



Gabriel Freitas Santos

**Detecção de Anomalias em Rolamentos de
Turbinas Eólicas utilizando Dados de CMMS e
Aprendizado de Máquina**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Mecânica Aplicada, do Departamento de Engenharia Mecânica da PUC-Rio.

Orientador : Prof. Dr. Florian Alain Yannick Pradelle
Co-orientador: Prof. Dr. Helon Vicente Hultmann Ayala
Co-orientador: Dra. Paula Aida Sesini

Rio de Janeiro
Outubro de 2024



Gabriel Freitas Santos

**Deteccção de Anomalias em Rolamentos de
Turbinas Eólicas utilizando Dados de CMMS e
Aprendizado de Máquina**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Mecânica Aplicada da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Prof. Dr. Florian Alain Yannick Pradelle

Orientador

Departamento de Engenharia Mecânica – PUC-Rio

Prof. Dr. Helon Vicente Hultmann Ayala

Coorientador

PUC PR

Dra. Paula Aida Sesini

Coorientador

PUC Rio

Prof. Dr. Igor Braga de Paula

PUC Rio

Dr. Sinvaldo Rodrigues Moreno

EDF Renewables do Brasil

Rio de Janeiro, 30 de Outubro de 2024

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Gabriel Freitas Santos

Possui título de Engenheiro Eletricista concedido pelo Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ), com 8 anos de experiência no mercado de energia renovável brasileiro, atuando majoritariamente no desenvolvimento de monitoramentos e predição de falhas em turbinas eólicas e sistemas fotovoltaicos.

Ficha Catalográfica

Santos, Gabriel Freitas

Detecção de Anomalias em Rolamentos de Turbinas Eólicas utilizando Dados de CMMS e Aprendizado de Máquina / Gabriel Freitas Santos; orientador: Florian Alain Yannick Pradelle; co-orientadores: Helon Vicente Hultmann Ayala, Paula Aida Sesini. – 2024.

66 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2024.

Inclui bibliografia

1. Engenharia Mecânica – Teses. 2. Análise de Componentes Principais. 3. Sistema de Gerenciamento e Monitoramento de Condição. 4. Turbina Eólica. 5. Rolamento. 6. Aprendizado de Máquina. I. Pradelle, Florian Alain Yannick . II. Ayala, Helon Vicente Hultmann. III. Sesini, Paula Aida . IV. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Mecânica. V. Título.

CDD: 621

*Dedico este trabalho à memória
do meu querido avô, Paulo Freitas.(In Memoriam)*

Agradecimentos

Primeiramente, quero agradecer a Deus, pois sem Ele nada seria possível. Ele é a minha força e inspiração, e eu realmente acredito que tudo o que faço é para Ele e por Ele. Que minhas conquistas contribuam para o Reino dEle e apontem para a Sua glória.

Um agradecimento especial à minha esposa, Mariana Freitas. Ao longo de todos esses anos, ela tem sido meu maior apoio e suporte. Seus sacrifícios, compreensão e paciência fizeram toda a diferença até aqui. Sem o seu incentivo e conselhos, eu não teria conseguido concluir este trabalho.

Agradeço também aos meus pais e ao meu irmão, que sempre vibram com cada uma das minhas conquistas. O apoio de vocês é fundamental para eu seguir em frente, mesmo nos momentos mais desafiadores.

Um agradecimento sincero aos meus orientadores, Prof. Dr. Helon Ayala e Prof. Dr. Florian Pradelle. Nos momentos em que pensei em desistir do mestrado, vocês me incentivaram e me orientaram, mostrando sempre o melhor caminho. Agradeço por entenderem as dificuldades que enfrentei e por acreditarem em mim.

Por último, quero agradecer à PUC Rio pela oportunidade de estudar em uma instituição tão respeitada. E um agradecimento à CAPES pela bolsa que proporcionou o acesso a este programa de mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Resumo

Santos, Gabriel Freitas; Pradelle, Florian Alain Yannick ; Ayala, Helon Vicente Hultmann; Sesini, Paula Aida . **Detecção de Anomalias em Rolamentos de Turbinas Eólicas utilizando Dados de CMMS e Aprendizado de Máquina**. Rio de Janeiro, 2024. 66p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

A energia eólica tem se destacado no Brasil como uma fonte importante para diversificar a matriz energética, gerando energia limpa e sustentável. Com os altos investimentos nessa área, cresce a necessidade de antecipar falhas em turbinas eólicas. O aumento exponencial no número de turbinas instaladas, juntamente com o envelhecimento da frota de geração, intensificou a competição para reduzir os custos de operação e manutenção, o que significa minimizar paradas não planejadas e minimizar grandes e caras manutenções corretivas. O objetivo deste estudo é utilizar os dados de vibração disponíveis nos Sistemas de Monitoramento e Gestão de Condição para identificar turbinas com desvios significativos de condição que apresentem alto risco de falha. A abordagem de processamento de dados usando algoritmos como Vizinho Mais Próximo Condensado e a Análise de Componentes Principais na etapa de pré-processamento, juntamente com Máquinas de Vetor de Suporte para classificação do estado de saúde, demonstrou excelente precisão, acima de 90% de acurácia tanto para testes de uma única turbina quanto para testes de múltiplas turbinas, tornando-a adequada para a gestão de parques eólicos com um grande número de turbinas. Os experimentos realizados em combinação de 5 turbinas, permitiu identificar os cenários de melhor performance, mantendo resultados acima dos 90% de acurácia do modelo proposto de acordo com o objetivo de identificação antecipada de falhas em uma frota considerando dados reduzidos para treino do modelo aplicado. Nestes cenários, destaquem-se situações que não houve performance adequada também, impactando tanto a acurácia quanto a razão de falsos positivos avaliados.

Palavras-chave

Análise de Componentes Principais; Sistema de Gerenciamento e Monitoramento de Condição; Turbina Eólica; Rolamento; Aprendizado de Máquina.

Abstract

Santos, Gabriel Freitas; Pradelle, Florian Alain Yannick (Advisor); Ayala, Helon Vicente Hultmann (Co-Advisor); Sesini, Paula Aida (Co-Advisor). **Anomaly Detection in Wind Turbine Bearings Using CMMS Data and Machine Learning**. Rio de Janeiro, 2024. 66p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Wind energy has become a key source for diversifying Brazil's energy matrix, significantly contributing to the generation of clean and sustainable energy. Due to its importance and the large investments being made in this sector, there is an increasing need to anticipate failures in wind turbines. The exponential increase in the number of installed turbines, along with the aging of the generation fleet, has intensified the competition to reduce operation and maintenance costs, which means minimizing unplanned downtime and reducing large, costly corrective maintenance. The objective of this study is to use vibration data available from Condition Monitoring and Management Systems (CMMS) to identify turbines with significant condition deviations that present a high risk of failure. The data processing approach, using algorithms such as Condensed Nearest Neighbor (CNN) and Principal Component Analysis (PCA) in the pre-processing stage, along with Support Vector Machines (SVM) for health state classification, has demonstrated excellent accuracy, above 90%, both in single-turbine tests and in multiple-turbine tests, making it suitable for managing wind farms with a large number of turbines. The experiments conducted with a combination of five different turbines allowed the identification of the best performance scenarios, maintaining results with over 90% accuracy in the proposed model according to the goal of early fault detection in a fleet, even when using reduced training data for the applied model. It is also important to highlight scenarios where performance was not adequate, impacting both accuracy and the rate of evaluated false positives.

Keywords

Principal Component Analysis; Condition Management Monitoring System; Wind Turbine; Bearing; Machine Learning.

Sumário

1	Introdução	15
1.1	Contextualização do Problema	15
1.2	Motivação e Relevância da Pesquisa	18
1.3	Objetivos do Estudo	19
1.4	Estrutura da Dissertação	19
2	Fundamentação da Pesquisa	21
2.1	Turbinas Eólicas	21
2.2	Importância da detecção de anomalias em turbinas eólicas	23
2.3	Estado da arte em detecção de anomalias em turbinas eólicas e manutenção preditiva	25
2.4	Uso de dados do sistema de monitoramento de condições (CMS) em turbinas eólicas	26
2.5	Técnicas de aprendizado de máquina aplicadas à detecção de anomalias	27
3	Metodologia	30
3.1	Coleta e preparação dos dados do CMS (Condition Monitoring System)	30
3.2	Premissas dos experimentos realizados	32
3.3	Estratégias de treinamento, validação e teste dos modelos	35
3.4	Métricas para Avaliação	36
3.5	Metodologia Experimento Inicial	38
3.6	Metodologia Experimento Completo	41
3.7	Descrição estratégia de processamento de dados	41
4	Discussão e Resultados	44
4.1	Resultados Modelagem Inicial	44
4.2	Resultados Modelagem com Validação Cruzada	44
4.3	Apresentação, Comparação de Métricas e Considerações sobre o Experimento Inicial	46
4.4	Expansão do Experimento para 5 turbinas	48
4.5	Hipótese de 2 Turbinas no Conjunto de treinamento	52
4.6	Discussão dos resultados obtidos	53
5	Considerações Finais	58
5.1	Recapitulação dos principais resultados e contribuições do estudo	58
5.2	Limitações e sugestões para trabalhos futuros	60
6	Referências bibliográficas	61
A	Exemplo Relatório Fabricante Aero gerador	64

Lista de figuras

Figura 1.1	Evolução Custos de Energia no Brasil até 2020 - Fonte: (CCEE, 2020)	16
Figura 2.1	Primeiro Modelo de Turbina Eólica conhecido e datado de 1887. (GIPE; MöLLERSTRöM, 2022)	22
Figura 2.2	Diagrama ilustrativo Turbina Eólica Eixo Horizontal " <i>Geared</i> " - (BADUREK, 2024)	23
Figura 3.1	Diagrama de Instalação Sensores de Vibração (Retirado Do- cumentação do Fabricante)	32
Figura 3.2	Relatório fabricante com medidas globais brutas	33
Figura 3.3	Relatório fabricante com interpretação de especialista indi- cando diagnóstico de anomalia	34
Figura 3.4	Registro Fotográfico de um Rolamento Danificado do Lado Acoplado do Gerador.	34
Figura 3.5	Detalhes do desgaste na pista interna do rolamento danificado.	35
Figura 3.6	Trend de variáveis de harmônica da velocidade de rotação determinantes para definição das categorias	36
Figura 3.7	Exemplo Utilização de Matriz de Confusão para interpretação de resultados de métodos de classificação.(MINH, 2024)	37
Figura 3.8	Representação esquemática do experimento inicial proposto	39
Figura 3.9	Pairplot Análise de Principais Componentes escolhidas dife- renciadas pelas categorias - WTG 22	40
Figura 3.10	Pairplot Análise de Principais Componentes escolhidas dife- renciadas pelas categorias - WTG 48	40
Figura 3.11	Representação esquemática do experimento completo	42
Figura 4.1	Matriz de Confusão Modelo Puro - WTG22	45
Figura 4.2	Matriz de Confusão Modelo PCA - WTG22	45
Figura 4.3	Matriz de Confusão Modelo Puro - WTG22 e WTG48	46
Figura 4.4	Matriz de Confusão Modelo PCA - WTG22 e WTG48	47
Figura 4.5	Gráfico de Barras Comparação Acuracidade entre Cenários Preliminares	47
Figura 4.6	Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo Puro	49
Figura 4.7	Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo PCA	50
Figura 4.8	Gráfico retirado do relatório de falha de sensor entregue pelo fabricante da turbina	50
Figura 4.9	Resultados de FPR para modelo puro considerando amostrais com categoria Normal	52
Figura 4.10	Resultados de FPR para modelo PCA considerando amostrais com categoria Normal	53
Figura 4.11	Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo Puro	54

Figura 4.12	Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo PCA	55
Figura 4.13	Resultados de FPR para modelo puro considerando amostrais com categoria Normal	56
Figura 4.14	Resultados de FPR para modelo PCA considerando amostrais com categoria Normal	57

Lista de tabelas

Tabela 3.1	Descrição de Variáveis disponíveis para o desenvolvimento da modelagem de dados	31
Tabela 3.2	Descrição de Turbinas e Anomalias Escolhidas	33

Lista de algoritmos

Algoritmo 1	Fluxo Completo de Dados para Classificação SVM	43
-------------	--	----

Lista de Abreviaturas

CCEE – Câmara de Comercialização de Energia Elétrica

CMMS – *Condition Monitoring and Management Systems* (Sistemas de Monitoramento e Gestão de Condição)

CNN – *Condensed Nearest Neighbor* (Vizinho mais Próximo Condensado)

FFT – *Fast Fourier Transform* (Transformada Rápida de Fourier)

FN – *False Negative* (Falso Negativo)

FP – *False Positive* (Falso Positivo)

FPR – *False Positive Ratio* (Razão de Falsos Positivos)

HSS – *High Speed Shaft* (Eixo de Velocidade Rápida)

IA – Inteligência Artificial

LCOE – *Levelized Cost of Energy* (Custo Nivelado de Energia)

NN – *Nearest Neighbour* (Vizinho Mais Próximo)

OM – Operação e Manutenção

OPEX – *Operational Expenditure* (Despesas Operacionais)

PCA – *Principal Component Analysis* (Análise de Componentes Principais)

RF – *Random Forest* (Árvore de Decisões Aleatórias)

SCADA – *Supervisory Control and Data Acquisition*

SVM – *Support Vector Machines* (Máquinas de Vetores de Suporte)

TN – *True Negative* (Verdadeiro Negativo)

TP – *True Positive* (Verdadeiro Positivo)

WTG – *Wind Turbine Generator* (Turbina Eólica)

*Porque dele e por ele, e para ele são todas as
coisas; glória, pois, a ele eternamente.
Amém!*

Romanos 11:36, *Bíblia Sagrada*.

1

Introdução

O capítulo de introdução desta dissertação tem como objetivo contextualizar o tema abordado, destacando a relevância da pesquisa e delineando com clareza a motivação e os objetivos do estudo. Além disso, busca compartilhar conceitos essenciais para o entendimento aprofundado dos resultados apresentados nos capítulos subsequentes.

1.1

Contextualização do Problema

A energia eólica tem se consolidado como uma das principais fontes de energia renovável no cenário mundial, impulsionada pelo avanço tecnológico e pela necessidade de descarbonizar as matrizes energéticas. Globalmente, países como a China, os Estados Unidos e a Alemanha têm investido significativamente em parques eólicos, contribuindo para a expansão da capacidade instalada e a redução de emissões de carbono.(HASSAN, 2024) No Brasil, a energia eólica desempenha um papel crescente, com o país se destacando como um dos principais geradores da fonte na América Latina.(ABEEÓLICA, 2023) O potencial eólico brasileiro, especialmente na região Nordeste, tem sido amplamente explorado e se tornou uma das principais fontes de energia elétrica do país, ajudando a diversificar a matriz energética nacional.(GÓES, 2021)

No início da segunda década do século XXI, os empreendedores de energia eólica se beneficiavam de um dos modelos de negócio mais competitivos entre as fontes renováveis, conforme apresentado na Figura 1.1 do relatório publicado pela Câmara de Comercialização de Energia Elétrica (CCEE) em 2020 (CCEE, 2020). No entanto, essa vantagem competitiva tem enfrentado forte pressão econômica devido à redução nos custos de venda, impulsionada por novos investimentos e políticas de incentivo voltadas para a energia solar. Essa competição tem levado os complexos eólicos a buscar a redução de custos como forma de aumentar a competitividade na comercialização de energia.

Com o aumento do número de turbinas em operação, ocorre também um crescimento nas falhas de componentes devido ao desgaste natural, exigindo uma atuação mais eficiente na gestão de operação e manutenção dos parques eólicos.(PENG, 2023) Esse cenário impacta desde a disponibilidade de técnicos em campo, passando pela necessidade de níveis adequados de estoque de sobressalentes, até investimentos em equipamentos de ensaio e equipes de engenharia especializadas para lidar com falhas e defeitos mais complexos.

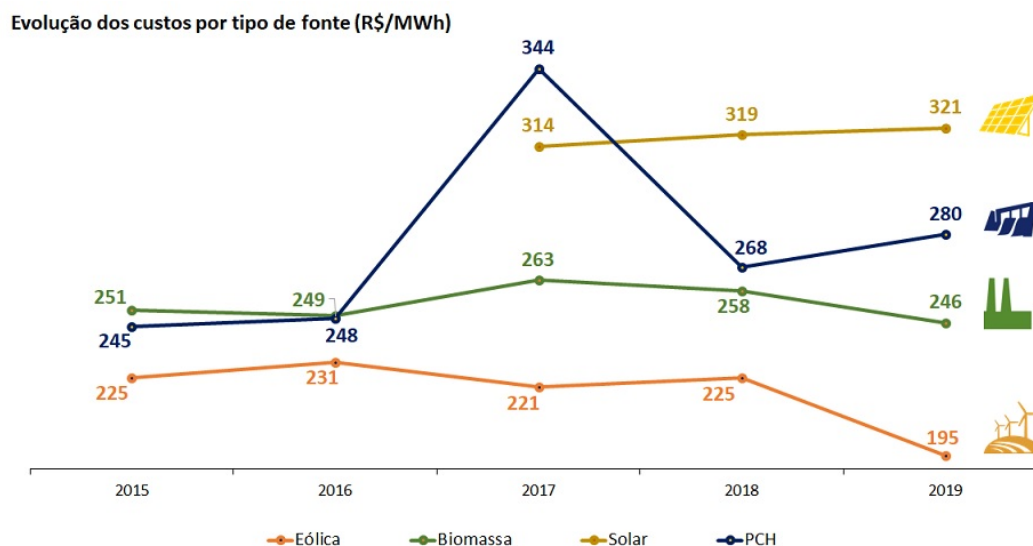


Figura 1.1: Evolução Custos de Energia no Brasil até 2020 - Fonte: (CCEE, 2020)

Essas demandas geram um impacto significativo nos custos de Operação e Manutenção (O&M) por MWh produzido, especialmente em projetos já pressionados pela alta competitividade do setor.(LEVITT, 2009)

Essa evolução levou a um aumento significativo no investimento em sistemas de monitoramento de condição para turbinas eólicas, incorporando diversos sensores, com foco particular em vibração, distribuídos pelos principais componentes do ativo.(KONG, 2023) A disponibilidade de um maior volume de dados possibilita maiores aplicações de métodos de identificação de falhas e permitiu novas oportunidades na gestão da operação de um portfólio de parques eólicos com múltiplas turbinas. Esses sistemas avançados de monitoramento não apenas melhoram a detecção precoce de problemas, mas também permitem uma manutenção mais proativa e eficiente, reduzindo o tempo de inatividade e os custos associados.(HYERS, 2006)

Embora a configuração de alarmes para valores fora da faixa normal de operação tenha se mostrado eficaz, há um potencial significativo para antecipar a degradação do equipamento bem antes de atingir os limites de alerta por meio do uso de modelos de aprendizado de máquina. A implementação dessas tecnologias pode não apenas reduzir os custos de manutenção, mas também aumentar a confiabilidade e a eficiência operacional das turbinas eólicas além da extensão da vida útil projetada para os componentes, contribuindo para a sustentabilidade econômica e ambiental dos projetos de energia eólica. A análise preditiva baseada em dados pode fornecer insights valiosos para a otimização do desempenho das turbinas, permitindo ajustes operacionais que

maximizem a produção de energia.(UDO, 2024)

Além dos benefícios diretos, como a redução de falhas e o aumento da eficiência operacional, a antecipação de problemas também pode trazer vantagens indiretas significativas. A diminuição da necessidade de intervenções emergenciais permite uma melhor gestão do portfólio de falhas, reduzindo a necessidade de manter grandes estoques de peças sobressalentes e diminuindo a quantidade de técnicos necessários em campo. Isso não só reduz os custos operacionais, mas também melhora a logística e a gestão de recursos, contribuindo para uma operação mais enxuta e eficiente.(VELMURUGAN; DHINGRA, 2021) Diante desse cenário, é essencial que os gestores de parques eólicos adotem uma abordagem integrada e baseada em dados para a operação e manutenção de suas instalações. Isso inclui a capacitação contínua das equipes técnicas, a atualização dos sistemas de monitoramento e a implementação de práticas de manutenção preditiva.

O cenário de consolidação da energia eólica como fonte de energia ao longo da última década tem pressionado fabricantes, operadores e a academia a avançarem nos processos de diagnóstico preditivo de falhas, visando reduzir os custos associados à manutenção corretiva e ao tempo de indisponibilidade não planejado. Como citado, a redução da manutenção corretiva impacta positivamente as despesas operacionais (OPEX) para parques eólicos operacionais, e a projeção desta economia no investimento diminuem o Custo Nivelado de Energia (LCOE) para novos projetos (DAO, 2019; ENEVOLDSEN; XYDIS, 2019).

As maiores oportunidades para a atuação preventiva em falhas de turbinas eólicas estão relacionadas aos componentes principais, como o eixo principal, a caixa multiplicadora (*Gearbox*) e o gerador.(SANTELO, 2022) Esses componentes são considerados críticos tanto por razões econômicas quanto técnicas. Por exemplo, são equipamentos geralmente importados, com alto custo de reposição devido às dificuldades de transporte e aos impostos associados. Além disso, o reparo ou substituição desses componentes apresenta desafios técnicos significativos, pois normalmente envolvem o uso de guindastes e a movimentação de peças grandes e pesadas em grandes alturas. Essas operações têm um impacto significativo na operação, não apenas pelos custos envolvidos, mas também pelo risco à segurança patrimonial e, principalmente, à segurança dos trabalhadores, que é extremamente complexa.

1.2

Motivação e Relevância da Pesquisa

A principal motivação desta pesquisa é impulsionar o desenvolvimento de tecnologias e ferramentas que auxiliem na tomada de decisão no processo de gestão de ativos eólicos, especialmente em um cenário de análise de portfólio. Gestores de ativos eólicos frequentemente lidam com diversos complexos contendo múltiplas turbinas, exigindo decisões constantes sobre as ações a serem tomadas, dado que os recursos materiais e humanos são limitados. Isso é particularmente importante quando se busca atuar preventivamente para evitar falhas. Adicionalmente, a pesquisa também visa demonstrar a utilização de uma base de dados ainda pouco explorada na academia, em um contexto mais amplo e relacionado a ferramentas de análise de dados mais robustas e atuais. Dados de sistemas de monitoramento de condição são comumente utilizados para análises de alarmes em valores absolutos e diagnósticos de falhas já confirmadas.

A comunidade acadêmica tem se concentrado cada vez mais em antecipar desvios na operação das turbinas. Inicialmente, o trabalho de Chang, Ding, Wang, Chai, Bhuiyan, Haw e Tan (2023) fez uma revisão dos principais avanços recentes na utilização de dados de sistemas de monitoramento de condição em turbinas eólicas, como proposto nessa pesquisa. No trabalho, os autores destacaram diferentes tipos de sistemas de monitoramento de condição existentes como os de sinais, focados em vibração e acústica por exemplo, sistemas de modelagens física que busca simular o comportamento dinâmico dos equipamentos e sistemas de dados que utilizam principalmente dados de SCADA no acompanhamento de anomalias.

Por exemplo, em Wen (2020), dados SCADA são utilizados para detecção de anomalias. Em Zhang, Zeng, Shen e Gao (2022), métodos de Máquina de Vetores de Suporte são empregados para estabelecer uma pontuação de anomalia e, subsequentemente, classificar a turbina eólica com base em um limiar predefinido. Outra contribuição significativa é apresentada em Peng, Kimmig, Niu, Wang, Liu, Wang e Ovtcharova (2022), demonstrando uma estratégia para identificação de anomalias como falhas em multiplicadora, gerador, rolamentos etc. através da avaliação de janelas de dados SCADA associadas à análise de discrepâncias e alarmes registrados. No entanto, apesar das referências apresentadas, há uma falta de contribuições científicas substanciais voltadas para a definição de um pipeline de dados sob a perspectiva da gestão de ativos. Grande parte das contribuições existentes estão focadas no desenvolvimento de ferramentas computacionais utilizadas ou no estudo do método de monitoramento do componente. As contribuições existentes buscam evoluir

e propor diagnósticos específicos de vibração por exemplo. O pipeline proposto não busca fornecer um diagnóstico exato de falhas, mas focar na identificação de desvios e falhas com o objetivo de otimizar custos e melhorar o desempenho operacional, ou seja, a pesquisa não busca diagnosticar uma falha, mas sim sinalizar um desvio operacional que identifique uma falha.

1.3

Objetivos do Estudo

Este estudo busca avançar na área de monitoramento e diagnóstico de falhas em turbinas eólicas, utilizando dados dos Sistemas de Monitoramento e Gestão de Condição (CMMS) fornecidos pelos fabricantes. O foco da pesquisa está em através da análise de dados como vibração e operacionais em identificar anomalias nos rolamentos do gerador da turbina. Serão utilizados dados globais de aceleração e velocidade de sensores instalados nos principais equipamentos, além de dados pré-processados, como amplitudes em bandas de frequência específicas e harmônicos da frequência de rotação.

O objetivo principal é demonstrar a efetividade da utilização deste conjunto de dados em uma abordagem inovadora e eficaz para o acompanhamento e monitoramento das turbinas eólicas. Esta abordagem visa identificar mudanças no comportamento operacional que possam indicar anomalias. A detecção precoce dessas anomalias é crucial, pois permite a realização de estudos aprofundados de diagnóstico por especialistas, prevenindo falhas catastróficas e otimizando a manutenção preventiva.

Além disso, a pesquisa pretende explorar uma base de dados ainda pouco utilizada na academia, aplicando ferramentas de análise de dados mais robustas e atuais. Ao integrar esses dados em um sistema de monitoramento contínuo, espera-se proporcionar uma visão mais abrangente e detalhada do estado de operação das turbinas eólicas. Isso não apenas melhora a eficiência operacional, mas também contribui para a segurança dos trabalhadores e a redução de custos associados a reparos e substituições.

1.4

Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos principais. No Capítulo 1, é apresentada a introdução ao tema, incluindo a motivação, os objetivos e a relevância da pesquisa. O Capítulo 2 aborda a revisão da literatura, onde são discutidos os principais conceitos, teorias e estudos relacionados ao tema além de citar os trabalhos mais recentes relacionados ao tema da pesquisa. No Capítulo 3, são detalhados os métodos e procedimentos utilizados na pesquisa,

incluindo a descrição do desenho experimental, a coleta de dados e as técnicas de análise. O Capítulo 4 apresenta os resultados obtidos nos experimentos iniciais e a discussão desses resultados e suas contribuições. O Capítulo 5 evidencia a evolução dos experimentos propostos e discute os resultados suas vantagens e desvantagens. Finalmente, no Capítulo 6, são apresentadas as conclusões da pesquisa, bem como as implicações práticas, limitações do estudo e sugestões para pesquisas futuras.

2

Fundamentação da Pesquisa

Neste capítulo, será apresentada a fundamentação teórica necessária para o desenvolvimento desta pesquisa, focada na detecção de anomalias em turbinas eólicas. Inicialmente, serão abordados os conceitos básicos e a importância das turbinas eólicas no contexto energético atual. Em seguida, será discutida a relevância da detecção de anomalias para a manutenção eficiente dessas turbinas. O estado da arte em detecção de anomalias e manutenção preditiva será explorado, destacando as principais técnicas e avanços na área. Além disso, será analisado o uso de dados provenientes dos sistemas de monitoramento de condições (CMS) em turbinas eólicas. Por fim, serão apresentadas as técnicas de aprendizado de máquina aplicadas à detecção de anomalias, incluindo Análise de Componentes Principais (PCA), Vizinho mais Próximo Condensado (CNN) e Máquinas de Vetores de Suporte (SVM).

2.1

Turbinas Eólicas

Existem vários tipos de turbinas eólicas, dispositivos projetados para converter a energia cinética do vento em energia elétrica. Embora a transformação da energia cinética do vento seja uma prática milenar, o design moderno das turbinas, que serve de base para esta pesquisa, data do final do século XIX. Desde então, como descrito em duas publicações recentes (GIPE; MÖLLERSTRÖM, 2022; GIPE; MÖLLERSTRÖM, 2023), as turbinas eólicas passaram por uma evolução significativa. Diversas topologias e componentes foram testados ao longo dos anos, começando com turbinas de eixo vertical, como ilustrado na Figura 2.1. Após a década de 1970, as turbinas eólicas modernas, turbinas de eixo horizontal e velocidade variável, tornaram-se populares, especialmente dois principais tipos de topologia de turbinas de eixo horizontal: as turbinas *Direct Drive*, que não possuem multiplicação de velocidade entre o eixo de rotação das pás e o gerador, e as turbinas *Geared*, que utilizam uma multiplicadora conectando o eixo das pás ao eixo do gerador, aumentando significativamente a velocidade de rotação deste último.

A principal diferença entre essas topologias é o tamanho estrutural do gerador. Como a rede elétrica que receberá a energia produzida opera em frequências entre 50 Hz e 60 Hz, há uma necessidade de muitos polos de magnetização quando a velocidade de rotação é baixa, como no caso das turbinas *Direct Drive*. Isso resulta em um gerador maior. Por outro lado, nas

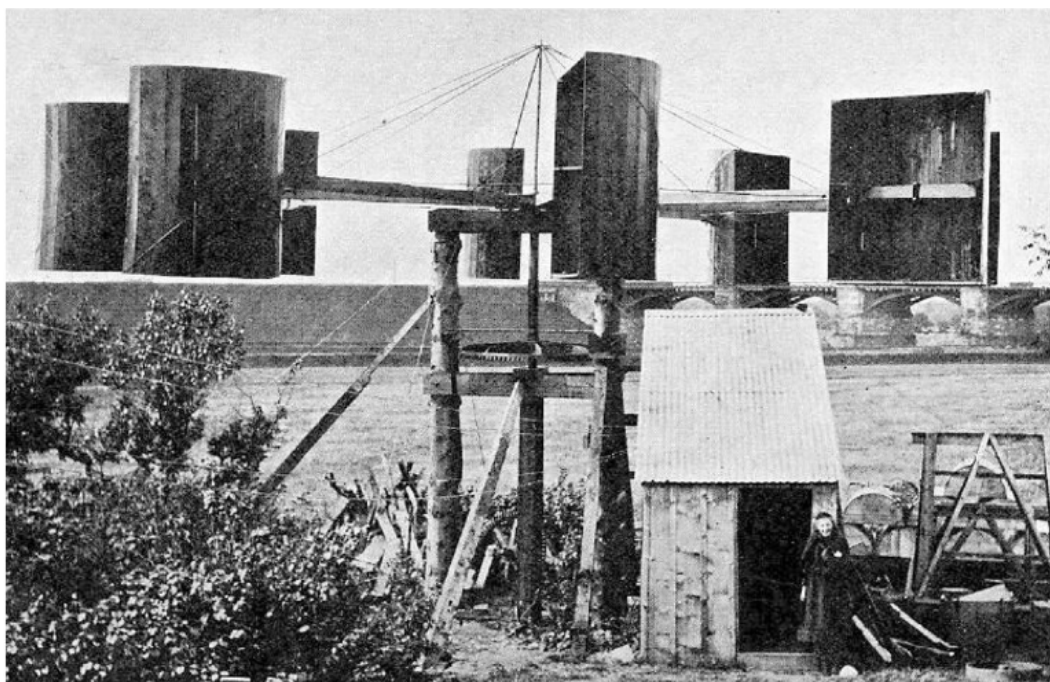
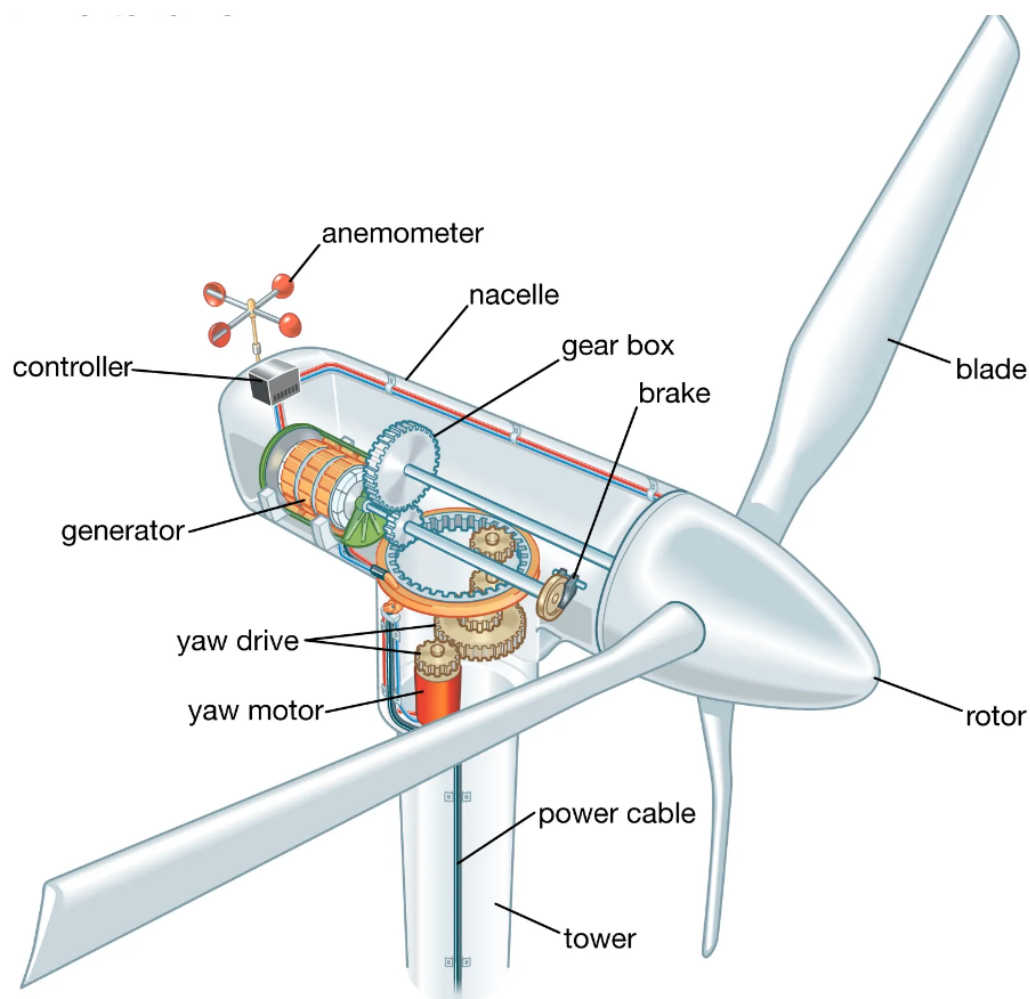


Figura 2.1: Primeiro Modelo de Turbina Eólica conhecido e datado de 1887. (GIPE; MöLLERSTRÖM, 2022)

turbinas *Geared*, a multiplicação da velocidade de rotação reduz a necessidade de polos, permitindo um gerador menor.(NEJAD, 2022)

A Figura 2.2 ilustra uma turbina eólica amplamente utilizada atualmente. É possível observar as pás, responsáveis por capturar a energia do vento e girar mecanicamente o eixo principal, que está acoplado a uma multiplicadora planetária (*Gearbox*). (BADUREK, 2024) Esta multiplicadora aumenta a velocidade de rotação no eixo do gerador, permitindo a conversão da energia mecânica em energia elétrica. Outros dispositivos importantes que influenciam a performance e a produção de energia incluem o sistema de yaw, que direciona a ponta do rotor paralelamente à direção do vento, aumentando a eficiência da conversão, além de sistemas passivos como mancais e rolamentos.(NEJAD, 2022)

Esta pesquisa está relacionada a dados de turbinas eólicas instaladas *onshore*, ou seja, instaladas no território, mas também há possibilidade de teste e validação em turbinas *offshore*, instaladas em mares profundos. A grande maioria da potência instalada no Brasil é *onshore* mas existem estudos e regulação sendo desenvolvidos pela Agência Nacional de Energia Elétrica (ANEEL) para disponibilização de incentivos à instalação desta modalidade.(ABEEÓLICA, 2023)



© Encyclopædia Britannica, Inc.

Figura 2.2: Diagrama ilustrativo Turbina Eólica Eixo Horizontal "Geared" - (BADUREK, 2024)

2.2

Importância da detecção de anomalias em turbinas eólicas

A era moderna das turbinas eólicas começou com o uso de sistemas SCADA para monitoramento histórico, permitindo análises simplificadas. Inicialmente, o foco estava no monitoramento reativo, com alarmes de falha e desligamentos por proteções. Posteriormente, surgiu o monitoramento ativo de variáveis de processo durante a operação. Esse monitoramento ativo utilizava gatilhos baseados em valores medidos fora de um envelope de operação aceitável, definido por engenheiros especialistas. Com o crescimento do conhecimento e da experiência adquirida através da investigação de falhas, juntamente com a evolução tecnológica, novos sistemas foram incorporados. Esses avanços permitiram não apenas um detalhamento e uma amostragem mais precisa dos dados já monitorados, mas também a adição de novos tipos de dados. (PANDIT, 2023) Exemplos incluem a medição de vibrações em diver-

so locais, o uso de câmeras térmicas e até microfones para monitoramento de ruídos audíveis. Essas inovações têm contribuído significativamente para a melhoria da eficiência e da confiabilidade das turbinas eólicas.

Empresas como GE e Siemens têm direcionado significativos investimentos em pesquisa e desenvolvimento na área de manutenção preditiva de aerogeradores. O foco principal dessas iniciativas é a prevenção de falhas através da aplicação de técnicas de Machine Learning, que permitem a análise de grandes volumes de dados operacionais para identificar padrões e prever possíveis falhas antes que ocorram (VESELY, 2017). De acordo com Vesely (2017), os custos de O&M são responsáveis por cerca de 30% do LCOE de um projeto. Além disso, essas empresas estão explorando a integração de sistemas de inteligência artificial (IA) para aprimorar ainda mais a eficiência e a confiabilidade dos aerogeradores, contribuindo para a sustentabilidade e a redução de custos operacionais.

Além da redução dos custos operacionais, o monitoramento avançado das turbinas eólicas tem um impacto significativo na segurança operacional desses ativos. A prevenção de falhas catastróficas ou mais danosas não apenas reduz os custos de reparo dos componentes, mas também evita falhas maiores e de alto risco, como a quebra de pás ou até mesmo da torre. Tais falhas podem resultar na queda da nacelle, causando uma perda substancial, pois quase todos os componentes da turbina são destruídos. Nesses casos, a perda financeira é significativa, mas o impacto na reputação dos fabricantes também é crucial. Muitos fabricantes de turbinas eólicas possuem modelos de negócios que incluem a prestação de serviços de operação e manutenção, além do fornecimento das turbinas e peças sobressalentes. Portanto, a confiabilidade e a segurança operacional são essenciais não apenas para a sustentabilidade financeira, mas também para manter a confiança e a satisfação dos clientes. Por exemplo, na publicação de Igba, Alemzadeh, Henningsen e Durugbo (2015), os autores comparam a razão econômica entre a execução de manutenções preventivas e manutenções corretivas em multiplicadoras de turbinas eólicas. Eles analisaram o impacto do custo de diversas ocorrências no longo prazo e concluem que o custo de uma ocorrência de manutenção preventiva é de aproximadamente 5% do valor de uma corretiva realizada no mesmo equipamento, evitando o impacto da disponibilidade e de custos muito maiores de reparos e substituições do componente principal. O autor propõe definir a viabilidade econômica baseada em uma função definida com o tempo médio entre falhas histórico e a frequência definida de manutenções preventivas realizadas. Em termos de ordem de grandeza para referência, uma troca de multiplicadora, segundo os autores, custa ao proprietário cerca de 150 mil

euros.

2.3

Estado da arte em detecção de anomalias em turbinas eólicas e manutenção preditiva

Diferente de outros temas na engenharia, quando a academia avança na evolução de uma determinada tecnologia e então o mercado e a indústria absorvem e aplicam a ferramenta, no caso da detecção de anomalias em aerogeradores a situação é um pouco distinta. Neste caso, devido ao modelo de negócio de grandes fabricantes, no qual eles dividem o risco da operação e manutenção com os investidores, os próprios fabricantes investem em pesquisa e desenvolvimento na busca pela detecção antecipada de anomalias.(SUMMERVILLE, 2024) Esta detecção, neste caso, permite a recuperação de muitos componentes que terão a vida útil estendida e, conseqüentemente, um retorno muito maior para o projeto.(ABADIE; GOICOECHEA, 2021)

Um aspecto importante na detecção de anomalias é que raramente uma nova tecnologia substitui completamente a anterior. Em vez disso, novos métodos, análises e ferramentas são continuamente agregados. Conseqüentemente, a prática de detecção de anomalias torna-se mais robusta e eficaz a cada ano, proporcionando resultados cada vez melhores para os profissionais da área.(JANKAUSKAS, 2023)

Dentre as ferramentas mais antigas que ainda são utilizadas, podemos destacar algumas como a prática de termografia em painéis elétricos, acompanhamento de sensores de temperatura e de nível de reservatórios de fluidos que indicam estados anormais que necessitam de uma ação do operador quanto à sua condição. Nestes casos, apesar de eficazes, possuem uma desvantagem por em muitos casos somente sinalizar a anomalia em estado mais avançado de falha. Outra prática antiga e muito comum que ainda é utilizada são as inspeções visuais durante manutenções preventivas que permitem ao mantenedor identificar condições de anormalidade.(PANDIT, 2023)

Entretanto, sem substituir as atividades mencionadas, atualmente com a evolução dos métodos matemáticos aplicados e com a disponibilidade de mais equipamentos sensoreados, ocorreu uma enorme evolução nas ferramentas à disposição da engenharia de confiabilidade dos equipamentos para determinar uma anomalia de comportamento antes da falha de fato acontecer. Além dos acelerômetros posicionados para medir as ondas vibratórias da maioria dos equipamentos essenciais envolvidos no processo de geração de energia, é também possível historiar com mais detalhes e por longo prazo dados elétricos da rede, dados ambientais, variáveis de processo, ruído audível dentro do

ambiente de produção por exemplo. A capacidade atual de lidar computacionalmente com uma grande quantidade de amostras, juntamente com ferramentas como Análise de Componentes Principais (PCA) , Árvores de Decisão Aleatórias(RFT) e Redes Neurais(NN), permite a gestão eficiente de um vasto número de características (features). (PANDIT, 2023)

O foco principal dos fabricantes de turbinas tem sido no desenvolvimento de algoritmos que permitam a classificação de anomalias na condição sem preocupação extrema com o diagnóstico, este sendo de responsabilidade da engenharia de manutenção posteriormente. Mas o grande ganho é na gestão da manutenção relacionada à tomada de decisão de atuação, principalmente que o portfólio de operação destas empresas é composto por centenas, se não milhares, de turbinas.(CARTELLI, 2023)

2.4

Uso de dados do sistema de monitoramento de condições (CMS) em turbinas eólicas

O sistema de monitoramento de condição é um termo abrangente que define todo o conjunto de sensores e dados utilizados para acompanhar a operação dos componentes. Este sistema pode incluir variáveis monitoradas como temperaturas, vibrações, pressões hidráulicas, além de variáveis de processo como velocidades, ângulos, potências, correntes e tensões elétricas.(CHANG, 2023)

O processamento desses dados é, na maioria das vezes, realizado localmente, especialmente para medidas de alta frequência, como vibrações. Essas variáveis frequentemente necessitam de um certo nível de processamento e análise para potencializar a informação contida, realizando transformações para o domínio da frequência, por exemplo. Outro ponto importante do processamento local é a otimização do transporte e envio dessas informações por meios digitais de comunicação, uma vez que, já processadas, as informações se tornam mais simples de serem transferidas.(QIN, 2022)

Os sistemas de monitoramento de condição são, em sua maioria, integrados a sistemas SCADA ou sistemas historiadores, que armazenam as informações em bases de dados. Esses sistemas coletam uma ampla variedade de dados, incluindo corrente, tensão, potências, diversas temperaturas, velocidade do vento, ângulos das pás e velocidade de rotação do rotor. No que diz respeito à vibração, além dos dados globais de aceleração e velocidade, são também processados dados que contabilizam harmônicos e o acúmulo de bandas de frequência, entre outros. Esses dados são disponibilizados para análises e visualizações, ou para sistemas mais sofisticados que conseguem avaliar e

notificar anomalias básicas ou diagnósticos mais simples.(CHEN, 2022)

2.5

Técnicas de aprendizado de máquina aplicadas à detecção de anomalias

2.5.1

Análise de Componentes Principais - PCA

A Análise de Componentes Principais (PCA) transforma dados para comparar descobertas entre conjuntos de dados e determinar a importância dos componentes. Ela oferece vantagens na análise de dados, visualização e eficiência computacional, tornando-se valiosa para acelerar algoritmos de aprendizado de máquina e lidar com diversos tipos de dados (HASAN; ABDULAZEEZ, 2021). O PCA extrai informações importantes de conjuntos de dados complexos, expressa-as como componentes principais e fornece uma ferramenta flexível para resumir dados e superar a duplicação de características, aumentando a interpretabilidade e a eficiência da análise. Em alguns casos, o PCA permite a simplificação de grandes conjuntos de dados com um alto número de características, mantendo uma quantidade mínima de informação e alcançando bons resultados de classificação com complexidade reduzida. O PCA utiliza a matriz de covariância e seus autovalores e autovetores para extrair informações não correlacionadas. Ao analisar os autovetores, ele identifica a extensão em que as informações variam e fornece insights sobre os padrões subjacentes e a estrutura dos dados (HIRA; GILLIES, 2015). Além disso, o PCA pode ser aplicado em diversas áreas, como reconhecimento de padrões, compressão de dados e redução de dimensionalidade, tornando-se uma ferramenta essencial em ciência de dados e aprendizado de máquina.

2.5.2

Vizinho mais Próximo Condensado - CNN

Para abordar essa questão, a metodologia Condensed Nearest Neighbor (CNN) pode ser empregada para selecionar aleatoriamente instâncias do conjunto de dados disponível, garantindo uma representação equilibrada de todas as categorias. Baseado no algoritmo de agrupamento chamado Nearest Neighbor (NN), o CNN identifica clusters e reconhece que todas as instâncias dentro de um cluster compartilham as mesmas características. Em seguida, descarta as instâncias redundantes, resultando em um conjunto de dados condensado que mantém as mesmas características e constrói uma representação equilibrada de todas as categorias.

Como mostrado em Chou, Kuo e Chang (2006), o algoritmo começa selecionando o primeiro elemento, denotado como x_0 , de um conjunto de instâncias agrupadas como X_n . Essas instâncias são escolhidas aleatoriamente. O método Condensed Nearest Neighbor (CNN) então itera por todos os membros de X_n , adicionando um membro x ao grupo inicial se seu vizinho mais próximo não pertencer à mesma categoria que x . Esse processo continua até que o grupo contenha um número igual de amostras para cada categoria, respeitando ainda o número total de amostras na categoria minoritária conforme definido pelo parâmetro especificado pelo usuário. Este método proposto por Hand e Batchelort (1978) tem sido utilizado tanto para classificação quanto para redução e balanceamento de conjuntos de dados, visando o melhor desempenho e precisão possíveis.

Além disso, a metodologia CNN tem se mostrado eficaz em diversas aplicações práticas, como na detecção de anomalias e na análise preditiva, onde a precisão e o balanceamento dos dados são cruciais para a obtenção de resultados confiáveis. (CHAPLOT, 2019) A capacidade de reduzir a redundância e manter a representatividade das categorias faz do CNN uma ferramenta valiosa em cenários onde a qualidade dos dados é fundamental para o sucesso das análises. (BATISTA, 2004)

2.5.3

Máquinas de Vetores de Suporte - SVM

A Máquina de Vetores de Suporte (SVM) é um método popular usado para algoritmos de classificação ou regressão, baseado na teoria de aprendizado estatístico, conforme descrito em Berk (2008). É um algoritmo de aprendizado de máquina que utiliza um conjunto de treinamento composto por vetores de dados com rótulos de classe conhecidos. Esses vetores de dados, caracterizados por características únicas, são usados para projetar um hiperplano linear que separa diferentes classes. O objetivo do SVM é encontrar o hiperplano ótimo que maximize a margem entre as classes, permitindo a classificação eficaz de novos dados não vistos. Primeiro, precisamos abordar o problema de duas classes. Vamos considerar um conjunto de treinamento composto por l vetores de características $x_i \in R^n$, então i ($= 1, 2, \dots, n$) representa o número de amostras. Cada amostra é atribuída a um rótulo de classe y_i , que pode assumir o valor de 1 para uma classe ou -1 para a outra classe (*i.e.*, $y_i \in -1, 1$). Se as duas classes são linearmente separáveis, isso significa que existem separadores lineares chamados hiperplanos separadores. O objetivo do SVM é encontrar o hiperplano ótimo que maximize a margem entre as classes, permitindo a classificação eficaz de novos dados não vistos (KAVZOGLU; COLKESEN,

2009). Além disso, o SVM pode ser estendido para problemas de múltiplas classes e para casos onde as classes não são linearmente separáveis, utilizando técnicas como o kernel trick, que transforma os dados em um espaço de maior dimensão onde um hiperplano linear pode ser utilizado para a separação. Isso torna o SVM uma ferramenta poderosa e versátil em diversas aplicações de aprendizado de máquina.(ABDULLAH; ABDULAZEEZ, 2021)

3

Metodologia

Neste capítulo, será apresentada a metodologia adotada para o desenvolvimento desta pesquisa, detalhando as características e premissas estabelecidas ao longo do estudo. Além disso, será descrito o contexto do estudo de caso analisado, juntamente com os experimentos realizados, incluindo as bases de dados, softwares, linguagens de programação, algoritmos e ferramentas empregadas.

Vale ressaltar que os cenários selecionados para os experimentos foram baseados em experiências reais de uma empresa gestora de ativos, onde a equipe de engenharia enfrentou uma situação de decisão tardia devido à ausência de ferramentas capazes de identificar antecipadamente uma anomalia. As falhas discutidas no capítulo poderiam ter sido tratadas de maneira mais ágil caso existissem ferramentas que permitissem a detecção prévia das falhas, proporcionando uma resposta mais rápida e eficaz na gestão dos ativos.

Ao final do capítulo, será discutida a estratégia computacional aplicada, acompanhada de um pseudo-código que representa o código original em Python, utilizado como principal ferramenta de apoio à pesquisa.

3.1

Coleta e preparação dos dados do CMS (Condition Monitoring System)

Conforme explicado anteriormente, os fabricantes de aerogeradores costumam usar os dados disponíveis de forma tradicional, focando apenas em identificar problemas com base em limites e tendências gerais. Porém, descobriu-se que é possível ir além disso com o uso de ferramentas de aprendizado de máquina. Essas ferramentas conseguem analisar os dados de forma mais detalhada, o que permite um monitoramento mais amplo e eficiente, ajudando a entender melhor as condições de operação do equipamento.

A Tabela 3.1, descreve em maiores detalhes todos as *features* disponíveis para o desenvolvimento. Relacionado as medidas de vibração, foram considerados acelerômetros piezoelétricos com sensibilidade de 100 mV g^{-1} , frequência abaixo de 20 kHz que foram montados no *Drive Train* da turbina. Formas de onda amostradas de forma síncrona acompanharam as mudanças de velocidade, produzindo linhas espectrais para máquinas de velocidade variável, com processamento Transformada Rápida de Fourier (FFT) usando uma janela de Hanning.

Além dos dados mencionados, a Figura 3.1 exemplifica a localização dos sensores de vibração disponíveis além das demais variáveis de processo

Tabela 3.1: Descrição de Variáveis disponíveis para o desenvolvimento da modelagem de dados

Tipo	Categoria	Variável
Medidas de Processo	Temperaturas	Interna Nacelle Óleo da Multiplicadora Rolamentos do Gerador
	Gerais	Potência Velocidade do Rotor Velocidade do Vento Intervalos de Potência Direção do Vento Ângulo do Pitch Velocidade de Rotação do Eixo de Alta Velocidade (HSS)
Medidas de Vibração	Rolamento Acoplado e Não Acoplado ao Gerador	ISOA e ISOV 1 a 6x Frequência da Velocidade Rotacional Pico de Alta Frequência Fator de Crista de Alta Frequência Banda de Passagem de Alta Frequência Banda de Passagem de 300 a 700Hz
	Rolamento Principal Frontal e Traseiro	ISOA Pico de Alta Frequência Fator de Crista de Alta Frequência Banda de Passagem de Alta Frequência Banda de Passagem de 1 -10Hz Banda de Passagem de 10 -100Hz Banda de Passagem de 10 -1000Hz
	Planetária da Multiplicadora	ISOA e ISOV Pico de Alta Frequência Fator de Crista de Alta Frequência Passa Banda de Alta Frequência Frequências Relacionadas aos 5 Dentes de Engrenagem Banda de Passagem de 1 -20Hz Banda de Passagem de 20 -100Hz Banda de Passagem de 200 -500Hz
	Estágio Intermediário Multiplicadora	ISOA e ISOV Pico de Alta Frequência Fator de Crista de Alta Frequência Passa Banda de Alta Frequência Frequências Relacionadas aos 3 Dentes de Engrenagem 1 e 2x Frequência da Velocidade Rotacional
	Rolamentos Frontal e Traseiro do Gerador e Eixo Rápido da Multiplicadora.	ISOA e ISOV Pico de Alta Frequência Fator de Crista de Alta Frequência Banda de Passagem de Alta Frequência Frequências Relacionadas aos 3 Dentes de Engrenagem 1 e 2x Frequência da Rotação Nominal
	Nacelle Eixos X e Z	Banda de Passagem de 0.1-100Hz 1 a 3x Frequência da Velocidade Rotacional

descritas.

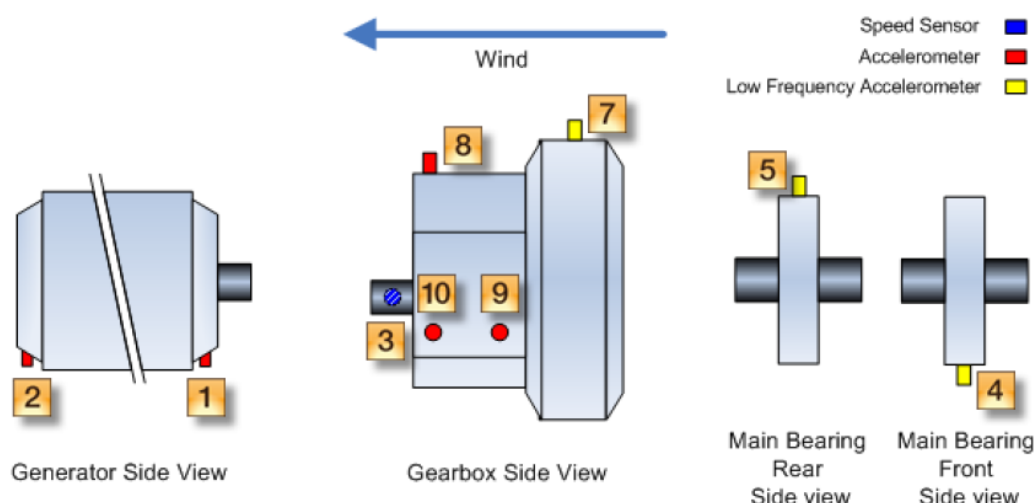


Figura 3.1: Diagrama de Instalação Sensores de Vibração (Retirado Documentação do Fabricante)

Baseados nas medidas expostas na Tabela 3.1, foram extraídos registros a cada 2 minutos de relacionados ao período de operação com e sem anomalia detectada para turbina selecionada.

3.2

Premissas dos experimentos realizados

O princípio da pesquisa passou pela seleção de 5 turbinas com resultados de falha similares avaliados a partir das inspeções em campo e dos laudos de análises do sistema de vibração realizados pelo fabricante do aerogerador.

As Figuras 3.2 e 3.3 mostram exemplo de um trecho de laudos recebidos do fabricante determinando a característica da falha e principalmente as razões do diagnóstico. O Anexo A mostra na íntegra um exemplo de relatório compartilhado pelo fabricante.

A Figura 3.2, em particular, ilustra uma análise de tendência global de uma medida de vibração, permitindo a identificação de padrões e comportamentos ao longo do tempo. Observando o gráfico, é possível notar variações na amplitude das vibrações, que podem indicar mudanças nas condições operacionais ou o início de falhas mecânicas. A análise detalhada desses dados, incluindo a decomposição em componentes harmônicos e a avaliação de bandas de frequência, fornece insights valiosos para a manutenção preditiva e a otimização do desempenho do equipamento.

A análise espectral de vibração é uma ferramenta crucial para o diagnóstico de falhas em componentes mecânicos. Na Figura 3.3 retirado de um

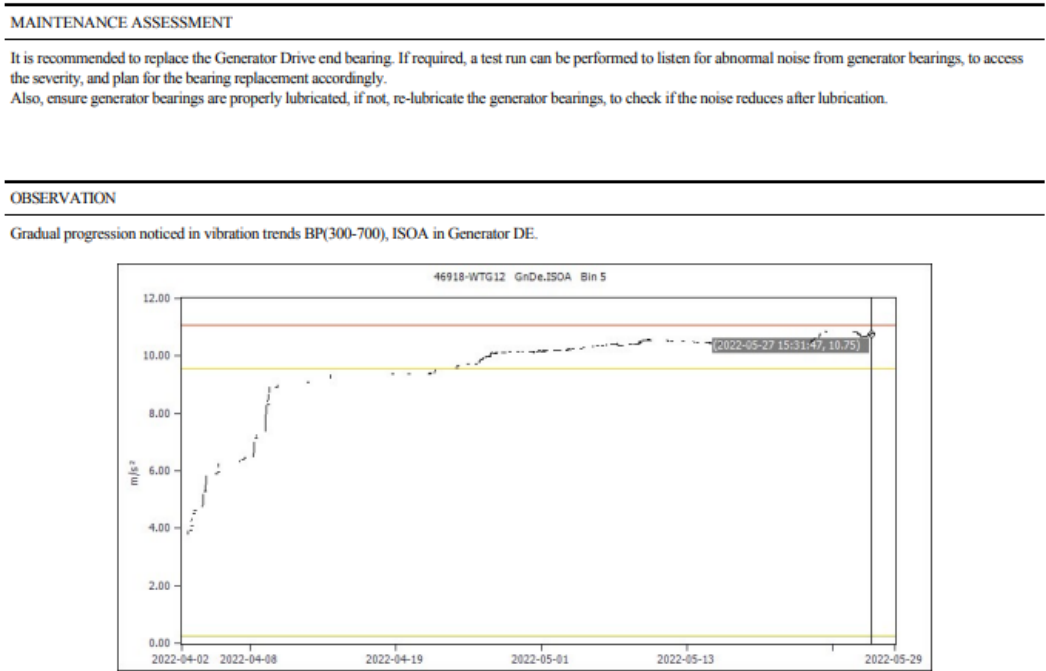


Figura 3.2: Relatório fabricante com medidas globais brutas

relatório enviado pelo fabricante da turbina, observa-se frequências de interesse que são indicativas de possíveis falhas. Especificamente, há um indicativo de falha na pista interna do rolamento, evidenciado pelos picos em harmônicos específicos na frequência de 106 Hz. Esses picos são sinais claros de que a pista interna está sofrendo desgaste ou dano, permitindo uma intervenção preventiva antes que ocorra uma falha catastrófica.

Portanto a partir das análises dos relatórios e da base de dados disponíveis, foram escolhidas 5 turbinas que sabidamente possuíam problemas relacionados a pista interna do rolamento acoplado do gerador. A Tabela 3.2 descreve as turbinas e os problemas conhecidos.

Tabela 3.2: Descrição de Turbinas e Anomalias Escolhidas

Turbina	Anomalia Base	Anomalia Secundária
WTG12	Pista Interna Rolamento - Lado Acoplado	Pista Interna Rolamento - Lado Não-Acoplado
WTG22	Pista Interna Rolamento - Lado Acoplado	Acoplamento - Eixo Rápido Gearbox
WTG24	Pista Interna Rolamento - Lado Acoplado	Sensor Ruim - Eixo Rápido Gearbox
WTG44	Pista Interna Rolamento - Lado Acoplado	Baixa Lubrificação - Eixo Principal
WTG48	Pista Interna Rolamento - Lado Acoplado	-

As Figuras 3.4 e 3.5 ilustram um exemplo de rolamento retirado da turbina com registro de anomalia e sinais claros desgaste.

INTERPRETATION

FFT Spectral analysis reveals inner race bearing fault frequency 106.0Hz and its harmonics at generator DE which indicates that fault is associated with the inner race of bearing.

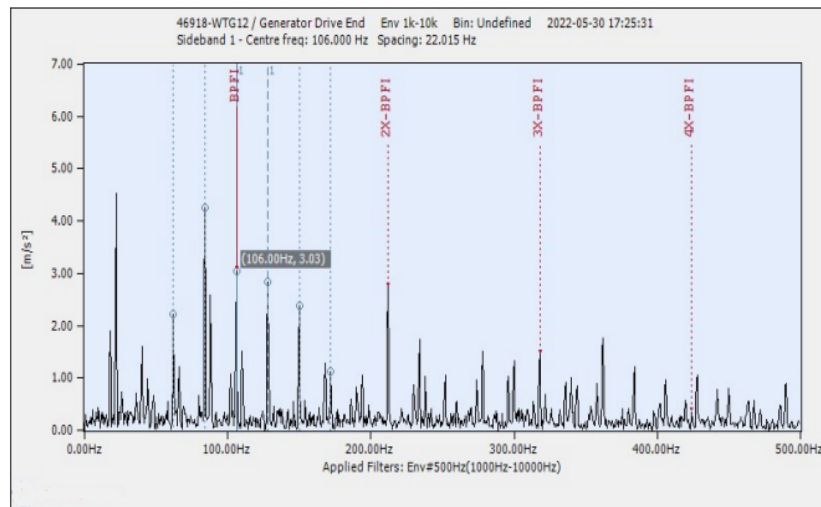


Figura 3.3: Relatório fabricante com interpretação de especialista indicando diagnóstico de anomalia

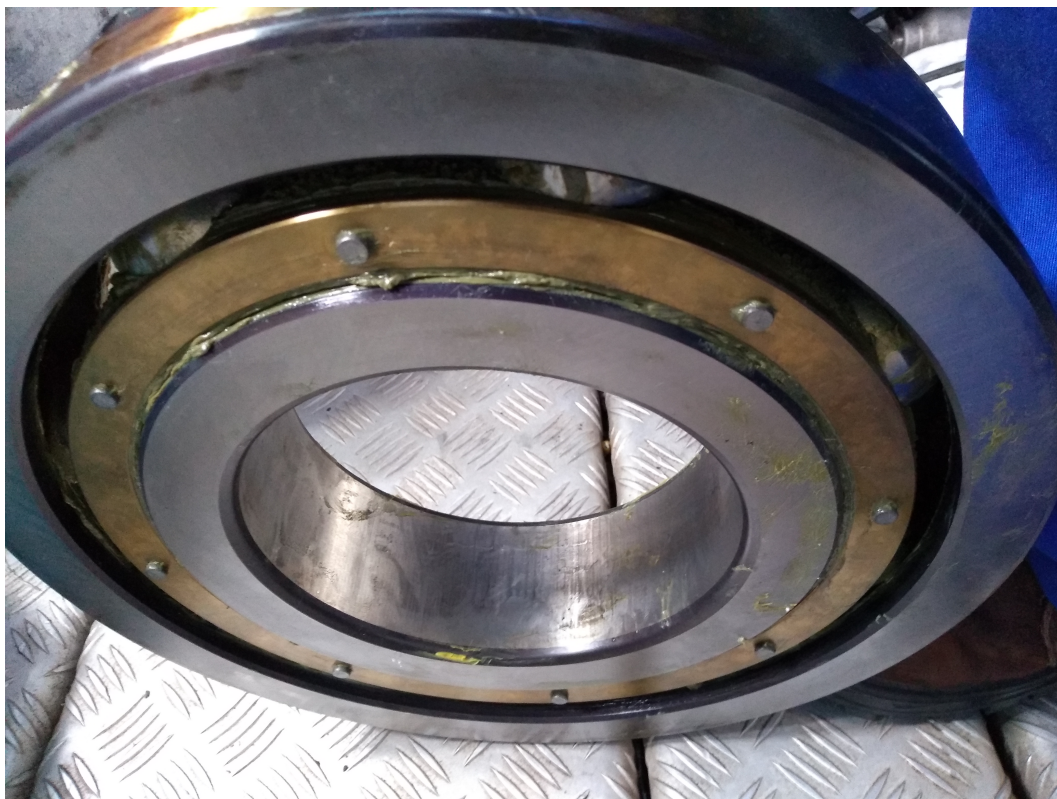


Figura 3.4: Registro Fotográfico de um Rolamento Danificado do Lado Acolado do Gerador.

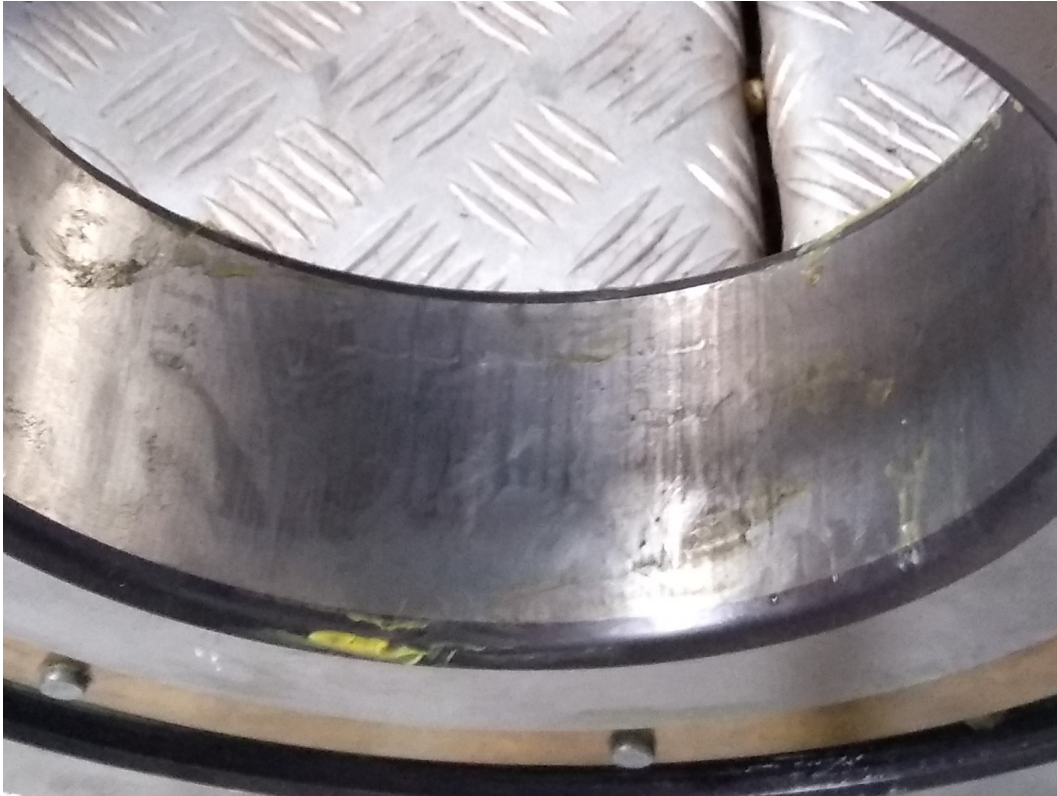


Figura 3.5: Detalhes do desgaste na pista interna do rolamento danificado.

3.3

Estratégias de treinamento, validação e teste dos modelos

3.3.1

Premissas de Aprendizado Supervisionado

Uma das premissas mais importantes da pesquisa é a questão da supervisão do algoritmo a ser utilizado. Os *labels* de cada amostra impactam diretamente os resultados obtidos uma vez que o algoritmo depende destas informações de entradas para extrapolar as características e portanto classificar outras amostras. A Figura 3.6 mostra um exemplo, referente a dados de medidas harmônicas da frequência de rotação da Turbina WTG12, do método utilizado na definição manual dos *labels* considerados. Esta definição baseou-se nos relatórios e recomendações do fabricante quanto ao equipamento monitorado e os diagnósticos propostos.

Por isso, as categorias foram determinadas de forma manual para cada turbina baseado na série temporal considerando informações exógenas como os relatórios do fabricante, assim como os valores absolutos de ISOA e ISOV para os equipamentos monitorados.

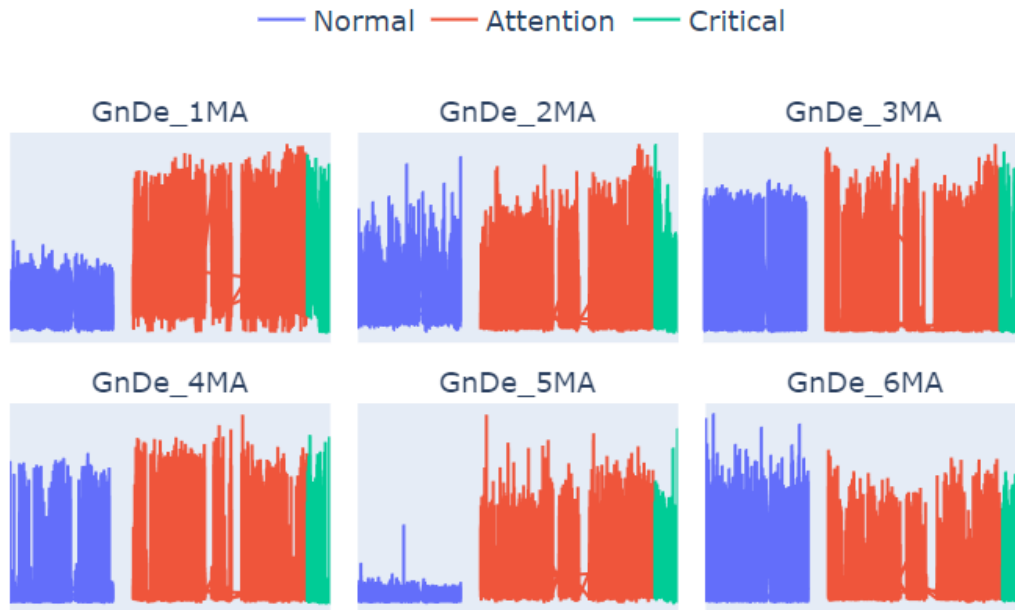


Figura 3.6: Trend de variáveis de harmônica da velocidade de rotação determinantes para definição das categorias

3.3.2

Premissas de Validação entre Turbinas

Aspecto muito importante para avaliação dos modelos de classificação é o formato de validação utilizado durante a pesquisa. Neste trabalho, a proposta é emular uma visão clássica de validação cruzada, em vez de usar uma única divisão dos dados, a validação cruzada divide os dados em várias partes, treinando e testando o modelo em diferentes subconjuntos. Isso assegura que o modelo não esteja ajustado apenas a um conjunto específico de dados e possa generalizar bem para novos dados. Neste trabalho, são utilizadas bases de dados de mais de um aerogerador para diferenciar os conjuntos de treinamento e validação. Dessa forma, ajuda a evitar que o modelo seja interpretado como utilizável somente ao mesmo equipamento de treino e portanto direcionando a aplicação em qualquer equipamento de características de operação semelhante. Assim, a validação reforça a confiabilidade da avaliação do modelo e oferece informações valiosas para aprimorar seu desempenho em aplicações reais.

3.4

Métricas para Avaliação

O formato da matriz de confusão é utilizada para avaliar os resultados da classificação experimental. Esta matriz possui eixos que representam os Classificações reais e os Classificações previstas, e o valor em cada quadrante

refere-se ao número percentual de registros de uma categoria específica que são classificados em uma certa categoria pelo total de amostras, seja corretamente ou não. (MINH, 2024) Esses quadrantes, demonstrados na Figura 3.7, têm nomes específicos de acordo com sua interpretação qualitativa descrita abaixo:

- Verdadeiros Positivos (TP) - o número de instâncias positivas corretamente classificadas;
- Verdadeiros Negativos (TN) - o número de instâncias negativas corretamente classificadas;
- Falsos Positivos (FP) - o número de instâncias negativas incorretamente classificadas como positivas;
- Falsos Negativos (FN) - o número de instâncias positivas incorretamente classificadas como negativas.

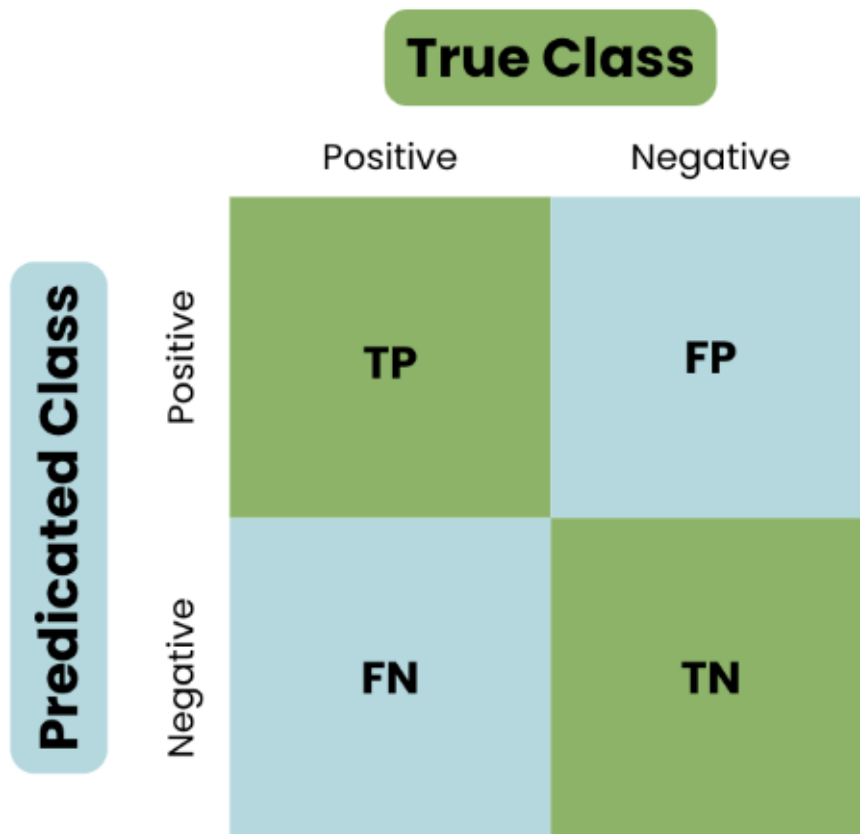


Figura 3.7: Exemplo Utilização de Matriz de Confusão para interpretação de resultados de métodos de classificação.(MINH, 2024)

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3-1)$$

É possível calcular KPIs que auxiliam na interpretação desta matriz e também reforçam o foco no que deve ser avaliado ou é mais importante para a avaliação da solução do problema proposto. Por exemplo, a Acurácia definida na Equação 3-1, um KPI que será utilizado para avaliação neste trabalho, representa a porcentagem de classificações corretas em relação ao número total de amostras testadas.

Além da Acurácia que será o principal indicador avaliado será a Razão de Falsos Positivos (FPR), este indicador mede a razão percentual de amostras com estado negativo que foram classificadas incorretamente como positivo. A Equação 3-2 expõe o método de cálculo do indicador que busca preservar o erro, valorizando o modelo que não classifica incorretamente uma amostra de sem anormalidade.

$$FPR = \frac{FP}{TN + FP} \quad (3-2)$$

3.5

Metodologia Experimento Inicial

Nesta seção, será exposto o primeiro experimento realizado dentro da hipótese proposta e apresentadas as principais métricas e resultados dos testes. Será, através de um resumo gráfico, apresentada o fluxo de trabalho em cada experimento, principalmente relacionado aos casos base, sem a utilização do PCA descrito no Capítulo anterior deste trabalho e com a aplicação da ferramenta.

3.5.1

Descrição Experimento com 2 Turbinas

Inicialmente, o experimento proposto é de apenas uma turbina compondo o conjunto de treino do algoritmo e a validação do modelo sendo realizado em um segundo aerogerador. A Figura 3.8 detalha o fluxo do experimento. Foi aplicado o CNN na base de dados com intenção de balanceamento das categorias das amostras, principalmente com intenção de diminuição da quantidade de registros. Isto foi necessário devido a capacidade computacional limitada que mesmo utilizando clusterização local para o treinamento não conseguiu ultrapassar a barreira de processamento. A solução foi utilizar o método de clusterização para seleção de amostras representativas para treinamento sem perder a capacidade de extrapolação de características.

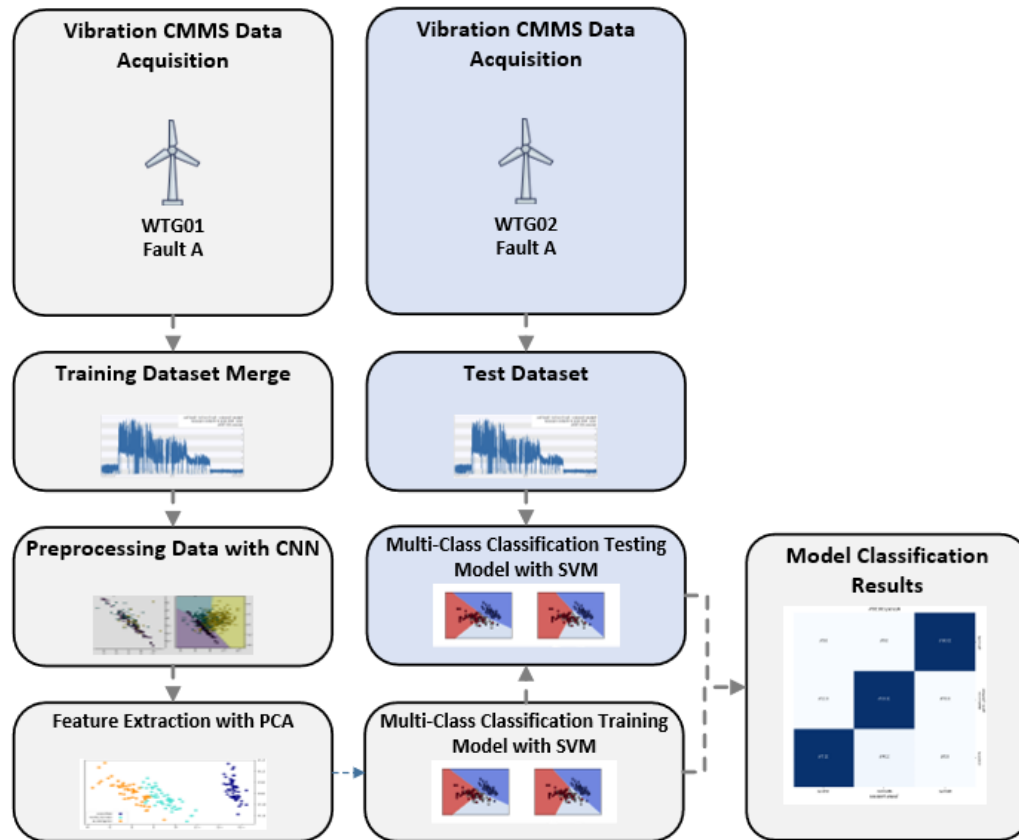


Figura 3.8: Representação esquemática do experimento inicial proposto

Após aplicação do CNN, há então o desenvolvimento de dois modelos paralelos para comparação: 1- Modelo com todas as medidas disponíveis do equipamento monitorado como características do modelo; e 2- Aplicação do PCA com objetivo de redução de características a partir da análise de principais componentes. Dessa forma, os primeiros resultados foram obtidos, caracterizando o aspecto promissor da hipótese levantada.

3.5.2

Premissas aplicação PCA

Aspecto necessário quando é abordada a aplicação do PCA é o threshold de variância escolhido para determinação da quantidade de componentes serão utilizadas e definidas no modelo. Neste caso, foi definido a escolha de 99 % da variância devido intenção de manter o conservadorismo e buscar perder o mínimo de características possíveis para o algoritmo de classificação.

As Figuras 3.10 e 3.9 demonstram através do gráfico do tipo *pair plot* a dispersão gráfica entre todas as componentes selecionadas a partir do limite definido.

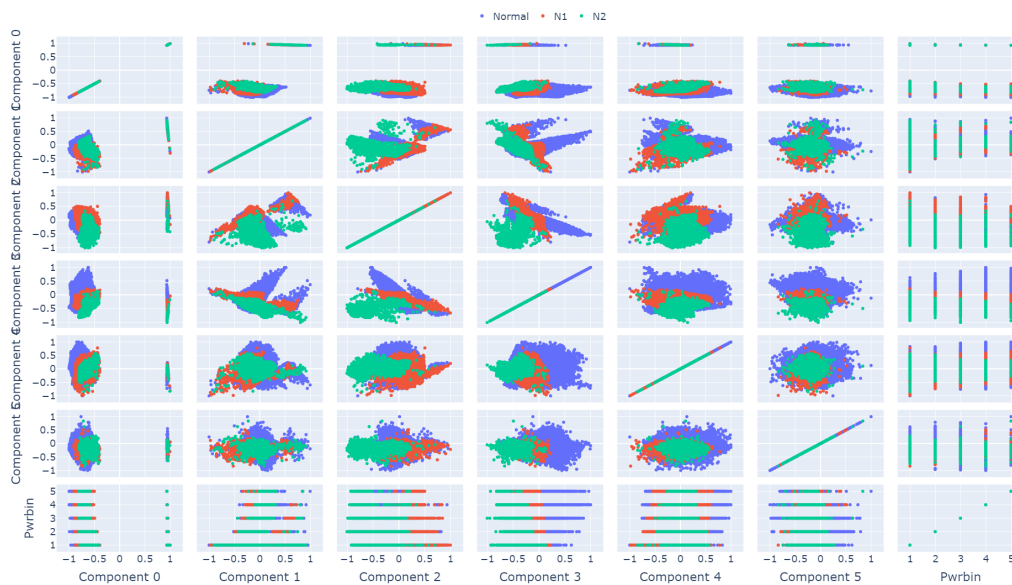


Figura 3.9: Pairplot Análise de Principais Componentes escolhidas diferenciadas pelas categorias - WTG 22

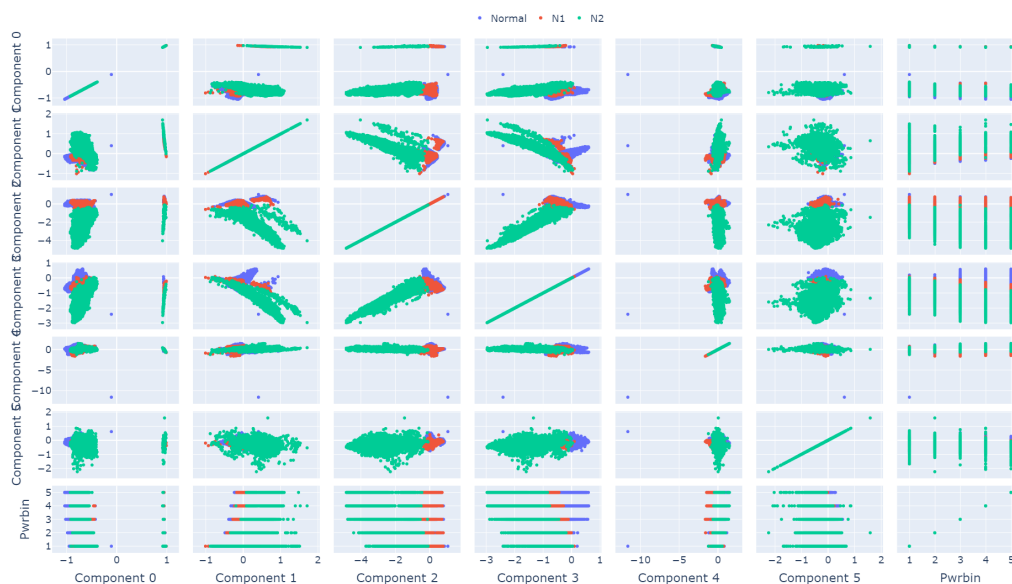


Figura 3.10: Pairplot Análise de Principais Componentes escolhidas diferenciadas pelas categorias - WTG 48

3.6

Metodologia Experimento Completo

3.6.1

Validação com 5 Turbinas

O próximo passo da pesquisa é expansão e avaliação da modelagem para mais turbinas com objetivo de avaliar como o modelo se comporta variando os indivíduos utilizados no dataset de treino e de validação. Foram utilizadas todas as turbinas mencionadas na Tabela 3.2.

3.6.2

Experimento com 2 Turbinas no Conjunto de Treinamento

A evolução da pesquisa na busca de um modelo capaz de identificar anomalias e principalmente possibilitar a escalabilidade da utilização da ferramenta para outros equipamentos de características semelhantes passa por buscar aumentar a robustez e expor o algoritmo ao máximo de situações aspectos operacionais possíveis. Portanto, a proposta é desenvolver um conjunto de amostras para treinos que seja uma fusão de bases de dados de mais de uma turbina. Esta proposição visa aumentar a heterogeneidade das amostras considerando que turbinas distintas, apesar de possuírem mesmas características de instalação e equipamentos, possuem exposição a fenômenos naturais ligeiramente distintos por estarem posicionadas em localizações diferentes.

A Figura 3.11 demonstra visualmente os passos propostos na aplicação do experimento até a avaliação das métricas. A fusão das bases de dados das duas turbinas propostas ocorre no momento inicial antes da aplicação das ferramentas computacionais de tratamento de dados mencionadas no experimento anterior.

Mantendo os parâmetros para extração de características e treinamento do algoritmo, as métricas foram calculadas considerando todas as combinações entre as 5 turbinas destacadas. O objetivo é verificar a robustez do algoritmo e a dependência dos resultados em relação à base de dados de treinamento.

3.7

Descrição estratégia de processamento de dados

O grande desafio do teste de diversos cenários com uma disponibilidade muito grande de amostras em cada base de dados, aproximadamente 2 anos de dados utilizados, é o tempo de processamento para os treinamentos e testes realizados.

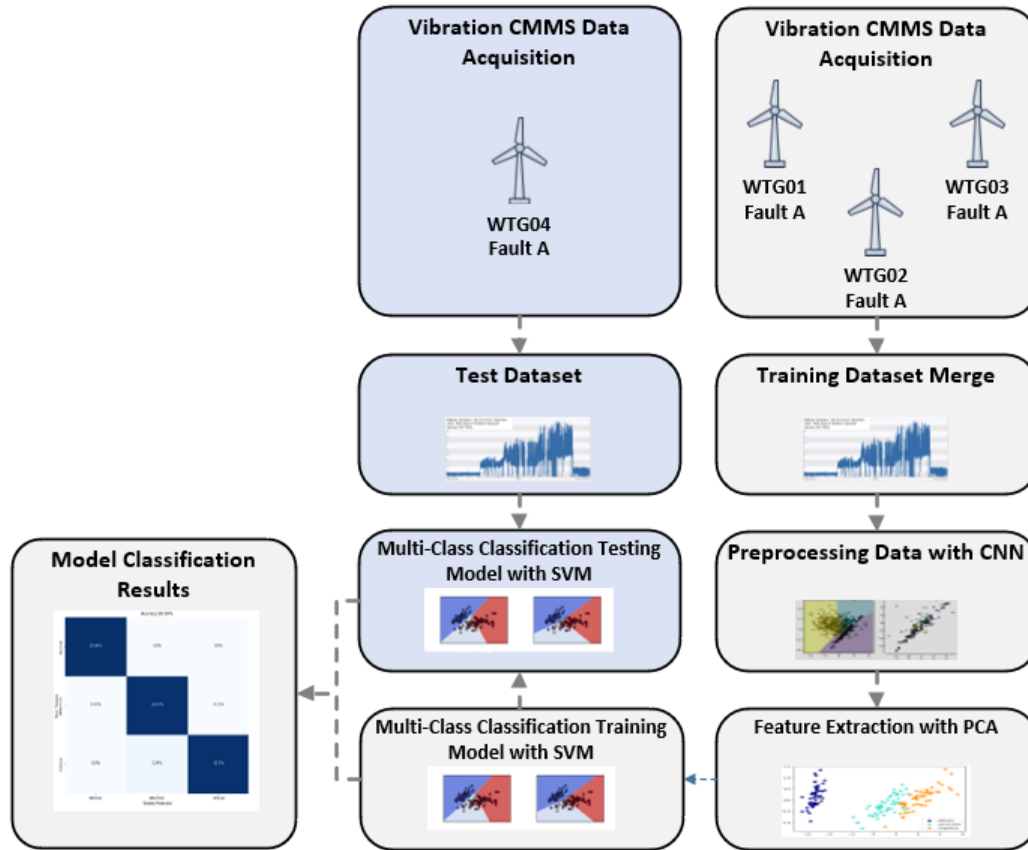


Figura 3.11: Representação esquemática do experimento completo

A estratégia utilizada foi de paralelizar o processamento de todos os cenários utilizando a biblioteca *Concurrent* na linguagem *Python*. Assim, foi possível realizar 3 iterações com 20 treinamentos realizados de forma simultânea e um tempo total aproximado de processamento em 6 horas.

Cada cenário consiste em um fluxo completo de dados como apresentado nas sessões anteriores, ou seja, desde a aquisição dos dados disponíveis, fusão dos dados de mais turbinas a depender do experimento em questão, tratamentos de pré-processamento quanto a qualidade de dados, CNN e PCA quando aplicável. Somente então é realizado o treinamento do classificador SVM para teste e validação conforme proposto. O Pseudo-código proposto é exposto no Algoritmo 1.

Algoritmo 1: Fluxo Completo de Dados para Classificação SVM

```
1  para cada cenário em conjunto_de_cenários faça
2  Aquisição dos Dados: adquirir_dados(cenário)
3  Fusão dos Dados de Múltiplas Turbinas (se aplicável): se
    múltiplas_turbinas(cenário) então
4  dados_fusionados = fundir_dados(dados) senão
5  dados_fusionados = dados
6  Pré-Processamento dos Dados:
    preprocessar_dados(dados_fusionados)
7  Aplicar CNN (se aplicável): se cnn_aplicável(cenário) então
8  dados_cnn = aplicar_CNN(dados_preprocessados) senão
9  dados_cnn = dados_preprocessados
10 Aplicar PCA (se aplicável): se pca_aplicável(cenário) então
11 dados_pca = aplicar_PCA(dados_cnn) senão
12 dados_pca = dados_cnn
13 Dividir os Dados em Conjuntos de Treinamento e Teste:
    treino_X, teste_X, treino_y, teste_y = dividir_dados(dados_pca)
14 Treinar o Classificador SVM: treinar_SVM(treino_X, treino_y)
15 Testar o Classificador SVM: previsões =
    testar_SVM(classificador_SVM, teste_X)
16 Validar os Resultados: resultado_validação =
    validar_resultados(previsões, teste_y)
17 Exibir ou Salvar os Resultados:
    exibir_resultados(resultado_validação)
```

4

Discussão e Resultados

Neste capítulo, serão apresentados e discutidos os resultados obtidos a partir da aplicação dos métodos de aprendizado de máquina para a detecção de anomalias em turbinas eólicas dispostos na metodologia do capítulo anterior. Inicialmente, descrevemos o conjunto de dados utilizado e as etapas de pré-processamento realizadas para garantir a qualidade dos dados. Em seguida, detalhamos o desempenho dos modelos de aprendizado de máquina empregados, como Agrupamento por Vizinho mais próximo (CNN), Análise de Componentes Principais (PCA) e Máquinas de Vetores de Suporte (SVM) na detecção de anomalias na operação de turbinas eólicas.

Por último nesse capítulo, será apresentado a expansão do experimento inicial considerando 5 turbinas para os conjuntos de treino e teste avaliando a robustez do modelo e a capacidade de aplicação em larga escala, principalmente aplicado a grandes frotas de turbina. Será mantido a métrica principal de avaliação como a Acurácia definida no Capítulo 3, entretanto o formato gráfico de apresentação será distinto. O Mapa de Calor baseado no indicador mencionado proporcionará uma melhor comparação entre o resultado dos modelos e de todos os cenários e testes realizados.

4.1

Resultados Modelagem Inicial

Preliminarmente, como forma de teste da metodologia, foi feita a classificação do modelo inicialmente dividindo os datasets individuais na proporção de 70%/30% para conjuntos de treino e teste, respectivamente. Os resultados das matrizes de confusão são apresentados nas Figuras 4.1. A Figura 4.2 demonstra os resultados com a aplicação do PCA no conjunto de dados originais.

É importante destacar que foram utilizadas as variáveis de processo e as variáveis relacionadas ao gerador, conforme apresentado na Tabela 3.1. Esses dados foram consolidados em uma base de dados ordenada temporalmente, com amostras coletadas de forma sincronizada.

4.2

Resultados Modelagem com Validação Cruzada

Portanto, a partir da validação inicial do modelo, o próximo passo é a expansão para a validação cruzada, ou seja, o treino e o teste do algoritmo é feito com o dataset da turbina original mas a validação é feito com dados de

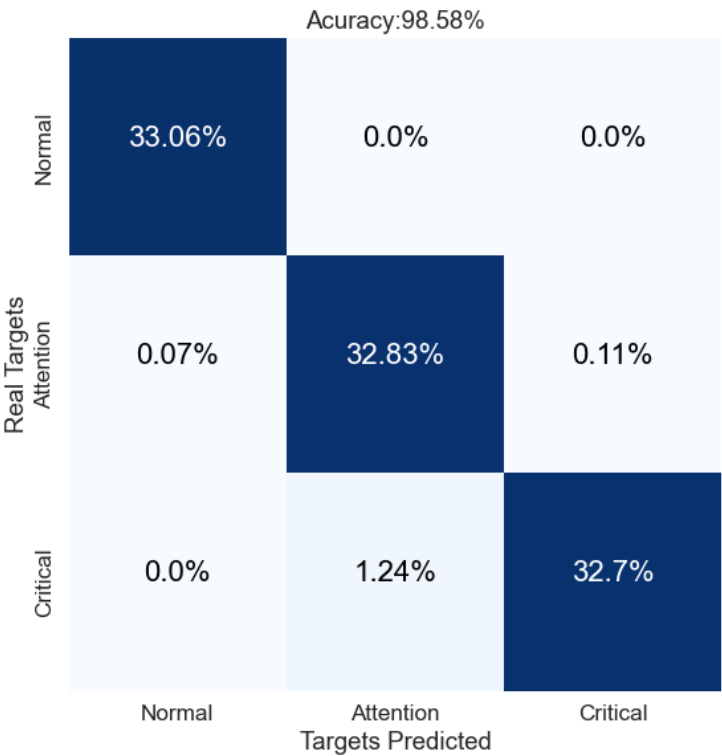


Figura 4.1: Matriz de Confusão Modelo Puro - WTG22

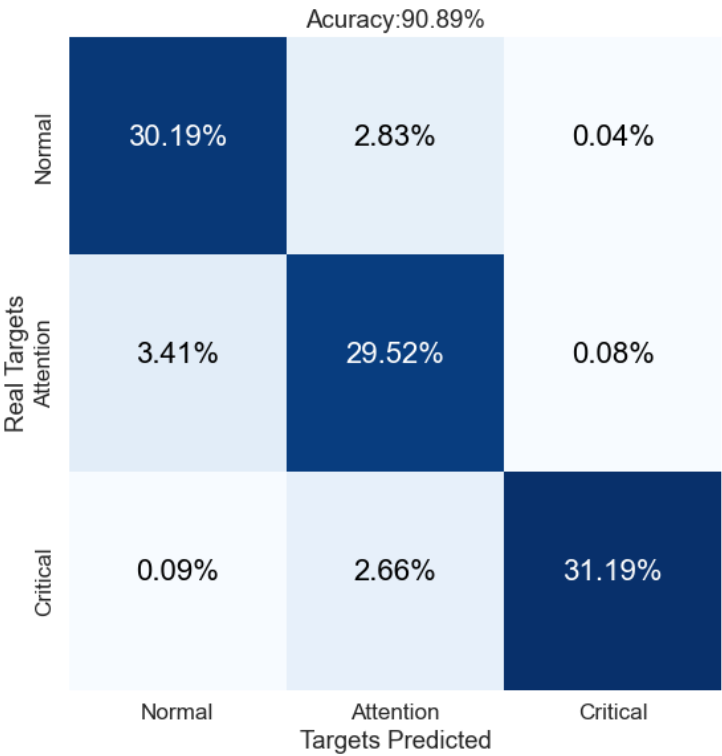


Figura 4.2: Matriz de Confusão Modelo PCA - WTG22

uma segunda turbina que o algoritmo nunca teve contato. O objetivo é testar capacidade de generalizar os resultados e de avaliar outros indivíduos que não tenham participado do dataset de treinamento. Dessa forma, são expostos nas Figuras 4.3 e 4.4 os resultados da classificação cruzada ao considerar o cenário exposto.

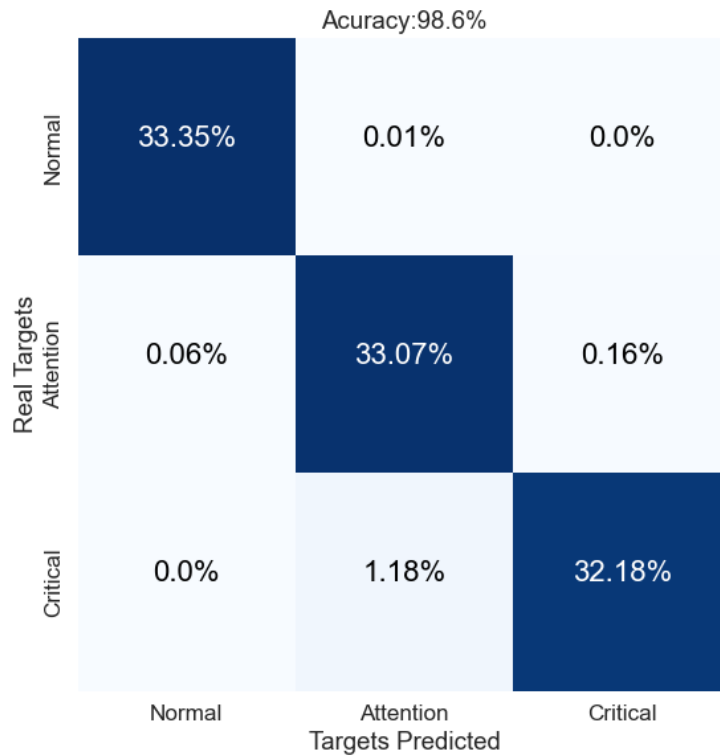


Figura 4.3: Matriz de Confusão Modelo Puro - WTG22 e WTG48

4.3

Apresentação, Comparação de Métricas e Considerações sobre o Experimento Inicial

De forma resumida, a Figura 4.5 mostra um gráfico de barras com a medida de acuracidade comparativa entre os resultados com o chamado modelo puro, ou seja, com todas as entradas originais *versus* o modelo com a análise de componentes principais e também os resultados para aplicação do modelo de validação cruzada em ambos os cenários.

Este experimento demonstrou que, com uma precisão superior a 90 %, é possível classificar a degradação do conjunto de rolamentos do gerador em uma turbina eólica. O uso combinado de CNN e PCA resultou em uma redução na complexidade de entrada sem comprometer a identificação de situações de falha, pois conforme a Figura 4.2 demonstra, somente 3.5 % das amostras

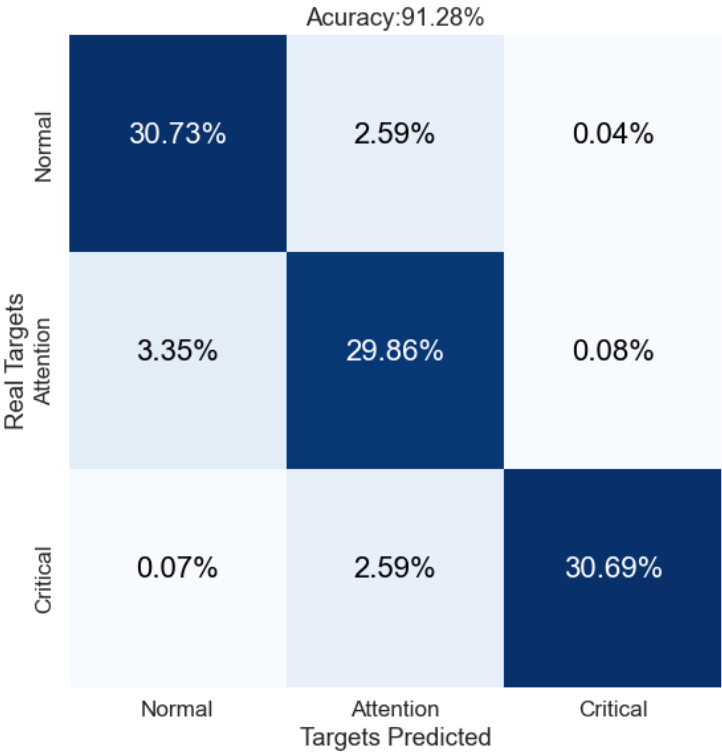


Figura 4.4: Matriz de Confusão Modelo PCA - WTG22 e WTG48

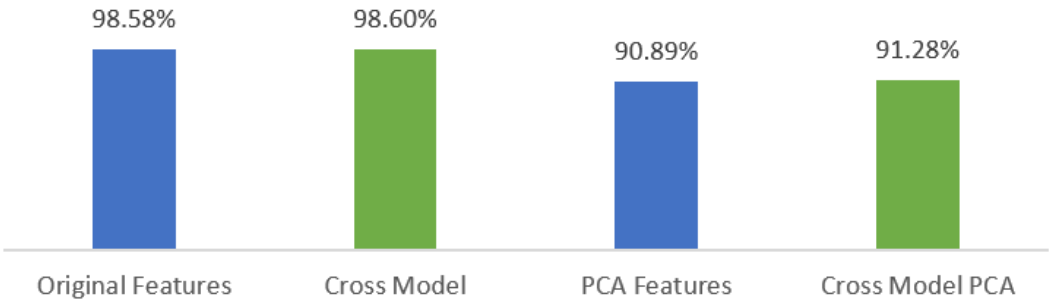


Figura 4.5: Gráfico de Barras Comparação Acuracidade entre Cenários Preliminares

de falha foram classificadas como normal. É importante destacar que, como o objetivo é implementar uma aplicação de avaliação em tempo real, o impacto de falsos positivos, é significativo na gestão da manutenção de ativos, por isso, um número muito baixo de falsos positivos afeta positivamente o processo e evita custos e tempo desnecessários associados a substituições desnecessárias.

Os resultados do teste de validação cruzada também são importantes para demonstrar que o comportamento de falha é bastante semelhante, independentemente do indivíduo avaliado. Portanto, em situações onde a disponibilidade de dados CMS para treinamento é curta, é possível usar dados rotulados de outra turbina eólica similar, mantendo a precisão. A progressão deste trabalho envolveria aumentar o nível de detalhamento na avaliação da criticidade da falha e até mesmo diagnosticar o tipo específico de falha.

4.4

Expansão do Experimento para 5 turbinas

O próximo passo da pesquisa é a expansão e avaliação da modelagem para mais turbinas com objetivo de avaliar como o modelo se comporta variando os indivíduos utilizados no conjunto de treino e de validação. Foram utilizadas todas as turbinas mencionadas na Tabela 3.2 para dois formatos de avaliação: Expansão do Experimento inicial para teste em todas as turbinas selecionadas; e Experimento Completo proposto determinando o conjunto de treino com duas turbinas.

A implementação do pipeline proposto, conforme descrito na metodologia desta pesquisa, no conjunto de dados das turbinas estudadas permitiu aprofundar a compreensão da aplicação, especialmente em relação às condições operacionais das turbinas. As condições apresentadas na Tabela 3.2 expõem um contexto de falha principal similar, porém com condições secundárias distintas. Os resultados de acuracidade dos modelos, apresentados nas Figuras 4.6 e 4.7 para os modelos puro e com PCA, respectivamente, indicam alguns aspectos críticos. Em particular, destacam-se negativamente os resultados referentes às Turbinas WTG24 e WTG44, ambas associadas a problemas específicos, como falha de sensor e má lubrificação do eixo principal, respectivamente.

4.4.1

Análise Resultados de Acuracidade Baixa

No caso da WTG24, destaca-se a necessidade crucial de avaliação e verificação da qualidade dos dados para assegurar a correta interpretação dos

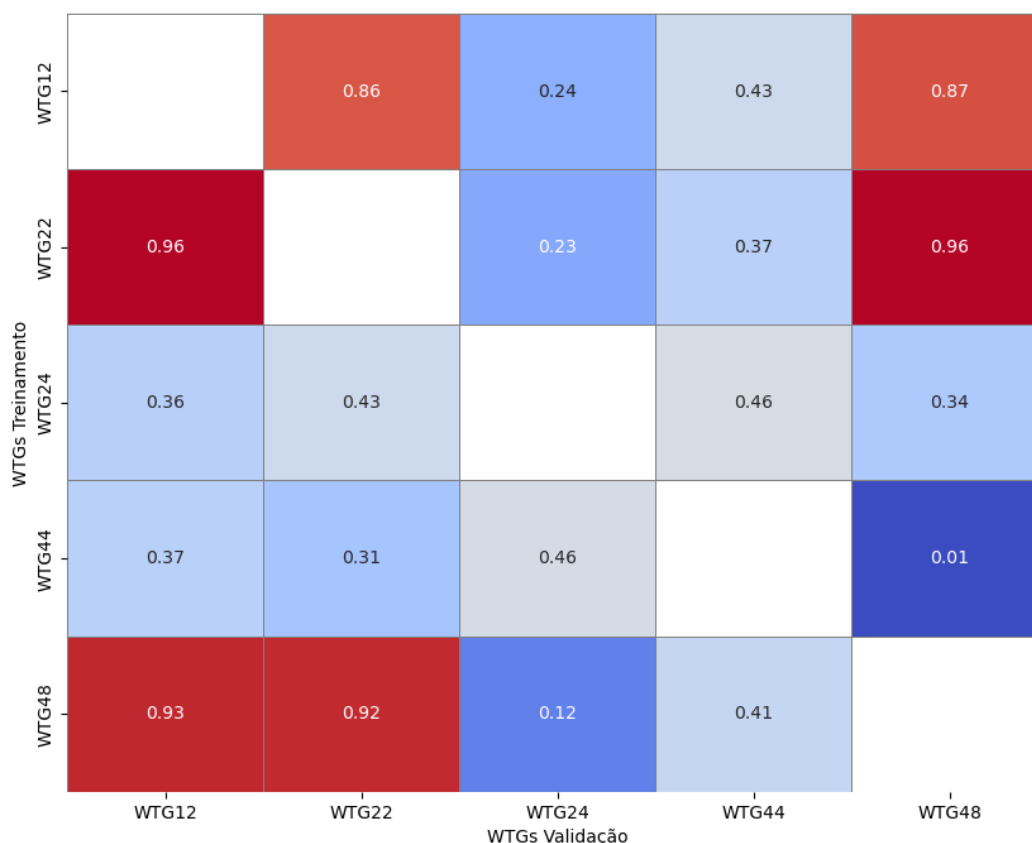


Figura 4.6: Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo Puro

resultados dos modelos, uma vez que medições incorretas impactam diretamente a performance e a confiabilidade do trabalho realizado. A Figura 4.8 apresenta um gráfico fornecido pelo fabricante após uma investigação técnica da falha. Mesmo assim, por meio da análise técnica de outros sensores disponíveis e da inspeção local, o fabricante conseguiu realizar o diagnóstico correto após a parada total do aerogerador causando um atraso significativo na solução do impacto da perda energética devido indisponibilidade da turbina.

Embora por razões distintas, a WTG44 também apresenta um padrão de acuracidade baixo nos modelos avaliados. Buscando mais informações sobre a condição operacional da turbina, conforme relatado em relatório pelo fabricante após inspeções e ensaios realizados, foi identificada uma condição de má lubrificação no eixo principal que, aliada à falha existente no rolamento analisado, alterou o padrão vibração esperado. Essa situação evidencia uma baixa capacidade de desassociação de falhas, especialmente quando ocorrem em sistemas distintos do aerogerador.

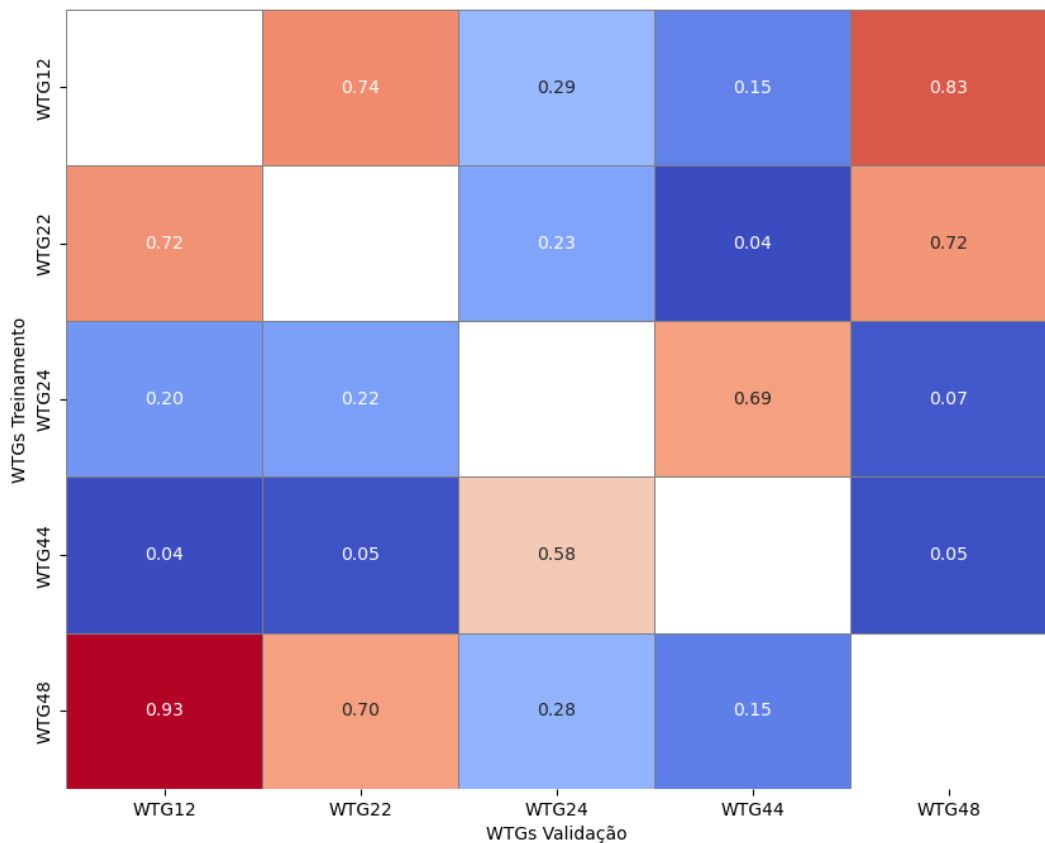


Figura 4.7: Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo PCA

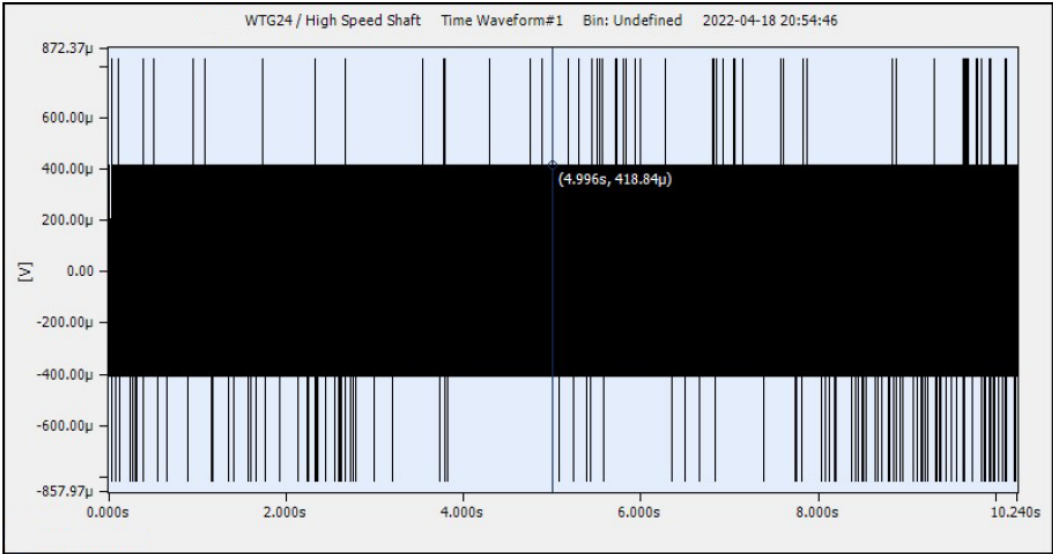


Figura 4.8: Gráfico retirado do relatório de falha de sensor entregue pelo fabricante da turbina

4.4.2

Considerações acerca dos resultados

Além dos bons resultados exibidos na Figura 4.6, é relevante ressaltar que especialmente para as turbinas WTG12, WTG22 e WTG48, a robustez do modelo ao ser testado em turbinas que apresentam problemas exclusivamente relacionados ao sistema em estudo. Por exemplo, a WTG12, que sofre degradação em ambos os rolamentos do gerador, e a WTG22, com um problema secundário no eixo rápido acoplado entre a *gearbox* e o gerador, demonstraram um desempenho consistente. A WTG48, que não apresentava nenhuma condição secundária conhecida, também mostrou uma boa capacidade de identificação das falhas nas demais turbinas.

Adicionalmente, uma contribuição significativa para o objetivo da pesquisa foi a confirmação dos resultados após a aplicação da análise de componentes principais, como mostrado na Figura 4.7. Embora se esperasse uma queda na performance dos modelos devido à redução das *features* disponíveis, observou-se uma melhora em alguns casos. Uma possível explicação para essa melhora pode estar associada à simplificação dos componentes, preservando apenas aqueles que carregam as principais assinaturas das falhas estudadas.

4.4.3

Análise Razão de Falsos Positivos

A acuracidade foi escolhida como o principal indicador de investigação nesta pesquisa, por ser uma medida que reflete a capacidade do modelo em acertar as classificações *versus* o total de classificações realizadas. Além disso, um indicador complementar importante é o percentual de falsos positivos (FPR), que resulta da aplicação do modelo. O FPR é calculado pela razão entre a quantidade de amostras classificadas como anormais (Atenção ou Críticas) e o total de amostras cuja classificação real é normal. Com isso, busca-se avaliar a robustez do modelo em evitar a indicação de falhas em momentos e/ou turbinas que não apresentam degradação. Esse objetivo é crucial devido ao custo de oportunidade associado à atuação desnecessária da equipe de manutenção em um complexo eólico, o que pode resultar na alocação de recursos em uma turbina que não requer manutenção, deixando de atender outra que demanda intervenção imediata.

Os mapas de calor apresentados nas Figuras 4.9 e 4.10 mostram os valores de FPR calculados para todos os modelos testados. Os resultados do modelo puro confirmam a robustez dos testes realizados nas turbinas WTG12, WTG22 e WTG48. No entanto, é importante destacar que os modelos validados na

WTG22 apresentam um valor maior de FPR, o que merece atenção.

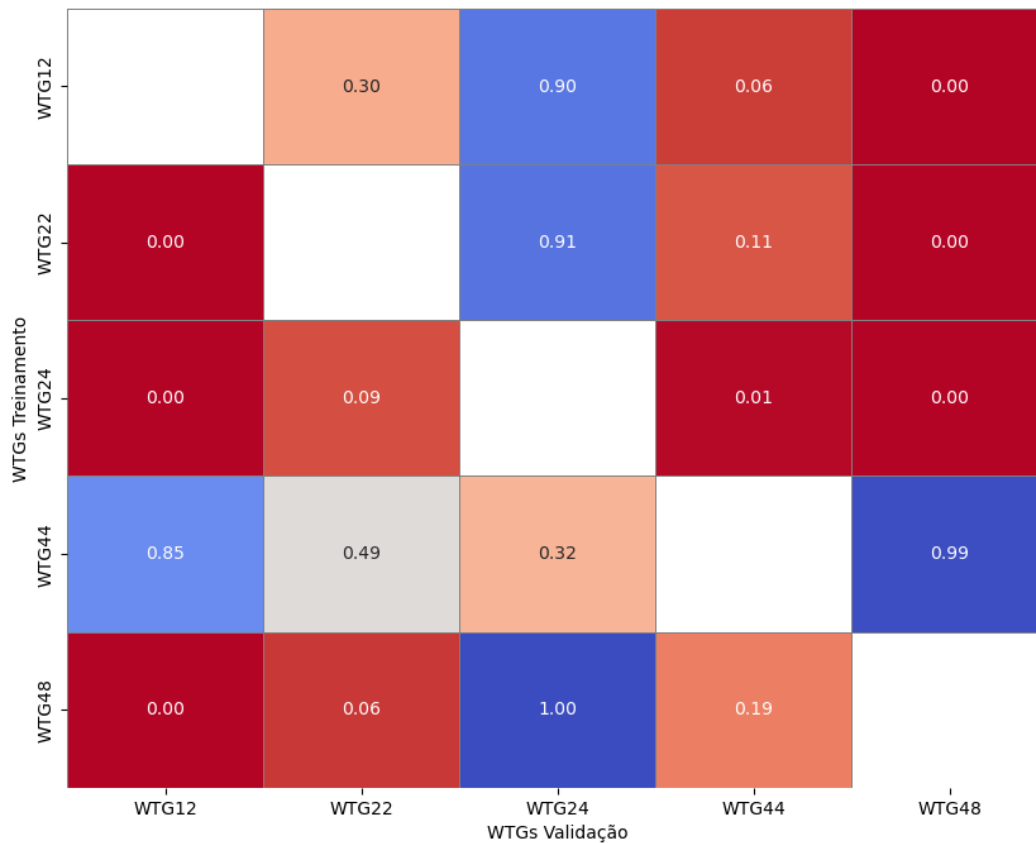


Figura 4.9: Resultados de FPR para modelo puro considerando amostrais com categoria Normal

Mais especificamente, os resultados obtidos após a aplicação do PCA, como exposto na Figura 4.10, mostram um aumento considerável do valor de FPR em comparação com o modelo puro. Embora a acuracidade desse modelo permaneça em bons padrões, o aumento no FPR sugere que esses modelos podem introduzir impactos negativos na operação, resultando em possíveis falsas indicações de falhas durante o funcionamento normal das turbinas.

4.5

Hipótese de 2 Turbinas no Conjunto de treinamento

A próxima hipótese testada envolveu a possibilidade de utilizar um número maior de turbinas no conjunto de treinamento. O objetivo principal dessa abordagem é fornecer ao classificador uma quantidade mais ampla e diversificada de características e possibilidades durante o processo de treinamento. Ao aumentar a variedade de dados de entrada, espera-se que o classificador possa absorver uma gama mais extensa de padrões e comportamentos, o que, por sua vez, pode expandir significativamente sua capacidade de identificar e generalizar essas características no conjunto de validação. Essa estratégia

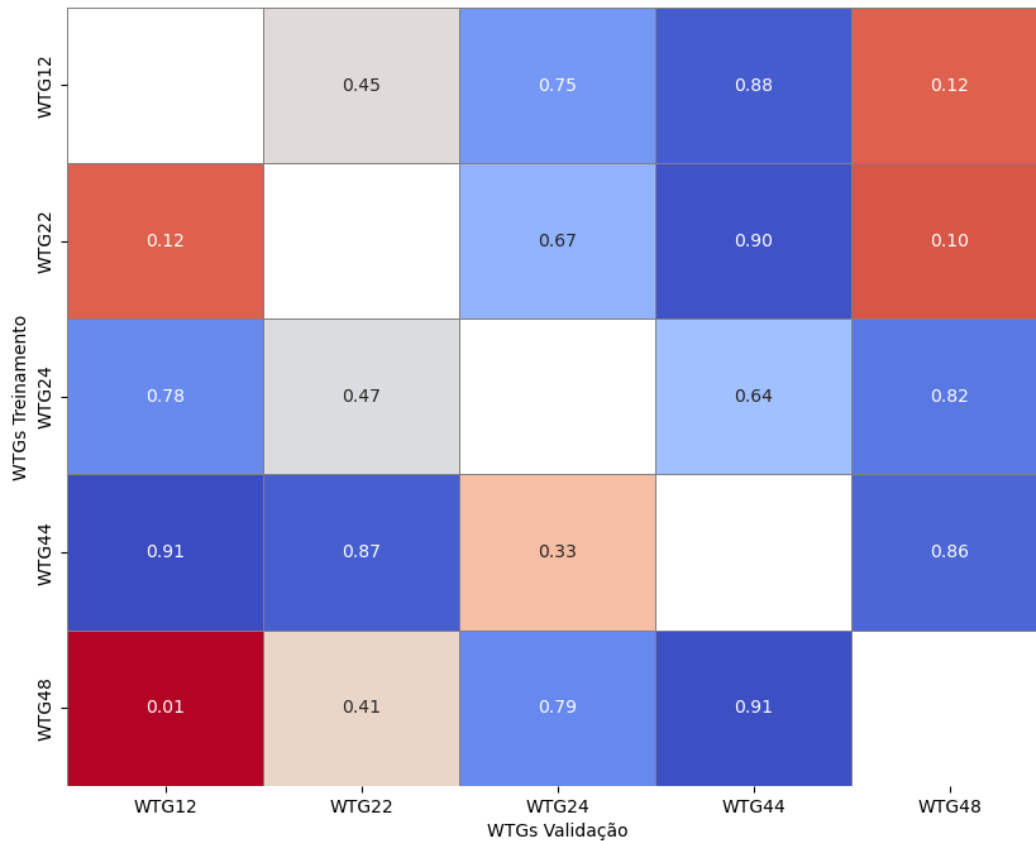


Figura 4.10: Resultados de FPR para modelo PCA considerando amostrais com categoria Normal

visa não apenas melhorar a precisão do modelo, mas também aumentar sua robustez e adaptabilidade a diferentes cenários operacionais. O pipeline proposto para essa abordagem está detalhadamente ilustrado na Figura 3.11, onde cada etapa do processo é descrita, desde a coleta e preparação dos dados até a fase de validação e avaliação do desempenho do classificador. Essa figura serve como um guia visual para entender a complexidade e a abrangência do método implementado, destacando a importância de cada componente no sucesso do modelo de classificação.

4.6

Discussão dos resultados obtidos

Baseado no mesmo formato de análise realizado nas seções anteriores, é possível avaliar os resultados obtidos pelo classificador treinado com amostras de duas turbinas distintas, conforme ilustrado nas Figuras 4.11 e 4.12 para os modelos puros e com PCA, respectivamente. Inicialmente, destaca-se que os resultados mais elevados de acurácia estão associados às turbinas WTG12, WTG22 e WTG48, seguindo o padrão observado no experimento anterior. Esse comportamento sugere uma consistência na capacidade do classificador

em identificar corretamente as características dessas turbinas específicas. Diferentemente do experimento anterior, os resultados dos testes utilizando a ferramenta de PCA não se mostraram eficazes. De modo geral, os valores de acurácia foram inferiores quando comparados aos obtidos sem a utilização do PCA. Essa observação pode indicar que, para este conjunto específico de dados e configuração, a aplicação do PCA não contribuiu para a melhoria do desempenho do modelo. Isso ressalta a importância de avaliar criticamente a aplicação de técnicas de pré-processamento de dados, uma vez que sua eficácia pode variar significativamente dependendo do contexto e das características dos dados utilizados.

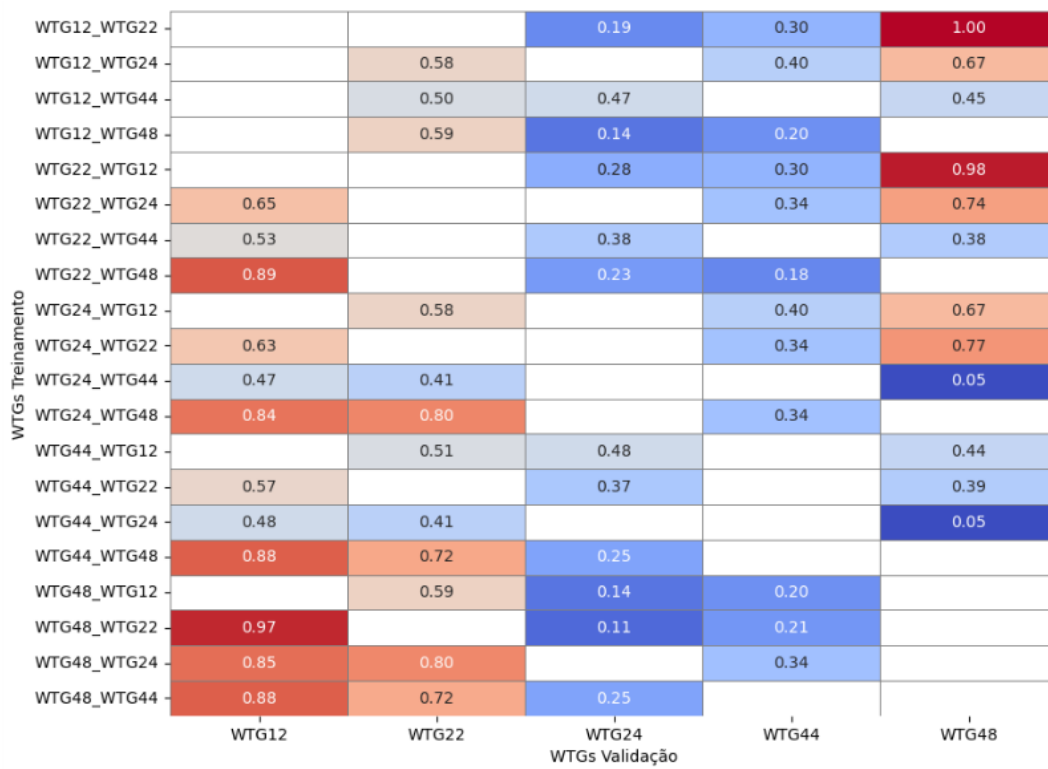


Figura 4.11: Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo Puro

Desde o início da exposição dos resultados desta pesquisa, a acurácia tem se destacado como o principal indicador de avaliação dos modelos e das hipóteses apresentadas. Entretanto, a avaliação adicional relacionada à razão percentual de falsos positivos (FPR) também é determinante para definir a validade do modelo em relação ao objetivo buscado. A escolha da avaliação de falsos positivos é crucial porque uma falsa sinalização de falha pode desviar o foco do gestor de outra turbina que realmente está em falha, tornando a gestão do ativo ineficiente. Por isso, a certeza da falha é muito importante. A minimização de falsos positivos garante que os recursos de manutenção sejam

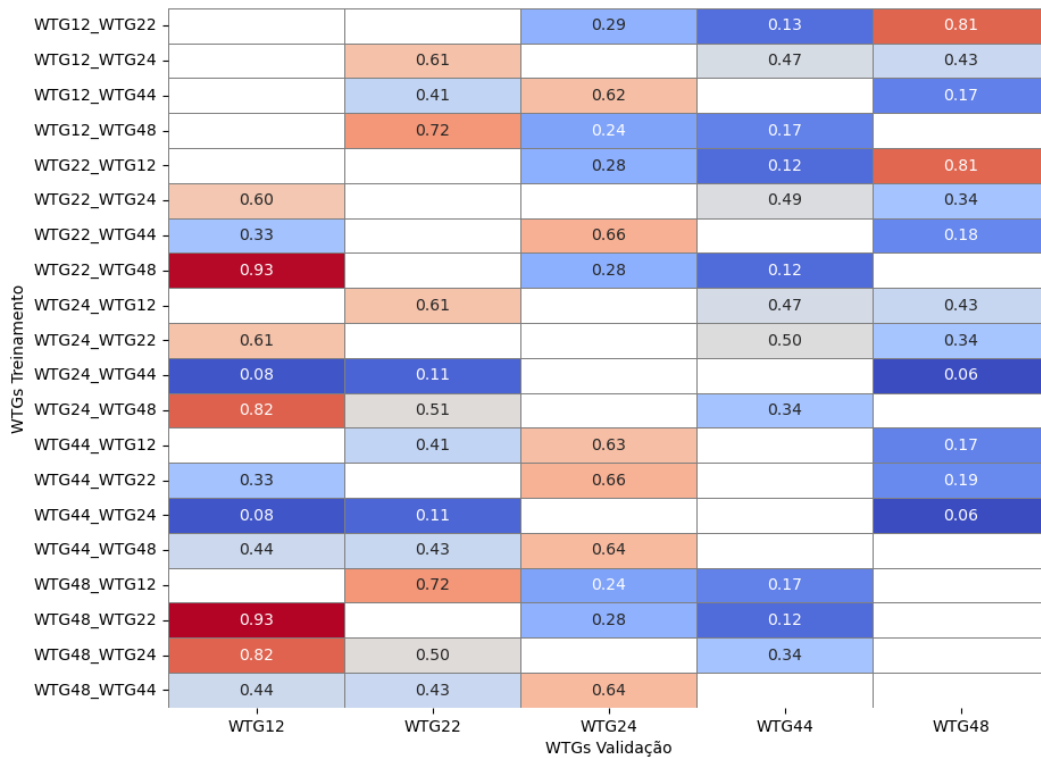


Figura 4.12: Mapa de Calor com valores de acuracidade por modelo avaliado - Modelo PCA

direcionados corretamente, evitando intervenções desnecessárias e assegurando que as turbinas que realmente necessitam de atenção sejam priorizadas.

A Figura 4.13 apresenta os indicadores de FPR relacionados aos mesmos testes realizados nos resultados da Figura 4.11, com foco principal no percentual de classificações erradas entre a operação da turbina com anomalias ou em condição normal. Em alguns casos, como nas classificações para validação da turbina WTG12, diversos modelos apresentaram acurácia inferior a 70 %, porém com valor de FPR zerado ou muito próximo de zero. Isso indica que, embora o modelo tenha tido dificuldade em determinar o nível de criticidade da anomalia, ele conseguiu identificar a presença de uma anomalia na operação da turbina.

Destaca-se uma mudança de comportamento nas classificações de validação da turbina WTG22, que apresentou desempenho positivo nos modelos treinados com as turbinas WTG12 e WTG48, mas não obteve nenhuma melhora nos modelos com dataset de treinamento combinado. Pelo contrário, houve uma degradação da performance. Essa degradação foi reforçada pela avaliação do indicador de FPR, demonstrando uma baixa capacidade de generalização desses modelos nessas condições. Em contrapartida, modelos treinados com a WTG22 obtiveram ótimos resultados nas validações com as demais turbinas.

Ao analisarmos os resultados de FPR relacionados aos experimentos trei-

WTGs Treinamento	WTG12_WTG22			1.00	0.67	0.00
	WTG12_WTG24		0.75		0.09	0.02
	WTG12_WTG44		0.70	0.88		0.53
	WTG12_WTG48		0.99	1.00	1.00	
	WTG22_WTG12			0.41	0.37	0.00
	WTG22_WTG24	0.00			0.00	0.00
	WTG22_WTG44	0.41		0.77		0.83
	WTG22_WTG48	0.00		0.61	0.60	
	WTG24_WTG12		0.75		0.07	0.02
	WTG24_WTG22	0.00			0.00	0.00
	WTG24_WTG44	0.53	0.66			0.96
	WTG24_WTG48	0.00	0.44		0.13	
	WTG44_WTG12		0.68	0.87		0.52
	WTG44_WTG22	0.35		0.86		0.83
	WTG44_WTG24	0.51	0.65			0.96
	WTG44_WTG48	0.06	0.58	0.92		
	WTG48_WTG12		0.99	1.00	1.00	
	WTG48_WTG22	0.00		1.00	0.63	
	WTG48_WTG24	0.00	0.45		0.14	
	WTG48_WTG44	0.07	0.60	0.93		
		WTG12	WTG22	WTG24	WTG44	WTG48
		WTGs Validação				

Figura 4.13: Resultados de FPR para modelo puro considerando amostrais com categoria Normal

nados com o dataset após a aplicação da ferramenta de PCA, observamos que o desempenho é inferior quando comparado ao modelo puro e, especialmente, em relação aos testes realizados com apenas uma turbina no dataset de treino. Ainda assim, alguns conjuntos de treino apresentaram resultados dentro do esperado, principalmente aqueles que envolveram as turbinas WTG12, WTG22 e WTG48.

A PCA busca, através da transformação do espaço vetorial, identificar as principais componentes que carregam a maior variância possível do conjunto de dados. No entanto, a fusão de dois conjuntos de dados distintos pode introduzir uma quantidade significativa de ruído, o que dificulta a correta identificação das características relacionadas às anomalias estudadas. Esse ruído adicional pode mascarar as variações importantes que a PCA deveria capturar, resultando em uma performance inferior dos modelos treinados com esses dados.

Por outro lado, os conjuntos de treino que envolveram as turbinas WTG12, WTG22 e WTG48 conseguiram manter um desempenho satisfatório, indicando que, nesses casos específicos, a PCA foi capaz de extrair componentes relevantes que ajudaram na identificação das anomalias. Isso sugere que a eficácia da PCA pode variar significativamente dependendo da natureza dos dados e da combinação de turbinas utilizadas no treinamento.

Em resumo, enquanto a aplicação da PCA pode trazer benefícios na

redução dimensional e na captura de variâncias significativas, sua eficácia pode ser comprometida pela introdução de ruído ao combinar conjuntos de dados distintos. Portanto, é crucial considerar cuidadosamente a composição dos datasets de treino ao utilizar PCA para garantir que as características essenciais das anomalias não sejam obscurecidas.

WTGs Treinamento	WTG12_WTG22			0.71	0.84	0.08
	WTG12_WTG24		0.16		0.65	0.09
	WTG12_WTG44		0.73	0.61		0.67
	WTG12_WTG48		0.51	0.88	0.90	
	WTG22_WTG12			0.71	0.85	0.08
	WTG22_WTG24	0.37			0.54	0.35
	WTG22_WTG44	0.65		0.43		0.70
	WTG22_WTG48	0.03		0.71	0.84	
	WTG24_WTG12		0.16		0.65	0.09
	WTG24_WTG22	0.37			0.54	0.35
	WTG24_WTG44	0.82	0.73			0.82
	WTG24_WTG48	0.04	0.06		0.59	
	WTG44_WTG12		0.73	0.61		0.67
	WTG44_WTG22	0.65		0.43		0.70
	WTG44_WTG24	0.81	0.72			0.82
	WTG44_WTG48	0.57	0.55	0.42		
	WTG48_WTG12		0.52	0.89	0.90	
	WTG48_WTG22	0.03		0.72	0.85	
	WTG48_WTG24	0.05	0.06		0.59	
	WTG48_WTG44	0.57	0.55	0.42		
		WTG12	WTG22	WTG24	WTG44	WTG48
		WTGs Validação				

Figura 4.14: Resultados de FPR para modelo PCA considerando amostrais com categoria Normal

5

Considerações Finais

Este capítulo finaliza o estudo, oferecendo uma visão geral dos principais resultados e contribuições. Primeiramente, serão destacados os achados mais significativos e suas implicações para a detecção de anomalias em turbinas eólicas. Em seguida, serão discutidas as limitações encontradas ao longo da pesquisa, proporcionando uma análise crítica dos desafios enfrentados. Por fim, serão apresentadas sugestões para pesquisas futuras, indicando direções promissoras para o aprofundamento dos estudos nesta área.

5.1

Recapitulação dos principais resultados e contribuições do estudo

É fundamental destacar o cumprimento do objetivo principal deste trabalho, que foi desenvolver uma ferramenta capaz de auxiliar na gestão de ativos eólicos. Esta ferramenta visa identificar turbinas com anomalias durante a operação e direcionar as equipes de operação e manutenção para atuar nessas turbinas antes que ocorra uma falha catastrófica. Tal abordagem não apenas resultará em uma significativa economia financeira, relacionada aos custos de ferramentas, horas de trabalho, engenharia e sobressalentes, mas também proporcionará uma extensão da vida útil dos equipamentos. Isso, por sua vez, impulsionará o retorno dos projetos e oferecerá uma grande contribuição para a sustentabilidade no futuro.

Este estudo apresentou uma análise de proposta de aplicação da PCA em conjunto com um modelo de classificação SVM para a detecção de anomalias nos rolamentos dos geradores de turbinas eólicas em comparação ao mesmo classificador treinado sem a ferramenta de redução de variáveis, utilizando dados de vibração do sistema de monitoramento de condição (CMS). Foram realizados experimentos com cinco turbinas diferentes, incluindo testes onde uma turbina foi utilizada para treinamento e outra para validação, além de experimentos onde os conjuntos de dados de duas turbinas foram combinados para o treinamento e posteriormente validados em uma terceira turbina. Estas turbinas foram selecionadas após análise exploratória dos dados disponíveis além da consideração das informações compartilhada pelo fabricante das turbinas através de laudos técnicos após as falhas confirmadas.

Os principais resultados indicam que a aplicação de PCA pode, em alguns casos, melhorar a capacidade do modelo de identificar anomalias, especialmente quando os dados de treinamento são limitados a uma única turbina como

apresentado na comparação dos cenários onde modelos que tinham baixa performance (abaixo de 80%) melhoram alguns pontos percentuais. Vale destacar que de fato, o modelo puro, na maior parte das vezes teve performance superior ao PCA aplicado, mantendo na maior parte dos casos resultados acima de 90% para turbinas sem falhas secundárias que impactaram a identificação.

Paralelamente, a combinação de conjuntos de dados de diferentes turbinas para o treinamento apresentou desafios adicionais que pode dificultar a correta identificação das características relacionadas às anomalias. Em particular, os modelos treinados com dados combinados mostraram uma performance inferior em comparação aos modelos treinados com dados de uma única turbina. Por outro lado, é muito importante destacar a simplicidade do modelo com aplicação do PCA em comparação ao modelo puro, permitindo uma aplicação mais simples e leve, ao considerar a aplicação destes modelos de configuração em algum sistema embarcado para alarmes.

De qualquer modo, a aplicação do pipeline de dados proposto mostrou-se promissor e eficaz sob determinadas condições para auxiliar a gestão de ativos durante sua operação e manutenção apontando as turbinas que possuem maior desvio de condição operacional dentro do portfólio avaliado.

As contribuições deste estudo incluem a demonstração da eficácia e das limitações da PCA na redução dimensional e na captura de variâncias significativas em dados de vibração de turbinas eólicas. Além disso, o estudo destaca a importância de considerar a composição dos datasets de treinamento ao aplicar técnicas de redução dimensional, para garantir que as características essenciais das anomalias não sejam obscurecidas.

Em conclusão, este estudo contribui para a compreensão da aplicação de PCA e SVM na detecção de anomalias em turbinas eólicas, destacando tanto os benefícios quanto as limitações dessas abordagens. Entretanto, a principal contribuição do trabalho não é relacionado as ferramentas utilizadas e sim com relação ao pipeline proposto, uma vez que isto pode direcionar a equipe de engenharia de operação e manutenção de turbinas eólicas para uma atuação antecipada e mais ágil na correção de anomalias encontradas permitindo assim uma melhor visão e uma melhor medida de risco associado. Por isso, futuros trabalhos devem continuar a explorar e aprimorar essas técnicas, visando a criação de modelos mais robustos e precisos para a manutenção preditiva de turbinas eólicas.

5.2

Limitações e sugestões para trabalhos futuros

Apesar dos resultados promissores, este estudo apresenta algumas limitações. Uma das principais é a dificuldade de identificação quando existem anomalias secundárias, como falhas de sensores ou anomalias em outros componentes que possam influenciar nas medições e que não foram consideradas durante o treinamento do modelo. Essas limitações sugerem que futuros trabalhos devem considerar de modo intencional a inclusão de dados adicionais que possam capturar essas variáveis, melhorando assim a robustez e a generalização dos modelos.

Atualmente, uma limitação que também se apresenta como uma oportunidade para estudos futuros é o fato de que a combinação de conjuntos de dados de diferentes turbinas na aplicação do PCA não produziu resultados satisfatórios, comprometendo a escalabilidade do modelo. Portanto, trabalhos futuros poderiam explorar melhor as componentes escolhidas e principalmente a formação dos datasets de treino, propondo diferentes premissas e até mesmo datasets mais completos ou com mais turbinas. Outras propostas de estudo consistem em avaliar técnicas diferentes de pré-processamento de dados para mitigar essa limitação, bem como investigar outras abordagens de redução dimensional que possam ser mais robustas a essas variações.

Além disso, a aplicação de técnicas de aprendizado profundo poderia ser explorada para melhorar a detecção de anomalias em dados de vibração de turbinas. Essas técnicas apresentam bons resultados em diversas áreas de detecção de falhas e poderiam oferecer melhorias significativas em relação aos métodos tradicionais de classificação, principalmente com a quantidade de variáveis e informações disponíveis.

Referências bibliográficas

ABADIE, L. M.; GOICOECHEA, N. Old wind farm life extension vs. full repowering: a review of economic issues and a stochastic application for Spain. **Energies**, v. 14, n. 12, p. 3678, 2021.

ABDULLAH, D. M.; ABDULAZEEZ, A. M. Machine learning applications based on svm classification a review. **Qubahan Academic Journal**, v. 1, n. 2, p. 81–90, 2021.

ABEEÓLICA. **Relatório Anual 2023**. [S.l.], 2023. Disponível em: https://abeeolica.org.br/wp-content/uploads/2024/11/424_ABEEOLICA_RELATORIO-ANUAL_2023_DIGITAL_V3.pdf.

BADUREK, C. A. **wind turbine**. 2024. Encyclopedia Britannica. Accessed: 2024-09-08.

BATISTA, G. E. A. P. A.; PRATI, R. C.; MONARD, M. C. A study of the behavior of several methods for balancing machine learning training data. **ACM SIGKDD Explorations Newsletter**, v. 6, n. 1, p. 20–29, 2004.

BERK, R. A. Support vector machines. In: _____. **Statistical Learning from a Regression Perspective**. [S.l.: s.n.], 2008. p. 1–28.

CARTELLI, F. **A Novel Framework for Condition-Based Maintenance and Performance Analysis Using Data-Driven Approaches**. Tese (Doutorado) — Politecnico di Torino, 2023.

CCEE, C. de Comercialização de E. E. **Nota Técnica NT CCEE - 0046/2020: Apresentação do Relatório de Custo das Fontes Incentivadas**. São Paulo, Brasil, 2020.

CHANG, C. C. W. et al. Recent advancements in condition monitoring systems for wind turbines: A review. **Energy Reports**, v. 9, p. 22–27, 2023.

CHANG, C. C. W. et al. Recent advancements in condition monitoring systems for wind turbines: A review. **Energy Reports**, v. 9, p. 22–27, 2023.

CHAPLOT, A.; CHOUDHARY, N.; JAIN, K. A review on data level approaches for managing imbalanced classification problem. **International Journal of Data Science**, v. 6, n. 2, p. 91–97, 2019.

CHEN, H. et al. Early fault warning method of wind turbine main transmission system based on SCADA and CMS data. **Machines**, v. 10, n. 11, p. 1018, 2022.

CHOU, C.-H.; KUO, B.-H.; CHANG, F. The generalized condensed nearest neighbor rule as a data reduction method. In: **18th International Conference on Pattern Recognition (ICPR'06)**. Hong Kong, China: [s.n.], 2006. p. 556–559.

- DAO, C.; KAZEMTABRIZI, B.; CRABTREE, C. Wind turbine reliability data review and impacts on levelised cost of energy. **Wind Energy**, v. 22, p. 1848–1871, 2019.
- ENEVOLDSEN, P.; XYDIS, G. Examining the trends of 35 years growth of key wind turbine components. **Energy for Sustainable Development**, v. 50, p. 18–26, 2019.
- GIPE, P.; MÖLLERSTRÖM, E. An overview of the history of wind turbine development: Part i—the early wind turbines until the 1960s. **Wind Engineering**, v. 46, n. 6, p. 1973–2004, 2022.
- GIPE, P.; MÖLLERSTRÖM, E. An overview of the history of wind turbine development: Part ii—the 1970s onward. **Wind Engineering**, v. 47, n. 1, p. 220–248, 2023.
- GÓES, M. et al. Wind power projects in brazil: challenges and opportunities increasing co-benefits and implications for climate and energy policies. **Environment, Development and Sustainability**, v. 23, p. 15341–15367, 2021.
- HAND, D. J.; BATCHELORT, B. G. Experiments on the edited condensed nearest neighbor rule. **Information Sciences**, v. 14, p. 171–180, 1978.
- HASAN, B. M. S.; ABDULAZEEZ, A. M. A review of principal component analysis algorithm for dimensionality reduction. **Journal of Soft Computing and Data Mining**, v. 2, n. 1, p. 20–30, 2021.
- HASSAN, Q. et al. The renewable energy role in the global energy transformations. **Renewable Energy Focus**, v. 48, p. 100545, 2024.
- HIRA, Z. M.; GILLIES, D. F. A review of feature selection and feature extraction methods applied on microarray data. **Advances in Bioinformatics**, v. 2015, n. 1, p. 1–13, 2015.
- HYERS, R. et al. Condition monitoring and prognosis of utility scale wind turbines. **Energy materials**, v. 1, n. 3, p. 187–203, 2006.
- IGBA, J. et al. Effect of preventive maintenance intervals on reliability and maintenance costs of wind turbine gearboxes. **Wind Energy**, v. 18, n. 11, p. 2013–2024, 2015.
- JANKAUSKAS, M. et al. Exploring the limits of early predictive maintenance in wind turbines applying an anomaly detection technique. **Sensors**, v. 23, n. 12, p. 5695, 2023.
- KAVZOGLU, T.; COLKESEN, I. A kernel functions analysis for support vector machines for land cover classification. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, p. 352–359, 2009.
- KONG, K. et al. Progress and trends in damage detection methods, maintenance, and data-driven monitoring of wind turbine blades—a review. **Renewable Energy Focus**, v. 44, p. 390–412, 2023.

- LEVITT, J. **The handbook of maintenance management**. [S.l.]: Industrial Press Inc., 2009.
- MINH, N. Q.; KHIEM, N. T.; GIANG, V. H. Fault classification and localization in power transmission line based on machine learning and combined cnn-lstm models. **Energy Reports**, v. 12, p. 5610–5622, 2024.
- NEJAD, A. R. et al. Wind turbine drivetrains: state-of-the-art technologies and future development trends. **Wind Energy Science**, v. 7, n. 1, p. 387–411, 2022.
- PANDIT, R. et al. Scada data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. **Wind Engineering**, v. 47, n. 2, p. 422–441, 2023.
- PENG, H. et al. Analysis of wind turbine equipment failure and intelligent operation and maintenance research. **Sustainability**, v. 15, n. 10, p. 8333, 2023.
- PENG, J. et al. Wind turbine failure prediction and health assessment based on adaptive maximum mean discrepancy. **International Journal of Electrical Power & Energy Systems**, v. 134, p. 107391, 2022.
- QIN, S.; WANG, R.; FU, D. Research on wind turbines fault diagnosis technology based on cms data feature extraction. **Wireless Personal Communications**, v. 127, p. 1–21, 2022.
- SANTELO, T. N. et al. Wind turbine failures review and trends. **Journal of Control, Automation and Electrical Systems**, p. 1–17, 2022.
- SUMMERVILLE, B. **NREL Technical Report - Distributed Wind Monitoring Best Practices**. Golden, CO (United States), 2024. 6-28 p.
- UDO, W. S. et al. Optimizing wind energy systems using machine learning for predictive maintenance and efficiency enhancement. **Journal of Renewable Energy Technology**, v. 28, n. 3, p. 312–330, 2024.
- VELMURUGAN, R. S.; DHINGRA, T. Maintenance strategy, practices and spare parts management. In: **Asset Maintenance Management in Industry: A Comprehensive Guide to Strategies, Practices and Benchmarking**. [S.l.: s.n.], 2021. p. 61–82.
- VESELY, E. Consider automated machine learning for wind turbine asset maintenance. **Power**, v. 161, n. 9, p. 1–15, 2017.
- WEN, X. Modeling and performance evaluation of wind turbine based on ant colony optimization extreme learning machine. **Applied Soft Computing Journal**, v. 94, 2020.
- ZHANG, J. et al. A one class shapelet dictionary learning method for wind turbine bearing anomaly detection. **Measurement**, v. 197, p. 111318, 2022.

A

Exemplo Relatório Fabricante Aerogerador

CMS Alarm Report

TURBINE DETAILS

Unit ID	[REDACTED]	CMS Vendor	[REDACTED]
Unit Location	WTG12	Report ID	[REDACTED]
Park	[REDACTED]	Linked Report ID	[REDACTED]
Short Unit Type	[REDACTED]	Issued By	[REDACTED]
SBU	LAT	Alarm Type	[REDACTED]

COMPONENTS

Gearbox	[REDACTED]	MnBrgFr	[REDACTED]
Generator	[REDACTED]	MnBrgRr	[REDACTED]

ALARM INDICATIONS

Severity	2 ■ Considerable progression
Estimated Lead Time (days)	7
Component	GnDe
Alarm Indications	GnDe.ISOA,GnDe.ECU[3k-10k,1-360],GnDe.BP[300-700]
VGA	

MAINTENANCE ASSESSMENT

It is recommended to replace the Generator Drive end bearing. If required, a test run can be performed to listen for abnormal noise from generator bearings, to access the severity, and plan for the bearing replacement accordingly.
 Also, ensure generator bearings are properly lubricated, if not, re-lubricate the generator bearings, to check if the noise reduces after lubrication.

[REDACTED]

OBSERVATION

Gradual progression noticed in vibration trends BP(300-700), ISOA in Generator DE.

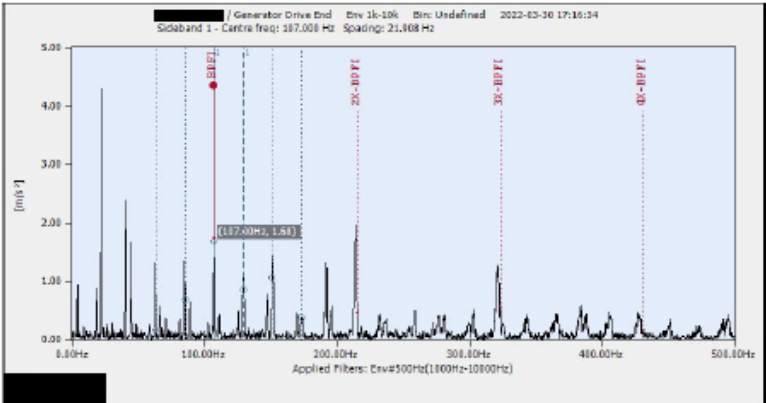
46928-WTG12 GnDe.ISOA Bin 4

2022-04-04 11:50:00, 7.14

CMS
Alarm Report

INTERPRETATION

FFT Spectral analysis reveals inner race bearing fault frequency 107.0Hz and its harmonics at generator DE which indicates that fault is associated with the inner race of bearing.



ALARM CLASSIFICATION

Severity	Description	Recommended Action (For lead time to action Refer above estimated lead time)
1 Red	Severe progression	Immediate action recommended
2 Orange	Considerable progression	Prioritized action recommended
3 Yellow	Small progression	Action recommended
4 Green	OK	OK - Alarm closing report
6 Blue	System	Hardware problem - Prioritized action recommended