



INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Recife

Engenharia Mecânica

LUCAS LIRA DE PAIVA

**DETECÇÃO DE FALHAS EM AEROGERADORES COM ANÁLISE DE
TEMPERATURAS E MODELOS ESTATÍSTICOS**

Recife - PE

2024

LUCAS LIRA DE PAIVA

**DETECÇÃO DE FALHAS EM AEROGERADORES COM ANÁLISE DE
TEMPERATURAS E MODELOS ESTATÍSTICOS**

Trabalho de conclusão de curso apresentada ao Departamento de Engenharia Mecânica do Instituto Federal de Ciência e Tecnologia de Pernambuco, como requisito para a obtenção do título de Bacharel em Engenharia Mecânica.

Orientador: Prof. Dr. Gustavo de Novaes Pires Leite

Coordenador: José Ângelo Peixoto da Costa

Recife – PE

2024

**Ficha elaborada pela bibliotecária Maria do Perpétuo Socorro
Cavalcante Fernandes CRB4/1666**

P149d
2024

Paiva, Lucas Lira de

Detecção de falhas em aerogeradores com análise de temperaturas e modelos estatísticos. / Lucas Lira de Paiva. --- Recife: O autor, 2024.
48f. il. Color.

TCC (Curso Superior de Bacharelado em Engenharia Mecânica) – Instituto Federal de Pernambuco, Departamento Acadêmico de Cursos Superiores - DACS, 2024.
Inclui Referências.

Orientador: Professor Dr. Gustavo de Novaes Pires Leite.
Coorientador: Professor Dr. José Ângelo Peixoto da Costa.

1. Engenharia mecânica. 2. Análises de temperatura. 3. Aerogerador. 4. Modelos estatísticos-falha. I. Título. II. Leite, Gustavo de Novaes Pires. (orientador). III. Costa, José Ângelo Peixoto da (Coorientador) IV. Instituto Federal de Pernambuco.

CDD620.1(21ed.)

**DETECÇÃO DE FALHAS EM AEROGERADORES COM ANÁLISE DE
TEMPERATURAS E MODELOS ESTATÍSTICOS**

Trabalho aprovado. Recife - PE, 09/09/2024.

Gustavo de Novaes Pires Leite

José Ângelo Peixoto da Costa

Kamila Fernanda Ferreira da Cunha Queiroz.

Recife - PE

2024

Dedico este trabalho aos meus familiares e amigos que sempre me ouviram, incentivaram e apoiaram fielmente na conclusão dele. Dedico também a todos que contribuíram de qualquer forma para a finalização do trabalho.

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão a todos que contribuíram para a realização deste trabalho.

Primeiramente, agradeço a Deus por permitir que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitário, mas que em todos os momentos é o maior mestre, doutor e inspiração que alguém um dia poderá conhecer.

Agradeço aos meus pais, Eduardo e Tereza, e familiares, por todo o amor e carinho a mim fornecido, por sempre acreditarem em mim, me incentivaram a seguir neste curso e estarem ao meu lado com conselhos e instruções no âmbito profissional.

Agradeço também à minha namorada, Beatriz, que sempre me apoiou em todas as minhas decisões ao longo da minha formação profissional, acadêmica e pessoal, e hoje se faz presente todos os dias da minha vida.

Ao meu professor orientador, prof. Dr. Gustavo de Novaes Pires Leite, pela orientação valiosa, dedicação e paciência ao longo de todo esse processo. Sendo o seu conhecimento e conselhos foram fundamentais para o desenvolvimento deste trabalho.

A todos os professores do curso de Engenharia Mecânica, que compartilharam seus conhecimentos não apenas racionais, mas a manifestação do seu caráter e a afetividade para com a educação no meu processo formativo profissional, por tanto que se dedicaram a mim, não apenas por ter me ensinado, mas me feito aprender e refletir sobre os assuntos abordados. O meu muito obrigado também às grandes experiências profissionais e pessoais repassadas a mim, pois contribuíram diretamente à minha formação como pessoa e como um profissional da área de exatas.

Agradeço também aos meus colegas de classe e amigos, pela imensa troca de conhecimento, colaboração e amizade ao longo dessa jornada.

Também reconheço fielmente o apoio da RAESA/Multiner S.A., pelo auxílio no

âmbito do projeto de P&D ANEEL NEO.PROGFALHAS, que apoiou o desenvolvimento do projeto que disponibilizou as informações para o andamento do trabalho.

Por fim, agradeço a todos aqueles que, de alguma forma, contribuíram para a realização deste trabalho, direta ou indiretamente.

RESUMO

Em um mundo em que a demanda por energia está aumentando constantemente, a energia eólica é essencial devido à sua natureza sustentável, custos competitivos e excelente potencial de produção de energia. No entanto, os aerogeradores são máquinas gigantescas com altos custos de instalação, operação e manutenção. Uma estratégia de manutenção incorreta pode arruinar o fluxo de caixa de um parque eólico, comprometendo o investimento de décadas. Os aerogeradores têm sistemas de monitoramento para avaliar se estão funcionando dentro dos limites operacionais. Esses sistemas armazenam uma quantidade enorme de dados valiosos. Que podem fornecer uma indicação antecipada de falhas que ocorrem nas máquinas e, conseqüentemente, evitar custos de manutenção inesperados. A temperatura é comumente usada para monitorar a condição do maquinário mecânico. O sistema de supervisão do aerogerador monitora a temperatura dos componentes críticos e armazena essas informações em um banco de dados. Nesse sentido, o presente trabalho propõe o desenvolvimento de um modelo de aprendizado de máquina baseado na análise da temperatura dos componentes do aerogerador para determinar, com antecedência, se ela está operando em seu comportamento normal. A presente metodologia propõe a implementação de modelos de comportamento normal usando diferentes algoritmos de aprendizado de máquina (rede neural artificial, floresta aleatória e k-ésimo vizinho mais próximo) para detectar falhas em componentes de Aerogeradores como Gearbox, Mainbearing e gerador. São usados dados operacionais de aerogeradores instaladas no Brasil, e os desafios relacionados ao uso de dados reais não rotulados são discutidos ao longo do trabalho. Alguns desafios são a filtragem de dados, a seleção de variáveis, janelas de dados para treinamento e a validação dos modelos. Os modelos podem identificar desvios do comportamento normal, caracterizando uma falha, mesmo antes de o sistema de supervisão acionar os alarmes. Os resultados também apresentam quais métodos têm desempenho eficiente e se há diferenças com relação ao componente analisado. As ações antecipadas da equipe de manutenção para corrigir as falhas podem ser executadas de forma planejada e eficiente, o que não só preserva o aerogerador, mas também aumenta os principais indicadores de desempenho do parque. Foram identificados diferentes períodos de detecção, dependendo da dinâmica de cada componente e das particularidades do modelo. Os modelos com melhor desempenho

foram as redes neurais artificiais e as de floresta aleatória, detectando falhas com 80 a 100 dias de antecedência para o Gearbox e com 90 a 120 dias de antecedência para o Mainbearing.

Palavras-chave: Análise de temperatura; Aerogerador; Modelos estatísticos, Falha.

ABSTRACT

In a world where the energy demand is constantly increasing, wind power is essential because of its sustainable nature, competitive costs, and outstanding potential to produce energy. However, wind turbines are giant machines with high installation, operation, and maintenance costs. A wrong maintenance strategy could ruin the cash flow of a wind farm, compromising the investment of decades. Wind turbines have monitoring systems for evaluating whether they work within operational limits. These systems store a massive amount of valuable data, which could provide an early indication of faults occurring in the machines and, consequently, avoid unexpected maintenance costs. Temperature is commonly used to monitor the condition of mechanical machinery. The wind turbine supervisory system monitors critical components' temperature and stores this information in a database. In this sense, the present work proposes developing a machine-learning model based on the analysis of the temperature of wind turbine components to determine, in advance, whether it is operating in its normal behavior. The present methodology proposes implementing normal-behavior models using different machine learning algorithms (artificial neural network, random forest, and K-nearest neighbor) to detect faults in wind turbine components such as the gearbox, main bearing, and generator. Operational data from wind turbines installed in Brazil are used, and challenges about using unlabeled real data are discussed throughout the paper. Some challenges are filtering data, selecting variables and data windows for training, and validating the models. The models can identify deviation from normal behavior, characterizing a fault, even before the supervisory system triggers the alarms. Results also present what methods perform efficiently and if there are differences regarding the analyzed component. Anticipated actions from the maintenance staff to correct the faults can be carried out in a planned and efficient way, which not only preserves the wind turbine but also increases the wind farm's key performance indicators. Different detection periods were identified depending on each component's dynamics and the model's particularities. The better-performing models were artificial neural networks and decision trees, detecting faults from 80 to 100 days in advance for the gearbox and 90 to 120 days before for the main bearing.

Keywords: Temperature analysis; Wind turbine; Statistical model.

LISTA DE ILUSTRAÇÕES

Quadro 1 - Capacidade de energia eólica no Brasil ao longo dos anos.....	13
Quadro 2 - Aerogeradores de energia eólica em São Miguel do Gostoso (RN).....	16
Quadro 3 - Esquema do Gearbox	18
Quadro 4 - Main Bearing do Aerogerador	18
Quadro 5 - Gerador de um aerogerador.....	19
Quadro 6 - Descrição simples do processo de aprendizado de máquina.	23
Quadro 7 - Fluxograma da metodologia proposta.	25
Quadro 8 - Fluxograma da metodologia do modelo computacional.	25
Quadro 9 - Curva de potência do aerogerador sem o pré-processamento e filtragem de dados.....	27
Quadro 10 - Curva de potência do aerogerador com o pré-processamento e filtragem de dados.....	28
Quadro 11 - Matriz de correlação das variáveis de entrada para o Gearbox.	30
Quadro 12 - Matriz de correlação das variáveis de entrada para o Mainbearing.	31
Quadro 13 - Matriz de correlação das variáveis de entrada para o gerador.	32
Quadro 14 - Layout de um modelo de aprendizado de máquina no estilo de rede neural.	35
Quadro 15 - Layout de um modelo de aprendizado de máquina estilo Floresta aleatória.....	37
Quadro 16 - Layout de um modelo de aprendizado de máquina K-Vizinho mais próximo.....	38
Quadro 17 - Resíduo de temperatura para o aerogerador 1.	42
Quadro 18 - Resíduo de temperatura para o aerogerador 2.	42
Quadro 19 - Resíduo de temperatura para o aerogerador 3.	43
Quadro 20 - 1º aerogerador exemplificando a visualização de sequências fora dos limites.	43

LISTA DE TABELAS

Tabela 1: Exemplos de sensores presentes no sistema SCADA.....	26
Tabela 2 - Variáveis de entrada para o Gearbox.....	30
Tabela 3: Variáveis de entrada para o Mainbearing.....	31
Tabela 4: Variáveis de entrada para o Gerador.	32
Tabela 5: Métricas do modelo computacional para o componente do Gearbox com e sem anomalias.	40
Tabela 6: Métricas do modelo computacional para o componente do Mainbearing com e sem anomalias.	41
Tabela 7: Métricas do modelo computacional para o componente do gerador com e sem anomalias.	41
Tabela 8: Sequência de dados fora dos limites (data, hora e valor do resíduo de temperatura).....	44
Tabela 9: Datas de detecção e ocorrência da falha	44

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Aerogeradores	16
2.2	Inteligência Artificial	20
2.2.1	Impactos e desafios da IA	21
2.3	Aprendizado de máquina	22
3	METODOLOGIA, MODELOS E DADOS DE ENTRADA	24
3.1	Metodologia	24
3.2	Modelo e sua metodologia	25
3.2.1	Sistema SCADA	26
3.2.2	Filtragem e pré-processamento de dados	27
3.3	Dados de entrada	29
3.3.1	Seleção de variáveis	29
3.3.2	Normalização de variáveis	32
3.3.3	Treinamento e teste do modelo	34
3.3.3.1	Rede Neural (RNe):	35
3.3.3.2	Floresta aleatória (RFo)	36
3.3.3.3	K-Vizinhos mais próximos (KNN):	37
3.3.4	Determinação de limites gráficos	38
4	RESULTADOS	40
4.1	Métricas do modelo	40
4.1.1	Gearbox	40
4.1.2	Mainbearing	40
4.1.3	Gerador	41
4.2	Gráfico de séries temporais de temperaturas residuais	41
5	CONCLUSÕES	45
	REFERÊNCIAS	47

1 INTRODUÇÃO

Com a crescente demanda por energia elétrica, os dispositivos de geração de energia renovável estão recebendo mais atenção devido à necessidade de sua implementação. Isso se deve ao fato de que a produção de energia a partir de fontes poluentes e esgotáveis, como petróleo, gás natural e carvão, tem se mostrado prejudicial ao meio ambiente, por exemplo, contribuindo para o aumento do efeito estufa.

Os aerogeradores são opções altamente sustentáveis, mas há despesas de manutenção significativas que poderiam ser mitigadas se as falhas nos aerogeradores pudessem ser "previstas" em um estágio inicial. Conforme citado por Manwell, (Manwell, 2010, p.352.), "a manutenção preditiva permite que os operadores identifiquem falhas incipientes, tomando medidas corretivas antes que o problema se agrave, otimizando assim a eficiência operacional e reduzindo os custos". Com a redução dos custos de manutenção, seria possível um maior investimento em energia renovável em todo o mundo, já que se trata de uma fonte de energia mais barata e ecologicamente correta. A manutenção preditiva baseia-se no comportamento da máquina e em saber o momento exato de realizar paradas estratégicas, minimizando os custos de manutenção corretiva e preventiva que não seriam necessários. Há várias maneiras de "prever" essas falhas, e uma delas é observar e analisar o comportamento de uma variável de temperatura diretamente ligada a um componente importante de um aerogerador (por exemplo, um componente com altos custos de substituição/reparo; um componente com altos períodos de inatividade após uma falha; um componente com uma alta taxa de falhas).

A capacidade de energia eólica no Brasil tem aumentado a cada ano, como é possível visualizar no **Quadro 1**, onde também mostra uma previsão feita pela Abeeólica para o ano de 2024, esperando que cresça exponencialmente para substituir os combustíveis fósseis por energia renovável, os aerogeradores são essenciais para a conversão da energia eólica em energia elétrica. No entanto, esses aerogeradores estão sujeitos a desafios operacionais e falhas que podem levar à redução do desempenho e a reparos dispendiosos. A detecção precoce de falhas é fundamental para garantir a operação ideal e a longevidade dos aerogeradores.

Este trabalho propõe uma abordagem que utiliza análise de temperatura relacionada aos principais componentes do aerogerador e modelos estatísticos. Por exemplo, o monitoramento das temperaturas do Mainbearing, do Gearbox e do gerador podendo detectar desvios das condições normais de operação, permitindo ações de manutenção preditiva e minimizando o tempo de inatividade.

Quadro 1 - Capacidade de energia eólica no Brasil ao longo dos anos.



Fonte: Abeeolica, 2024.

1.1 Objetivos

A crescente busca por fontes de energia sustentável tem impulsionado o desenvolvimento e a implementação de aerogeradores como uma alternativa viável aos combustíveis fósseis. No entanto, a operação e manutenção desses dispositivos representam desafios que podem afetar tanto a eficiência quanto a viabilidade econômica de projetos de energia eólica. Para enfrentar esses desafios, este trabalho tem como foco principal a otimização do plano de manutenção de aerogeradores, com o objetivo de maximizar a vida útil dos equipamentos e minimizar os custos associados a falhas e paradas inesperadas.

Um dos maiores desafios em sistemas de geração eólica é a detecção precoce de falhas. Identificar falhas antecipadamente permite que ações de manutenção preditiva sejam realizadas de forma eficiente, reduzindo o tempo de inatividade e evitando reparos dispendiosos. Isso se alinha à necessidade de um monitoramento

contínuo e ao uso de tecnologias avançadas de detecção, como a análise de dados de temperatura e vibração de componentes críticos, como o Gearbox e o Main bearing.

Neste contexto, também se torna crucial determinar quais algoritmos oferecem melhor desempenho na detecção precoce de falhas. O uso de inteligência artificial e algoritmos de aprendizado de máquina está ganhando destaque por sua capacidade de analisar grandes volumes de dados operacionais e prever padrões de falhas com precisão. O trabalho investiga quais métodos são mais eficazes para melhorar a precisão das previsões e garantir que as intervenções de manutenção ocorram no momento certo.

Por fim, com a crescente conscientização global sobre os impactos ambientais da energia tradicional, este trabalho também busca atrair a demanda por energias sustentáveis, reforçando o papel da energia eólica como uma solução ecológica e economicamente competitiva. Ao otimizar os processos de manutenção e melhorar a confiabilidade dos aerogeradores, o trabalho visa contribuir para o aumento da atratividade da energia eólica e seu potencial para substituir fontes de energia poluentes.

2 FUNDAMENTAÇÃO TEÓRICA

Conforme mencionado anteriormente, a demanda por energia elétrica em todo o mundo tem aumentado ao longo do tempo. No entanto, os impactos ambientais são agora uma preocupação significativa e devem ser evitados por todos os meios. O uso de fontes de energia renováveis está se tornando cada vez mais atraente para todos, e a energia eólica é uma alternativa bem-sucedida para países com condições climáticas favoráveis, no Brasil, a energia eólica pode ser amplamente utilizada devido às suas condições climáticas ideais. No entanto, há desafios a serem enfrentados, como o alto custo dos equipamentos e da manutenção, conforme proposto por Rosa da Silva (2013). Portanto, é necessário minimizar os gastos desnecessários, principalmente em termos de manutenção preventiva. Este trabalho apresenta uma maneira de reduzir e substituir esses custos de manutenção preventiva pela manutenção preditiva, que é mais econômica. Segundo Murgia e Verbeke (A. Murgia, 2021, p.3.), “A seleção dos recursos mais importantes para diagnosticar falhas no componente grande em estudo depende do tempo de amostragem de dados disponível e do componente grande específico que está sendo investigado”. Nesse caso, o foco foi apenas o Gearbox, enquanto Turnbull (2021) discute o mesmo, mas com um componente grande diferente, o gerador. Dito isso, este trabalho não se limita a um componente grande específico, mas se concentra em vários componentes grandes do aerogerador, com a análise sendo determinada por variáveis de temperatura diretamente relacionadas a esses componentes, onde será explorado o diagnóstico de falhas em aerogeradores por meio da análise de dados de temperatura. O estudo se concentra no desenvolvimento de um modelo para o comportamento normal da temperatura do óleo do Gearbox, da temperatura do rolamento do Gearbox e da temperatura do enrolamento do gerador usando redes neurais artificiais (ANN). O modelo leva em conta variáveis de entrada, como geração de energia, temperatura ambiente e valores históricos de temperaturas-alvo em uma e duas instâncias de tempo anteriores. O objetivo dos pesquisadores era aproveitar os padrões de temperatura para melhorar a compreensão e a detecção de possíveis falhas em aerogeradores usando essa abordagem. Este documento usa as mesmas variáveis de entrada de dados de temperatura, mas aborda outros modelos computacionais, como K-nearest neighbors (KNN) e Random Forest (RF), que é um recurso utilizado

por Turnbull (2021), onde se baseia em um modelo de regressão linear para analisar a variabilidade da temperatura em aerogeradores. Assim, a implementação de modelos estatísticos torna-se necessária para realizar análises gráficas quando o modelo computacional está em operação para selecionar os melhores métodos para aplicação em aerogeradores. Isso se baseia em um estudo que compara métodos de determinação de limites, conforme mencionado por Reimann (Reimann, 2005), “Esses estimadores são robustos contra valores extremos. Outra solução seria usar o boxplot para a identificação de valores extremos (Tukey, 1977)”. Uma vez escolhido o método de Tukey como método de determinação de limiares, foi iniciado o processo descrito por Zaher (2009), que apresenta técnicas estatísticas para determinar não apenas limiares, mas também entropias relativas estatísticas. Além disso, ele estabelece uma relação e aplica essas técnicas aos dados coletados, usando o sistema SCADA no caso deste trabalho.

2.1 Aerogeradores

Aerogeradores (**Quadro 2**) são máquinas projetadas para converter a energia cinética do vento em energia mecânica ou elétrica. Eles são elementos-chave na geração de energia eólica, uma das formas mais promissoras e sustentáveis de produção de energia, considerando o impacto ambiental mínimo e o aproveitamento de um recurso natural e abundante.

Quadro 2 - Aerogeradores de energia eólica em São Miguel do Gostoso (RN).



Fonte: G1, 2021.

O funcionamento de um aerogerador é baseado no conceito de conversão de

energia do vento, que, ao passar pelas pás do aerogerador, provoca um movimento de rotação no rotor. Esse rotor está conectado a um gerador, que converte a energia mecânica rotacional em energia elétrica por meio de indução eletromagnética.

O Teorema de Betz estabelece um limite teórico de eficiência para os aerogeradores, mostrando que apenas 59,3% da energia cinética do vento pode ser convertida em energia utilizável. Esse limite é conhecido como o "limite de Betz", e nenhum aerogerador, independentemente de seu design, pode superar esse valor. A potência disponível no vento depende da velocidade do vento, da massa específica do ar e do diâmetro do rotor. A equação da potência disponível no vento (1) ilustra que a potência extraída do vento aumenta com o cubo da velocidade do vento, o que significa que pequenas variações na velocidade podem impactar consideravelmente a quantidade de energia gerada. Além disso, a área de varredura das pás também desempenha um papel essencial na eficiência da captura de energia.

$$E = \frac{1}{2} * \rho * A * m * V^3 \quad (1)$$

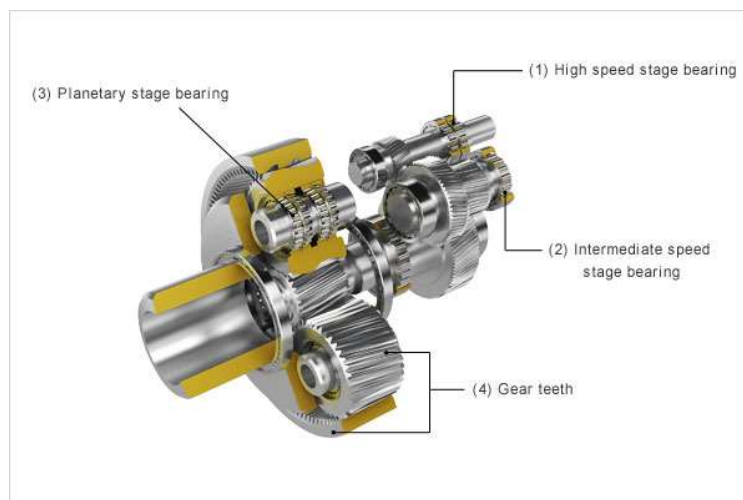
Onde:

- "P" é o valor da potência disponível no vento;
- "m" é o valor da massa do objeto;
- A corresponde à área varrida pelo rotor;
- "V" é o valor da velocidade do vento

Os aerogeradores são compostos por diversos componentes que trabalham em conjunto para capturar a energia do vento e convertê-la em eletricidade de forma eficiente, para esse trabalho em específico, serão citados 3 grandes componentes que serão estudados, analisados e testados, tais componentes são:

(i) - Gearbox: De acordo com Manwell et al. (2010, p.303) "A função da caixa de engrenagens é aumentar a velocidade de rotação do rotor para níveis compatíveis com o gerador, que normalmente requer velocidades superiores a 1.000 rpm, enquanto o rotor gira a cerca de 10-20 rpm". Essa peça é crucial para o funcionamento eficiente da maioria dos aerogeradores, embora alguns designs modernos eliminem a necessidade dela ao usar geradores de baixa rotação.

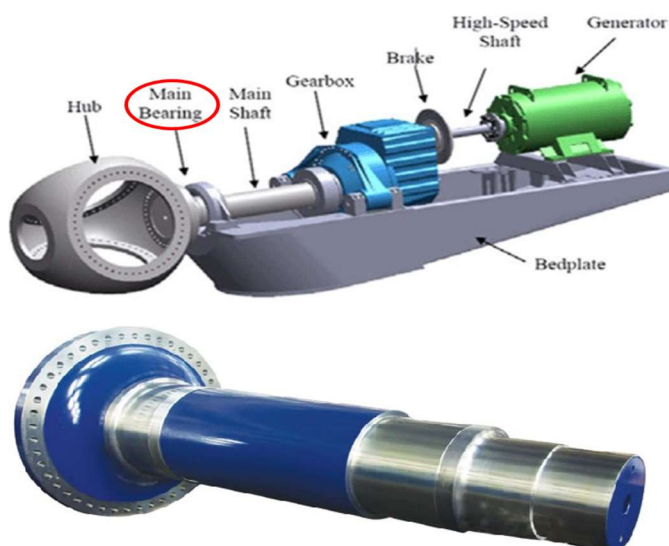
Quadro 3 - Esquema do Gearbox



Fonte: Evident, 2024.

(ii) - Main Bearing: Conforme mencionado por Hau (2013, p.251.) “O Main Bearing (rolamento principal) é responsável por suportar as forças radiais e axiais geradas pelo rotor do aerogerador e garantir a transmissão eficiente dessas forças para o sistema de transmissão”. Ele suporta o eixo do rotor, permitindo que gire suavemente com o mínimo de fricção, suportando as forças combinadas do vento sobre as pás e o peso do próprio rotor, onde garante que a energia seja transferida de forma eficiente para o restante do sistema, incluindo o gearbox e o gerador.

Quadro 4 - Main Bearing do Aerogerador



Fonte: Energês, 2020.

(iii) - Gerador: Este componente converte a energia mecânica rotacional em energia elétrica. Os geradores podem ser síncronos ou assíncronos, e muitas turbinas

modernas utilizam geradores de ímã permanente, devido à sua maior eficiência. Já mencionado por Burton et al. (2011, p.412.) em “Os geradores de ímã permanente são amplamente utilizados em turbinas eólicas modernas devido à sua alta eficiência e à capacidade de operar em baixas velocidades de rotação, eliminando a necessidade de uma caixa de engrenagens em alguns designs”.

Quadro 5 - Gerador de um aerogerador



Fonte: Leroy Somer, 2024.

Os aerogeradores são classificados em duas categorias principais: de eixo horizontal e de eixo vertical. A escolha entre essas duas tipologias depende do local de instalação e das características do vento disponível. Os aerogeradores de Eixo Horizontal (HAWT) são os mais comuns e eficientes. Nesse modelo, o eixo de rotação do rotor é paralelo ao solo. Esses aerogeradores são amplamente utilizados em parques eólicos devido à sua alta eficiência, mas precisam estar sempre voltados para a direção do vento, o que exige um sistema de orientação. Já os aerogeradores de Eixo Vertical (VAWT) são menos comuns, esses modelos têm o eixo de rotação perpendicular ao solo. A vantagem dos VAWTs é que eles podem captar vento de qualquer direção, o que elimina a necessidade de um sistema de orientação. No entanto, sua eficiência é geralmente menor em comparação com os de eixo horizontal.

Os aerogeradores são um dos pilares da energia renovável, possuindo diversas vantagens, como o vento ser uma fonte inesgotável e limpa de energia, contribuindo significativamente para a redução das emissões de carbono e o combate às mudanças climáticas, a redução da Dependência de Combustíveis Fósseis, visto que a energia eólica pode substituir combustíveis fósseis na matriz energética, proporcionando maior independência energética para muitos países e também a descentralização da geração de energia, uma vez que os aerogeradores podem ser instalados em diversas

regiões, incluindo áreas rurais e costeiras, levando eletricidade para locais remotos e promovendo a descentralização do sistema elétrico.

Porém, há também os desafios e limitações que dificultam o uso dos aerogeradores, como a intermitência do vento, na qual a produção da energia depende das condições climáticas, onde a variação do vento pode causar flutuações na geração de eletricidade. A integração de sistemas de armazenamento de energia, como baterias, é uma solução potencial para esse problema, mas ainda enfrenta limitações de custo e eficiência. O impacto visual e acústico pode ser um fator negativo, dependendo do local que o aerogerador é instalado, uma vez que eles emitem altos ruídos devido o movimento das pás, o custo da instalação e manutenção, pois a construção de altas torres, transporte de equipamentos e alto investimento na infraestrutura dos parques, quando somado com a manutenção dos componentes é um fator que deve ser levado em consideração.

2.2 Inteligência Artificial

A Inteligência Artificial (IA) é um campo da ciência da computação que busca desenvolver sistemas capazes de executar tarefas que, quando realizadas por seres humanos, exigem inteligência. Isso inclui habilidades como o raciocínio lógico, aprendizado, percepção visual, compreensão de linguagem natural e tomada de decisão. Segundo Peter e Stuart (Russell; Norvig, 2010, p. 25) "A Inteligência Artificial (IA) é a capacidade de uma máquina imitar o comportamento humano inteligente, como o aprendizado, o raciocínio, a percepção, a interação com o ambiente e a tomada de decisões autônomas.". O objetivo principal é criar sistemas que possam funcionar de forma autônoma e aprender a partir da experiência, sem intervenção constante dos humanos. Essa área abrange diversos subcampos, como o aprendizado de máquina (Machine Learning), que estão sendo utilizado fortemente nesse trabalho, redes neurais artificiais, processamento de linguagem natural (PLN) e visão computacional.

O aprendizado de máquina é uma das abordagens mais importantes dentro da IA. Ele permite que os sistemas de IA melhorem seu desempenho com base em dados, sem serem explicitamente programados para cada tarefa. Existem três tipos principais de aprendizado de máquina, são eles o Aprendizado Supervisionado, onde

o sistema é treinado com dados rotulados, ou seja, os dados de entrada estão associados às respostas corretas na qual o objetivo é que o sistema aprenda a mapear entradas para saídas corretas; o Aprendizado Não Supervisionado, onde o sistema aprende padrões nos dados sem respostas corretas fornecidas, sendo capaz de identificar agrupamentos ou regularidades nos dados; e finalmente, o Aprendizado por Reforço, onde o sistema aprende a tomar decisões interagindo com o ambiente, recebendo recompensas ou penalidades com base em suas ações, tal esse método é amplamente utilizado em jogos e robótica.

Já as redes neurais artificiais são modelos inspirados no funcionamento dos neurônios do cérebro humano. Elas são compostas de várias camadas de "neurônios" artificiais, que processam dados de maneira hierárquica. Uma aplicação notável dessas redes é o Deep Learning, que utiliza redes neurais profundas (com muitas camadas) para processar grandes volumes de dados e resolver problemas complexos, como reconhecimento de fala, tradução automática e diagnósticos médicos.

2.2.1 Impactos e desafios da IA

A IA tem transformado diversos setores de extrema importância para a sociedade, no setor da saúde a IA vem sendo usada para diagnósticos médicos mais precisos, interpretação de exames e desenvolvimento de medicamentos. Modelos baseados em aprendizado de máquina conseguem detectar padrões em imagens, como exames de ressonância magnética, e identificar doenças com precisão superior à dos médicos humanos em alguns casos. Já na indústria, a IA se faz presente na automação industrial em robôs, que realizam atividades de fabricação e logística, como também na parte de sistemas de previsão e otimização de processos baseados em IA são usados para reduzir desperdícios e aumentar a eficiência. Além de setores públicos como transporte e educação, onde veículos autônomos utilizam a IA para navegar e tomar decisões em tempo real e em plataformas educacionais, que a utilizam para a personalização da experiência de aprendizagem dos alunos, adaptando o conteúdo às necessidades de cada um.

Apesar dos avanços, a IA enfrenta desafios éticos e técnicos significativos. As preocupações éticas variam desde os sistemas de IA, que podem herdar e amplificar

preconceitos presentes nos dados de treinamento, levando a discriminação, especialmente em áreas como reconhecimento facial e decisões de crédito; a transparência, onde muitos sistemas de IA, especialmente aqueles baseados em aprendizado profundo, são "caixas pretas", onde até mesmo seus criadores têm dificuldade em explicar como chegaram a determinadas conclusões; e, principalmente a autonomia e controle da IA, pois existem diversas discussões sobre o grau de controle que os humanos devem manter sobre sistemas de IA, especialmente em áreas críticas como saúde, justiça e forças militares.

2.3 Aprendizado de máquina

O aprendizado de máquina é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos estatísticos que permitem que os sistemas computacionais aprendam padrões e tomem decisões com base nos dados de treinamento fornecidos. O objetivo é capacitar as máquinas a aprender padrões e tomar decisões ou fazer previsões com base nesses padrões, melhorando seu desempenho ao longo do tempo. Neste trabalho, foi usada a linguagem de programação Python, que é facilmente acessível e compreensível, com bibliotecas ricas para plotagem de gráficos que foram de grande importância para o desenvolvimento do projeto. A aprendizagem automática tem muitas aplicações em diferentes campos, incluindo classificação e regressão, que foram as aplicações usadas neste estudo:

I. **Classificação:** O aprendizado de máquina pode ser usado para classificar dados em diferentes categorias com base em características ou atributos específicos. Por exemplo, ele pode classificar e-mails como spam ou não spam, identificar se uma transação financeira é fraudulenta ou não, ou até mesmo classificar imagens em diferentes categorias.

II. **Regressão:** O aprendizado de máquina também pode ser usado para análise de regressão, em que o objetivo é prever um valor contínuo de base em um conjunto de variáveis de entrada. Isso pode ser aplicado para prever preços de imóveis com base em suas características, prever a demanda futura de um produto com base em dados de histórico de vendas, entre outros.

Essas são apenas algumas de muitas aplicações do aprendizado de máquina. A amplitude e a versatilidade desse campo possibilitam sua aplicação em vários domínios, como medicina, finanças, manufatura, agricultura, transporte, entre outros. O avanço do aprendizado de máquina também impulsionou a pesquisa em inteligência artificial e possibilitou o desenvolvimento de sistemas mais inteligentes e autônomos. Abaixo no **Quadro 6** é possível entender melhor como funciona o método de aprendizado de máquina utilizado nesse trabalho.

Quadro 6 - Descrição simples do processo de aprendizado de máquina.



O aprendizado de máquina tradicional usa recursos feitos a mão, que são tediosos e custam bastante para desenvolver

Fonte: Merkle, 2023.

3 METODOLOGIA, MODELOS E DADOS DE ENTRADA

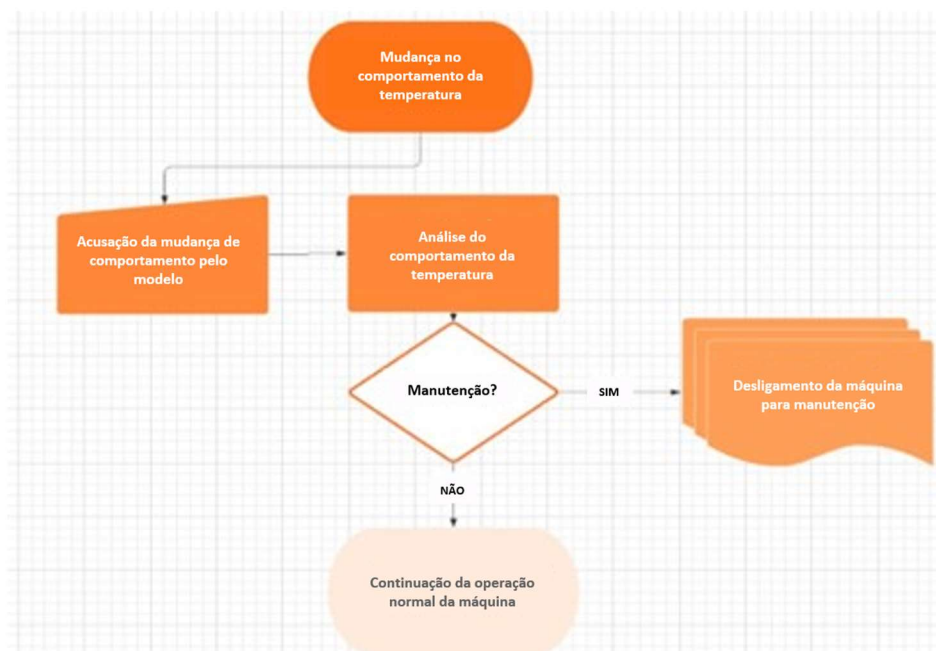
A metodologia proposta neste documento é baseada na detecção de falhas em aerogeradores, envolvendo as seguintes etapas: coleta de dados, pré-processamento, extração de recursos, modelagem estatística e detecção de falhas. Os dados de temperatura do Mainbearing, do gerador e do Gearbox são coletados por meio de sensores incorporados ao sistema do aerogerador. As técnicas de pré-processamento são aplicadas para remover ruídos e exceções dos dados. Em seguida, os algoritmos de extração de recursos são usados para derivar recursos relevantes dos dados de temperatura da série temporal. Modelos estatísticos, como análise de regressão, redes neurais, florestas aleatórias e K-vizinhos mais próximos (K-nearest neighbors) são desenvolvidos usando os recursos extraídos. Por fim, a detecção de falhas é realizada comparando os padrões de temperatura observados com o comportamento normal esperado dos componentes.

3.1 Metodologia

No **Quadro 7** é possível observar um fluxograma que descreve os estágios do objetivo proposto com base no comportamento das variáveis de temperatura que serão analisadas.

No primeiro ponto, é mencionada a mudança de comportamento da variável de temperatura que está sendo analisada pelo modelo, em que as variáveis são influenciadas por fatores externos que fazem com que elas se comportem de forma diferente ao longo do tempo. Em seguida, no segundo ponto, é abordada a detecção dessa mudança no comportamento da temperatura, em que modelos estatísticos e computacionais identificam essa mudança no comportamento das variáveis de temperatura. O tópico da análise do comportamento da temperatura é então discutido, onde, apesar da mudança de comportamento, é necessária uma análise minuciosa para determinar se essa alteração no desempenho da variável indica uma falha iminente ou se é devido a um fator externo emergente que não resultará em uma falha. Com base nessa análise, a decisão de intervir ou não intervir (dependendo do resultado da análise anterior) é tomada pela equipe de manutenção para reparar o item específico que não está funcionando conforme o esperado.

Quadro 7 - Fluxograma da metodologia proposta.

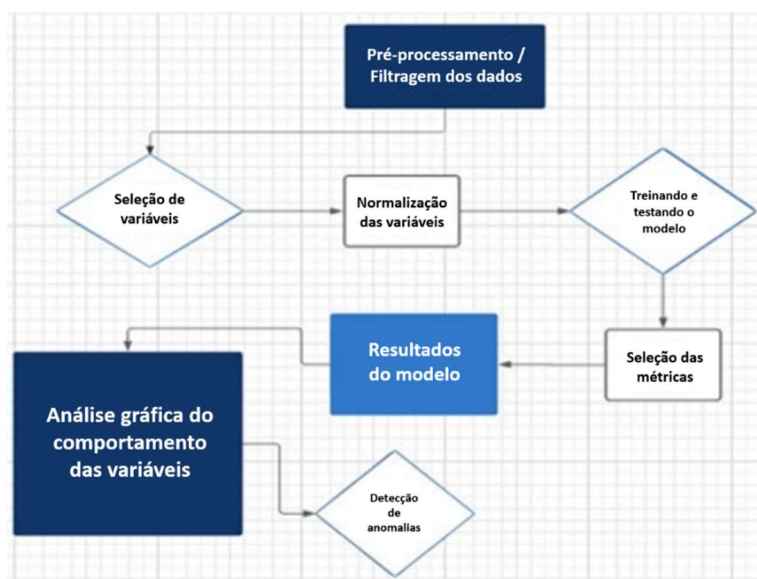


Fonte: Autor, 2024.

3.2 Modelo e sua metodologia

O **Quadro 8** mostra um fluxograma que consiste em 8 etapas principais, e cada uma dessas etapas envolve diferentes áreas de conhecimento para a criação de modelos, desde estatística e programação até a análise da própria engenharia mecânica.

Quadro 8 - Fluxograma da metodologia do modelo computacional.



Fonte: Autor, 2024.

3.2.1 Sistema SCADA

O SCADA (Supervisory Control and Data Acquisition) é um sistema de controle e aquisição de dados usado em vários setores industriais, inclusive na geração de energia eólica. No contexto dos aerogeradores, o sistema SCADA é responsável por coletar, monitorar e controlar remotamente as operações das turbinas analisando variáveis que estão relacionadas com o sistema, como exemplificado na **Tabela 1**. O sistema SCADA em aerogeradores coleta uma ampla gama de informações em tempo real, que são exemplificadas na **Tabela 1**:

Tabela 1: Exemplos de sensores presentes no sistema SCADA.

Coleta de dados SCADA
Temperatura do óleo do Gearbox
Temperatura do Mainbearing
Temperatura do rolamento do gerador
Velocidade do vento
Direção do vento
Energia gerada
Velocidade do rotor
Temperatura ambiente

Fonte: Autor, 2024.

Para o presente estudo, os dados do SCADA são coletados a cada 10 minutos de operação do aerogerador durante 10 anos. A média dos dados, e os valores de variância, máximo e mínimo são registrados para cada intervalo de 10 minutos. Essas médias são documentadas em relatórios e utilizadas no modelo computacional atual.

Em resumo, as aplicações do sistema SCADA para aerogeradores são diversas e incluem:

I. Monitoramento do desempenho: O sistema SCADA permite o monitoramento contínuo do desempenho da turbina, fornecendo informações sobre o produto de energia, velocidade do vento, vibração e outros parâmetros. Isso ajuda a identificar as possíveis falhas.

II. Controle remoto: Os operadores podem controlar remotamente as turbinas por meio do sistema SCADA. Isso inclui o ajuste da velocidade de rotação das pás, o controle do ângulo de inclinação das pás, com a finalidade de otimizar a captação de energia e a realização de outras operações de controle.

III. Diagnóstico de falhas: O sistema SCADA registra dados históricos e em tempo real, possibilitando o desenvolvimento de modelos de detecção de falhas ou

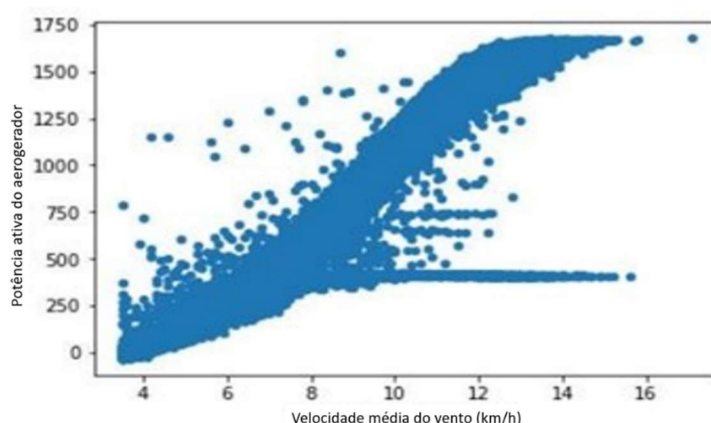
comportamentos anormais nas turbinas. Isso permite uma manutenção preventiva ou corretiva mais eficiente, reduzindo o tempo de inatividade e os custos associados.

IV. Análise de dados: O sistema SCADA fornece dados significativos, entre outros, dados de temperatura. A análise desses dados pode revelar padrões, tendências e relações entre variáveis. Por exemplo, a análise da temperatura pode ajudar a identificar anomalias, como superaquecimento de componentes ou variações fora da faixa esperada.

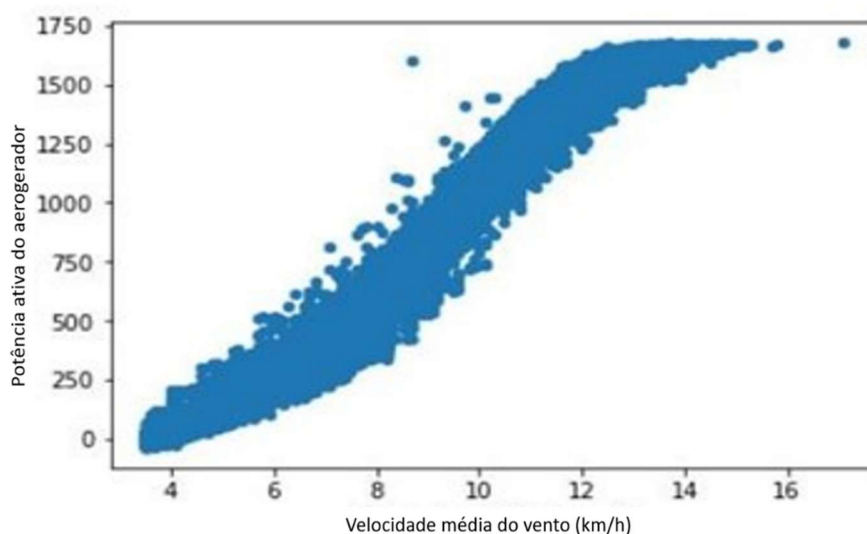
3.2.2 Filtragem e pré-processamento de dados

O **Quadro 8** identifica a primeira etapa como o estágio de pré-processamento e filtragem de dados. Nessa etapa, vários filtros são aplicados aos dados brutos coletados do sistema SCADA para eliminar dados prestes a falhar ou em estado de falha, como a velocidade do vento e a potência gerada, observando sempre a curva de potência. O objetivo é remover os dados que não apresentam o comportamento ideal esperado para o aerogerador em sua operação normal e inserir apenas dados saudáveis no modelo computacional. Nos **Quadros 9 e 10**, é possível visualizar a curva de potência (variável de geração de potência do aerogerador) sem pré-processamento e com pré-processamento, respectivamente. No **Quadro 9**, a curva mostra pontos de dados dispersos e o aerogerador operando com uma limitação de potência. Embora os dados continuem a ser coletados, entende-se que eles devem ser eliminados para melhor análise e treinamento do modelo. No **Quadro 10**, a curva de potência representa o comportamento "ideal" em que a turbina opera conforme o esperado.

Quadro 9 - Curva de potência do aerogerador sem o pré-processamento e filtragem de dados.



Quadro 10 - Curva de potência do aerogerador com o pré-processamento e filtragem de dados.



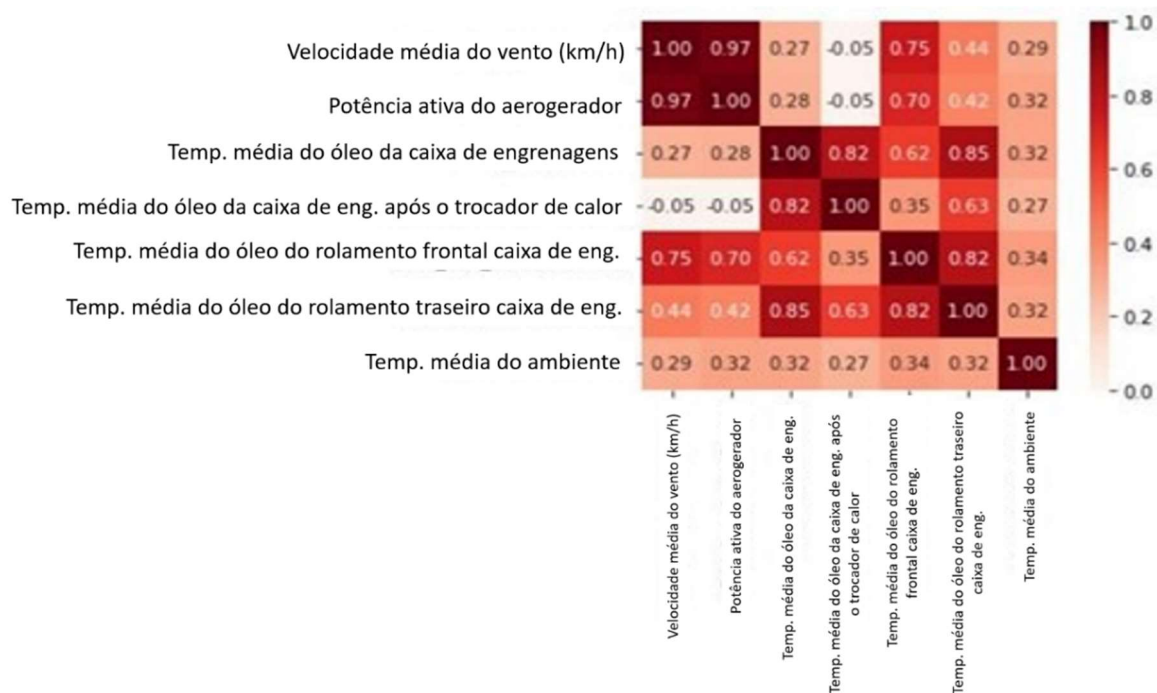
Fonte: Autor, 2024.

Os filtros aplicados são funções na linguagem de programação Python, definidos para identificar pontos considerados discrepantes por meio da análise de dados. O sistema SCADA (sistema de coleta de dados) não consegue identificar mudanças no comportamento da geração de energia em um aerogerador. Essas mudanças podem ocorrer devido às limitações de energia no aerogerador, como quando um componente defeituoso restringe a geração de energia para preservar a saúde da turbina. O sistema detecta apenas os alarmes, mas continua a operação da turbina e a coleta de dados. No entanto, esses dados são considerados "brutos", pois o aerogerador não está operando com 100% da capacidade. Portanto, é necessária uma técnica de análise para identificar essas limitações de energia (conforme mostrado no **Quadro 9**) e o objetivo principal é extrair os dados "brutos" por meio da referência cruzada dos relatórios de alarme com o histórico de falhas e manutenção da turbina. Isso garante que o modelo computacional seja treinado somente com dados "saudáveis". Conforme mencionado anteriormente, os dados de velocidade do vento fora dos limites operacionais para geração de energia são desconsiderados. Acima da nacela, há dois sensores de velocidade do vento no sistema e a diferença entre as velocidades do vento devem ser muito próximas, o que significa que, quando discrepantes, os dados também devem ser classificados como anômalos.

3.3 Dados de entrada

3.3.1 Seleção de variáveis

Para essa análise, o modelo se baseou nas turbinas que apresentaram falhas ou anomalias nos principais componentes observados durante sua operação em 10 anos de dados coletados pelo sistema SCADA. Para a seleção das variáveis de entrada, foram feitas avaliações para determinar quais variáveis seriam selecionadas dependendo do componente principal que estava sendo analisado. As matrizes de correlação ajudaram nessa análise, conforme mostrado nos **Quadros 11** (alta relação com a temperatura média dos rolamentos do Gearbox), **12** (alta relação com as temperaturas do rolamento do gerador) e **13** (alta relação com a temperatura média do gerador), mostrando claramente a alta relação das variáveis de temperatura entre si e com a velocidade média do vento e da potência. As variáveis que foram selecionadas têm as relações mais fortes com seu respectivo grande componente, apresentadas nas **Tabela 2** (Gearbox), **Tabela 3** (Mainbearing) e **Tabela 4** (Gerador). Além de selecionar as variáveis, é necessário selecionar os aerogeradores que apresentaram falhas nos componentes principais em estudo, para que seja comparado o período em que os aerogeradores estavam em falha, com o que estava em perfeito funcionamento, com o intuito de validar a metodologia. Muitos aerogeradores foram analisadas, e, as selecionadas tinham o componente principal apresentando anomalias, visto que o objetivo era diminuir as perdas de energia gerada pelos parques eólicos, as turbinas selecionadas foram aquelas que apresentaram as maiores taxas de downtime, e também as que possuíam grandes componentes danificados, que são os mais complexos e caros de se realizar a manutenção, no geral foram analisadas 15 aerogeradores, sendo 7 para o Gearbox, 4 para o Mainbearing e 4 para o gerador.

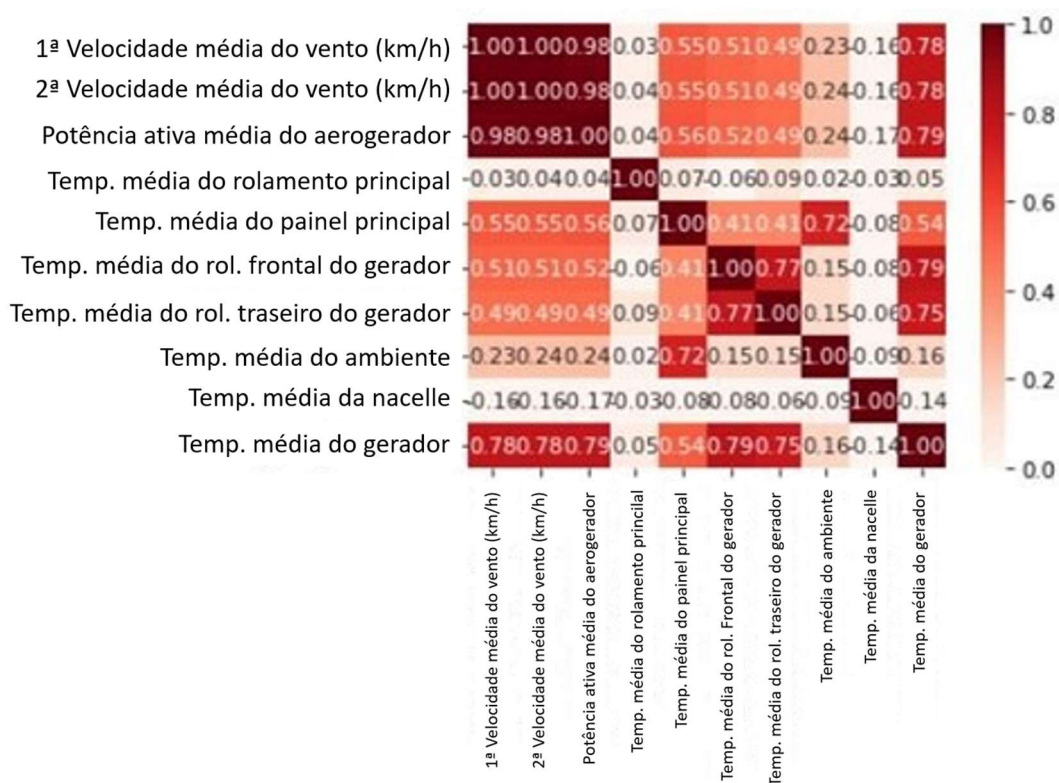
Quadro 11 - Matriz de correlação das variáveis de entrada para o Gearbox.

Fonte: Autor, 2024.

Tabela 2 - Variáveis de entrada para o Gearbox.

Gearbox - Variáveis-alvo = temperaturas do óleo do Gearbox
Velocidade do vento
Potência ativa gerada
Temperatura do óleo do Gearbox
Temperatura do óleo do Gearbox após o trocador de calor
Temperatura do rolamento dianteiro do Gearbox
Temperatura do rolamento traseiro do Gearbox
Temperatura ambiente

Fonte: Autor, 2024.

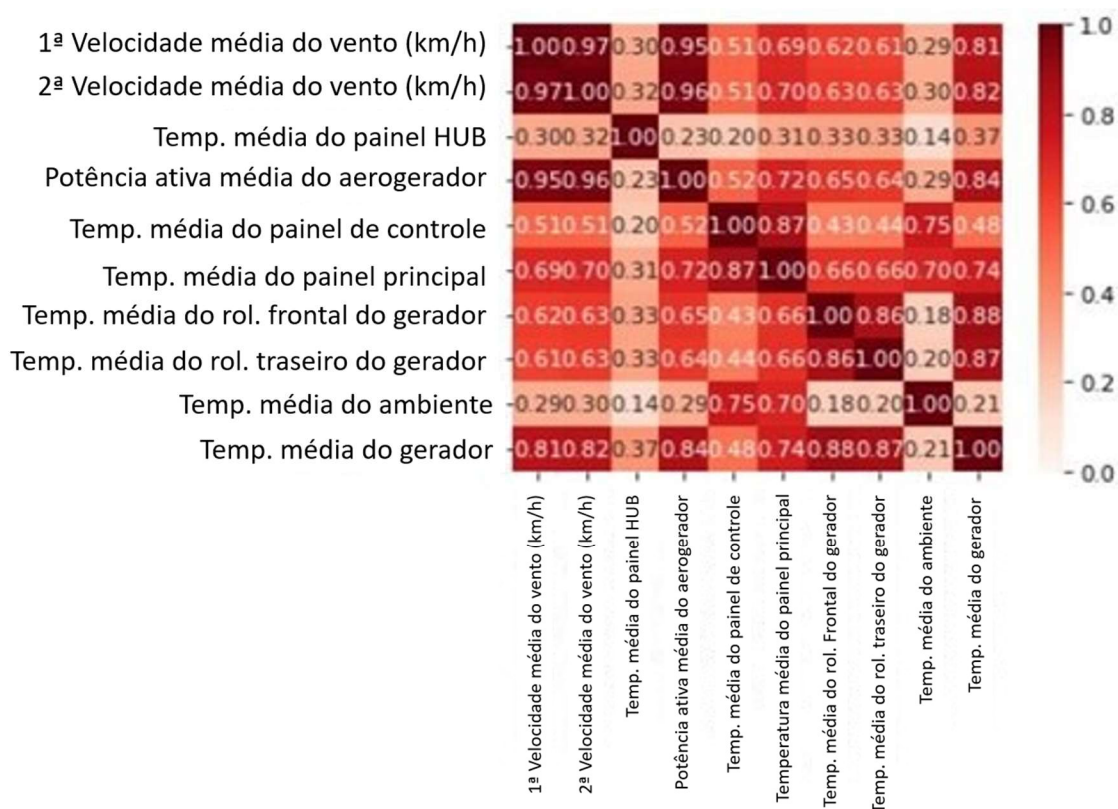
Quadro 12 - Matriz de correlação das variáveis de entrada para o Mainbearing.

Fonte: Autor, 2024.

Tabela 3: Variáveis de entrada para o Mainbearing.

Mainbearing - Variável-alvo = temperatura do Mainbearing
Velocidade do vento 1 e 2
Potência ativa gerada
Temperatura do Mainbearing
Temperatura do painel de controle principal
Temperatura do mancal dianteiro do gerador
Temperatura do rolamento traseiro do gerador
Temperatura da nacelle
Temperatura ambiente

Fonte: Autor, 2024.

Quadro 13 - Matriz de correlação das variáveis de entrada para o gerador.

Fonte: Autor, 2024.

Tabela 4: Variáveis de entrada para o Gerador.

GERADOR - Variável-alvo = Temperatura do rolamento do gerador dianteiro
Velocidade do vento 1 e 2
Potência ativa gerada
Temperatura do painel de controle
Temperatura do painel principal
Temperatura do mancal dianteiro do gerador
Temperatura do rolamento traseiro do gerador
Temperatura do gerador
Temperatura ambiente

Fonte: Autor, 2024.

3.3.2 Normalização de variáveis

A normalização das variáveis é feita para padronizar a escala das variáveis de entrada antes de alimentá-las em um modelo computacional para análise de dados. Essa etapa é importante porque muitos algoritmos e modelos de aprendizado de máquina são sensíveis à escala das variáveis. A normalização ajuda a evitar que variáveis com escalas muito diferentes afetem desproporcionalmente a análise ou o modelo.

Alguns motivos pelos quais a normalização de variáveis é o equilíbrio de pesos, em que os algoritmos de aprendizado de máquina que usam medidas de

distância ou similaridade entre variáveis, como regressão linear, k-mean, SVM (Support Vector Machine) e redes neurais, podem ser influenciados pela escala das variáveis. Uma variável com uma escala muito maior do que as outras pode dominar o processo de modelagem e resultar em pesos ou coeficientes tendenciosos. A normalização ajuda a equilibrar a influência de cada variável, atribuindo-lhes uma escala comparável; a convergência mais rápida, ou seja, os algoritmos de otimização usados nos modelos de aprendizado de máquina, como a descida de gradiente, convergem mais rapidamente quando as variáveis estão na mesma escala. A normalização das variáveis acelera o processo de otimização ao reduzir o número de iterações necessárias para chegar a um resultado satisfatório; impacto reduzido de outliers, nessa expressão, os outliers em uma variável podem afetar significativamente determinados algoritmos. A normalização das variáveis pode reduzir o impacto desses outliers, tornando o modelo mais robusto e menos sensível a valores extremos; interpretação dos coeficientes, onde, em alguns casos, a normalização das variáveis também facilita a interpretação dos coeficientes ou pesos atribuídos a cada variável. Quando as variáveis estão na mesma escala, é mais fácil comparar a magnitude dos coeficientes e entender sua importância relativa para o modelo; preservação da privacidade, em alguns cenários em que a privacidade dos dados é uma preocupação, a normalização pode ser aplicada como uma técnica de anonimização, em que os valores originais são substituídos por valores normalizados que ocultam informações confidenciais.

Em resumo, a normalização de variáveis é uma etapa importante na análise de dados e na modelagem de aprendizado de máquina, pois ajuda a melhorar o desempenho, a estabilidade e a interpretabilidade dos modelos e facilita a comparação entre as variáveis. Para essa tarefa, foi usada a normalização "StandardScaler" ou "Z-score", que tem como a base do cálculo a equação (2). Essa técnica é amplamente usada na análise de dados e na modelagem de aprendizado de máquina para normalizar as variáveis em uma escala com média igual a zero e desvio padrão igual a um.

A fórmula para padronizar uma variável "X" é dada pela equação (2):

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

Onde:

- “Z” é o valor padronizado da variável “X”;
- “X” é o valor original da variável;
- “μ” é a média da variável;
- “σ” é o desvio padrão da variável.

No presente trabalho, foi utilizada a função “StandardScaler” da biblioteca “sklearn” que implementa essa padronização calculando a média e o desvio padrão dos dados de treinamento e, em seguida, aplicando a transformação para normalizar os dados.

3.3.3 Treinamento e teste do modelo

Um modelo de regressão linear foi usado para treinar e testar o modelo. Na aprendizagem automática, a regressão linear é amplamente empregada como uma técnica de aprendizagem supervisionada para estimar valores numéricos contínuos com base em um conjunto de variáveis de entrada. A regressão linear tem como objetivo encontrar uma equação matemática que represente a relação linear entre as variáveis independentes e a variável dependente. Essa equação é usada para fazer estimativas dos valores da variável dependente com base nos valores conhecidos das variáveis independentes. A regressão linear é representada pela equação (3).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (3)$$

onde Y é a variável dependente, X_1, X_2, \dots, X_n são as variáveis independentes, β_0 , é o coeficiente que representa a interceptação com o eixo das ordenadas e $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes de regressão que representam a inclinação da regressão.

A regressão linear pode ser aplicada a problemas como previsão de preços, análise de tendências, avaliação do impacto de variáveis e modelagem de relações entre variáveis. É uma ferramenta versátil e amplamente utilizada no aprendizado de máquina, permitindo a compreensão e a previsão de fenômenos complexos com base nos dados disponíveis.

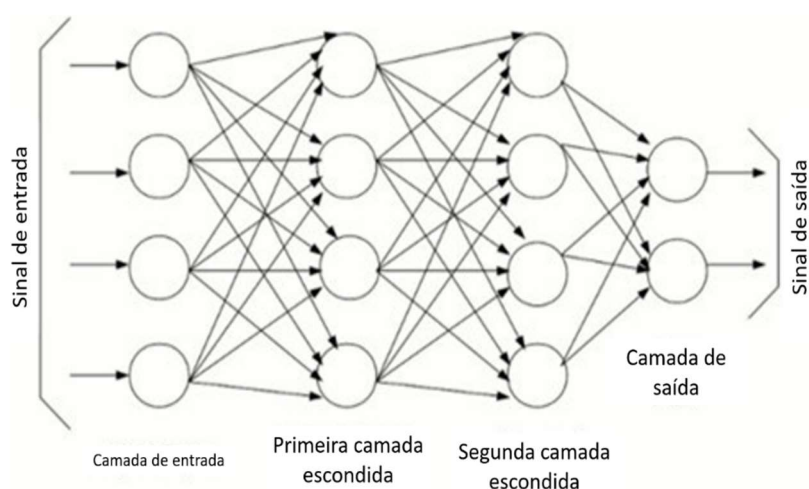
Para o estudo em questão, também feito por Ángel Dávila (2021), 70% dos

dados foram usados para treinamento e 30% para testar um modelo de regressão linear. A validação foi realizada em um período com dados completos e não filtrados. Os períodos de treinamento, teste e validação do modelo variam de acordo com o componente mecânico que está sendo estudado e o aerogerador selecionada. Para exemplificar, foi selecionada uma turbina que teve seu Gearbox comprometida devido a uma quebra no rolamento do Gearbox. Para essa turbina, o período de treinamento foi de 1º de janeiro de 2016 a 5 de junho de 2017, e o período de teste foi de 5 de junho de 2017 a 31 de dezembro de 2019. Foram usados três tipos de algoritmos de aprendizado de máquina, descritos nas seções a seguir.

3.3.3.1. Rede Neural (RNe):

O objetivo do aprendizado de máquina é um método de análise de dados que automatiza a construção de modelos analíticos. As redes neurais são um dos métodos de aprendizado de máquina mais proeminentes, altamente eficazes em problemas de classificação, regressão, processamento de imagens e reconhecimento de fala, entre outros. As redes neurais são compostas de unidades interconectadas chamadas neurônios artificiais ou unidades de processamento. Esses neurônios são organizados em camadas, conforme mostrado no **Quadro 14**, sendo que a camada de entrada é responsável por receber os dados e as camadas escondidas e de saída realizam o processamento e produzem resultados.

Quadro 14 - Layout de um modelo de aprendizado de máquina no estilo de rede neural.



Fonte: Monolito Nimbus, 2017.

Cada conexão entre os neurônios tem um peso que determina a importância

dessa conexão na transmissão do sinal. Há várias arquiteturas de redes neurais usadas no aprendizado de máquina. Uma das mais comuns é a Rede Neural Artificial (ANN) Feedforward, também conhecida como Perceptron Multicamadas (MLP). Nessa arquitetura, os neurônios são organizados em camadas sucessivas, em que cada neurônio recebe informações dos neurônios da camada anterior e passa sua saída para os neurônios da camada seguinte.

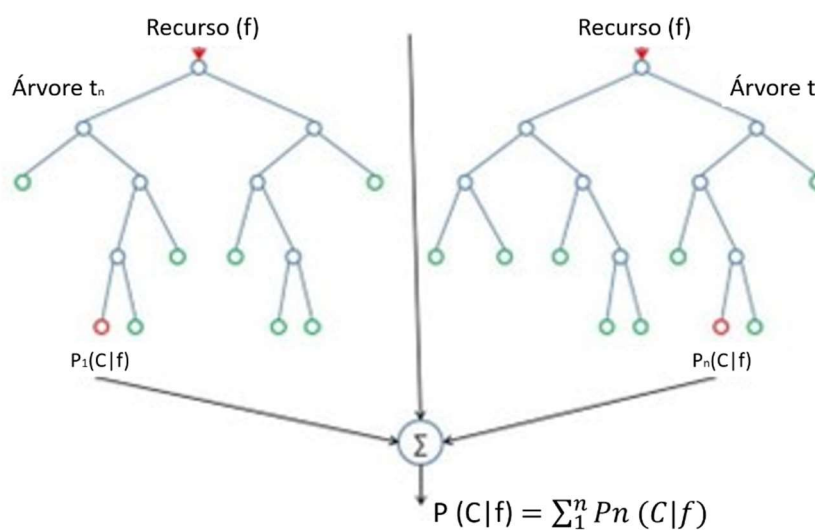
As redes neurais são um método poderoso de aprendizado de máquina, com a capacidade de lidar com problemas complexos e não lineares. Entretanto, a escolha da arquitetura correta, o ajuste adequado dos hiperparâmetros e a disponibilidade de dados de treinamento de alta qualidade são fatores essenciais para a obtenção de bons resultados. As redes neurais continuam a ser objeto de estudo e pesquisa, com avanços contínuos que expandem seu potencial e aplicabilidade em vários domínios.

3.3.3.2. Floresta aleatória (RFo)

O Random Forest (Floresta aleatória) é um algoritmo de aprendizado de máquina que combinam várias árvores de decisão para tarefas de classificação e regressão. Essa técnica se destaca por sua capacidade de lidar com dados complexos e variáveis de entrada altamente correlacionadas. O Random Forest é construído a partir de um conjunto de árvores de decisão, tal conjunto é representado pelo **Quadro 15**, em que cada árvore é treinada em uma amostra aleatória do conjunto de dados de treinamento. Durante o treinamento, cada árvore é desenvolvida de forma independente, dividindo os dados em diferentes nós com base no ganho de informações ou nos critérios do índice de Gini, que é uma medida de desigualdade desenvolvida por um estatístico italiano chamado Corrado Gini. No processo de classificação, o Random Forest combina as previsões de cada árvore para determinar a classe mais provável. Na regressão, é calculada a média das previsões de todas as árvores.

Essa combinação de previsões reduz a tendência de sobreajuste, tornando o Random Forest um modelo robusto e capaz de generalizar bem o treinamento dos modelos.

Quadro 15 - Layout de um modelo de aprendizado de máquina estilo Floresta aleatória.



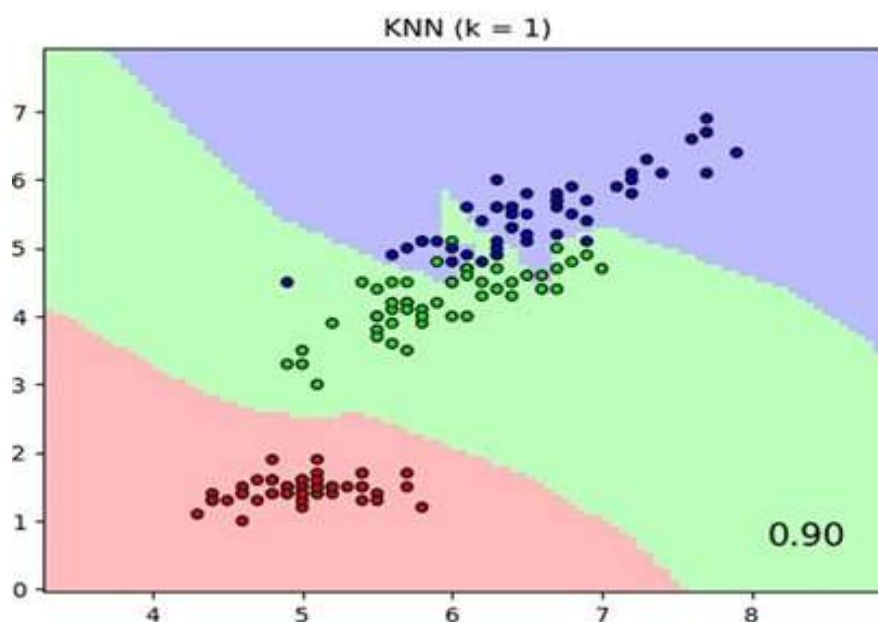
Fonte: Medium, 2018.

O Random Forest é um algoritmo versátil e eficaz que lida com problemas de classificação e regressão em diferentes domínios. Sua capacidade de lidar com variáveis correlacionadas e sua robustez contra o ajuste excessivo fazem dele uma escolha popular no aprendizado de máquina.

3.3.3.3. K-Vizinhos mais próximos (KNN):

O K-Nearest Neighbors (KNN) é um algoritmo de aprendizado de máquina baseado no princípio de que instâncias semelhantes tendem a pertencer à mesma classe ou a ter valores semelhantes. Esse algoritmo não é paramétrico, o que significa que ele não faz suposições específicas sobre a distribuição de dados. O KNN determina a classe ou o valor de um ponto de dados desconhecido com base nas classes ou valores de seus pontos de dados vizinhos mais próximos. A distância entre os pontos é calculada usando métricas como a distância euclidiana. O parâmetro K define o número de vizinhos considerados para a tomada de decisões. Ele atribui a classe mais frequente entre os K vizinhos mais próximos ao ponto de dados desconhecido, conforme apresentado no **Quadro 16**. Para a regressão, o valor previsto é calculado como a média dos valores dos K vizinhos mais próximos. O valor de K pode variar de acordo com o problema e deve ser escolhido com cuidado, considerando o equilíbrio entre a tendência e a variação.

Quadro 16 - Layout de um modelo de aprendizado de máquina K-Vizinho mais próximo.



Fonte: MQL5, 2024.

O K-Nearest Neighbor é um algoritmo de aprendizado de máquina simples, porém eficaz, adequado para problemas com estrutura bem definida e dados rotulados. Entretanto, sua eficácia pode ser afetada por dados desequilibrados ou de alta dimensão. É importante considerar esses fatores ao aplicar o algoritmo KNN e ajustar o valor de K de acordo com as características do problema.

3.3.4 Determinação de limites gráficos

A definição de limites superior e inferior nos gráficos de temperatura residual é uma prática comum para auxiliar na análise de resíduos e identificar padrões ou comportamentos anômalos nos dados. Os gráficos de resíduos são usados para avaliar a qualidade do ajuste do modelo de regressão e verificar se há padrões nos erros residuais, ou seja, se os erros apresentam alguma estrutura sistemática. Eles são construídos plotando os resíduos (a diferença entre os valores observados e os valores previstos pelo modelo) em relação às variáveis independentes ou em relação ao tempo, dependendo do contexto do problema. A definição dos limites superior e inferior nos gráficos residuais permite identificar as regiões em que os resíduos se desviam além do esperado, indicando possíveis problemas ou padrões incomuns nos dados. Esses limites podem ser baseados em critérios estatísticos, como desvios padrão dos resíduos ou intervalos de confiança. É possível identificar pontos que

estão fora dos limites estabelecidos, definindo esses limites, indicando resíduos incomuns ou discrepantes. Esses pontos podem fornecer *insights* sobre possíveis problemas com o modelo, como suposições não atendidas, presença de outliers ou influência de observações atípicas. A definição de limites superiores e inferiores nos gráficos de temperatura residual ajuda a detectar comportamentos anômalos em relação ao modelo ajustado. Isso permite uma análise mais aprofundada e ajuda a identificar possíveis melhorias ou ajustes necessários no modelo de regressão. A identificação desses comportamentos incomuns pode fornecer informações valiosas sobre o sistema ou fenômeno estudado.

Para o presente estudo, foi usado o método de Tukey que se baseia no cálculo do intervalo interquartil (IQR), que é a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1) de um conjunto de dados. O primeiro quartil (Q1) é o valor que divide os dados em 25% dos valores mais baixos, e o terceiro quartil (Q3) divide os dados em 25% dos valores mais altos. O IQR é apresentado na Equação (4).

$$IQR = Q3 - Q1 \quad (4)$$

Para identificar outliers usando o método Tukey, os limites superior e inferior são normalmente definidos como múltiplos do IQR acima e abaixo do Q3 e Q1, respectivamente, utilizando as equações (5) e (6). Qualquer valor além desses limites é considerado um outlier.

$$\text{Limite superior} = Q1 - (n * IQR) \quad (5)$$

$$\text{Limite inferior} = Q3 + (n * IQR) \quad (6)$$

onde o parâmetro "n" é um fator que controla a sensibilidade na detecção de outliers. Os valores comuns para "n" são 1,5 ou 3. Os valores além dos limites superior e inferior são considerados outliers.

4 RESULTADOS

4.1 Métricas do modelo

As métricas dos modelos de análise computacional têm a função principal de avaliar e quantificar o desempenho dos modelos em relação aos dados de teste ou validação. Essas métricas fornecem medidas objetivas que permitem a comparação e a seleção de modelos e identificam áreas em que o modelo pode precisar de aprimoramento. As métricas variam de acordo com o tipo de problema (classificação, regressão etc.) e o objetivo específico da análise. Quanto à análise específica de modelos de regressão linear, foram selecionadas as métricas mais comuns para esses modelos, também utilizado na dissertação do mestrado de Magalhães (2011), que incluem R^2 (coeficiente de determinação), MAE (erro absoluto médio) e RMSE (erro quadrático médio).

4.1.1 Gearbox

Na **Tabela 5**, é perceptível que o modelo de regressão linear com aprendizado e máquinas, sendo o de Random Forest com o melhor desempenho para o Gearbox, mostrando bons valores métricos quando testado em turbinas saudáveis. Isso indica que o modelo pode prever com precisão o comportamento ideal de um aerogerador. Entretanto, as métricas não são tão boas quando testadas em turbinas com anomalias, o que é exatamente o esperado. Como o modelo espera valores semelhantes àqueles em que foi treinado, quando ocorrem valores anômalos, as métricas produzirão resultados ruins devido ao desvio da temperatura prevista em relação à temperatura real (aumento do resíduo de temperatura).

Tabela 5: Métricas do modelo computacional para o componente do Gearbox com e sem anomalias.

GEARBOX (COM/SEM ANOMALIAS)			
PARÂMETRO	RNE (com/sem anomalias)	RFO (com/sem anomalias)	KNN (com/sem anomalias)
R^2	0,490 / 0,95	0,308 / 0,97	0,228 / 0,92
Erro médio absoluto (MAE)	0,670 / 0,320	0,89 / 0,370	1,0 / 0,4
Erro quadrático médio (RMSE)	1,053 / 0,497	1,226 / 0,61	1,294 / 0,66

Fonte: Autor, 2024.

4.1.2 Mainbearing

A mesma análise que foi feita para o componente do Gearbox também pode

ser feita para o Mainbearing. Portanto, foram obtidos bons resultados para esse componente principal, quando são analisadas as turbinas saudáveis e resultados ruins quando inseridas no modelo aerogeradores com anomalias, mostrando o bom funcionamento do modelo computacional, como pode ser visto na Tabela 6 abaixo.

Tabela 6: Métricas do modelo computacional para o componente do Mainbearing com e sem anomalias.

MAINBEARING (COM/SEM ANOMALIAS)			
PARÂMETRO	RNE (com/sem anomalias)	RFO (com/sem anomalias)	KNN (com/sem anomalias)
R ²	0.703 / 0.999	0.736 / 0.913	0.655 / 0.920
Erro médio absoluto (MAE)	3.29 / 2.425	3.41 / 2.62	3.2 / 2.6
Erro quadrático médio (RMSE)	10.054 / 4.193	10.153 / 4.294	9.911 / 4.278

Fonte: Autor, 2024.

4.1.3 Gerador

A mesma análise que foi feita para o componente do Gearbox e o Mainbearing também pode e foi feita para o componente do gerador. Enfim, foram obtidos resultados já esperados, entendendo que o modelo também funciona para esse componente principal, como pode ser visto na Tabela 7 abaixo.

Tabela 7: Métricas do modelo computacional para o componente do gerador com e sem anomalias.

GERADOR (COM/SEM ANOMALIAS)			
PARÂMETRO	RNE (com/sem anomalias)	RFO (com/sem anomalias)	KNN (com/sem anomalias)
R ²	0.488 / 0.77	0.468 / 0.722	0.449 / 0.67
Erro médio absoluto (MAE)	2.00 / 1.916	2.21 / 2.24	2.4 / 2.4
Erro quadrático médio (RMSE)	7.023 / 4.603	7.156 / 5.05	7.281 / 5.505

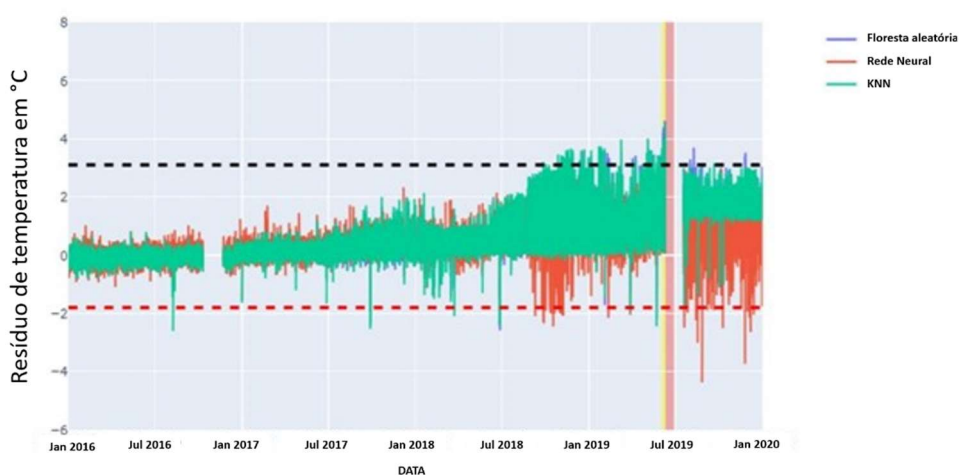
Fonte: Autor, 2024.

4.2 Gráfico de séries temporais de temperaturas residuais

Três aerogeradores, representando cada componente principal estudado foram selecionadas para mostrar o comportamento da temperatura residual ao longo do tempo e a tendência da curva residual até que haja uma intervenção real da equipe de manutenção. Em primeiro lugar, será apresentada o aerogerador que representa anomalias no componente do Gearbox, seguida pelo aerogerador que representa anomalias no Mainbearing e, por fim, o aerogerador que mostra o comportamento anomalias no gerador.

Nos **Quadros 17, 18 e 19** é possível observar a tendência de aumento dos resíduos à medida que eles se aproximam do ponto de falha, o que é favorável para o analista, pois mostra que o modelo está bem treinado e é capaz de revelar tendências de aumento de temperatura. Embora as visualizações de cada componente sejam diferentes, elas apresentam o mesmo comportamento de aumento dos resíduos, à medida que o período de falha se aproxima. No **Quadro 17**, há um leve crescimento do resíduo.

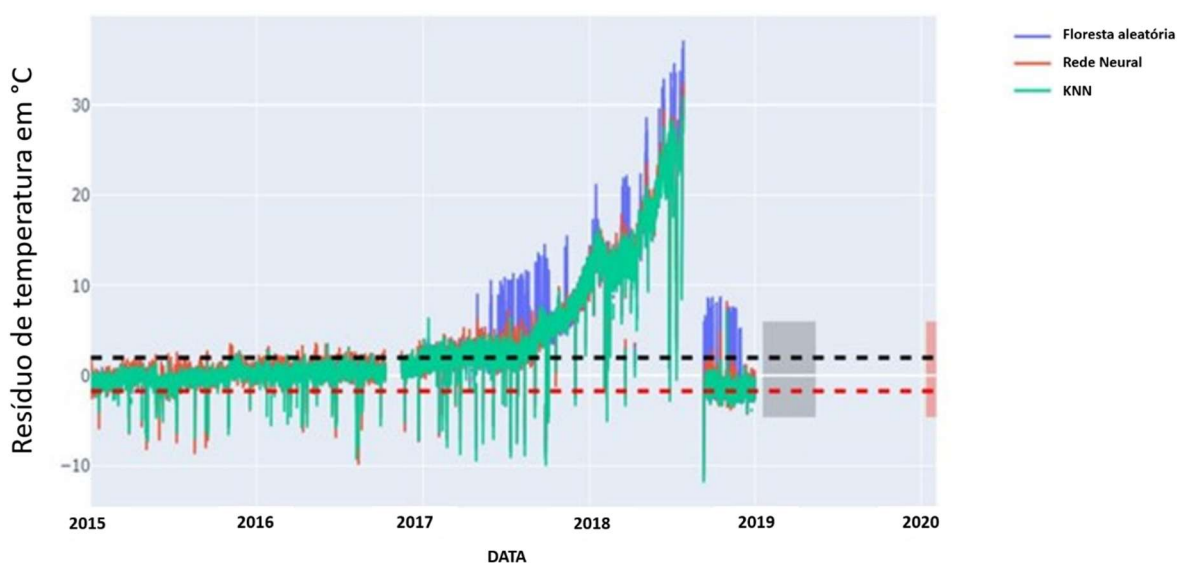
Quadro 17 - Resíduo de temperatura para o aerogerador 1.



Fonte: Autor, 2024.

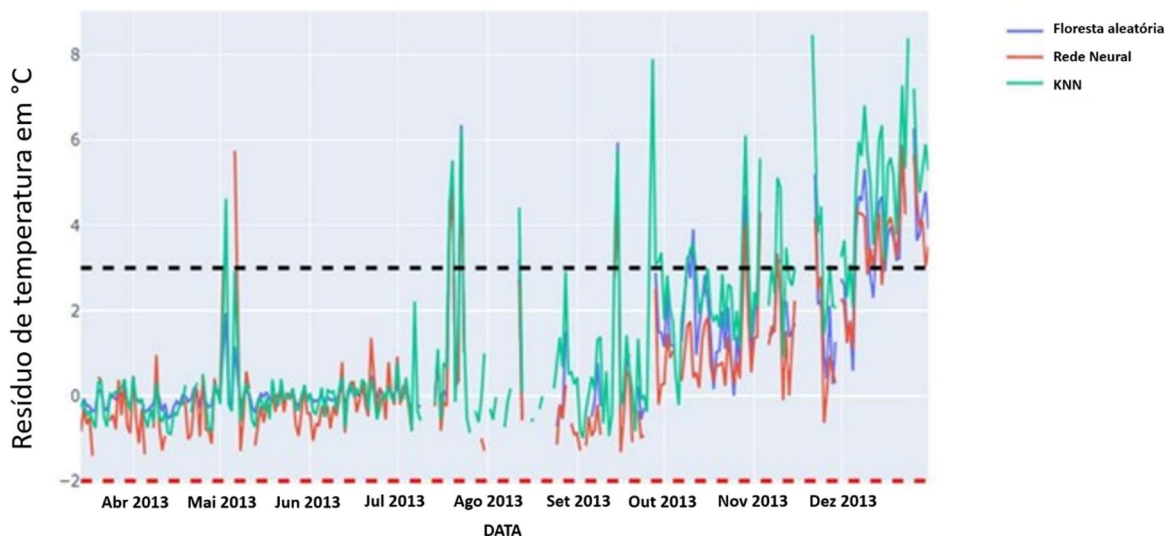
Já não é tão evidente quanto nos **Quadros 18 e 19**. No **Quadro 18**, há uma visualização clara do crescimento do resíduo à medida que se aproxima da falha, mas há uma forte presença de falsos positivos no início da série.

Quadro 18 - Resíduo de temperatura para o aerogerador 2.



Fonte: Autor, 2024.

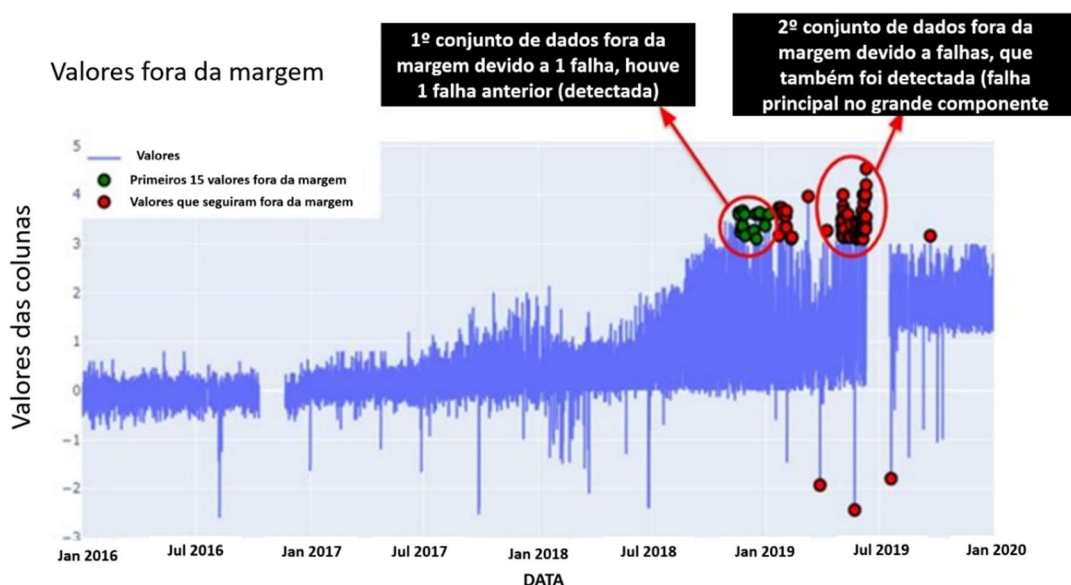
Quadro 19 - Resíduo de temperatura para o aerogerador 3.



Fonte: Autor, 2024.

Além de observar o desempenho do modelo por meio do gráfico residual, o modelo também retorna à sequência de pontos de dados que estavam fora dos limites, indicando a data e a hora em que os valores permaneceram fora dos limites. Essas informações são úteis, pois estimam uma curva de crescimento com base na data em que os resíduos começaram a exceder os limites superiores ou inferiores determinados. O **Quadro 20** e a **Tabela 8** destacam as sequências fora dos limites, já a **Tabela 9** apresenta as datas que as falhas foram detectadas e o momento em que as mesmas ocorreram.

Quadro 20 - 1º aerogerador exemplificando a visualização de sequências fora dos limites.



Fonte: Autor, 2024.

Tabela 8: Sequência de dados fora dos limites (data, hora e valor do resíduo de temperatura)

DATA	HORÁRIO	VALOR DO RESÍDUO DE TEMP.
09/06/2019	09:00:00	3,067
09/06/2019	10:00:00	3,333
09/06/2019	11:00:00	3,1
09/06/2019	12:00:00	3,033
09/06/2019	13:00:00	3,4
09/06/2019	14:00:00	4
09/06/2019	15:00:00	3,967
09/06/2019	16:00:00	4
09/06/2019	17:00:00	3,3

Fonte: Autor, 2024.

Tabela 9: Datas de detecção e ocorrência da falha

AEROGERADOR	GRANDE COMPONENTE	DATA DA DETECÇÃO	DATA DA FALHA
1° Aerogerador	Gearbox	Março/2019	Junho/2019
2° Aerogerador	Mainbearing	Abri/2018	Agosto/2018
3° Aerogerador	Gerador	Novembro/2013	Janeiro/2014

Fonte: Autor, 2024.

5 CONCLUSÕES

Em conclusão, o estudo apresentou uma abordagem abrangente para a detecção de falhas em aerogeradores usando análise de temperatura e modelos estatísticos. Ao monitorar a temperatura dos principais componentes, é possível detectar desvios das condições normais de operação, possibilitando a manutenção proativa e preditiva, minimizando o tempo de inatividade. A integração de técnicas de aprendizado de máquina, como redes neurais, florestas aleatórias e K-Nearest Neighbors, aumenta a precisão da detecção de falhas. A correlação com o estado atual da geração de energia eólica destaca a importância de técnicas eficazes de detecção de falhas para garantir a confiabilidade e a eficiência dos aerogeradores. O uso da linguagem de programação Python, juntamente com bibliotecas relevantes para análise e visualização de dados, facilita a implementação da metodologia proposta. Pesquisas futuras podem se concentrar no refinamento dos modelos estatísticos e na exploração de fontes de dados adicionais para sistemas abrangentes de detecção de falhas em aerogeradores. O desempenho dos modelos e o comportamento dos resíduos indicam padrões anormais durante as falhas, embora permaneçam lineares e dentro dos limites predeterminados durante os períodos saudáveis dos aerogeradores selecionadas. Os gráficos residuais demonstram que, para cada componente principal, há um intervalo de tempo em que há uma variação nas variáveis de temperatura. Os modelos computacionais (rede neural e floresta aleatória) são capazes de prever falhas com um prazo de 80 a 100 dias para o Gearbox, 90 a 120 dias para o Mainbearing e 60 a 80 dias para o gerador.

Independentemente do componente principal específico estudado, o modelo de aprendizado de máquina mostra variações nos resultados dos aerogeradores selecionadas e das variáveis de entrada e de destino. No entanto, o modelo foi projetado para detectar anomalias nos componentes principais (Gearbox, Mainbearing e gerador) usando três técnicas de aprendizado de máquina e as variáveis de entrada mais relevantes para o componente específico, na qual foi possível perceber nesse estudo os padrões anormais durante períodos de falhas e comportamento linear dentro dos limites predeterminados durante períodos saudáveis, onde o mesmo deve ser implementado em plataformas que o operador ou os especialistas de manutenção tenham acesso para realizar as análises e possam coletar informações sobre o comportamento do aerogerador que está sendo

observado.

Para trabalhos futuros, tem-se como objetivo o refino do modelo, para que um dia ele também possa realizar a classificação das falhas, não apenas a detecção dela, utilizando posteriormente o modelo de classificação. Também pode se ampliar a gama de aplicação do modelo, podendo não só ser restrito a aerogeradores, como também outros componentes bastante utilizados na engenharia que tenham comportamentos de temperatura similares.

REFERÊNCIAS

Batista, B. C. F. **Soluções de Equações Diferenciais Utilizando Redes Neurais Multicamadas com os Métodos Steepest Descent e Levenberg-Marquardt** (Dissertação de Mestrado). UFPA. 2012.

Bhavsar, D. **Dissipando mitos: Aprendizado profundo vs. aprendizado de máquina**. Blog da Merkle, 2020. Disponível em: <https://www.merkle.com/blog/dispelling-myths-deep-learning-vs-machine-learning>. Acesso em: 12 abr. 2024.

Burton, T., Sharpe, D., Jenkins, N., & Bossanyi, E. **Wind Energy Handbook**. John Wiley & Sons, 2011, p.412.

Donges, N. **Random Forest: Um guia completo para aprendizado de máquina**. BuiltIn, 2024. Disponível em: <https://builtin.com/data-science/random-forest-algorithm>. Acesso em: 13 jun. 2024.

Encalada-Dávila, Á.; Puruncajas, B.; Tutivén, C.; Vidal, Y. **Wind turbine main bearing fault prognosis based solely on scada data**. Sensors 2021.

ENERGÊS, **7 Componentes do Aerogerador que você precisa conhecer**, 2024. Disponível em: <https://energes.com.br/componentes-aerogerador/>. Acesso em: 17/09/2024.

EVIDENT. **Inspections of Wind Turbine Gearboxes**. 2024. Disponível em: <https://www.olympus-ims.com/en/applications/rvi-wind-turbine/>. Acesso em: 17/09/2024.

González, M. O. A., Gonçalves, J. S., & Vasconcelos, R. M. **Desenvolvimento sustentável: Estudo de caso na implementação de energia renovável no Brasil**. ELSEVIER. 2017.

GUERRA, Jonathas. **Linguagens de programação: Líderes de mercado em 2021**. 2021. Disponível em: https://www.tecnoveste.com.br/linguagens-de-programacao-lideres-de-mercado-em-2021/#google_vignette. Acesso em: 26 mar. 2024.

Han, B., Xie, H., Shan, Y., Liu, R., & Cao, S. **Método de ajuste de curva característica da velocidade do vento e saída do aerogerador com base na limpeza de dados anormais**. ATAMI. 2021

Hau, E. **Wind Turbines: Fundamentals, Technologies, Application, Economics**. Springer Science & Business Media, 2013. p.251.

HEAVEN, W. **What is AI?** 2024. Disponível em:

<https://www.technologyreview.com/2024/07/10/1094475/what-is-artificial-intelligence-ai-definitive-guide/>. Acesso em: 20 set. 2024.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ - IPECE.

Entendendo o índice de Gini. 2015. Disponível em:

https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2015/02/Entendendo_Indice_GINI.pdf Acesso em: 20 ago. 2024.

Loca, A. L. S. **Uma metodologia para avaliação experimental de abordagens de aprendizado de máquina para diagnóstico de falhas com base em sinais de vibração**. Defesa de mestrado, Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brasil, 2020.

Magalhães, N. A. F. (2011). **Sistema de Apoio à Análise de Desempenho de Parques Eólicos** (Dissertação de mestrado). Faculdade de Engenharia da Universidade do Porto, 2011.

Manwell, J. F., McGowan, J. G., & Rogers, A. L. **Wind Energy Explained: Theory, Design and Application**. John Wiley & Sons, 2010, p.303.

Msigwa, O. J. **Ciência de dados e aprendizado de máquina (Parte 09): O algoritmo K-Nearest Neighbors (KNN)**. MQL5. 2022. Disponível em:

<https://www.mql5.com/en/papers/11678>. Acesso em: 02 fev. 2024.

Murgia, A; Verbeke, R; Tsiorkova, E; Terzi, L.; Astolfi, D. **Discussão sobre a adequação do monitoramento de condições baseado em SCADA para diagnóstico de falhas em aerogeradores por meio da análise de dados de temperatura**. Energies, 2023.

NANDA, Anita. **Multiple comparison test by Tukey's honestly significant difference (HSD): Do the confident level control type I error**. 2021. Disponível em:

<https://doi.org/10.22271/math.2021.v6.i1a.636>. Acesso em: 07 jul. 2022.

PADOAN, Mariana. **Os melhores cursos de Python em 2022**. Geekhunter,

2022. Disponível em: <https://blog.geekhunter.com.br/cursos-python/>. Acesso em: 26 mar. 2024.

Reimann, C., Filzmoser, P., & Garret, R. G. **Background e threshold:** comparação crítica de métodos de determinação. ELSEVIER, 2005.

Russel, Norvig. **Artificial Intelligence:** A modern approach. 2010, p.25.

Silva, N. F. da, Rosa, L. P., Freitas, M. A. V., & Pereira, M. G. **Energia eólica no Brasil:** Do modelo de crise de expansão do setor elétrico ao ambiente favorável. ELSEVIER, 2013.

Sommer. Leroy, **Wind turbines generator**, 2024. Disponível em:

<https://acim.nidec.com/generators/leroy-somer/products/power-alternators/alternators-for-windturbines>. Acesso em 17/09/2024.

Terra, N. Energia eólica no Brasil bate recordes e gera empregos, 2024.

Disponível em: <https://www.airswift.com/blog/wind-energy-brazil>. Acesso em: 02 jun. 2024.

Turnbull, A.; Carrol, J.; McDonal, A. **Uma análise comparativa da variabilidade dos limites de temperatura por meio do tempo para geradores de aerogeradores usando modelagem de comportamento normal.** Energies 2021, 14, 3211.

Vinicius. **ArtificialNeuralNetworks.** Monolito Nimbus, 2017. Disponível em: <https://www.monolitonimbus.com.br/redes-neurais-artificiais/>. Acesso em: 17 abr. 2023.

Wang, C., Viswanathan, K., Choudur, L., Talwar, V., Satterfield, W., & Schwan, K. **Statistical techniques for online anomaly detection in data centers** (Técnicas estatísticas para detecção de anomalias on-line em data centers). IEEE, 2011.

Zaher, A.; McArthur, S.; Infield, D.; Patel, Y. **Online wind turbine fault detection through automated SCADA data analysis.** Energia eólica: Int. J. Prog. Appl. Wind Power Convers. Technol. 2009, 12, 574-593.