um exemplo ocorreu em 2021 na região Nordeste, quando um apagão afetou milhões de pessoas e prejudicou diversos sistemas públicos, como o de saúde, evidenciando fragilidades no sistema energético regional.

[Assireu et al. \(2024\)](#) investigaram as causas dos fenômenos na dinâmica do vento que geram quedas inesperadas na geração de energia eólica no litoral do Rio Grande do Norte. Os autores nomearam o fenômeno de *Vento-Sul*, que é caracterizado por mudanças rápidas na direção do vento e consequente

redução na disponibilidade do vento que chega às turbinas instaladas no Complexo Delta Maranhão.

A imprevisibilidade desse fenômeno, como descrito em [Assireu et al. \(2024\)](#), dificulta a implementação de medidas preventivas. Como consequência, os operadores de parques eólicos enfrentam dificuldades na manutenção da continuidade da geração, o que pode levar a impactos econômicos e sociais, como ocorreu durante o apagão de 2021.

Com o intuito de estudar o comportamento do vento e suas anomalias, sensores que utilizam tecnologia LiDAR (*Light Detection and Ranging*), foram instalados próximos às turbinas eólicas no Complexo Delta Maranhão. Esses equipamentos permitem medições precisas de velocidade, direção e outros parâmetros do vento em alturas de 40 m a 200 m ([Dong & Chen, 2017](#)). Com o LiDAR, é possível gerar séries temporais detalhadas do perfil do vento, proporcionando dados de alta qualidade para estudos relacionados.

O objetivo desse trabalho foi utilizar redes neurais profundas para prever a velocidade do vento a partir dos dados LiDAR e com uma antecedência de 6 (seis) horas ([Sarker, 2021](#)). Essa previsão visa oferecer subsídios ao Operador Nacional de Energia Elétrica (ONS) para se preparar diante de fenômenos como o *Vento-Sul*, mitigando possíveis impactos no sistema elétrico.

Para alcançar o objetivo proposto, a metodologia adotada foi a CRISP-DM (*Cross-Industry Standard Process for Data Mining*), que prevê um conjunto de etapas para realizar análise de dados ([Saltz, 2021](#)). Uma das etapas é a definição da tarefa de mineração. Nesse trabalho, a tarefa foi preditiva, uma vez que o objetivo foi prever a velocidade do vento.

Segundo [Ibrahim & Altun \(2024\)](#), métodos de previsão tradicionais, como modelos estatísticos e físicos, frequentemente falham em captar as dinâmicas complexas do comportamento do vento. Por outro lado, técnicas recentes de aprendizado profundo, especialmente as redes LSTM (*Long Short-Term Memory*), têm se destacado no campo da energia eólica, uma vez que essas redes aprendem automaticamente relações não lineares na série temporal.

Nesse contexto, esse trabalho propõe um modelo de aprendizado profundo baseado em uma arquitetura seq2seq (*sequence-to-sequence*) com redes LSTM bidirecionais combinadas a uma camada de atenção (seção 2.4.3). Essa abordagem mostrou-se adequada para prever, de forma confiável, a velocidade do vento com seis horas de antecedência.

A principal contribuição deste trabalho é o uso do modelo seq2seq na previsão da velocidade do vento, diferenciando-se da maioria dos estudos na literatura, que utilizam exclusivamente a arquitetura LSTM para prever a geração de energia eólica ([Lei et al., 2024](#); [Ibrahim & Altun, 2024](#)).

Este trabalho está organizado da seguinte forma: a Seção 2 aborda o fenômeno do vento, introduz o LIDAR e detalha a análise e preparação dos dados, além da lógica do modelo proposto. A Seção 3 descreve o desenvolvimento da solução seguindo as etapas da metodologia CRISP-DM, incluindo entendimento do negócio, preparação dos dados, modelagem e avaliação. A Seção 4 apresenta os resultados obtidos, comparando previsões e métricas de erro, e discute o desempenho do modelo. Por fim, a Seção 5 traz considerações sobre o modelo desenvolvido, aponta possibilidades de aprimoramento e sugere direções para investigações futuras.

2 Fundamentação Teórica

A energia eólica no Brasil, com destaque para a região Nordeste, representa uma parcela significativa da matriz elétrica nacional, consolidando-se como uma fonte estratégica para o suprimento energético. Atualmente, 89,3% das usinas eólicas em operação estão concentradas nessa região, somando 17,7 GW de potência instalada, com os estados do Rio Grande do Norte e da Bahia liderando a produção com 30,5% e 26,5% do total, respectivamente. Projeções da Empresa de Pesquisa Energética indicam que a participação da energia eólica na matriz elétrica nacional crescerá de 9%, em 2021, para 14% até 2030, demandando investimentos superiores a R\$ 50 bilhões, com o Nordeste sendo o maior beneficiário devido à sua competitividade nesse setor ([Banco do Nordeste do Brasil, 2021](#)).

Enquanto a hidrelétrica depende de reservatórios de água e sofre com períodos de seca, a energia eólica enfrenta a intermitência do vento, que dificulta sua integração estável à rede elétrica. Essa limitação exige tecnologias avançadas de armazenamento ou geração complementar para mitigar flutuações, garantindo a continuidade do fornecimento energético. O Nordeste é especialmente impactado por esses desafios devido à sua alta dependência dessa fonte ([ABEEólica, 2024b](#)).

Entender o comportamento do vento é, portanto, um elemento-chave para lidar com essas dificuldades. Entre os principais fatores que afetam diretamente a geração de energia, destacam-se eventos caracterizados por mudanças abruptas na direção e redução na velocidade do vento. Esses fenômenos evidenciam a necessidade de modelos preditivos robustos

que antecipem tais variações, auxiliando na operação e planejamento dos parques eólicos.

2.1 Dinâmica do Vento no Nordeste Brasileiro e evento Vento-Sul

O estudo de [Assireu et al. \(2024\)](#) investigou um evento no Complexo Eólico Delta Maranhão, localizado no litoral do Rio Grande do Norte. Esse evento é caracterizado por quedas abruptas na velocidade do vento e mudanças significativas na direção, geralmente ocorrendo no início da manhã e causadas por interações complexas entre fatores meteorológicos locais e regionais. Essas variações comprometem a eficiência das turbinas e resultam em perdas financeiras significativas, estimadas em milhões de dólares ao longo de dois anos. O estudo utilizou tecnologias avançadas, como o sensor LIDAR, descrito em detalhes na Seção 2.2, que mede a velocidade e direção do vento em diversas altitudes com alta precisão, e SODAR, fornecendo perfis verticais detalhados do vento. Além disso, foram utilizados dados de reanálise ERA5 e imagens térmicas e visíveis do satélite GOES-16, permitindo mapear as condições meteorológicas antes, durante e após os eventos, denominados pelos autores como *Vento-Sul*.

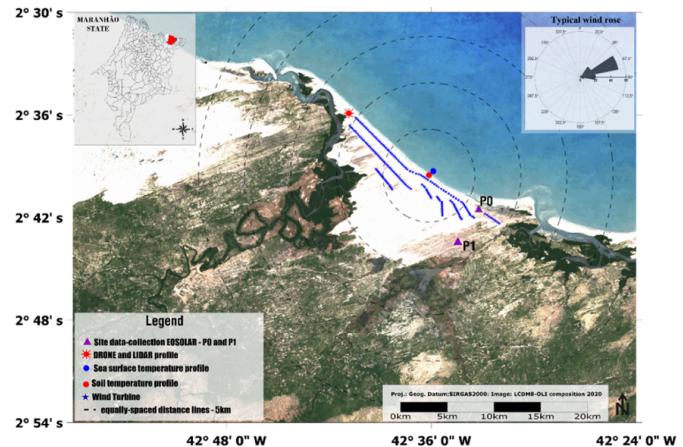


Figura 1: Região de estudo. As localizações das turbinas são indicadas por pontos azuis. O mirante PO está localizado a 1,5 km da praia e o P1 fica cerca de 5 km para o interior. Os marcadores vermelhos e azuis representam locais para estações base de UAV e medições de temperatura terrestre-mar.

Fonte: Extraído de [Assireu et al. \(2024\)](#)

A análise detalhou que a interação entre brisas marítimas e ventos predominantes interfere nos padrões de vento observados nas alturas de operação das turbinas. Diferenças de temperatura entre terra e mar geram gradientes térmicos que, combinados à circulação costeira, influenciam diretamente a formação e a evolução desses fenômenos. O estudo também apontou que, em situações em que a brisa não se forma ou é bloqueada por fatores como nuvens matinais, o vento muda de direção, passando a soprar predominantemente de terra para o mar. Isso resulta em uma redução significativa da força do vento que alcança as turbinas, diminuindo sua eficiência operacional ([Assireu et al., 2024](#)).

Além disso, os resultados indicaram que a ocorrência dos eventos *Vento-Sul* é mais frequente durante a estação seca, especialmente no período da manhã, quando os gradientes térmicos são mais intensos. Esses eventos também foram associ-

ados a formações de linhas de nuvens sobre a costa, que desempenham um papel no atraso ou intensificação da circulação atmosférica local. A utilização de modelos conceituais no estudo ajudou a compreender os mecanismos por trás da dinâmica do vento e sua relação com fatores locais e sinóticos.

A figura 1 no estudo ilustra a área de análise, incluindo a localização das turbinas eólicas e das estações meteorológicas. Embora a imagem seja apenas uma representação espacial, ela fornece um contexto visual claro do local investigado e da infraestrutura associada ao estudo, ajudando a compreender a complexidade regional das variações de vento e sua influência na geração de energia.

O objetivo dos estudos de [Assireu et al. \(2024\)](#) foi investigar a causa do evento *Vento-Sul* e, para tanto, utilizaram dados de vento de alta qualidade, obtidos por sensoriamento remoto a laser.

2.2 LiDAR

Sensoriamento Remoto pode ser definido como a aquisição de informação sobre alguma propriedade de um objeto ou fenômeno sem contato físico com ele. A informação sobre um alvo é obtida pela energia eletromagnética refletida, emitida ou retroespelhada por esse alvo, coletada por meio de sensores instalados a bordo de plataformas em altitude, tais como balões, foguetes, aviões ou satélites. Os sensores podem ser classificados como passivos ou ativos. Sensores passivos detectam a energia refletida ou emitida pelos alvos, geralmente proveniente de uma fonte externa, como o Sol. Por outro lado, sensores ativos emitem sua própria energia e captam a fração refletida ou retroespelhada pelos alvos. Na categoria dos sensores ativos, estão incluídos os radares e *lasers* ([Lorenzetti, 2015](#)).

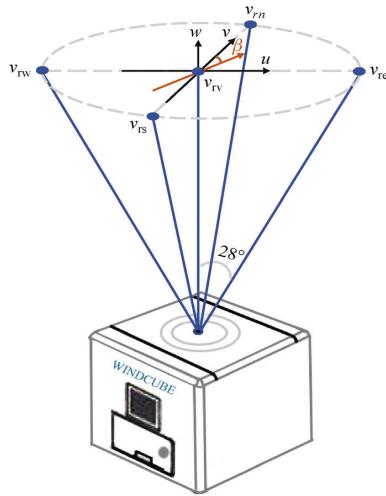


Figura 2: Apresenta um esquema ilustrativo do funcionamento de um sensor LiDAR WindCube V2. Ela mostra o princípio de medição baseado na emissão e recepção de feixes de laser em diferentes direções angulares. **Fonte:** ([ResearchGate, 2024](#))

O LiDAR (*Light Detection and Ranging*) é uma tecnologia de sensoriamento remoto ativo que mede distâncias com base no tempo decorrido entre o envio e o recebimento de um sinal. Ele utiliza luz na forma de um laser pulsado para medir distâncias. No espectro eletromagnético, o LiDAR geralmente opera na faixa do infravermelho próximo (NIR), mas também

podem incluir o ultravioleta (UV) e o visível, dependendo do tipo de sensor e do objetivo ([Dong & Chen, 2017](#)).

A compreensão da dinâmica do vento no Nordeste brasileiro depende de ferramentas capazes de capturar medições precisas e contínuas das condições atmosféricas. Entre essas tecnologias, destaca-se o WindCube V2, um sensor baseado na tecnologia LiDAR projetado para medir componentes horizontais e verticais do vento em diferentes altitudes. Ou seja, com o WindCube é possível gerar uma série temporal de perfil do vento e, por isso, ele é classificado como um perfilador vertical. Ou, LiDAR vertical ([Yang et al., 2021](#)).

Ao contrário da maioria dos sistemas LiDAR, o WindCube fica fixo em um local (*onshore* ou *offshore*) e emite feixes de laser na vertical, como ilustrado na Figura 2. Esses feixes, direcionados em ângulos fixos de 28°, interagem com partículas em suspensão no ar, capturando informações sobre a velocidade do vento radial (v_r) ao longo de cada linha de visão. A partir dessas medições, o sensor combina os dados dos feixes direcionais para calcular os componentes horizontais (u e v) e verticais (w) do vento. Isso permite a geração de um perfil tridimensional detalhado do vento em diversas altitudes. A Figura 2 ilustra o princípio de funcionamento do WindCube V2, destacando como os feixes de laser são emitidos em diferentes direções e retornam ao dispositivo, permitindo a reconstrução do vetor de vento em várias camadas verticais.

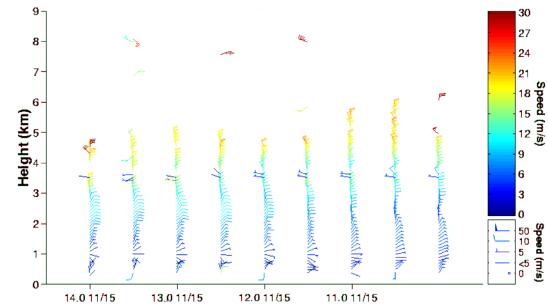


Figura 3: Exemplo de uma série temporal de perfil de velocidade do vento. **Fonte:** ([Wikipedia, 2024](#))

Os dados coletados pelo WindCube V2 incluem velocidade do vento horizontal, velocidade do vento vertical e direção do vento, cobrindo altitudes que variam de 40 a 260 metros. As medições são realizadas com uma resolução de 10 metros até 200 metros e, acima disso, em intervalos de 20 metros. Essa configuração possibilita a construção de um perfil detalhado das condições do vento em diferentes camadas atmosféricas, sendo ideal para aplicações em estudos meteorológicos e operações de parques eólicos.

Além disso, o sensor opera com alta frequência temporal, registrando dados a cada minuto. Essa capacidade de coleta contínua e em alta resolução permite monitorar mudanças rápidas nos padrões do vento.

A Figura 3 apresenta um exemplo de série temporal de perfil de vento. É possível notar que para cada momento no tempo, dados do vento são coletados em diferentes alturas.

2.3 Análise de Séries Temporais

No contexto da ciência de dados, *forecasting* (ou previsão) refere-se à tarefa de estimar valores futuros com base em observações anteriores de uma série temporal. Essa técnica é am-

plamente utilizada em áreas como meteorologia, economia e energia renovável, permitindo a antecipação de eventos críticos e a tomada de decisões informadas (Lyra et al., 2014).

Neste trabalho, a análise de séries temporais foi aplicada para modelar o comportamento do fenômeno *Vento-Sul* e realizar previsões da velocidade do vento. Modelos baseados em aprendizado profundo têm demonstrado maior eficácia em capturar relações complexas e não lineares presentes nos dados, uma vez que técnicas tradicionais muitas vezes não conseguem representar essas dinâmicas de forma satisfatória (Ibrahim & Altun, 2024).

Para aprimorar a modelagem e a previsão, foram utilizadas diversas ferramentas estatísticas e técnicas de processamento de sinais. A seguir, são detalhadas algumas das ferramentas empregadas na análise dos dados de vento neste trabalho: Matriz de Correlação, Função de Autocorrelação Parcial (PACF), Função de Correlação Cruzada (CCF) e Transformada de Wavelet.

2.3.1 Matriz de Correlação

A Matriz de Correlação é uma ferramenta estatística utilizada para medir o grau de dependência linear entre pares de variáveis em um conjunto de dados. Cada elemento da matriz representa o coeficiente de correlação entre duas variáveis, variando de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita). Valores próximos a zero indicam ausência de correlação linear. Essa matriz é útil para identificar multicolinearidade, um problema comum em modelos de regressão e aprendizado de máquina, onde variáveis altamente correlacionadas podem distorcer os resultados do modelo (Dürre et al., 2015).

Além disso, a Matriz de Correlação auxilia na seleção de atributos relevantes para análise ou modelagem preditiva, destacando quais variáveis possuem relações significativas que podem influenciar o comportamento das séries temporais. No contexto deste estudo, a análise de correlação permite identificar quais fatores meteorológicos estão mais intimamente relacionados à velocidade do vento, facilitando a construção de modelos preditivos mais robustos.

2.3.2 Função de Autocorrelação Parcial (*Partial Autocorrelation Function - PACF*)

A Função de Autocorrelação Parcial (PACF) mede a correlação entre uma variável e seus valores defasados, ajustando os efeitos das correlações intermediárias. Diferentemente da autocorrelação simples, que considera todas as defasagens anteriores, a PACF foca nas relações diretas entre a variável atual e cada defasagem específica. Isso é eficaz na determinação da ordem de modelos autorregressivos (AR), onde a PACF ajuda a identificar quantos termos defasados devem ser incluídos no modelo (Dürre et al., 2015).

Além de sua aplicação em modelos AR, a PACF é útil para identificar padrões de estacionalidade e entender melhor as dinâmicas internas das séries temporais, facilitando a construção de modelos mais precisos e interpretáveis. No presente estudo, a PACF é utilizada para determinar a estrutura apropriada dos modelos preditivos, garantindo que as dependências temporais sejam devidamente capturadas.

2.3.3 Função de Correlação Cruzada (*Cross Correlation Function - CCF*)

A Função de Correlação Cruzada (CCF) analisa como variações em uma variável influenciam outra ao longo de diferentes defasagens temporais. Essa ferramenta permite explorar interações entre variáveis, como por exemplo, como as velocidades do vento em diferentes alturas afetam umas às outras ao longo do tempo. A CCF é empregada para detectar relações potenciais que não são imediatamente aparentes, auxiliando na modelagem de entradas para modelos multivariados (Dürre et al., 2015).

Ao identificar defasagens específicas onde a correlação cruzada é significativa, a CCF garante que essas relações sejam adequadamente capturadas nos modelos preditivos, melhorando a precisão das previsões e a compreensão das interdependências entre as variáveis analisadas. Neste estudo, a CCF é aplicada para explorar como diferentes medições meteorológicas interagem com a velocidade do vento, contribuindo para a construção de modelos mais completos e eficazes.

2.3.4 Transformada de Wavelet

A Transformada de Wavelet é uma técnica matemática que permite decompor sinais em diferentes escalas de frequência, identificando variações localizadas em componentes de alta e baixa frequência (Liu et al., 2019). Para sinais como os dados de vento, que possuem alta variância e flutuações associadas a diferentes regimes temporais, a Transformada de Wavelet facilita a análise e o processamento. Essa decomposição ajuda a isolar padrões específicos e a remover informações menos relevantes, como ruídos ou tendências de longo prazo. A separação dos componentes também melhora a análise de fenômenos de curta duração, possibilitando maior precisão na interpretação de eventos dinâmicos.

Neste trabalho, a Transformada de Wavelet é utilizada para pré-processar os dados de velocidade do vento, aprimorando a capacidade dos modelos preditivos de capturar comportamentos complexos e não lineares.

2.4 Aprendizado Profundo

O aprendizado profundo (*Deep Learning*) refere-se a uma subárea do aprendizado de máquina que explora redes neurais artificiais com múltiplas camadas. Essas redes são projetadas para aprender representações hierárquicas a partir de dados brutos, permitindo que as camadas superiores capturem padrões mais complexos e abstratos, enquanto as inferiores lidam com características mais elementares. Essa abordagem pode ser aplicada em tarefas como processamento de linguagem natural, classificação de imagens, tradução automática e previsão de séries temporais (Sarker, 2021).

As redes neurais profundas diferenciam-se das redes neurais rasas (ou *shallow networks*) pelo número de camadas e pela capacidade de representação (Chen et al., 2021). Enquanto redes rasas, como o *Multi-Layer Perceptron* (MLP), geralmente possuem uma ou poucas camadas escondidas, redes profundas incluem múltiplas camadas, o que aumenta sua capacidade de capturar relações mais complexas nos dados. Em redes rasas, a representação dos dados é limitada à capacidade das camadas intermediárias, restringindo a modelagem de padrões não

lineares intrincados. Redes profundas, por outro lado, exploram a profundidade para decompor o aprendizado em níveis hierárquicos, o que permite detecção de padrões de alta complexidade. A figura 4 representa a diferença entre redes rasas e profundas.

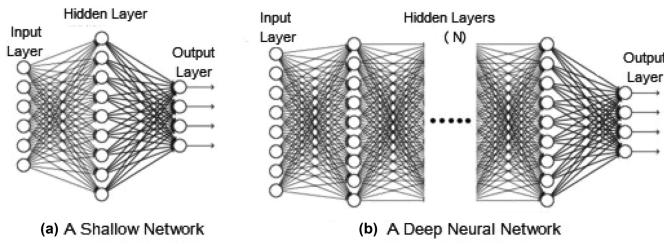


Figura 4: Diferença entre uma rede de aprendizado rasa e uma rede de aprendizado profundo .

Fonte: (Chen et al., 2021), em Inglês.

A Figura 5 apresenta um panorama das diferentes técnicas mencionadas, destacando sua aplicação e interconexões. Dentre as principais técnicas utilizadas no aprendizado profundo, destacam-se:

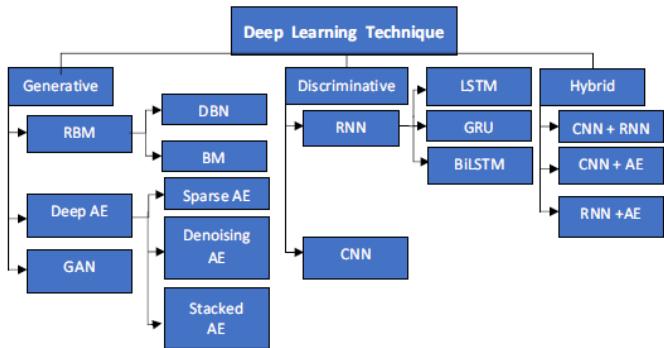


Figura 5: Classificação das arquiteturas de Aprendizado Profundo, em Inglês.

Fonte: (Dau & Salim, 2020).

- **Redes Neurais Convolucionais (Convolutional Neural Networks - CNNs):** Muito utilizadas em aplicações de visão computacional, como reconhecimento de objetos, segmentação de imagens e detecção de padrões visuais complexos (Sarker, 2021).
- **Redes Neurais Recorrentes (Recurrent Neural Networks - RNNs):** Especializadas no processamento de dados sequenciais, sendo amplamente empregadas em tarefas como tradução automática, análise de séries temporais e reconhecimento de fala. Suas variantes, como LSTMs e GRUs, oferecem soluções para problemas relacionados a longas dependências temporais (Sarker, 2021).
- **Autoencoders:** Ferramentas versáteis aplicadas em aprendizado não supervisionado, com utilidade em compressão de dados, redução de dimensionalidade e detecção de anomalias (Sarker, 2021).
- **Redes Adversárias Generativas (Generative Adversarial Networks - GANs):** Projetadas para gerar novos

dados sintéticos a partir da aprendizagem da distribuição dos dados originais, com impacto significativo em áreas como geração de imagens, síntese de voz e criação de conteúdo (Sarker, 2021).

- **Transformers:** Modelos que revolucionaram o processamento de linguagem natural, sendo amplamente utilizados em tarefas de tradução, resumo de textos e análise de sentimentos, além de sua expansão para multimodalidade e outras áreas (Sarker, 2021).

2.4.1 LSTM

As redes LSTM (*Long Short-Term Memory*) são projetadas para modelar dados sequenciais com dependências de longo prazo, superando limitações de redes recorrentes simples, que enfrentam problemas como o desaparecimento e explosão do gradiente (Hochreiter & Schmidhuber, 1997).

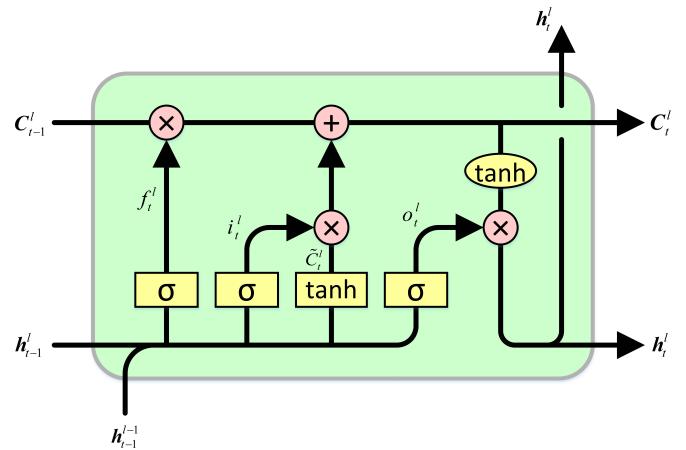


Figura 6: Apresenta a arquitetura interna de uma célula LSTM (Long Short-Term Memory), destacando os componentes principais responsáveis pelo processamento de informações em uma sequência temporal.

Fonte: (Stack Overflow, 2024)

Uma LSTM é composta por células que armazenam informações de forma controlada por três portas principais: entrada, esquecimento e saída. Essas portas são ativadas por funções como a sigmoide e a *tanh*, que operam como "chaves" permitindo abrir ou fechar o fluxo de informações. A combinação dessas funções possibilita que a rede decida quais informações manter, atualizar ou descartar, promovendo um aprendizado eficaz em contextos de longa dependência temporal (Olah, 2015).

Segundo Li et al. (2022), as LSTMs são eficazes em tarefas como previsão de séries temporais devido à sua capacidade de capturar tanto dependências locais quanto globais nos dados. Essa flexibilidade as torna adequadas para integração em modelos mais avançados. Portanto, a LSTM mostrou-se adequada para os objetivos desse trabalho.

A Figura 6 ilustra a arquitetura detalhada de uma célula LSTM, que contém três componentes principais:

1. **Porta de Entrada:** Controla quais informações da entrada atual serão armazenadas no estado da célula.
2. **Porta de Esquecimento:** Decide quais informações antigas no estado da célula devem ser descartadas.

3. **Porta de Saída:** Determina quais informações do estado da célula serão propagadas como saída.

2.4.2 LSTM Bidirecional

Uma LSTM pode ser unidirecional ou bidirecional. As redes LSTM bidirecionais (*Bidirectional Long Short-Term Memory*) estendem o conceito das LSTMs convencionais ao processar sequências em ambas as direções — da entrada para a saída e vice-versa. Essa arquitetura combina duas LSTMs independentes, uma que processa os dados em ordem direta e outra em ordem reversa. A saída final é uma concatenação das representações geradas por ambas as direções (Huang et al., 2015).

O principal benefício das LSTMs bidirecionais é a capacidade de capturar dependências passadas e futuras dentro da sequência, tornando-as úteis em tarefas que exigem o contexto completo da entrada. Em séries temporais, como a previsão da velocidade do vento, essa abordagem permite que o modelo considere tanto os padrões anteriores quanto as tendências futuras ao realizar previsões (Huang et al., 2015).

A Figura 7 apresenta a arquitetura de uma rede LSTM bidirecional, ilustrando como a informação é processada em duas direções e combinada para gerar uma representação mais robusta.

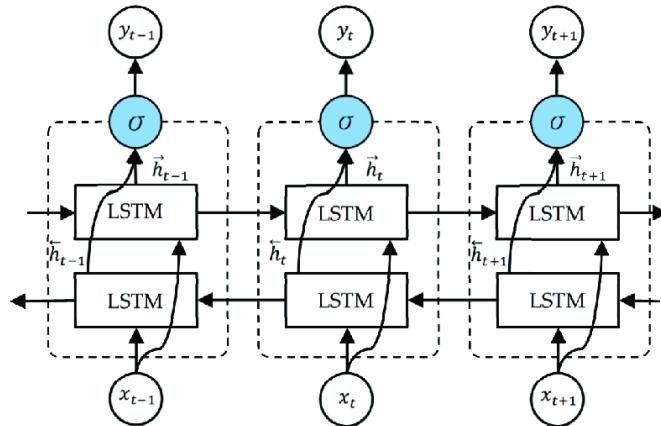


Figura 7: Ilustra a arquitetura de uma rede neural bidirecional baseada em LSTMs (*Long Short-Term Memory*), com camadas de entrada, bidirecional e de saída.

Fonte: (Yang et al., 2021)

2.4.3 seq2seq

O modelo seq2seq foi inicialmente desenvolvido para problemas de tradução automática e, posteriormente, adaptado para diversas aplicações, incluindo previsão de séries temporais multivariadas. Essa arquitetura utiliza redes neurais recorrentes, como LSTMs, para processar sequências de entrada e gerar saídas correspondentes (Sutskever et al., 2014). O modelo é composto por dois elementos principais:

1. **Codificador (Encoder):** Converte a sequência de entrada em um vetor de contexto que encapsula suas informações mais relevantes.
2. **Decodificador (Decoder):** Gera a sequência de saída a partir do vetor de contexto, estados ocultos internos e entradas adicionais.

O codificador consiste em camadas de LSTMs que processam a sequência de entrada, armazenando informações contextuais nos estados ocultos. No decodificador, essas informações são utilizadas para prever os próximos elementos da saída. Ele também pode receber variáveis conhecidas, como hora do dia e outros atributos temporais, que são integrados por meio de camadas de *embeddings*. Sua arquitetura base é mostrada na Figura 2.4.3.

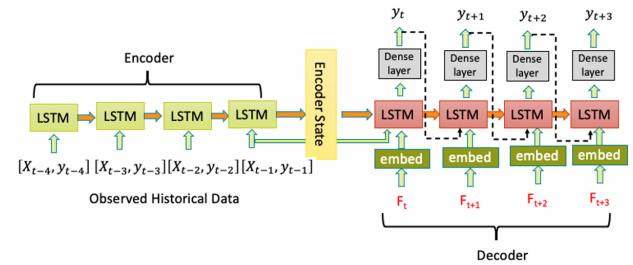


Figura 8: Ilustra a arquitetura de uma seq2seq baseada em LSTMs (*Long Short-Term Memory*), com camadas de encode, decode e atenção.

Fonte: (Wang et al., 2021), em Inglês.

Essa arquitetura foi aplicada para previsão de séries temporais, demonstrando maior flexibilidade e precisão em comparação com métodos estatísticos tradicionais.

2.4.4 Camada de Dropout

O *Dropout* é uma técnica que reduz o risco de *overfitting* em modelos de redes neurais ao desativar aleatoriamente uma fração dos neurônios durante o treinamento. Essa abordagem força o modelo a aprender representações mais robustas, promovendo maior generalização.

O *Dropout* funciona desativando, de forma aleatória, uma porcentagem dos neurônios de uma camada durante o treinamento do modelo. Essa desativação impede que o modelo dependa excessivamente de neurônios específicos, forçando-o a aprender representações mais distribuídas e robustas, o que ajuda a reduzir o *overfitting* e melhora a capacidade de generalização (Srivastava et al., 2014).

2.4.5 Camada de atenção

Apesar de sua eficácia, o modelo seq2seq enfrenta um desafio significativo relacionado à dependência do vetor de contexto único (*context vector*). Esse vetor é uma única representação fixa gerada pelo codificador, que condensa todas as informações da sequência de entrada em um único estado. Essa abordagem pode levar à perda de informações, especialmente em sequências mais longas, onde detalhes importantes podem ser descartados ou suavizados durante o processo de compressão. Como resultado, o desempenho do modelo pode ser prejudicado em tarefas que exigem compreensão detalhada de padrões temporais complexos (Vaswani et al., 2017).

Para mitigar essa limitação, o seq2seq pode ser aprimorado com mecanismos de atenção. Esses mecanismos permitem que o decodificador acesse diretamente todos os estados ocultos gerados pelo codificador, em vez de depender exclusivamente do vetor de contexto. O mecanismo de atenção cal-

cula pesos dinâmicos para cada elemento da sequência de entrada, identificando quais partes são mais relevantes em cada etapa do decodificador. Esses pesos são usados para gerar uma representação ponderada da entrada, permitindo ao modelo focar em informações específicas conforme necessário, melhorando significativamente sua capacidade de lidar com sequências longas e capturar dependências temporais (Vaswani et al., 2017).

O cálculo da atenção é realizado por:

1. **Pontuação (Score):** Mede a relevância de cada elemento da entrada em relação ao estado atual do decoder.
2. **Normalização Softmax:** Converte as pontuações em probabilidades.
3. **Combinação Ponderada:** Gera uma representação ponderada da entrada, com base nas probabilidades calculadas.

Essa abordagem melhora o desempenho do modelo em tarefas com dependências longas ou relações complexas na sequência. Na Figura 9, observa-se como o mecanismo de atenção é integrado ao modelo seq2seq.

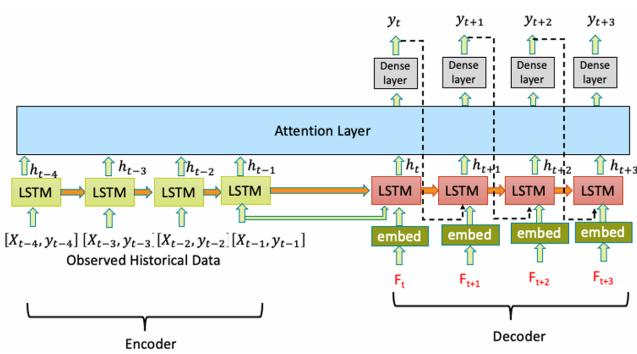


Figura 9: Ilustração base de uma rede seq2seq com camada de atenção.

Fonte: (Wang et al., 2021), em inglês.

Com a inclusão do mecanismo de atenção, o modelo é capaz de capturar nuances e dependências intrínsecas na sequência, aprimorando a precisão em aplicações preditivas.

2.5 Trabalhos Relacionados

A previsão de velocidade do vento é um tema mundialmente estudado devido à sua relevância para o desempenho de parques eólicos e a gestão eficiente de recursos renováveis. Diversos trabalhos exploraram abordagens avançadas, combinando aprendizado profundo e métodos estatísticos para melhorar a precisão das previsões.

Li et al. (2020) propuseram um modelo híbrido baseado em aprendizado profundo, denominado *Hybrid Deep Interval Prediction Model*. Este modelo utiliza redes neurais profundas em conjunto com intervalos preditivos para lidar com a incerteza inerente das séries temporais de vento, apresentando resultados robustos em termos de confiabilidade e precisão. A consideração de intervalos de previsão é particularmente importante em ambientes de alta variabilidade, como parques eólicos.

Ibrahim & Altun (2024) fizeram uma revisão das técnicas de aprendizado profundo para previsão de energia eólica, com foco especial nas redes LSTM para previsão de curto prazo da produção de energia eólica. Segundo os autores, as redes LSTM são eficazes em capturar dependências temporais nos dados de vento e melhoraram a precisão das previsões.

Estes trabalhos reforçam a necessidade de integrar técnicas de aprendizado profundo para abordar desafios específicos, como a não linearidade e a alta variabilidade dos dados de vento.

3 Desenvolvimento do Trabalho

Para a realização das análises e processamento dos dados, foi utilizado o ambiente Python, que oferece bibliotecas e frameworks para a análise e preparação dos dados, modelagem, treinamento de modelos preditivos e avaliação de desempenho. As principais tecnologias empregadas neste trabalho são descritas a seguir:

- **Pandas:** Utilizada para manipulação e análise de dados, permitindo trabalhar com estruturas de dados de alto desempenho e de fácil utilização.
 - **NumPy:** Fornece suporte para arrays e matrizes multidimensionais, além de uma coleção de funções matemáticas de alto nível para operações com esses arrays.
 - **PyWavelets:** Utilizada para realizar transformações wavelet, permitindo a análise de sinais em diferentes frequências. Essa técnica é particularmente útil para analisar dados não estacionários ou transientes, como os encontrados em séries temporais de velocidade do vento.
 - **Matplotlib e Seaborn:** Bibliotecas de visualização de dados, utilizadas para criar gráficos e visualizações estatísticas informativas.
 - **Statsmodels:** Fornece classes e funções para a estimativa de modelos estatísticos, realização de testes estatísticos e exploração de dados.
 - **Scikit-learn:** Utilizada para ferramentas de aprendizado de máquina, incluindo métricas de avaliação como *Mean Absolute Error* (MAE) e *Mean Squared Error* (MSE).
 - **TensorFlow e Keras:** Frameworks para construção e treinamento de modelos de aprendizagem profunda, incluindo arquiteturas como LSTM e modelos seq2seq com camadas de atenção.
- O trabalho foi desenvolvido utilizando a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Trata-se de uma abordagem estruturada e iterativa para execução de projetos de ciência de dados. Desde sua criação nos anos 1990, ela tem sido utilizada por equipes para organizar o ciclo de vida de projetos de análise de dados em fases bem definidas (Saltz, 2021). Neste trabalho, o CRISP-DM é aplicado para estruturar o estudo do *Vento-Sul*, desde a coleta e compreensão dos dados capturados pelo WindCube V2 até a modelagem e validação preditiva. Essa metodologia permite integrar as etapas de análise de séries temporais e previsão de padrões do vento com o planejamento e execução eficiente do projeto.

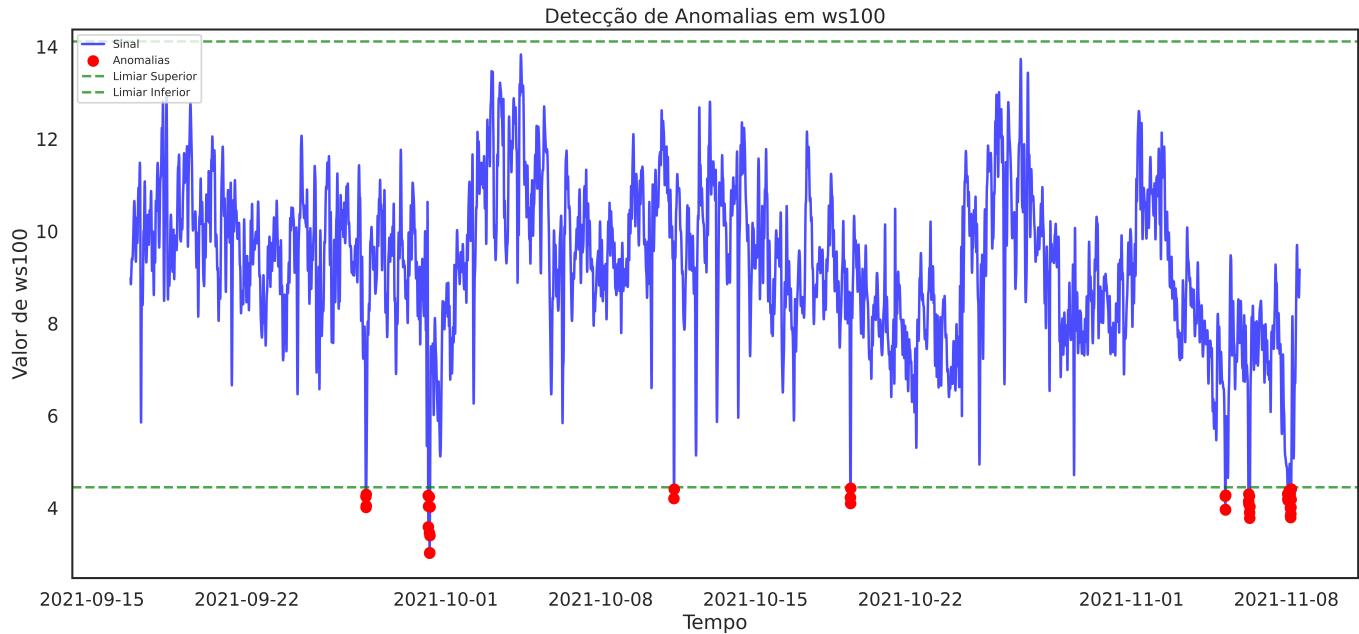


Figura 10: Anomalias na velocidade do vento ao longo do período estudado.

1. **Entendimento do Negócio:** Compreender os objetivos do projeto e os requisitos a partir de uma perspectiva de negócios.
2. **Entendimento dos Dados:** Coletar e familiarizar-se com os dados, identificando problemas de qualidade e explorando as principais características.
3. **Preparação dos Dados:** Limpar e preparar os dados para a modelagem, incluindo a seleção de variáveis, transformação de dados e manipulação de dados faltantes.
4. **Modelagem:** Aplicar técnicas de modelagem preditiva, selecionar algoritmos apropriados e ajustar os parâmetros dos modelos.
5. **Avaliação:** Avaliar a qualidade dos modelos desenvolvidos, verificando se eles atendem aos objetivos do negócio.
6. **Implantação:** Implementar os modelos em um ambiente de produção, onde eles podem ser utilizados para fazer previsões.

A Figura 11 ilustra o relacionamento entre as fases do CRISP-DM.

As próximas subseções descrevem os procedimentos executados em cada etapa do CRISP-DM.

3.1 Entendimento do Negócio

Essa etapa foi realizada a partir do estudo da literatura, especialmente o trabalho de Assireu et al. (2024), e também a partir de reuniões realizadas junto aos especialistas nos dados e no problema. Buscou-se entender o que é o evento *Vento-Sul*, seus efeitos e como mitigá-los.

A definição do problema passa pela dificuldade na previsão precisa da velocidade do vento no Complexo Eólico Delta Maranhão, especialmente durante a ocorrência do fenômeno conhecido como *Vento-Sul*. Esse evento é caracterizado

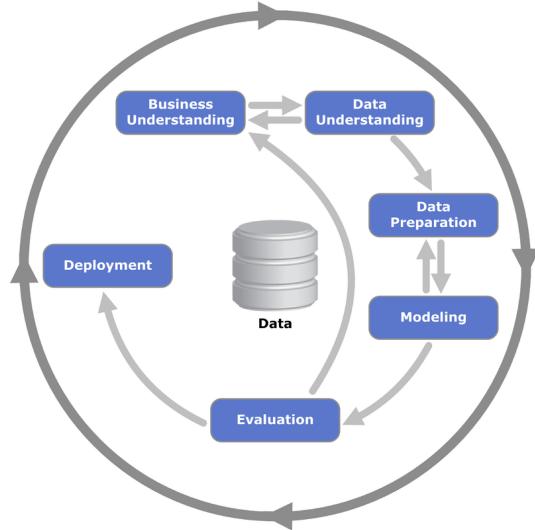


Figura 11: Diagrama de processo que mostra a relação entre as diferentes fases do CRISP-DM.

Fonte: Adaptado de ([Wikipedia, 2024](#))

por alterações repentinas na direção e intensidade do vento, podendo causar insuficiência energética na região assim que ocorre.

A Figura 10 apresenta a série temporal da velocidade do vento a 100 metros de altura. Os pontos vermelhos no gráfico são ocorrências do *Vento-Sul*. Tais eventos foram previamente mapeados por Assireu et al. (2024).

Também são objetivos dessa etapa definir a tarefa de mineração de dados e produzir o plano do projeto.

A tarefa de mineração de dados é preditiva. Objetiva-se prever a velocidade do vento com uma janela mínima de uma hora de antecedência. Apoiado pela literatura, o modelo utilizado na predição será um modelo de aprendizagem profunda. O modelo terá como entrada a série temporal com diferentes informações do vento na região de estudo.

O desenvolvimento da solução será realizado em etapas estruturadas:

1. **Coleta e Pré-processamento de Dados:** Os dados históricos coletados pelo Windcube serão consolidados, assegurando sua consistência e qualidade para análise posterior.
2. **Análise Exploratória de Dados:** Identificação de padrões, tendências e correlações que possam influenciar a previsão da velocidade do vento, considerando a especificidade dos dados do Windcube.
3. **Desenvolvimento do Modelo Preditivo:** Um modelo seq2seq com camada de atenção será implementado, com o objetivo de prever a velocidade do vento com antecedência mínima de uma hora.
4. **Validação e Teste do Modelo:** O modelo será avaliado utilizando dados de validação, buscando otimizar seu desempenho e alcançar um RMSE inferior a 1.0.
5. **Implementação:** Apesar de este trabalho focar no desenvolvimento do modelo, recomenda-se sua futura integração aos sistemas operacionais do parque para influenciar decisões em tempo real.
6. **Monitoramento e Manutenção:** Sugere-se que estudos futuros explorem estratégias para monitorar o desempenho do modelo e atualizá-lo com novos dados do Windcube, garantindo sua eficácia ao longo do tempo.

3.2 Entendimento dos Dados

Nesta fase, os dados devem ser descritos, explorados e avaliados quanto à sua qualidade. Deve haver um entendimento das variáveis envolvidas com o problema abordado.

Os dados utilizados neste estudo foram coletados por um sensor LiDAR Windcube V2, instalado próximo ao Complexo Eólico Delta Maranhão, nas coordenadas 2°43'29.6"S 42°34'30.7"W. As medições incluem velocidades do vento (horizontal e vertical) e direção do vento, capturadas em intervalos regulares de 1 minuto, no período de 16 de setembro a 8 de novembro de 2021. Para cada momento do tempo, os dados são coletados em 20 alturas diferentes, sendo a primeira 40 metros e a última 260 metros. Entre 40 e 200 metros, o intervalo é de 10 em 10 metros. A partir da altura 200, até 260, o intervalo é de 20 em 20 metros.

Além disso, variáveis atmosféricas como pressão, temperatura e umidade foram obtidas por uma torre micrometeorológica localizada próxima ao sensor LiDAR. Essas variáveis complementam a caracterização das condições atmosféricas locais. Outras variáveis utilizadas foram derivadas a partir do processamento dos dados originais capturados pelo Windcube e pela torre micrometeorológica.

A base de dados contém uma série temporal composta por 7561 registros. O conjunto de dados foi previamente processado por [Assireu et al. \(2024\)](#) para intervalos de 10 minutos, resultando em uma base de alta qualidade.

Para melhor entendimento do conjunto de dados, foi gerado um dicionário de dados. Este dicionário, apresentado na Tabela 1, descreve as variáveis presentes na base, suas unidades e respectivas descrições. As informações contidas no dicionário foram validadas por especialistas da área, assegurando

a correta interpretação dos dados. Posteriormente, para facilitar o acesso e manipulação dos dados, eles foram armazenados em um banco de dados relacional (PostgreSQL) e não relacional (Mongo).

Utilizando ferramentas de visualização de dados, como o Matplotlib, foram realizadas análises exploratórias para compreender o comportamento dos dados. A análise foi conduzida mediante a geração de gráficos que destacam as características do vento ao longo do tempo e em diferentes alturas.

A análise diária revelou padrões cílicos na velocidade média do vento. Observou-se que entre meia-noite e 6h, e entre meio-dia e 18h, há uma tendência de diminuição na velocidade do vento. Já das 6h ao meio-dia e das 18h à meia-noite, a velocidade do vento tende a aumentar.

Esses ciclos de seis horas estão relacionados ao diferencial de temperatura entre a terra e o mar, que é mais pronunciado ao meio-dia e à meia-noite. Esse diferencial térmico influencia a circulação de brisas marítimas e terrestres, afetando a velocidade e a direção do vento na região do parque eólico. Assim, a velocidade do vento está associada à ocorrência de brisas resultantes dessas variações de temperatura. A Figura 12 ilustra bem este padrão. O gráfico foi gerado computando a média diária do vento entre 16 de Setembro e 8 de Novembro.

Outra análise realizada foi sobre a presença de padrões cílicos e sazonalidade nos dados de velocidade do vento. Para isso, utilizou-se a Função de Autocorrelação Parcial (PACF), que mede a correlação entre uma série temporal e suas defasagens, eliminando a influência das defasagens intermediárias. Apesar de padrões cílicos superficiais serem observados visualmente em gráficos da série, a análise com uma janela de 144 amostras (equivalente a 24 horas) mostra que, ao suavizar a série, a tendência resultante é praticamente um curva muito suave ao longo do período analisado, quase uma linha reta, ou seja, não há tendência diária significativa do vento. Além disso, o gráfico da PACF (Figura 13) revela que a autocorrelação parcial dos dados diminui rapidamente para zero, confirmado a ausência de sazonalidade significativa. Indicando que a série temporal da velocidade do vento não apresenta sazonalidade consistente no período analisado.

3.3 Preparação dos Dados

Esta fase inclui todas as tarefas necessárias para gerar o conjunto de dados (*dataset*) final.

A partir dos dados originais e por *insights* dos especialistas na área, foram inseridas novas variáveis no *dataset*, a fim de enriquecer o conjunto de dados. Foram elas:

- **Cisalhamento Vertical do Vento (*Shear*):** Calculado para cada par de alturas adjacentes, o cisalhamento é dado pela diferença entre as velocidades do vento em duas alturas dividida pela diferença de altura entre elas. A fórmula do cisalhamento é apresentada na Equação 1, onde ws_x representa a velocidade do vento na altura x , e h_x , representa a altura x .

$$S = \frac{ws_j - ws_i}{h_j - h_i} \quad (1)$$

O cisalhamento refere-se à mudança na velocidade ou na direção do vento em uma curta distância, tanto na vertical quanto na horizontal. Em turbinas eólicas, o cisa-

Tabela 1: Dicionário de dados.

Variável	Tipo	Unidade	Descrição
mtime	double	Segundos	Série temporal com data e horário, medida a cada 10 minutos, representada em formato numérico. Capturada pelo Windcube em alturas de 40 a 260 metros.
press	double	hPa	Pressão atmosférica capturada pela torre micrometeorológica próxima ao LiDAR. Valores próximos ao nível do mar.
humid	double	%	Umidade relativa do ar capturada pela torre micrometeorológica próxima ao LiDAR. Valores típicos da região costeira.
temp	double	°C	Temperatura do ar capturada pela torre micrometeorológica próxima ao LiDAR. Indica a temperatura próxima ao mar.
lon	double	Graus decimais	Longitude do local onde o sensor está posicionado, em graus decimais. Valor fixo.
lat	double	Graus decimais	Latitude do local onde o sensor está posicionado, em graus decimais. Valor fixo.
heights	double	Metros	Alturas específicas de medição, variando de 40 a 260 metros. A partir da primeira coluna, a altura aumenta de 10 em 10 metros até 200 metros e, a partir daí, de 20 em 20 metros até 260 metros.
wspeed (40~260)	double	m/s	Velocidade média do vento para cada instância de tempo (mtime) e altura (heights), variando de 40 a 260 metros. Dados originais capturados pelo Windcube, formando o perfil de vento.
wdisp	double	m/s	Dispersão da velocidade do vento, representando o desvio padrão. Processada a partir dos dados originais.
wmin	double	m/s	Velocidade mínima do vento para cada instância de tempo (mtime) e altura (heights). Representa a menor velocidade em uma determinada altura. Processada a partir dos dados originais.
wmax	double	m/s	Velocidade máxima do vento para cada instância de tempo (mtime) e altura (heights). Representa a maior velocidade em uma determinada altura. Processada a partir dos dados originais.
wdir (40~260)	double	Graus	Direção do vento para cada instância de tempo (mtime) e altura (heights), variando de 40 a 260 metros. Dados originais capturados pelo Windcube. Indica de onde o vento vem; por exemplo, 60° significa vento chegando à turbina, enquanto 150° indica o evento "Vento Sul".
vertspeed (40~260)	double	m/s	Velocidade vertical do vento para cada instância de tempo (mtime) e altura (heights), variando de 40 a 260 metros. Dados originais capturados pelo Windcube. Indica se o vento está ascendente ou descendente, podendo estar associado ao surgimento de nuvens na região.
vertdisp	double	m/s	Dispersão da velocidade vertical do vento, representando o desvio padrão. Processada a partir dos dados originais.

lhamento pode afetar o desempenho e causar desgaste desigual nas pás.

- **Médias Móveis:** Foram calculadas as médias móveis de todas as variáveis, utilizando uma janela deslizante de tamanho 36, que representa um intervalo de 6 horas. A média móvel MA_t para uma série temporal X no instante t é definida pela seguinte função:

$$MA_t = \frac{1}{N} \sum_{k=0}^{N-1} X_{t-k} \quad (2)$$

onde N é o tamanho da janela.

O uso de médias móveis no processamento dos dados justifica-se pelo comportamento cíclico da velocidade do vento ao longo do dia. Observa-se que, geralmente, o vento apresenta maior intensidade durante a manhã e a noite, enquanto ao entardecer e ao amanhecer tende a ser mais fraco. Esse padrão permite dividir o ciclo diário em quatro quadrantes de 6 horas cada, refletindo diferentes características da velocidade do vento.

Nesse sentido, o *dataset* final contou com 203 atributos no total, sendo 98 apresentados na Tabela 1, mais 19 referentes ao cisalhamento das diferentes alturas e 95 referente às médias móveis.

Esse *dataset* completo foi utilizado nos modelos estatísticos e de *machine learning*. Posteriormente, para os modelos de aprendizado profundo, houve a necessidade de realizar mais

quatro etapas: filtragem dos dados, seleção dos atributos, Definição do Tamanho da Janela de Análise do Modelo *seq2seq* e Preparação dos Vetores de Treinamento, Validação e Teste. Essas etapas são descritas a seguir.

Filtragem de Dados

A filtragem dos dados foi realizada com o objetivo de preservar as componentes de alta frequência da velocidade do vento, eliminando frequências baixas indesejadas. Para isso, utilizou-se a Transformada de Wavelet (seção 2.3.4), que decompõe os sinais em componentes de diferentes frequências. Essa abordagem é útil para capturar tanto variações rápidas quanto tendências de longo prazo.

Para a aplicação da Transformada de Wavelet neste trabalho, foi utilizado o *wavelet sym18 (Symlet 18)*, uma família de wavelets simétricas conhecida por sua capacidade de capturar detalhes finos nos sinais, mantendo uma boa relação entre suavização e preservação de informações relevantes.

O nível 2 da decomposição foi configurado para alcançar um equilíbrio entre a redução de ruídos indesejados e a manutenção das flutuações importantes do sinal. Essa parametrização garante que os padrões temporais, incluindo picos e tendências locais, sejam preservados sem distorções significativas. Como resultado, foi possível identificar e destacar as características críticas do comportamento do vento.

A Transformada de Wavelet aplicada resultou em uma série suavizada que mantém as variações relevantes do sinal original. A Figura 14 apresenta a comparação entre os valo-

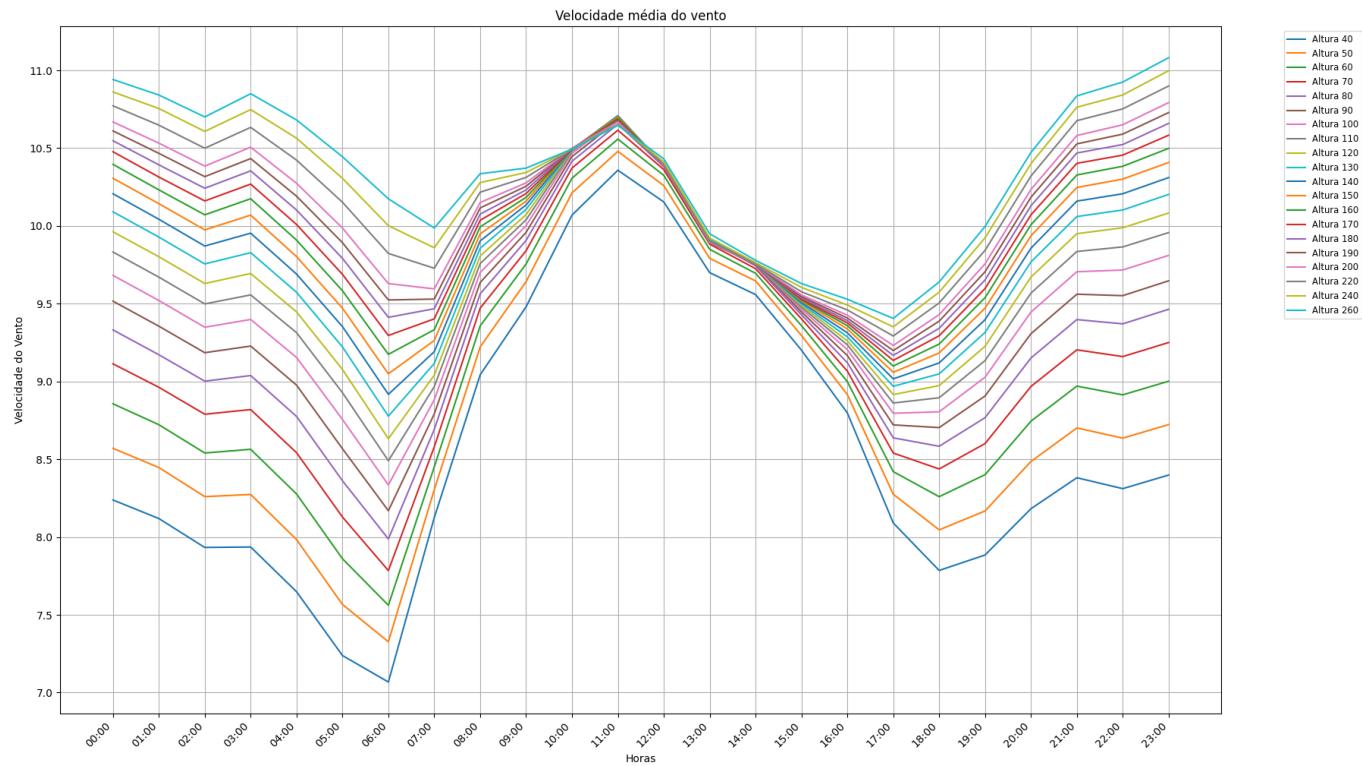


Figura 12: Velocidade média do vento para o período de 16 de Setembro a 8 de Novembro, para cada altura de 40 a 260 metros, ao longo das horas do dia.

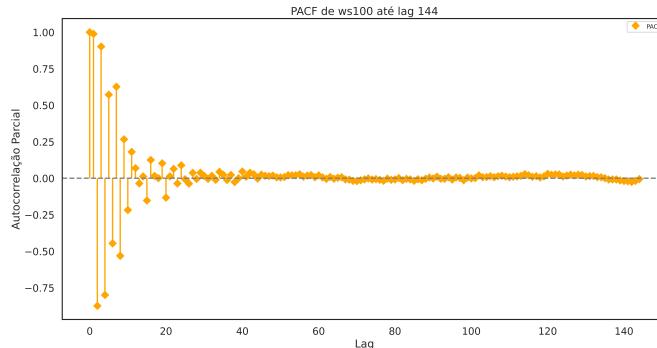


Figura 13: Função de Autocorrelação Parcial (PACF) da velocidade do vento.

res originais de velocidade do vento a 100 metros (*ws100*) e os valores após a aplicação do filtro. Observa-se que a série transformada reduz oscilações de alta frequência, enquanto preserva variações de baixa frequência, incluindo os valores de máximo e mínimo, essenciais para a análise preditiva.

Seleção de Atributos

A seleção de atributos foi realizada após a aplicação da Transformada Wavelet, garantindo que as variáveis analisadas já tivessem os ruídos reduzidos e as características relevantes preservadas. Foram consideradas as variáveis de velocidade do vento a 40, 100, 160, 220 e 260 metros de altura. As médias móveis foram desconsideradas, pois apresentam alta correlação com as variáveis originais, não adicionando informações relevantes.

A Figura 24 exibe a matriz de correlação em escopo re-

duzido, boa parte das variáveis foram retiradas para manter a legibilidade da tabela ao final deste trabalho. A análise dessa matriz justificou a escolha das variáveis para o modelo, considerando suas correlações e relevância prática.

As variáveis de velocidade do vento a 40 metros (*ws40*), 100 metros (*ws100*) e 260 metros (*ws260*) foram selecionadas com base em sua importância física e prática:

- ***ws40*:** Representa a menor altura capturada pelo *Wind-Cube*, a 40 metros, refletindo condições de vento próximas à superfície, influenciadas por barreiras como o terreno e a vegetação. Além disso, apresenta alta correlação com dispersão do vento (*wdisp*) e variáveis atmosféricas, como umidade (*humid*) e temperatura (*temp*).
 - ***ws260*:** Refere-se à maior altura registrada pelo *Wind-Cube*, a 260 metros, capturando padrões de vento em camadas superiores, como turbulência e cisalhamento. Sua correlação com direção do vento (*wdir*) e cisalhamento (*cis*) reforça sua relevância para análises verticais.
 - ***ws100*:** Variável alvo do modelo, a 100 metros, é de interesse prático para aplicações como turbinas eólicas. Sua previsão é o objetivo principal do estudo.
- Ademais, a análise da matriz de correlação evidenciou que:
- **Velocidades do vento (*ws40*, *ws100* e *ws260*):** Apresentam alta correlação entre si (coeficientes acima de 0.9), indicando forte dependência linear entre as alturas.
 - ***ws40*:** Mostra correlações exclusivas com dispersão do vento (*wdisp40*), dispersão vertical (*vertdisp40*), umidade

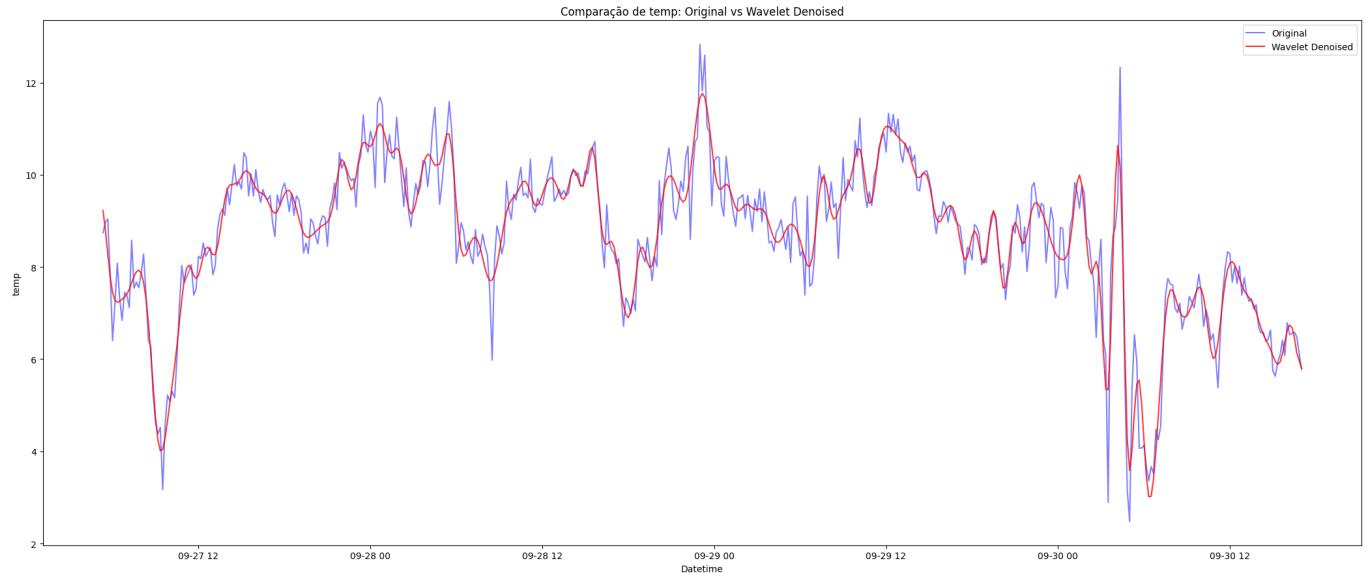


Figura 14: Comparação de *ws100_wavelet* antes e após a aplicação da Transformada Wavelet.

(*humid*) e temperatura (*temp*), o que não é observado para as demais variáveis.

- ***ws260***: Relaciona-se moderadamente com a direção do vento (*wdir260*) e cisalhamento (*cis1*), destacando sua importância para padrões atmosféricos em camadas superiores.
- ***ws100***: Funciona como elo intermediário entre *ws40* e *ws260*, sendo influenciada indiretamente pelas variáveis correlacionadas com estas alturas.

Com base nesses resultados, as variáveis selecionadas para o modelo foram:

- Velocidade do vento a 40, 100 e 260 metros: *ws40*, *ws100*, *ws260*.
- Dispersão da direção e da velocidade do vento a 40 metros: *wdisp40*, *vertdisp40*.
- Cisalhamento do vento entre 40 e 50 metros: *cis1*.
- Umidade e temperatura a 40 metros: *humid*, *temp*.
- Direção do vento a 260 metros: *wdir260*.

3.4 Modelagem

É nesta fase que as técnicas (algoritmos) de mineração serão aplicadas. A escolha da(s) técnica(s) depende dos objetivos desejados Hand (2007).

Tabela 2: Comparação das métricas de desempenho entre os modelos ARIMA e Random Forest para previsão de uma hora.

Modelo	MSE	MAE	RMSE
ARIMA	2.7023	1.8345	1.8924
Random Forest	1.9056	1.4657	1.4843

Inicialmente, foram testados modelos estatísticos tradicionais, como o ARIMA. Apesar de sua eficácia em séries temporais com tendências claras ou padrões sazonais bem definidos,

o ARIMA não apresentou bons resultados no caso dos dados de vento. Esse desempenho limitado deve-se à natureza imprevisível do vento em escala micro. A ausência de tendências lineares consistentes ou padrões sazonais evidentes dificultou a captura das dinâmicas do vento por um modelo baseado em relações lineares e autocorrelação.

Posteriormente, foi testado o *Random Forest*, um modelo de aprendizado de máquina reconhecido por sua robustez em problemas não lineares. Apesar de sua capacidade de lidar com grandes volumes de dados e interações complexas entre variáveis, o *Random Forest* também não conseguiu oferecer previsões satisfatórias. O principal fator limitante foi a sensibilidade dos dados de vento a pequenas mudanças temporais e espaciais, além da dificuldade do modelo em capturar sequências temporais complexas.

Destaca-se que nenhum dos modelos testados foi capaz de prever tendências de vento ou anomalias, mesmo com janela de previsão de uma hora, reforçando os desafios inerentes à previsão de séries temporais de vento.

Devido a isso, optou-se por modelos de aprendizado profundo (*deep learning*). A arquitetura seq2seq (seção 2.4.3) foi escolhida devido à sua capacidade de modelar dependências temporais de longo alcance e capturar padrões dinâmicos e não lineares presentes nos dados de vento. Essa abordagem mostrou-se mais adequada para lidar com as características desafiadoras dos dados, fornecendo previsões mais precisas e consistentes.

Durante as análises com a arquitetura seq2seq, foi necessário retornar à fase de pré-processamento dos dados para realizar ajustes no *dataset* de entrada. Essa conduta está alinhada com a metodologia CRISP-DM, que prevê revisões iterativas nas fases do processo para alcançar melhores resultados.

3.4.1 Ajuste de Hiperparâmetros do Modelo

Para otimizar o desempenho do modelo, foi realizada a seleção dos hiperparâmetros utilizando o *HyperBand* do framework Keras, que permite uma busca eficiente em espaços de hiperparâmetros de alta dimensionalidade. Os hiperparâmetros

ajustados incluem:

- **Número de unidades LSTM:** Foram testadas diferentes quantidades de unidades nas camadas LSTM do codificador e decodificador (por exemplo, 64, 128, 256, 512).
- **Taxa de aprendizagem:** Foram avaliadas taxas de aprendizagem variando entre 1×10^{-4} e 1×10^{-2} .
- **Função de perda:** Foram consideradas as funções de perda MSE (Erro Quadrático Médio) e MAE (Erro Absoluto Médio).
- **Taxa de dropout:** Foram experimentadas taxas de *dropout* entre 0,1 e 0,5 para regularizar o modelo e evitar sobreajuste.

Após a busca de hiperparâmetros, foi selecionada a configuração que apresentou o melhor desempenho no conjunto de validação, resultando nos seguintes valores:

- **Unidades LSTM no codificador:** 256 unidades bidirecionais.
- **Unidades LSTM no decodificador:** 512 unidades.
- **Taxa de aprendizagem:** 1×10^{-3} utilizando o otimizador Adam.
- **Função de perda:** MSE.
- **Taxa de dropout:** 0,1 nas saídas das camadas LSTM.

Para a seq2seq prever um valor, é preciso informar para ela qual o tamanho da janela de análise. Ou seja, o quanto do passado o modelo deve analisar para prever o futuro. A determinação do tamanho da janela de análise baseou-se na análise da Função de Correlação Cruzada (CCF) (seção 2.3.3) entre as variáveis *ws40_wavelet* e *ws260_wavelet* em relação à variável alvo *ws100_wavelet*.

As CCFs foram calculadas até um lag de 144, correspondendo a um intervalo de um dia de registros, considerando a frequência de amostragem de 10 minutos para os dados de velocidade do vento. Esse intervalo foi escolhido para capturar relações entre as variáveis ao longo de um ciclo diário completo, o que é relevante para analisar possíveis padrões ou dependências temporais. Os resultados estão apresentados nas Figuras 15 e 16, respectivamente.

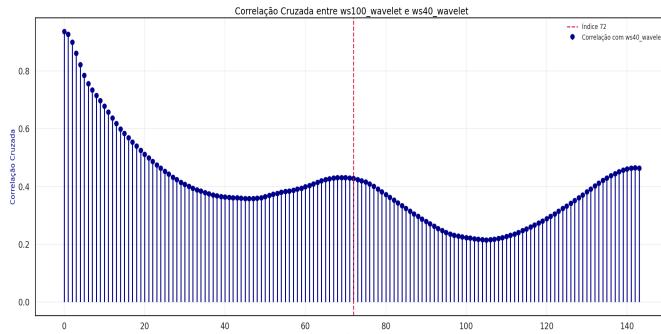


Figura 15: Função de Correlação Cruzada (CCF) entre *ws40_wavelet* e *ws100_wavelet*.

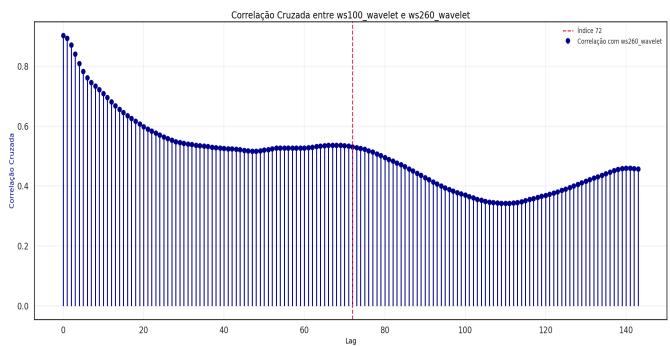


Figura 16: Função de Correlação Cruzada (CCF) entre *ws260_wavelet* e *ws100_wavelet*.

Analizando as curvas de correlação cruzada, observou-se que no *lag* 72 há uma redução acentuada na magnitude das correlações para ambas as variáveis em relação à variável alvo. Este comportamento reflete provavelmente o padrão diário intrínseco dos dados, dado que 72 horas correspondem a dois quadrantes do ciclo completo do dado. Além disso, é coerente que o terceiro quadrante apresente correlações mais fracas, uma vez que o impacto temporal das observações mais distantes tende a diminuir.

Com base nesses resultados, definiu-se que a janela de análise do modelo *seq2seq* deve ser de 72 passos temporais, o que equivale a um intervalo de 12 horas nos dados analisados. Essa escolha garante que o modelo capture adequadamente as dependências temporais relevantes, evitando incluir *lags* com baixa correlação, que podem adicionar ruído ao treinamento.

3.4.2 Arquitetura do Modelo

O modelo proposto é composto por um codificador (*encoder*) e um decodificador (*decoder*), integrados por um mecanismo de atenção, conforme ilustrado na Figura 17. A arquitetura combina LSTMs bidirecionais, mecanismos de atenção e camadas densas para capturar padrões temporais complexos e gerar previsões multivariadas.

- **Codificador (Encoder):** O codificador recebe como entrada sequências multivariadas de dimensão (N, T, F) , onde N é o número de amostras, T é o tamanho da janela temporal (72 passos, equivalentes a 12 horas), e F é o número de *features* selecionadas. Foi utilizada uma camada LSTM bidirecional com 256 unidades, configurada para retornar tanto as sequências completas quanto os estados finais. Isso significa que a camada LSTM bidirecional processa a sequência de entrada e, em cada passo temporal, gera uma representação correspondente (sequências completas). Além disso, uma camada de ativação e uma camada *Dropout* (Seção 2.4.4) foram aplicadas para evitar sobreajuste (*overfitting*) durante o treinamento.

- **Decodificador (Decoder):** O decodificador recebe como entrada as sequências alvo deslocadas, com dimensão $(N, T', 1)$, onde T' é o horizonte de previsão (36 passos, correspondentes a 6 horas). Foi utilizada uma camada LSTM unidirecional com 512 unidades (o dobro das unidades do codificador, devido à concatenação dos estados ocultos da LSTM bidirecional). Esta camada é configurada para retornar as sequências completas. Assim como

no codificador, o decodificador inclui camadas de ativação e *Dropout* para regularizar o aprendizado.

- **Mecanismo de Atenção:** Entre o codificador e o decodificador, foi empregada uma camada de atenção que permite ao modelo focar em diferentes partes da sequência de entrada ao gerar cada passo da saída. A atenção é calculada entre as saídas do decodificador e as saídas do codificador, permitindo que o modelo aprenda alinhamentos relevantes entre as entradas e as saídas. Esse mecanismo destaca-se por melhorar a capacidade do modelo de lidar com longas sequências e identificar padrões críticos.

- **Camadas de Concatenação e Saída:** As saídas do decodificador e do mecanismo de atenção são concatenadas e passam por uma camada *TimeDistributed Dense* com ativação linear. Essa camada processa cada passo temporal da sequência de saída, produzindo a previsão final da velocidade do vento para cada ponto do horizonte de previsão.

A Figura 17 ilustra como essas componentes interagem para processar as entradas e produzir a saída prevista.

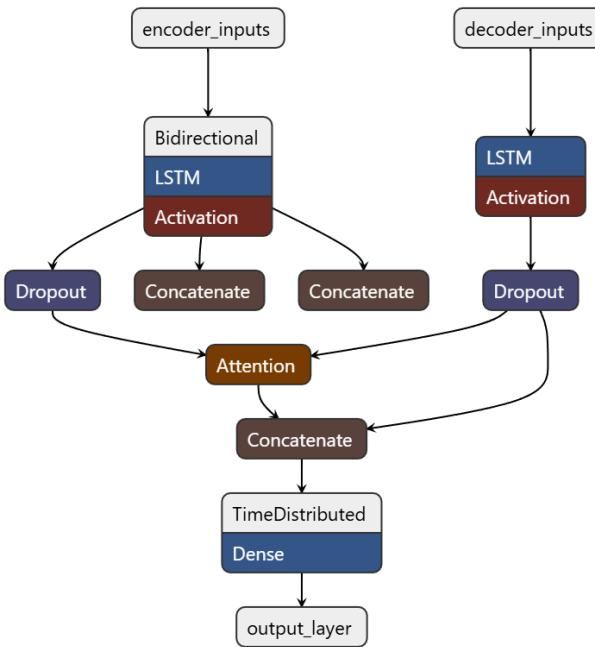


Figura 17: Visão geral da arquitetura do modelo preditivo desenvolvido neste trabalho. A imagem foi gerada utilizando a ferramenta Netron ([Roeder, 2024](#)).

3.4.3 Treinamento do Modelo

Para o treinamento do modelo *seq2seq*, os dados foram divididos em três conjuntos: 77% reservados para treinamento, 18% para validação e 5% para teste. A criação das sequências para entrada e saída do modelo foi realizada por meio de uma função que transforma o conjunto de dados original em vetores tridimensionais adequados para o modelo.

As sequências de entrada para o codificador (X_{encoder}) possuem o formato $(7454, 72, N)$, indicando 7454 amostras,

com janelas temporais de 72 passos e N variáveis por passo, sendo N o número de variáveis selecionadas para o treinamento como mencionado na Seção 3.3. As sequências de entrada para o decodificador (X_{decoder}) e as saídas esperadas do decodificador (y_{decoder}) têm o formato $(7454, 36, 1)$, onde 36 representa o número de passos de previsão e 1 corresponde à variável alvo, que é a velocidade do vento a 100 metros de altura.

A distribuição entre os conjuntos de treinamento, validação e teste manteve essas características. Para o conjunto de treinamento, as dimensões são: $X_{\text{encoder}} (5739, 72, N)$, $X_{\text{decoder}} (5739, 36, 1)$ e $y_{\text{decoder}} (5739, 36, 1)$. No conjunto de validação, os formatos são: $X_{\text{encoder}} (1342, 72, N)$, $X_{\text{decoder}} (1342, 36, 1)$ e $y_{\text{decoder}} (1342, 36, 1)$. Para o conjunto de teste, as dimensões são: $X_{\text{encoder}} (373, 72, N)$, $X_{\text{decoder}} (373, 36, 1)$ e $y_{\text{decoder}} (373, 36, 1)$.

Esse detalhamento ilustra a estrutura dos dados utilizados durante o treinamento, validação e teste, evidenciando o formato adequado para as entradas e saídas do modelo *seq2seq*. Durante o treinamento, foram adotadas as seguintes estratégias:

- **Callbacks:** Foi implementado *EarlyStopping* com *Patience* de 10 épocas para interromper o treinamento caso a perda de validação não apresentasse melhoria significativa. Também foi utilizado *ReduceLROnPlateau* para reduzir a taxa de aprendizagem em 50% caso a perda de validação estagnasse por 5 épocas consecutivas. Além disso, foi empregado *ModelCheckpoint* para salvar o modelo com melhor desempenho na validação.

- **Épocas e Lote:** O modelo foi treinado por até 100 épocas, com tamanho de lote (*batch size*) de 32 amostras.

Para a busca dos melhores hiperparâmetros do modelo, utilizando a configuração do PC(Ryzen 5 3600, 32GB de RAM, RTX 3070), o processo demorou 1 hora e 17 minutos. Durante essa etapa, o consumo médio foi de 4,340 MB de RAM, 6.600MB de VRAM e uso médio de 60W da GPU durante o processo. O treinamento completo do modelo, considerando 100 épocas, foi realizado em 21 minutos na mesma máquina e consumindo os mesmos recursos descritos para a etapa de busca de hiperparâmetros.

3.5 Avaliação

A avaliação é uma etapa essencial no processo de análise de dados, pois valida a qualidade dos modelos desenvolvidos e sua aplicabilidade prática ([Camilo & Silva, 2009](#)).

Neste trabalho, a avaliação foi exclusivamente quantitativa. Após o treinamento do modelo, utilizou-se os últimos 5% do *dataset* para validação, representando o período mais recente dos dados. Essa abordagem simula um cenário real, onde o modelo prevê a velocidade do vento a 100 metros de altura com antecedência de 6 horas, com base em dados continuamente atualizados.

Para avaliar o desempenho, foram utilizadas duas métricas quantitativas:

- **Mean Squared Error (MSE):** Utilizada como a função de perda durante o treinamento do modelo. O MSE é calcu-

lado como a média dos quadrados dos erros, sendo definido pela equação:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Essa métrica é sensível a grandes erros, o que incentiva o modelo a minimizar discrepâncias extremas (A., 2020).

- **Root Mean Squared Error (RMSE):** Uma métrica derivada do MSE, frequentemente utilizada para interpretar os resultados na mesma unidade das variáveis originais. O RMSE é definido como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Ele permite uma análise mais intuitiva do desempenho, especialmente em aplicações práticas (A., 2020).

- **Mean Absolute Error (MAE):** Métrica complementar ao MSE e ao RMSE, utilizada para avaliar a precisão geral das previsões. O MAE é definido como a média dos valores absolutos dos erros:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Por ser menos sensível a *outliers*, o MAE fornece uma visão robusta do erro médio (A., 2020).

Os resultados obtidos com o modelo *seq2seq* são apresentados na Seção 4. No entanto, a última fase da metodologia CRISP-DM, a aplicação, não foi contemplada neste trabalho. O objetivo principal foi validar o modelo em cenários simulados, sem integração direta em sistemas operacionais.

4 Resultados

Para simular um ambiente de previsão em tempo real, foi empregada a técnica de *rolling forecasting*. Essa abordagem consiste em realizar previsões sequenciais ao longo do conjunto de teste, atualizando a janela de entrada a cada nova previsão. A cada passo, o modelo incorpora novas observações, ajustando a janela de histórico e gerando previsões futuras. Esse método reflete condições práticas de operação, frequentemente encontradas em sistemas de previsão utilizados em parques eólicos.

No *rolling forecasting*, os dados são processados de forma contínua, utilizando uma janela deslizante que é atualizada com as medições mais recentes, enquanto os dados mais antigos são descartados. Essa abordagem mantém um histórico fixo para análise, permitindo que o modelo utilize informações atualizadas para realizar previsões imediatas. Esse método é amplamente utilizado em cenários operacionais que dependem de informações em tempo real, como na gestão de turbinas eólicas, onde as condições previstas para as próximas horas orientam decisões estratégicas e ajustes operacionais.

4.1 Lógica de Operação

A lógica do processo é descrita a seguir:

- **Inicialização:** A função recebe o modelo treinado, os dados escalonados, os parâmetros de entrada e saída (número de passos), o escalonador utilizado para a variável alvo e o índice da coluna correspondente à variável alvo no conjunto de dados.
- **Conjunto de Teste e Janela Inicial:** O conjunto de teste é definido como os últimos 5% dos dados escalonados, garantindo que a avaliação seja realizada sobre o período mais recente. Uma janela inicial de dados é extraída imediatamente antes do conjunto de teste, com tamanho igual ao número de passos de entrada do modelo (72 passos). Essa janela representa o histórico disponível no momento inicial da previsão.
- **Iteração e Atualização da Janela:** Durante cada iteração, o modelo gera previsões para os próximos 36 passos (6 horas), e os valores previstos são armazenados junto com os valores reais correspondentes. A janela de entrada é continuamente atualizada, simulando a obtenção de dados em tempo real.
- **Escalação Inversa dos Resultados:** As previsões e os valores reais são inversamente escalados para retornar aos valores originais, permitindo comparações diretas com os dados reais.

4.2 Resultado da simulação

Os modelos apresentados foram treinados com uma janela de 72 passos (12 horas) para prever os próximos 36 passos (6 horas). As variáveis de entrada incluem velocidade do vento em diferentes alturas (40, 100 e 260 m), dispersão da direção e da velocidade do vento a 40 m, cisalhamento entre 40 e 50 m, umidade e temperatura a 40 m, e direção do vento a 260 m. A variável alvo é a velocidade do vento a 100 m.

Os resultados são ilustrados nas Figuras 18 e 19. A Figura 18 compara as previsões do modelo *seq2seq* com a série temporal de velocidade do vento a 100 metros filtrada pela Transformada de Wavelet, evidenciando a alta precisão em capturar padrões suavizados. Já a Figura 19 mostra a comparação entre os valores reais da velocidade do vento a 100 metros e as previsões, destacando a robustez do modelo mesmo em condições mais ruidosas.

A Tabela 3 apresenta os resultados comparando a série filtrada pela Transformada de Wavelet e a série original. Observa-se que a suavização pelo uso da Wavelet contribui para um melhor desempenho, reduzindo os erros.

Tabela 3: Resultados do modelo *seq2seq* para previsão da velocidade do vento a 100 m (6 horas de antecedência).

Cenário	MAE	MSE	RMSE
Filtrada (Wavelet)	0,1589	0,0447	0,2115
Original	0,2921	0,1433	0,2220

Apesar de o RMSE ser ligeiramente maior para a série original (0,2220) do que para a filtrada (0,2115), ambos os valores

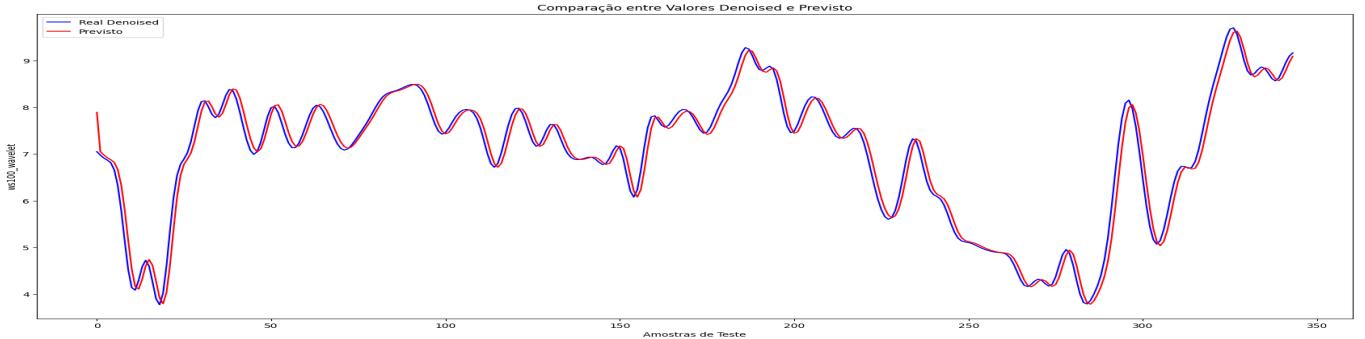


Figura 18: Comparação entre ws100_wavelet e as previsões da velocidade do vento a 100 metros de altura.

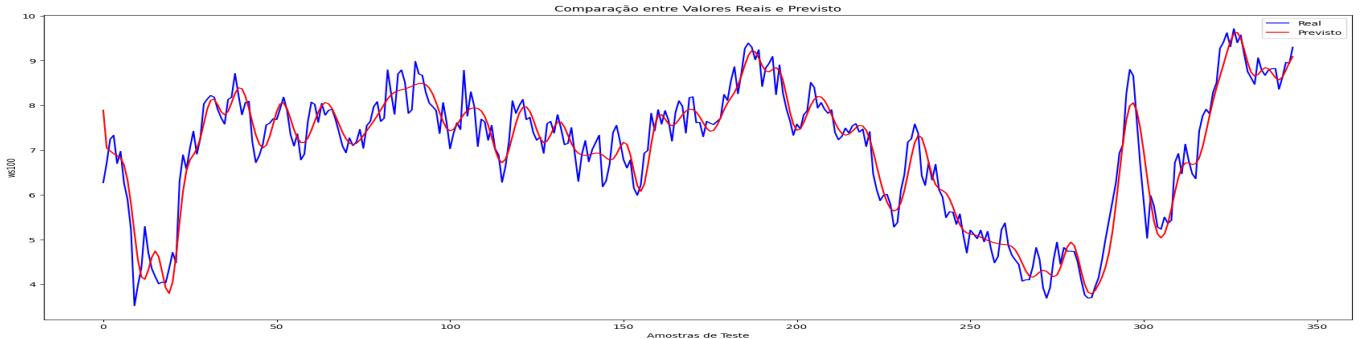


Figura 19: Comparação entre os valores reais de ws100 e as previsões da velocidade do vento a 100 metros de altura.

estão bem abaixo da meta de RMSE menor que 1,0. Isso evi-
dencia a robustez do modelo em lidar com ruídos e capturar
a dinâmica do vento. O histórico de treinamento (Figura 20)
mostra a convergência da função de perda e a estabilização
das métricas ao longo do tempo, reforçando a eficácia do
processo de treinamento.

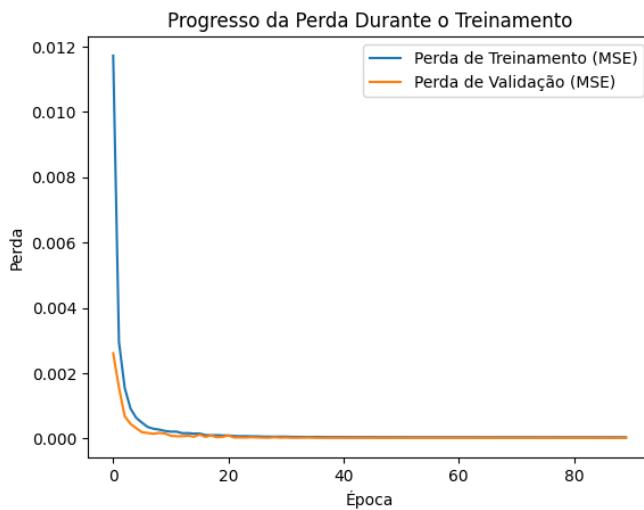


Figura 20: Histórico de treinamento do modelo: perda de treinamento e validação ao longo das épocas.

4.3 Resultado da simulação sem Umidade e Temperatura

Para avaliar o efeito da remoção da umidade e temperatura, foram utilizados apenas os dados do LIDAR. A abordagem

manteve o *rolling forecasting* com janelas de 72 passos (12 horas) para prever os próximos 36 passos (6 horas), considerando velocidade e direção do vento em diferentes alturas, cisalhamento e dispersão. O objetivo foi verificar se a ausência de variáveis não oriundas do LIDAR comprometeria o desempenho preditivo.

A Tabela 4 apresenta os resultados com e sem a Transformada de Wavelet. Nota-se que a remoção da umidade e temperatura não gera diferença significativa nos erros, mantendo valores próximos aos obtidos com todas as variáveis. A Figura 21 ilustra a comparação entre as previsões e a série filtrada pela Transformada de Wavelet, enquanto a Figura 22 apresenta a comparação com os valores originais.

Tabela 4: Resultados do modelo seq2seq sem umidade e temperatura, para a previsão da velocidade do vento a 100 m (6 horas de antecedência).

Cenário	MAE	MSE	RMSE
Filtrada (Wavelet)	0,1708	0,0505	0,2248
Original	0,2928	0,1445	0,2248

4.3.1 Considerações sobre os resultados

Um fator determinante para alcançar esses resultados foi a alta precisão dos dados fornecidos pelo WindCube, que é capaz de capturar medições detalhadas de velocidade e direção do vento em diversas alturas com excelente qualidade. Esses dados contribuíram significativamente para a qualidade do aprendizado do modelo, especialmente ao serem combinados com o pré-processamento baseado na Transformada de Wavelet.

Adicionalmente, o objetivo inicial do projeto era prever a velocidade do vento com um horizonte de uma hora de ante-

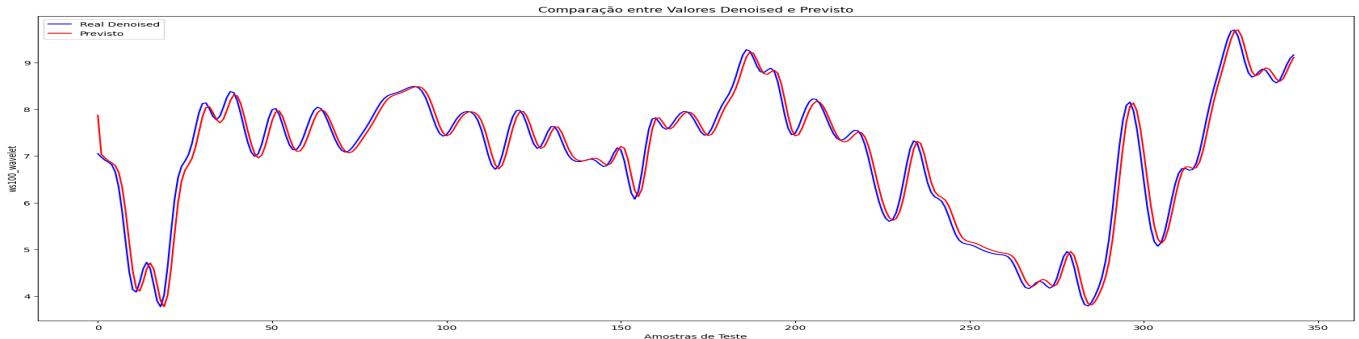


Figura 21: Comparação entre a série filtrada pela Transformada de Wavelet e as previsões da velocidade do vento a 100 metros de altura, sem considerar umidade e temperatura.

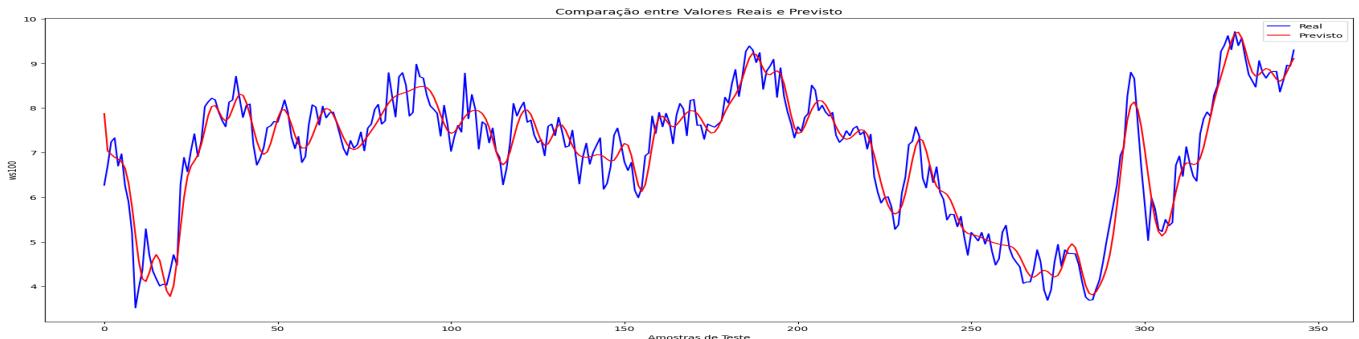


Figura 22: Comparação entre a série original e as previsões da velocidade do vento a 100 metros de altura, sem considerar umidade e temperatura.

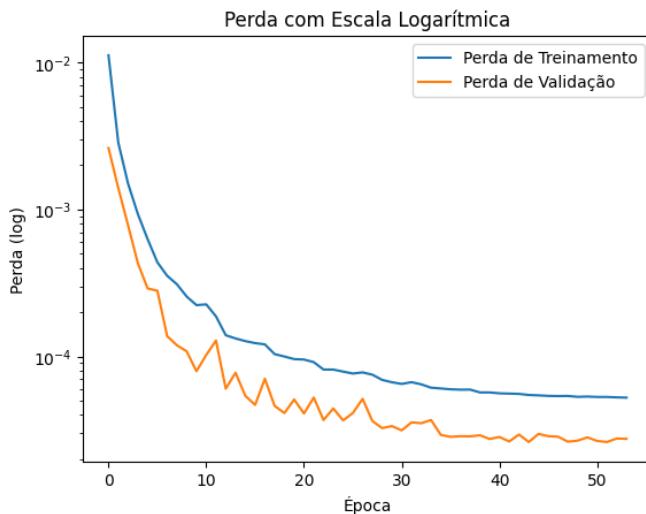


Figura 23: Histórico de treinamento do modelo sem umidade e temperatura: perda de treinamento e validação ao longo das épocas.

cedência. No entanto, o modelo *seq2seq* demonstrou ser capaz de realizar previsões precisas com até 6 horas de antecedência, ampliando consideravelmente o escopo de aplicação do sistema.

O modelo apresentou uma boa correspondência com os valores previstos tanto para a série filtrada pela Transformada de Wavelet quanto para os valores originais. A menor margem de erro observada na série filtrada reflete o impacto positivo do pré-processamento na melhoria das previsões. Ao mesmo tempo, o desempenho consistente mesmo nos dados originais

destaca a capacidade do modelo de lidar com a complexidade da dinâmica do vento. Esses resultados comprovam que o modelo é uma ferramenta viável para previsão de vento em ambientes dinâmicos e com condições de vento ocasionalmente anômalas.

A Tabela 5 apresenta uma comparação de todos os resultados obtidos nesta seção.

Tabela 5: Resultados do modelo *seq2seq* com e sem umidade e temperatura.

Condição	MAE	MSE	RMSE
Com Umid. e Temp. (Wavelet)	0,1589	0,0447	0,2115
Sem Umid. e Temp. (Wavelet)	0,1708	0,0505	0,2248
Com Umid. e Temp. (Original)	0,2921	0,1433	0,2220
Sem Umid. e Temp. (Original)	0,2928	0,1445	0,2248

5 Conclusões

Neste trabalho, foram analisados dados do Complexo Eólico Delta Maranhão com o objetivo de criar um modelo preditivo capaz de prever a velocidade do vento com pelo menos uma hora de antecedência, mantendo um RMSE menor que 1.0. Os resultados alcançados superaram essa meta: foi desenvolvido um modelo *seq2seq* que prevê a velocidade do vento com até 6 horas de antecedência, obtendo um RMSE de 0.2220, demonstrando alta precisão.

Além disso, realizou-se uma análise aprofundada dos dados, evidenciando padrões diários e comprovando a estacionariedade do comportamento do vento na região. Essa investigação permitiu a identificação das variáveis mais relevantes para

o modelo, incluindo direção, velocidade vertical, temperatura local e cisalhamento do vento. Esse refinamento foi essencial para aumentar a eficácia do modelo preditivo.

Os resultados obtidos confirmam que é possível desenvolver modelos de alta precisão para prever a velocidade do vento em um horizonte temporal significativo utilizando exclusivamente dados de sensores LiDAR. Este trabalho apresentou como principal contribuição a criação de um modelo seq2seq capaz de prever a velocidade do vento com até 6 horas de antecedência, utilizando exclusivamente dados de sensores LiDAR.

Uma limitação identificada foi o número de amostras disponíveis (7651), que, embora tenha sido suficiente para este estudo, pode ser ampliado em futuros experimentos. Com um volume maior de dados, seria possível aumentar ainda mais a precisão das previsões, especialmente em condições extremas ou eventos menos frequentes. Ademais, a quantidade de eventos *Vento-Sul* presentes na base era de 12, o que inviabilizou a previsão do evento em si. Para garantir um modelo que preveja a ocorrência do evento, é preciso que mais eventos estejam disponíveis para o treinamento.

Para aplicações reais, destaca-se a importância de implementar aprendizado *online* no modelo, permitindo o treinamento periódico com a chegada de novos dados. Isso garantiria que o modelo se adaptasse continuamente às condições mais recentes, mantendo uma precisão consistente ao longo do tempo e aumentando sua aplicabilidade prática em operações de parques eólicos.

Por fim, entende-se que os objetivos definidos para o trabalho foram alcançados. Como trabalhos futuros, sugere-se o cruzamento de dados atmosféricos com as informações dos sensores LiDAR, enriquecendo a base de dados e permitindo maior contextualização meteorológica na região. Além disso, é recomendada a adaptação do modelo para janelas de previsão mais flexíveis, explorando diferentes horizontes temporais para atender a necessidades específicas.

Agradecimentos

Gostaria de expressar minha gratidão à professora Vanessa Souza, pela orientação dedicada, ao Flávio Mota, pela coorientação, e ao professor Arcilan Assireu, pelo suporte e atenciosidade na explicação do problema e na obtenção dos dados. Suas contribuições foram essenciais para a realização deste Trabalho de Final de Graduação.

Referências

- A., T. (2020). Prevendo números: Entendendo métricas de regressão. Accessed: 2024-12-05.
- ABEEólica (2024a). Associação brasileira de energia eólica e novas tecnologias. <https://abeeolica.org.br/> [Accessed: Dezembro/2024].
- ABEEólica (2024b). *Energia Eólica no Brasil: Avanços e Desafios*. Relatório técnico, ABEEólica. <https://brsa.org.br/wp-content/uploads/wpcf7-submissions/32174/ENERGIA-EOLICA-NO-BRASIL-AVANCOS-E-DESAFIOS.pdf> [Accessed: Dezembro/2024].

ABEEólica (2024c). *INFOVENTO - Associação Brasileira de Energia Eólica e Novas Tecnologias*. Infográfico Edição 35 - Outubro/2024, ABEEólica.

Assireu, A. T., Fisch, G., Carvalho, V. S., Pimenta, F. M., de Freitas, R. M., Saavedra, O. R., Neto, F. L., Júnior, A. R., Oliveira, D. Q., Lopes, D. C., et al. (2024). Sea breeze-driven effects on wind down-ramps: Implications for wind farms along the north-east coast of brazil. *Energy*, 294, 130804.

Banco do Nordeste do Brasil (2021). *Caderno Setorial de Energias Renováveis*. Relatório técnico, Banco do Nordeste do Brasil (BNB).

Camilo, C. O. & Silva, J. C. (2009). *Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas*. Technical report, Universidade Federal de Goiás.

Chen, Z., Liu, X., Sun, X., & Jiang, Y. (2021). Deep learning for wind power prediction: Current trends and new perspectives. *SN Computer Science*, 2(4), 1–11.

Dau, A. & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53.

Dong, P. & Chen, Q. (2017). *LiDAR remote sensing and applications*. CRC Press.

Dürre, A., Fried, R., & Liboschik, T. (2015). Robust estimation of (partial) autocorrelation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3), 205–222.

Hand, D. J. (2007). *Principles of Data Mining*. New York, NY, USA: Springer.

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Ibrahim, A. B. A. & Altun, K. (2024). Lstm deep learning techniques for wind power generation forecasting. *Journal of Soft Computing and Artificial Intelligence*.

Lei, P., Ma, F., Zhu, C., & Li, T. (2024). Lstm short-term wind power prediction method based on data preprocessing and variational modal decomposition for soft sensors. *Sensors*, 24(8), 2521.

Li, J., Chen, R., & Huang, X. (2022). A sequence-to-sequence remaining useful life prediction method combining unsupervised lstm encoding-decoding and temporal convolutional network. *Measurement Science and Technology*, 33(8), 085013.

Li, Y., Chen, X., Li, C., Tang, G., Gan, Z., & An, X. (2020). A hybrid deep interval prediction model for wind speed forecasting. *IEEE Access*, 9, 7323–7335.

Liu, X., Wang, L., Xu, Y., & Li, J. (2019). Application of wavelet transform to time series analysis in wind energy. *Applied Sciences*, 9(7), 1345.

Lorenzzetti, J. A. (2015). *Princípios físicos de sensoriamento remoto*. Editora Blucher.

Lyra, G. B., Oliveira-Júnior, J. F., & Zeri, M. (2014). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in alagoas state, northeast of brazil. *Pesquisa Agropecuária Brasileira*, 49(12), 1035–1045.

Olah, C. (2015). Understanding lstm networks.

ResearchGate (2024). Schematic of the measurement principle of the windcube v2 dwl. https://www.researchgate.net/figure/Schematic-of-the-measurement-principle-of-the-WindCube-V2-DWL-v-re-v-rs-v-rw-and-v_fig3_348070017. Accessed: 2024-12-05.

Roeder, L. (2024). Netron - visualizer for neural network, deep learning, and machine learning models.

Saltz, J. S. (2021). Crisp-dm for data science: Strengths, weaknesses and potential next steps. Em *Proceedings of the IEEE International Conference on Big Data* (pp. 2337–2344). Chicago, IL, USA: IEEE.

Santos, P. E. L. & Araújo, F. J. C. (2023). O desenvolvimento da energia eólica no brasil: Uma revisão bibliográfica. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 9(6), 2978–2989.

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(420).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.

Stack Overflow (2024). Illustration from stack overflow. <https://i.stack.imgur.com/RHNrZ.jpg>. Accessed: 2024-12-05.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Kaiser, Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Wang, L., Mykityshyn, A., Johnson, C., & Cheng, J. (2021). Flight demand forecasting with transformers. Em *AIAA Sci-tech 2021 Forum: American Institute of Aeronautics and Astronautics (AIAA)*. Available for public release, PRS Case 21-3415.

Wikipedia (2024). Crisp-dm process diagram. https://pt.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining#/media/Ficheiro:CRISP-DM_Process_Diagram.png. Accessed: 2024-12-05.

Wikipedia (2024). Wind profiler. https://www.wikiwand.com/en/articles/Wind_profiler [Accessed: Dezembro/2024].

Yang, Y., Zhang, C., Dong, H., Xu, B., & Jin, S. (2021). The improved windcube v2 lidar: A case study on a coastal wind farm. *Remote Sensing*, 13(15), 2973.

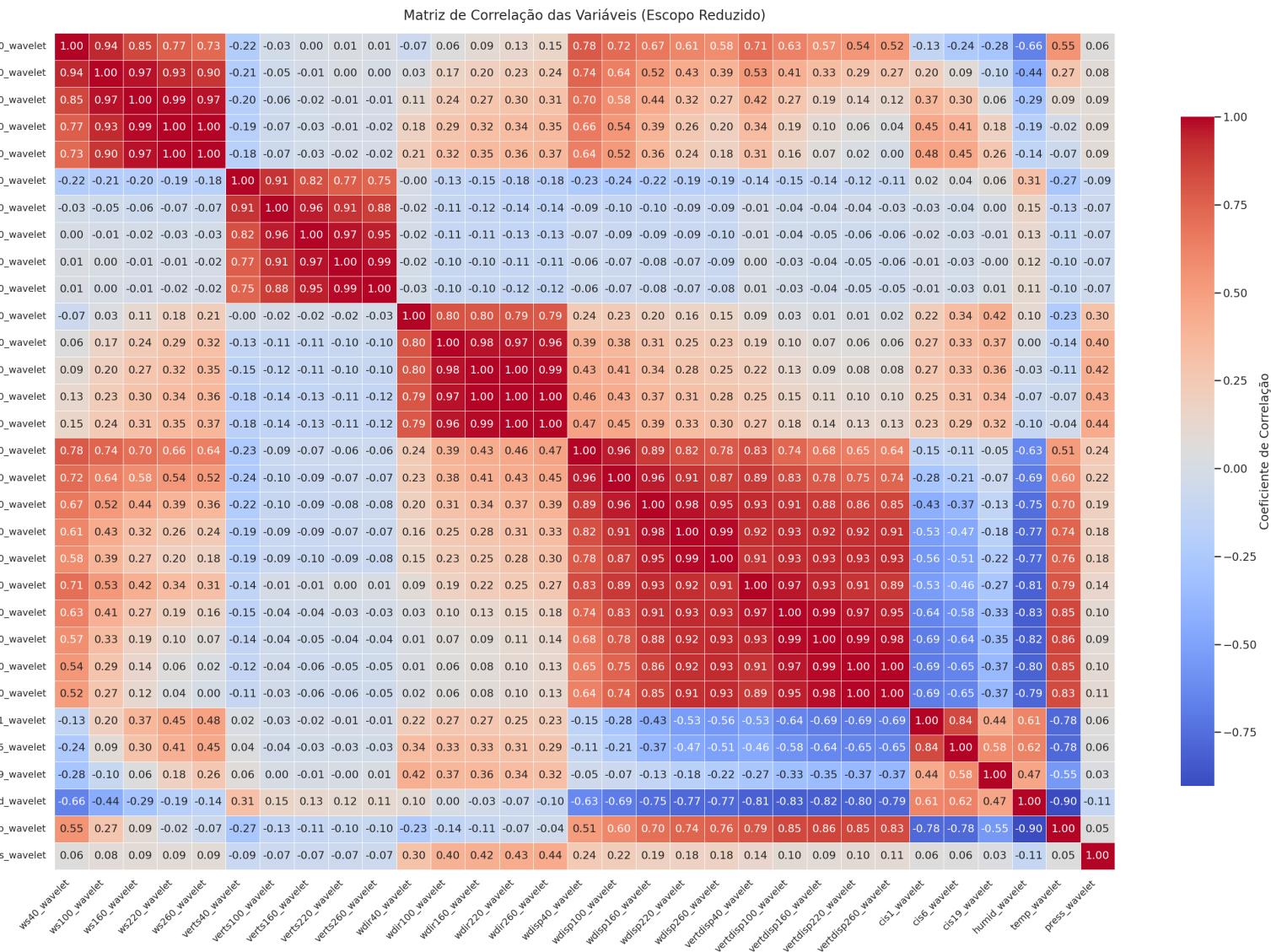


Figura 24: Matriz de correlação com escopo reduzido, ou seja apenas certas variáveis dos dados originais.

