UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

JANUÁRIO LEAL DE MORAES VIEIRA

FRAMEWORK PARA PROGNÓSTICO DE FALHAS DE AEROGERADORES: Uma proposta para uso eficaz de dados reais de operação

Recife
Agosto de 2024

JANUÁRIO LEAL DE MORAES VIEIRA

FRAMEWORK PARA PROGNÓSTICO DE FALHAS EM ROLAMENTOS PRINCIPAIS DE AEROGERADORES: Uma proposta para uso eficaz de dados reais de operação

Tese apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal de Pernambuco, como requisito para obtenção do título de doutor em Engenharia Mecânica. Área de concentração: Energia

Orientador (a): Álvaro Antonio Ochoa Villa
Coorientador (a): Gustavo de Novaes Pires Leite

Recife
Agosto de 2024

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Vieira, Januario Leal de Moraes.
Framework para prognóstico de falhas de aerogeradores: uma proposta para uso eficazde dados reais de operação / Januario Leal de Moraes Vieira. - Recife, 2024.
153f.: il.

Tese (Doutorado) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Programa de Pós-Graduação em Engenharia Mecânica, 2024.
Orientação: Álvaro Antonio Ochoa Villa.
Coorientação: Gustavo de Novaes Pires Leite.
Inclui referências e apêndices.

1. Aerogerador; 2. Transmissão; 3. Manutenção baseada na condição; 4. Vida útil remanescente; 5. Modelos de aprendizado de máquina; 6. Dados SCADA. I. Villa, Álvaro Antonio Ochoa. II. Leite, Gustavo de Novaes Pires. III. Título.

UFPE-Biblioteca Central

Ata da 116ª (centésima décima sexta) defesa de Tese de Doutorado do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco, no dia 29 de agosto de 2024.

Ao vigésimo nono dia do mês de agosto de dois mil e vinte e quatro, às oito horas, em sessão pública, teve início a defesa da Tese intitulada "*Framework para Prognóstico de Falhas de Aerogeradores: Uma proposta Para Uso Eficaz de Dados Reais de Operação*" do doutorando Januário Leal de Moraes Vieira, na área de concentração Energia sob a orientação do Professor Alvaro Antonio Ochoa Villa e coorientação do Professor Gustavo de Novaes Pires Leite. A Comissão Examinadora foi indicada pelo colegiado do programa de pós-graduação em 16 de agosto de 2024, na sua 8ª reunião ordinária, sendo composta pelos Professores: Alvaro Antonio Ochoa Villa, do Instituto Federal de Pernambuco; Isis Didier Lins, do Departamento de Engenharia de Produção da Universidade Federal de Pernambuco; Leandro Maciel Almeida, do Centro de Informática da Universidade Federal de Pernambuco; Pedro André Carvalho Rosas, do Departamento de Engenharia Elétrica e Sistemas de Potência da Universidade Federal de Pernambuco; e Tsang Ing Ren, do Centro de Informática da Universidade Federal de Pernambuco. Após cumpridas as formalidades, o candidato ao grau de Doutor foi convidado a discorrer sobre o conteúdo da Tese. Concluída a explanação, o candidato foi arguido pela Comissão Examinadora que, em seguida, reuniu-se para deliberar e conceder, ao mesmo, a menção **APROVADO**. Para a obtenção do grau de Doutor em Engenharia Mecânica, o concluinte deverá depositar versão final da Tese na Biblioteca Central. O depósito deverá ser feito, após realizadas as correções conforme recomendações feitas, nesta sessão, pela comissão examinadora, observando os prazos e procedimentos vigentes nas normas internas desta universidade. E, para constar, lavrei a presente Ata que vai por mim assinada, Jorge Rodrigo da Silva Oliveira, e pelos membros da Comissão Examinadora.

OBSERVAÇÃO

A defesa em epígrafe foi realizada integralmente, por videoconferência, envolvendo a Banca Examinadora e o (a) discente, através de recursos de videoconferência, que possibilitaram realizar a discussão acadêmica sobre o objeto de estudo, com som e imagem.

Recife, 29 de agosto de 2024.

_____

Jorge Rodrigo da Silva Oliveira

COMISSÃO EXAMINADORA

Prof. Alvaro Antonio Ochoa Villa
Profª Isis Didier Lins
Prof. Leandro Maciel Almeida
Prof. Pedro André Carvalho Rosas
Prof. Tsang Ing Ren

Documento assinado digitalmente
gov.br
ISIS DIDIER LINS
Data: 05/11/2024 12:03:23-0300
Verifique em https://validar.iti.gov.br

Documento assinado digitalmente
gov.br
ALVARO ANTONIO OCHOA VILLA
Data: 05/11/2024 18:10:12-0300
Verifique em https://validar.iti.gov.br

CANDIDATO AO GRAU DE DOUTOR
Januário Leal de Moraes Vieira

À Camilla, minha amada esposa e eterna companheira, ao Pedro e Francisco, meus filhos amados e fiéis escudeiros e aos meus pais, Angela e Cristovam.

AGRADECIMENTOS

A Deus, por sua presença silenciosa e constante, nos momentos de luz e escuridão, e por tudo que encontrei durante esta caminhada.

A minha mãe e meu pai, pelos valores ensinados, pela dedicação em minha formação e pelo esforço e apoio na minha educação.

A minha amada esposa Camilla e aos meus amados filhos, Pedro e Francisco, pelo apoio incondicional, pela compreensão nas ausências e por acreditarem na minha capacidade de vencer essa jornada.

Ao Prof. Dr. Álvaro Antônio Ochoa Villa e ao Prof. Dr. Gustavo de Novaes Pires Leite por terem me oferecido a oportunidade de realizar este trabalho e por terem me orientado de forma humana e empática durante todo o processo.

Às amizades que a vida me deu, que estiveram sempre presentes com incentivo e força, em especial a Jorge Antônio Chaves de Oliveira e Myrna Loreto.

Aos colegas do IFPE, Felipe Farias, Frederico Duarte de Menezes e José Ângelo Costa pelo apoio, contribuição e estímulo no desenvolvimento deste trabalho.

Aos colegas da Coordenação de Desenho do IFPE, campus Recife, pelo apoio que possibilitou que eu me afastasse das minhas atividades docentes. Especialmente a Alex Elton de Moura, Andréa Gonçalves Sousa, Gustavo Costa e Keylla Costa pelo incentivo e escuta nos momentos difíceis desse processo.

Ao IFPE pela concessão do meu afastamento para a finalização deste trabalho.

A Multiner S.A pelo fornecimento dos dados do aerogerador, da usina eólica e do sistema SCADA utilizados neste trabalho.

A Caio Munguba, colega da pós-graduação pela amizade, debates técnicos, incentivo e colaboração no desenvolvimento da parte computacional.

Aos colegas do projeto NEOPROG.FALHAS do Centro de Energias Renováveis da UFPE pelas discussões, provocações e contribuições nos seminários do projeto que ajudaram no desenvolvimento deste trabalho, especialmente o Prof. Dr. Alexandre Carlos Araújo da Costa e a Prof.ª. Dra. Olga de Castro Vilela.

Aos meus orientadores de outrora na graduação na UFPE, Prof. Mestre José Laurênio Accioly e no mestrado na Escola de Engenharia de São Carlos, Prof. Dr. Álvaro Cota Neto, pela escuta e incentivo que me motivaram a vencer essa etapa.

Aos professores e funcionários do Programa de Pós-Graduação em Engenharia Mecânica da UFPE pelos conhecimentos transmitidos e atendimento prestado.

Computadores fazem arte
Artistas fazem dinheiro
Computadores avançam
Artistas pegam carona
Cientistas criam o novo
Artistas levam a fama
(Chico Science e Nação Zumbi, “Computadores Fazem Arte”. In: *Da lama Ao Caos*, 1994)

RESUMO

Este trabalho aborda a lacuna significativa no prognóstico de falhas em aerogeradores, especialmente offshore, propondo um framework inovador baseado em uma abordagem de complexidade crescente para garantir a generalização de modelos de aprendizagem de máquina. Utilizando dados sintéticos, benchmarks públicos e dados reais de temperatura do sistema SCADA de três turbinas no nordeste do Brasil, foram desenvolvidos modelos regressivos (*support vector machine, gradient boosting, random forest* e *extra trees*) para estimar a Vida Útil Remanescente (VUR) dos rolamentos principais. Os modelos apresentaram erros médios de 20 dias na estimativa da VUR, com MAE de 0,047, MSE de 0,012 e R2 Score de 0,625. O framework proposto para estimar a VUR com dados reais de operação tem potencial para impactar a cultura organizacional em relação às tecnologias de Inteligência Artificial Industrial, oferecendo estimativas conservadoras e assertivas para subsidiar o planejamento de manutenção, evitar falhas catastróficas e aumentar a disponibilidade das usinas. Esta metodologia preenche uma lacuna importante na manutenção preditiva de aerogeradores, apresentando uma solução robusta e adaptável com implicações significativas para a indústria de energia eólica.

Palavras-chave: aerogerador; transmissão, manutenção baseada na condição, vida útil remanescente, modelagem, modelos de aprendizado de máquina, generalização, dados SCADA.

ABSTRACT

This work addresses the significant gap in wind turbine failure prognosis, especially offshore, by proposing an innovative framework based on an increasing complexity approach to ensure the generalization of machine learning models. Using synthetic data, public benchmarks, and real temperature data from the SCADA system of three turbines in northeastern Brazil, regressive models (support vector machine, gradient boosting, random forest, and extra trees) were developed to estimate the Remaining Useful Life (RUL) of the main bearings. The models presented average errors of 20 days in the RUL estimate, with MAE of 0.047, MSE of 0.012, and R2 Score of 0.625. The proposed framework to estimate RUL with real operating data has the potential to impact the organizational culture regarding Industrial Artificial Intelligence technologies, offering conservative and assertive estimates to support maintenance planning, avoid catastrophic failures, and increase plant availability. This methodology fills an important gap in predictive maintenance of wind turbines, presenting a robust and adaptable solution with significant implications for the wind energy industry.

Keywords: wind turbine; drivetrain, condition-based maintenance, remaining useful life, modeling, machine learning models, generalization, SCADA data.

LISTA DE ILUSTRAÇÕES

Figura 1 – Previsão de aumento de capacidade instalada em energia eólica (2024-2028) em valores estimados, marcados com *e*, após o ano. 20
Figura 2 – Capacidade instalada nos 31 países maiores produtores de energia eólica..... 21
Figura 3 – Subsistemas de um programa de MBC 23
Figura 4 – Componentes principais de um aerogerador de eixo horizontal. 24
Figura 5 – Média da velocidade do vento na cidade de Natal/RN (período de 1984 – 2012).. 25
Figura 6 – Média horária da velocidade do vento, representando as Plataformas de coleta de Dados (PCDs) do Litoral..... 26
Figura 7 – Principais categorias de modelos para predição da vida útil remanescente..... 36
Figura 8 – Contagem de registros de publicações de diferentes bases de dados científicas para *strings* de pesquisa com *failure* AND *prognostics* OR *prognosi** e adição do nome da indústria para todos os períodos disponíveis nas bases de dados..... 37
Figura 9 – Contagem de publicações dos últimos 5 anos de diferentes bases de dados científicas para *strings* de pesquisa com *failure* AND *prognostics* OR *prognosi** e adição do nome da indústria para todos os períodos disponíveis nas bases de dados 38
Figura 10 – Número de registros de publicações relacionadas a prognósticos de falhas de todos os componentes ou subsistemas de aerogeradores na Engineering Village (COMPENDEX). 38
Figura 11 - Artigos de revisão sobre prognóstico de aerogeradores nos últimos 7 anos e lacuna identificada 43
Figura 12 – Fluxograma do protocolo PRISMA utilizado para seleção de registros para revisão 45
Figura 13 – Número de registros de publicações selecionados por ano. 49
Figura 14 - Os 9 principais periódicos incluídos na pesquisa em bases de dados de generalização de modelos prognósticos antes da triagem. 50
Figura 15 – Classificação de tipos de documentos baseada em bases de dados científicas. 50
Figura 16 – Uso de conjuntos de dados abertos (públicos) utilizados em estudos encontrados em artigos. 51
Figura 17 – Fontes de conjuntos de dados abertos encontrados em artigos. 52
Figura 18 – Uso de conjuntos de dados de execução até falha no estudo. 53
Figura 19 – Porcentagem de uso de pré-processamento em conjuntos de dados de origem. ... 54
Figura 20 – Grupos de tipos de pré-processamento utilizados. 54

Figura 21 – Tipos de modelos para prognóstico de falhas de aerogeradores identificados nos artigos selecionados 55
Figura 22 – Modelos baseados em dados mais comuns para estimativa de VUR para prognósticos de aerogeradores. 56
Figura 23 – Modelos baseados em dados para estimativa de VUR para prognósticos de aerogeradores organizados por ano. 57
Figura 24 – Estimativa da VUR usando dados sobreajustados 58
Figura 25 – Estimativa da VUR usando dados superajustados com classificação SCImago... 58
Figura 26 – Intervalo de confiança ou cálculo de incerteza da estimativa da VUR com classificação SCImago 59
Figura 27 – Capacidade de generalização de modelos baseados em critérios desenvolvidos.. 61
Figura 28 – Capacidade de generalização de modelos baseados em critérios de desenvolvimento versus classificação de periódicos SCImago. 61
Figura 29 – Fonte de conjuntos de dados usados para cálculo da VUR por diferentes modelos baseados em dados. 63
Figura 30 – Cálculo da incerteza considerando a classificação do modelo..... 63
Figura 31 – Capacidade de generalização de diferentes modelos baseados em dados para computação da VUR. 65
Figura 32 – Os modos de falha foram identificados nos artigos selecionados..... 66
Figura 33 - Número de registros de publicações relacionadas a prognósticos de falhas de todos os componentes ou subsistemas e prognósticos de falhas do rolamento principal dos aerogeradores da Engineering Village (COMPENDEX). 71
Figura 34 - Diagrama do uso de diferentes bases de dados com aumento gradual da complexidade 76
Figura 35 – Série temporal do conjunto de dados sintéticos 78
Figura 36 – Anotação dos dados entre o intervalo 1 (estado saudável) até zero (falha do componente) para estimativa da VUR. 79
Figura 37 - Sinais brutos de vibração de rolamento até a falha na pista interna 80
Figura 38 -Características estatísticas do sinal adquirido por medições intensivas Mathworks/GPMS. (a) Média, (b) Desvio-padrão, (c) RMS, (d) Pico-a-Pico, (e) Assimetria (Skewness), (f) Curtose (Kurtosis) e (g) Fator de crista 83
Figura 39 - Série temporal do conjunto de dados Mathworks/GPMS e anotação do conjunto de dados para estimativa da VUR 84
Figura 40 – Pré-processamento do sinal..... 85

Figura 41 – Série de temperatura média dos rolamentos principais com valores discrepantes (*outliers*) (WT9, WT14 e WT29) 86
Figura 42 – Variação de temperatura do rolamento principal após remoção dos valores discrepantes (*outliers*) (WT9, WT14 e WT29) 86
Figura 43 – Variação de temperatura para o rolamento principal após decomposição sazonal (WT9, WT14 e WT29) 87
Figura 44 - Anotação dos dados estimados a partir das principais séries temporais de temperatura dos mancais para os aerogeradores: (a) WT9, (b) WT14 e (c) WT29 89
Figura 45 - Fluxograma de anotação, treinamento, validação, verificação do desempenho com uso de dados sintéticos e testes para estimativa da VUR. 94
Figura 46 – Exemplo de 4 divisões (splits) para validação cruzada usando a classe TimeSeriesSplit 95
Figura 47 - Fluxograma de anotação, treinamento, validação, verificação do desempenho com uso de dados de benchmark (Mathworks/GPMS) e testes para estimativa da VUR. 96
Figura 48 – Metodologia - Estimativa da VUR por modelos de aprendizado de máquina de regressão 98
Figura 49 - Métricas MSE, MAE e RMSE na etapa de teste dos modelos de regressão usando o conjunto de dados sintéticos elaborados 103
Figura 50 - Métrica R2 Score na etapa de teste dos modelos de regressão usando o conjunto de dados sintéticos 104
Figura 51 - Estimativa e VUR do modelo DTR utilizando conjunto de dados sintéticos 106
Figura 52 - Estimativa e VUR do modelo ETR utilizando conjunto de dados sintéticos 106
Figura 53 - Estimativa e VUR do modelo GBR utilizando conjunto de dados sintéticos 107
Figura 54 - Estimativa e VUR do modelo ISOR utilizando conjunto de dados sintéticos 107
Figura 55 - Estimativa e VUR do modelo RFR utilizando conjunto de dados sintéticos 108
Figura 56 - Estimativa e VUR do modelo SVR utilizando conjunto de dados sintéticos 108
Figura 57 - Métricas MSE, MAE e RMSE, na etapa de teste, dos modelos de regressão usando o conjunto de dados Mathworks/GPMS 109
Figura 58 - Métrica R2 Score, na etapa de teste, dos modelos de regressão usando o conjunto de dados Mathworks/GPMS 110
Figura 59 – Estimativa e VUR do rolamento do HSS utilizando modelo DTR 112
Figura 60 - Estimativa e VUR do rolamento do HSS utilizando modelo ETR 112
Figura 61 - Estimativa e VUR do rolamento do HSS utilizando modelo GBR 113
Figura 62 - Estimativa e VUR do rolamento do HSS utilizando modelo ISOR 113

Figura 63 - Estimativa e VUR do rolamento do HSS utilizando modelo RFR 114
Figura 64 - Estimativa e VUR do rolamento do HSS utilizando modelo SVR 114
Figura 65 - Resultados de MAE, MSE e RMSE para cada modelo de regressão e cada caso de simulação: (a) Vetor de suporte, (b) Isotônica, (c) Aumento de gradiente, (d) Árvore de decisão (e) Árvores extras e (f) Floresta aleatória 116
Figura 66 - Resultados do R2 Score para cada modelo de regressão e cada caso de simulação: (a) Vetor de suporte, (b) Isotônica, (c) Aumento de gradiente, (d) Árvore de decisão (e) Árvores extras e (f) Floresta aleatória 117
Figura 67 - Resultados de simulações com modelos SVR (MB-SC02 e MB-SC05) 120
Figura 68 - Visão geral do framework 123
Figura 69 - Etapas detalhadas do framework. 124

LISTA DE TABELAS

Tabela 1 - Questões de pesquisa 44
Tabela 2 – Grupos de palavras-chave para encontrar registros relacionados à generalização de métodos ou modelos em bases de dados selecionadas 46
Tabela 3 - Grupos de palavras-chave para localizar registros relacionados à manutenção prescritiva em bases de dados selecionadas 46
Tabela 4 - Critérios de inclusão usados na etapa de triagem 46
Tabela 5 - Critérios de inclusão para etapa de elegibilidade 47
Tabela 6 - As vantagens e desvantagens dos métodos mais utilizados conforme revisão da literatura. 64
Tabela 7 - Considerações sobre problemas e oportunidades de melhoria para prognósticos de aerogeradores 70
Tabela 8 – Subconjuntos usados para lidar com escassez dados de aerogeradores 90
Tabela 9 – Lista de parâmetros e intervalos de parâmetros usados para etapa de validação cruzada 101
Tabela 10 - Valores máximos para as métricas MAE, MSE e RMSE, na etapa de teste, para o conjunto de dados sintéticos de temperatura 104
Tabela 11 - Valores máximos para a métrica do coeficiente de determinação (R2 Score), na etapa de teste, para o conjunto de dados sintéticos de temperatura 105
Tabela 12 - Valores máximos para as métricas MAE, MSE e RMSE, na etapa de teste, para o conjunto de dados GPMS 110
Tabela 13 - Valores máximos para a métrica do coeficiente de determinação (R2 Score), na etapa de teste, para o conjunto de dados GPMS 111
Tabela 14 - Resultados do erro absoluto médio (MAE) para todos os casos de simulação, para cada modelo, para etapas de treinamento, validação e teste 118
Tabela 15 - Resultados do erro quadrático médio (MSE) para todos os casos de simulação, para cada modelo, para as etapas de treinamento, validação e teste 118
Tabela 16 - Resultados do erro quadrático médio (RMSE) para todos os casos de simulação, para cada modelo, para etapas de treinamento, validação e teste 118
Tabela 17 - Resultados do coeficiente de determinação (R2 Score) para todos os casos de simulação, para cada modelo, para as etapas de treinamento, validação e teste 119
Tabela 18 - Resultados de simulação com subconjuntos de teste 120
Tabela 19 - Lacunas identificadas e propostas deste trabalho 126

LISTA DE ABREVIATURAS E SIGLAS

AE	Aerogerador	
AM	Aprendizagem de máquina	
ANN	Artificial neural network	Rede neural artificial
CAGR	Compound Annual Growth Rate	Taxa de crescimento anual composta
CMS	Condition monitoring system	Sistema de monitoramento da condição
DOI	Digital Object Identifier	Identificador de Objeto Digital
DT	Digital Twin	Gêmeo digital
EMD	Empirical Mode Decomposition	Decomposição do Modo Empírico
FAIR	Findable, Accessible, Interoperable and Reusable	Encontrável, acessível, interoperável e reutilizável
FPT	First prediction time	Instante de detecção da falha
FTT	Failure threshold time	Instante do limiar de falha
HSS	High speed shatf	Eixo de alta velocidade
IA	Inteligência artificial	
IAI	Inteligência artificial industrial	
kNN	k-Nearest Neighbours	k-Vizinhos mais próximos
MAE	Erro médio absoluto	
MBC	Manutenção baseada na condição	
ML	Machine Learning	
MSE	Erro quadrático médio	
O&M	Operação e Manutenção	
PCA	Principal component analysis	Análise do componente principal
PDF	Power density function	Função de densidade de energia
PHM	Prognostics and Health Management	Prognósticos e Gestão da Saúde
QP	Questão de pesquisa	
RBF	Radial Basis Function	Função de Base Radial

R2 Score	Coeficiente de determinação	
RMSE	Raiz do erro quadrático médio	
RNA	Redes neurais artificiais	
SCADA	Supervisory Control and Data Acquisition	Sistema de Supervisão e Aquisição de Dados
SHM	Structural Health Management	Gerenciamento da saúde estrutural
SVM	Support Vector Machine	Máquina de vetores de suporte
VUR	Vida útil remanescente	

SUMÁRIO

1. INTRODUÇÃO 19

1.1. O AEROGERADOR 23
1.2. INTELIGÊNCIA ARTIFICIAL EM APLICAÇÕES INDUSTRIAIS 26
1.3. PROGNÓSTICO DE FALHAS EM AEROGERADORES 28
1.4. OBJETIVOS 30
1.4.1. OBJETIVO GERAL 30
1.4.2. OBJETIVOS ESPECÍFICOS 30
1.5. CONTRIBUIÇÃO CIENTÍFICA 31
1.6. ORGANIZAÇÃO DA TESE 31

2. ESTADO DA ARTE 33

2.1. PROGNÓSTICO DE FALHAS 33
2.2. PROGNÓSTICO DE FALHAS EM MÁQUINAS EM GERAL 34
2.3. PROGNÓSTICO E GERENCIAMENTO DA SAÚDE EM DIFERENTES INDÚSTRIAS 37
2.4. IMPORTÂNCIA DA AVALIAÇÃO DOS MODELOS E IMPLICAÇÕES OPERACIONAIS DO PROGNÓSTICO DE FALHAS EM AEROGERADORES 39
2.5. VIDA ÚTIL REMANESCENTE E CONEXÕES COM A MANUTENÇÃO PRESCRITIVA DE AEROGERADORES 40
2.6. GENERALIZAÇÃO DE MÉTODOS PARA ESTIMATIVA DE VIDA ÚTIL REMANESCENTE: REVISÃO SISTEMÁTICA COM METODOLOGIA INSPIRADA NO PROTOCOLO PRISMA 44
2.6.1. COLETA DE DADOS 48
2.6.2. RESULTADOS DA REVISÃO SISTEMÁTICA 48
2.6.2.1. Métodos e modelos para estimar a vida útil remanescente (VUR) 53
2.6.2.2. Resultados dos modelos e estimativa da vida útil remanescente (VUR) 55
2.6.2.3. Cálculo da incerteza do modelo (ou nível de confiança) 59
2.6.2.4. Capacidade de generalização baseada na minimização de overfitting 60
2.6.3. DESCOBERTAS ACERCA DOS PROGNÓSTICOS DE FALHAS E OS GANHOS POTENCIAIS 65
2.7. PROGNÓSTICO DE FALHAS EM ROLAMENTOS PRINCIPAIS DE AEROGERADORES 70
2.8. CONCLUSÕES DO CAPÍTULO ESTADO DA ARTE 72

3. Fundamentação teórica 73

3.1. MODELOS DE APRENDIZAGEM DE MÁQUINA 73
3.2. MODELOS REGRESSIVOS DE APRENDIZAGEM DE MÁQUINA 73
3.2.1. MODELO DE VETOR DE SUPORTE (SUPPORT VECTOR MACHINE (SVM)) 74
3.2.2. MODELO ISOTÔNICO (ISOTONIC REGRESSION) 74
3.2.3. MODELOS DE ÁRVORES E CONJUNTOS (ENSEMBLES) 74
3.2.3.1. Árvore de decisão (Decision tree) 75
3.2.3.2. Aumento de gradiente (Gradient boosting) 75
3.2.3.3. Floresta aleatória (Random forest) 75

3.2.3.4. Árvores extras (Extra trees) 75

4. DESENVOLVIMENTO DE MODELO COMPUTACIONAL PARA ESTIMATIVA DA VIDA ÚTIL REMANESCENTE 75

4.1. DADOS SINTÉTICOS 77
4.2. DADOS GPMS – VIBRAÇÃO DO ROLAMENTO DE ALTA VELOCIDADE 79
4.3. DADOS SCADA – CONJUNTO DE DADOS DE TEMPERATURA DO ROLAMENTO PRINCIPAL 84
4.3.1. LEITURA E PRÉ-PROCESSAMENTO DE DADOS DO SISTEMA SCADA 84
4.3.2. CRIAÇÃO DE ANOTAÇÃO DOS DADOS DE VARIAÇÃO DE TEMPERATURA 87
4.3.3. CRIAÇÃO DE SUBCONJUNTOS DE DADOS 89
4.4. MODELOS DE REGRESSÃO PARA ESTIMATIVA DA VIDA ÚTIL REMANESCENTE 90
4.4.1. ESTUDO DE CASO – USO DE DADOS SINTÉTICOS 93
4.4.2. ESTUDO DE CASO – USO DE DADOS GPMS 95
4.4.3. ESTUDO DE CASO – USO DE DADOS SCADA – TEMPERATURA ROLAMENTO PRINCIPAL 96
4.4.3.1. Estimativa da VUR de aerogeradores usando dados de temperatura do rolamento principal 97
4.5. CONCLUSÕES DO CAPÍTULO DESENVOLVIMENTO DE MODELO COMPUTACIONAL PARA ESTIMATIVA DA VIDA ÚTIL REMANESCENTE 102

5. RESULTADOS E DISCUSSÃO 103

5.1. RESULTADOS DO ESTUDO DE CASO – USO DE DADOS SINTÉTICOS 103
5.2. RESULTADOS DO ESTUDO DE CASO – USO DE DADOS MATHWORKS/GPMS 108
5.3. RESULTADOS DO ESTUDO DE CASO – USO DE DADOS SCADA – TEMPERATURA ROLAMENTO PRINCIPAL 115
5.3.1. ESTIMATIVA DA VIDA ÚTIL REMANESCENTE DO CONJUNTO DE AMOSTRAS DE TURBINAS COM FALHA 119
5.3.2. OBSERVAÇÕES SOBRE OS RESULTADOS DA ESTIMATIVA DA VUR 120
5.3.3. ESTIMATIVA DA VIDA ÚTIL RESTANTE DO CONJUNTO DE AMOSTRAS DE TURBINAS COM FALHA 121
5.4. CONCLUSÕES DO CAPÍTULO RESULTADOS E DISCUSSÕES 122
5.4.1. PROPOSTA DE FRAMEWORK PARA ESTIMATIVA DA VUR DE AEROGERADORES COM VISTAS A GENERALIZAÇÃO 123

6. CONCLUSÕES E TRABALHOS FUTUROS 126

6.1. TRABALHOS FUTUROS 128
6.2. ARTIGOS PUBLICADOS 128

REFERÊNCIAS BIBLIOGRÁFICAS 130

APÊNDICES 143

1. INTRODUÇÃO

Desde o início do século 21, a busca pelo desenvolvimento e melhoria da produção de energia renovável tem sido impulsionada por regulamentações ambientais, novas perspectivas de negócios e uma mudança de mentalidade global em relação à importância da utilização de fontes de energia com o mínimo de impacto no planeta. (STRIELKOWSKI *et al.*, 2021).

Esta crescente demanda tem estimulado esforços do meio acadêmico e do produtivo para viabilização técnica e econômica do uso de diferentes fontes de energia limpa, como a biomassa, eólica, hídrica, das ondas e marés e solar.

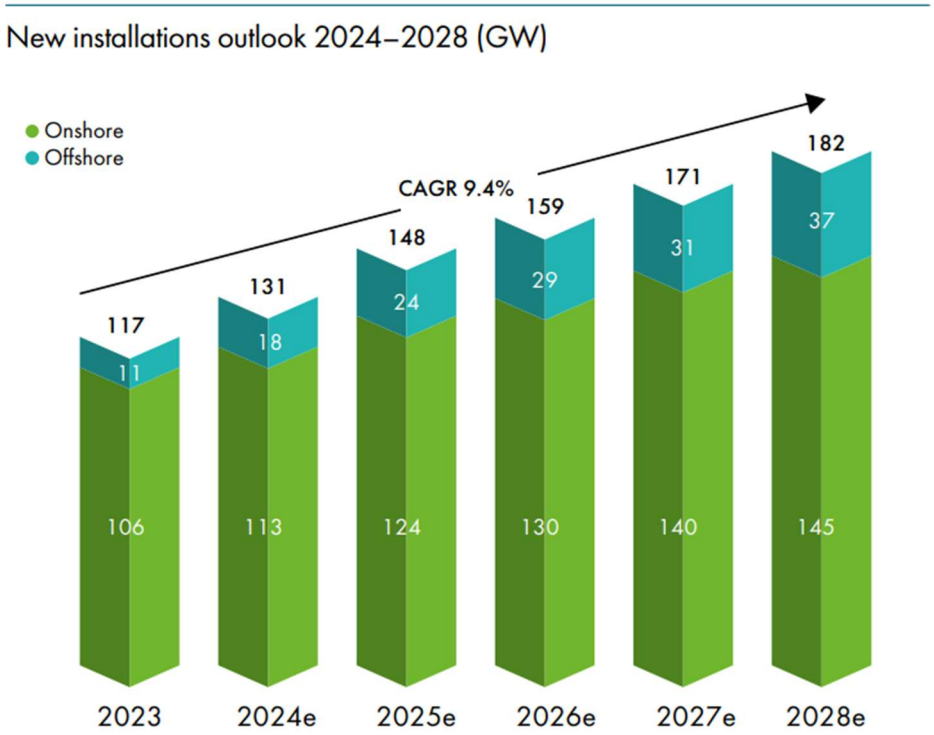
De acordo com o Conselho Global de Energia Eólica (*Global Wind Energy Council* - GWEC) (HUTCHINSON; ZHAO, 2023), a capacidade instalada de energia eólica atingiu 651 GW em 2019 globalmente. Em 2021, registrou-se um aumento global de 93,6 GW de novas instalações face ao ano anterior, fazendo com que a capacidade eólica subisse para 837 GW. Para 2022, a estimativa global era de 840 GW, mas chegou a 906 GW. Em 2023, o aumento registrado na capacidade foi de 116,6 GW, 50% maior que em 2022.(LEE; ZHAO, 2024).

Segundo os mesmos dados, até 2028 se prevê a adição de mais 791 GW, o que equivale um aumento médio da capacidade instalada em 158 GW por ano ou a uma taxa de crescimento anual composta (CAGR – *Compound Annual Growth Rate*) por ano de 9% (LEE; ZHAO, 2024), conforme mostrado na Figura 1.

Ainda de acordo com esse relatório, nos próximos 5 anos, as instalações offshore apresentarão uma taxa de crescimento da capacidade instalada de 28%, mais que 4 vezes a taxa de crescimento onshore que deve ser de 6,6% (LEE; ZHAO, 2024).

Também é importante destacar os avanços tecnológicos no desenvolvimento de aerogeradores de grande porte. Os fabricantes têm desenvolvido projetos de máquinas (DURAKOVIC, 2021; GENERAL ELETRIC, 2021; RICHARD, 2021), com rotores de até 236 m de diâmetro, altura do cubo de até 260 m e capacidade instalada entre 12 MW e 15 MW, que permitirão a implantação de usinas eólicas com maior produção de energia, podendo suprir até 6.000 casas com uma única turbina (AMELANG, SÖREN; WEHRMANN, 2020).

Figura 1 – Previsão de aumento de capacidade instalada em energia eólica (2024-2028) em valores estimados, marcados com *e*, após o ano.



Fonte: (LEE; ZHAO, 2024).

No Brasil, as energias renováveis mais utilizadas são a energia hídrica, a eólica e solar. A energia hídrica é a fonte com maior potência instalada. Não obstante, o crescimento do número de plantas de geração de energia eólica, de energia solar e híbrida (eólica em conjunto com a solar), tem se destacado em relação às outras fontes de energias renováveis (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA, 2021, 2024; PEREIRA, E. B.; MARTINS, F. R.; GONÇALVES, A. R.; COSTA, R. S.; LIMA, F. J. L.; RÜTHER, R.; ABREU, S. L.; TIEPOLO, G. M.; PEREIRA, S. V.; SOUZA, 2017). Esse crescimento, manteve o Brasil na 7ª posição no ranking de maiores capacidades instaladas, com adição de 4,9 GW e um total de 28,5 GW em 2023 (WORLD WIND ENERGY ASSOCIATION, 2021, 2024) conforme mostrado na Figura 2.

Figura 2 – Capacidade instalada nos 31 países maiores produtores de energia eólica

WWEA Annual Report 2023

Country/Region	2023	New Capacity 2023	Growth 2023/2022	2022	2021	2020
China	470'630	75'000	19,0%	395'630	346'670	290'750
United States	150'455	6'402	4,4%	144'053	135'177	121'510
Germany	69'475	3'233	4,9%	66'242	63'924	62'708
India	44'736	3'136	7,5%	41'600	39'800	38'625
Spain	30'748	590	2,0%	30'158	28'143	27'294
United Kingdom	30'215	1'452	5,0%	28'763	25'748	24'458
Brazil	28'580	4'919	20,8%	23'661	21'567	18'010
France	23'474	2'559	12,2%	20'915	19'084	17'949
Canada	16'986	1'774	11,7%	15'212	14'206	13'627
Sweden	16'251	1'973	13,8%	14'278	12'173	10'068
Australia	11'482	942	8,9%	10'540	11'100	9'305
Italy	12'012	365	3,1%	11'647	11'322	10'850
Turkey	11'697	292	2,6%	11'405	9'126	7'296
Netherlands	11'015	2'800	34,1%	8'215	6'570	5'953
Poland	9'383	1433	18,0%	7'950	7'846	6'784
Mexico	8'310	998	13,6%	7'312	7'262	6'789
Denmark	7'107	158	2,3%	6'949	6'181	6'102
Finland	6'946	1'269	22,4%	5'677	3'256	2'586
Portugal	5'804	74	1,3%	5'730	5'612	5'502
Belgium	5'315	274	5,4%	5'041	4'883	4'670
Greece	5'226	543	11,6%	4'683	5'105	3'866
Japan	5'214	487	10,3%	4'727	4'574	4'372
Norway	5'130	25	0,5%	5'105	4'452	4'113
Vietnam	4'910	944	23,8%	3'966	4'332	4'323
Ireland	4'713	186	4,1%	4'527	3'444	2'829
Chile	4'800	990	26,0%	3'810	3'291	3'120
Austria	3'885	313	8,8%	3'572	3'291	2'818
Argentina	3'705	396	12,0%	3'309	3'444	0'513
South Africa	3'560	0	0,0%	3'560	3'256	2'495
Romania	3'077	0	0,0%	3'077	3'029	3'029
Rest of the World*	31'939	2'537	8,6%	29'402	26'445	21'790
Total*	1046'781	116'065	12,5%	930'716	844'313	744'104

Fonte: (WORLD WIND ENERGY ASSOCIATION, 2024).

Uma questão relevante que tem impacto direto do resultado financeiro das usinas eólicas são os custos de operação e manutenção (O&M) que representam de 12% a 23%, dos custos operacionais (LEITE; ARAÚJO; ROSAS, 2018). Além disto as intervenções da manutenção em aerogeradores demandam esforços em termos de planejamento da intervenção, de disponibilidade e/ou aquisição de suprimentos, da disponibilidade de recursos humanos, máquinas e equipamentos, de acessibilidade ao local (HU; YOUN; WANG, 2018) e do atendimento a requisitos de segurança.

A manutenção de máquinas e equipamentos pode ser feita usando estratégias de manutenção corretiva, preventiva, preditiva e prescritiva. A manutenção corretiva é aquela na qual se realiza a intervenção para correção de um defeito ou falha de modo a corrigir e restabelecer a função do componente ou subsistema quando este apresenta falha ou perda da

função. Este tipo de manutenção pode resultar em paradas não programadas da máquina, implicar em riscos à segurança e um dano maior na máquina ou equipamento.

Já a manutenção preventiva se caracteriza pela realização da intervenção a partir de um plano de manutenção estabelecido pelo setor responsável, que leva em conta recomendações do fabricante e/ou histórico de falhas de um determinado componente ou subsistema. Na manutenção preventiva o componente ou subsistema é substituído antes da perda da função. Este tipo de manutenção pode gerar paradas desnecessárias nas máquinas e substituição de componentes que possuem vida útil remanescente implicando em custos indesejados por substituições desnecessárias.

A manutenção preditiva é uma estratégia baseada no monitoramento contínuo da condição dos componentes através de medições de grandezas físicas, como contaminação, temperatura, ruído e vibração. Essa abordagem exige investimentos em recursos técnicos e humanos, próprios ou terceirizados. Suas principais vantagens são a possibilidade de substituir componentes próximos ao fim de sua vida útil, facilitar o planejamento de intervenções, identificar causas-raiz de falhas e aumentar os intervalos entre substituições. No entanto, a implementação da manutenção preditiva pode gerar custos elevados e requer uma avaliação criteriosa para implementação maiores que a manutenção corretiva e preventiva.

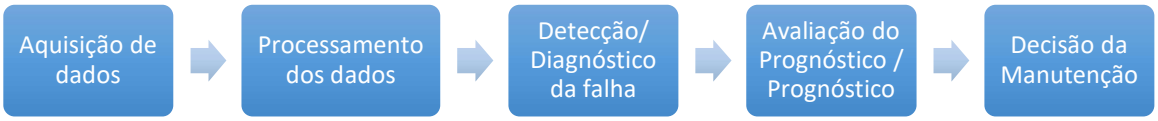
Já a manutenção prescritiva é uma abordagem proativa que se baseia na análise de medições de parâmetros operacionais da máquina e de seus componentes. Nesta estratégia, as intervenções são realizadas quando os parâmetros monitorados ultrapassam limites pré-estabelecidos. As principais vantagens da manutenção prescritiva são a utilização de análises preditivas e prescritivas para evitar falhas, além da redução de custos operacionais (FOX *et al.*, 2022; GORDON *et al.*, 2020). Ou seja, a manutenção prescritiva se diferencia da preditiva por se basear na análise de parâmetros operacionais específicos, com o objetivo de prescrever ações preventivas antes que os problemas ocorram.

A diferença fundamental entre as duas abordagens é que a manutenção preditiva monitora a condição dos componentes, enquanto a manutenção prescritiva analisa os parâmetros operacionais da máquina e de seus subsistemas para tomar ações proativas e evitar falhas.

Na indústria de energia eólica, as estratégias mais comumente usadas são a corretiva, preventiva e preditiva. O uso da estratégia prescritiva ainda é incipiente (FOX *et al.*, 2022; STRACK *et al.*, 2022). A manutenção dos aerogeradores apresenta grandes impactos econômicos, de segurança e ao meio-ambiente devido ao porte, localização e complexidade do sistema. Para o caso dos componentes críticos, a utilização de manutenção preventiva pode

implicar em paradas desnecessárias para verificação e manutenção. E a manutenção corretiva é indesejada devido aos riscos de acidentes, impactos ambientais e elevados custos de reparo. Neste contexto, a manutenção baseada condição (MBC) constitui-se como a estratégia mais adequada para estes componentes por viabilizar o uso de técnicas mais avançadas para detecção, diagnóstico e prognóstico das falhas em aerogeradores. Vale salientar que a implementação bem-sucedida da MBC depende da eficácia de todos os subsistemas que a compõe conforme destacado por Leite, Araújo e Rosas (LEITE; ARAÚJO; ROSAS, 2018), vide na Figura 3.

Figura 3 – Subsistemas de um programa de MBC



Fonte: Adaptado de (LEITE; ARAÚJO; ROSAS, 2018).

Neste contexto, a detecção de falha é o processo de identificação da redução parcial ou perda da função de um determinado componente ou sistema, gerada pela degradação do seu desempenho quando novo ou saudável. O diagnóstico de falhas é um processo que envolve a identificação dos modos e das causas-raiz das falhas mais comuns dos componentes ou sistemas principais utilizando técnicas de monitoramento, análise de sinais e/ou métodos de inteligência artificial (LIU; ZHANG, 2020; PENG *et al.*, 2022; QIAO; LU, 2015b) para realização de uma manutenção assertiva. E o prognóstico de falhas é o processo de identificar quantitativamente o intervalo de tempo entre a detecção e a ocorrência da falha ou perda definitiva da função do componente ou sistema. Este cálculo pode ser realizado utilizando-se um indicador de saúde e/ou computando a vida útil remanescente (VUR) com o respectivo intervalo de confiança.

1.1.O aerogerador

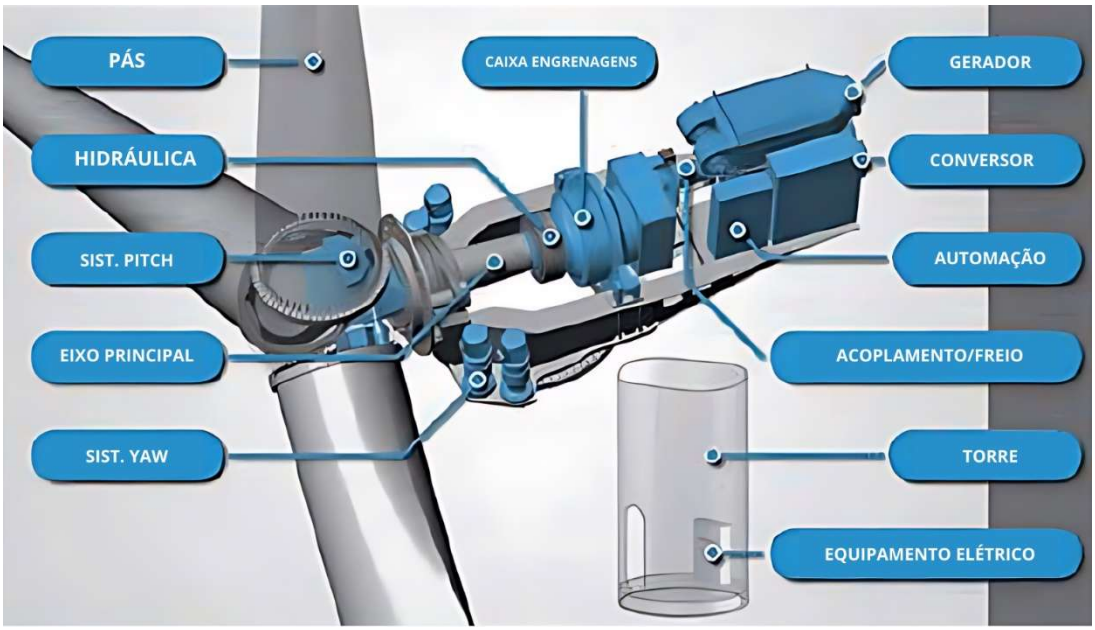
O aerogerador é um sistema eletromecânico complexo, com centenas de componentes (QIAO; LU, 2015a), que pode ser dividida nos seguintes subsistemas:

1. Fundação e torre;
2. Rotor;
3. Pás;
4. Nacele;
5. Eixos;
6. Rolamentos;

7. Caixa de engrenagens;
8. Gerador;
9. Sistema de condições meteorológicas (anemômetro e sensor de direção);
10. Eletrônica de potência.

A Figura 4 mostra os principais componentes de um aerogerador de eixo horizontal equipado com caixa de engrenagens (*gearbox*).

Figura 4 – Componentes principais de um aerogerador de eixo horizontal.



Fonte Adaptado de Qiao & Lu (QIAO; LU, 2015b)

Também existem modelos de aerogeradores não possuem a caixa de engrenagens e o movimento das pás é transmitido diretamente do eixo principal para o gerador. (QIAO; LU, 2015) mostram uma classificação a partir das configurações, princípios de operação e controle e métodos de conexão na rede. Estes autores destacam que a caixa de redução de turbinas com redutores de velocidade e o subsistema de eletrônica de potência, dos aerogeradores de acionamento direto devem ser monitorados com técnicas de MBC.

A transmissão de movimento no aerogerador se dá pela passagem do vento pelas pás. Nas turbinas de acionamento direto, o movimento das pás é transmitido pelo rotor para o eixo principal movimentando diretamente o gerador. Esse conjunto mecânico é empacotado dentro da nacele.

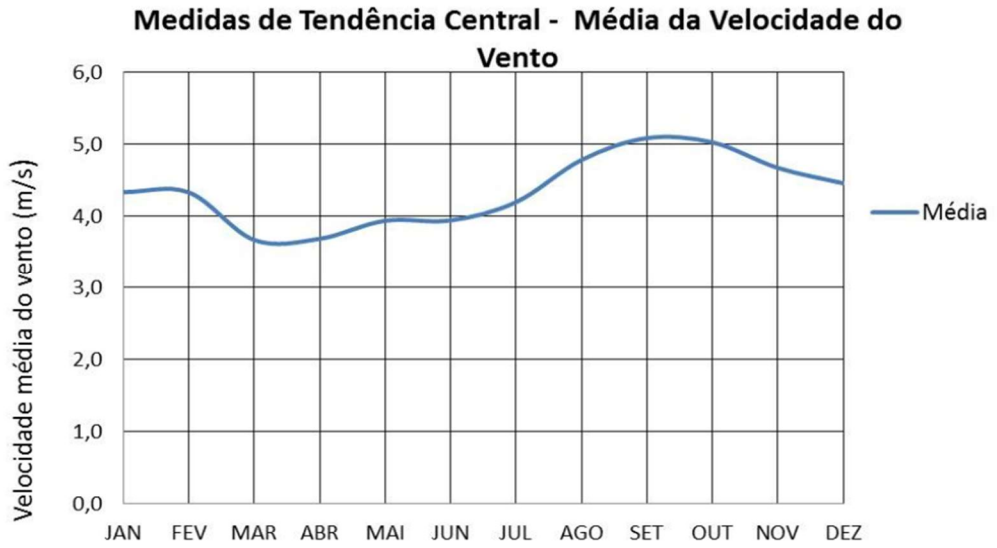
Nas turbinas com caixa de engrenagens, o movimento das pás é transmitido pelo rotor para o eixo principal e está conectado à caixa de engrenagens, que amplifica a rotação e

movimenta o gerador. A caixa de engrenagens, tipicamente, possui um ou dois redutores planetários conectados a um eixo de baixa rotação que transmite movimento para um eixo intermediário e depois para o eixo de alta rotação. Esse conjunto mecânico também é empacotado dentro da nacele.

Devido à grande quantidade de componentes tanto para as turbinas de acionamento direto como as com redutores de velocidade é comum que a vibração de um componente ou subsistema seja transmitida ou acople com outro componente o subsistema, reduzindo a razão de sinal/ruído dos sinais medidos pelos sensores de vibração.

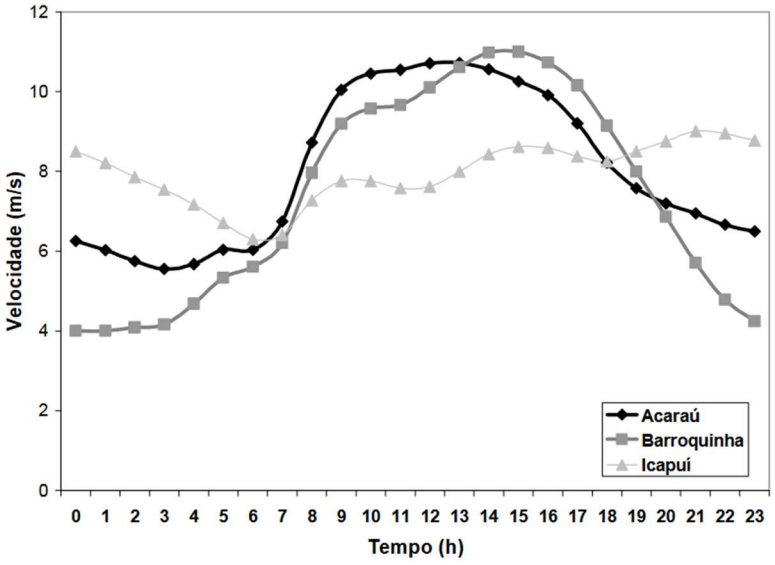
As características de variação da intensidade do vento ao longo do dia e sua sazonalidade ao longo do ano (DANTAS BARROS *et al.*, 2013) gera carregamentos que variam ao longo do dia e do ano podendo constituir-se como um desafio para análise das séries temporais dos dados coletados pelo sistema de Supervisão e Aquisição de Dados (SCADA) ou pelo sistema de MBC. A Figura 5 ilustra o comportamento sazonal da velocidade do vento ao longo dos 12 meses do ano na cidade de Natal/RN. E a Figura 6 apresenta a variação da velocidade do vento ao longo das 24 horas de um dia no litoral do estado do Ceará.

Figura 5 – Média da velocidade do vento na cidade de Natal/RN (período de 1984 – 2012).



Fonte: (DANTAS BARROS *et al.*, 2013)

Figura 6 – Média horária da velocidade do vento, representando as Plataformas de coleta de Dados (PCDs) do Litoral.



Fonte: (CAMELO *et al.*, 2008)

Cada uma das turbinas instaladas em usinas eólicas pode apresentar diferentes tempos de vida útil devido a fatores com o carregamento promovido pela incidência do vento que dependerão da sua localização e do efeito de esteira aerodinâmica pelas posições relativas com outras máquinas. Também como qualquer máquina industrial, os aerogeradores de uma mesma usina poderão apresentar diferentes tempos de vida útil devido a regime de operação, a degradação prematura de determinado componente ou substituição antecipada de um componente devido a uma parada realizada para manutenção de outro.

Em resumo, turbinas iguais operando numa mesma usina eólica terão tempos de VUR diferentes. Não obstante, quanto mais precisas e antecipadas as estimativas de VUR, mais assertivo será o planejamento para uma manutenção eficaz e maior será a disponibilidade de cada máquina.

1.2. Inteligência artificial em aplicações industriais

Com o advento da indústria 4.0, o setor industrial tem buscado implementar estratégias de digitalização que promovam flexibilidade, responsividade e inteligência para lidar com as demandas de produção variáveis e otimização em tempo real (PERES *et al.*, 2020). Neste

contexto a inteligência artificial (IA) funciona como um elo entre os pilares de big data, simulação, internet das coisas (IoT) da indústria 4.0 para viabilizar a capacidades de auto-otimização, autoconsciência e automonitoramento que viabilizem sistemas de produção cyber-físicos (PERES *et al.*, 2020) contribuindo para a geração de valor na quarta revolução industrial (WILLIAM DIAB *et al.*, 2022).

Desta forma, a inteligência artificial industrial (IAI) caracterizada pelo uso combinado de IA e IoT em aplicações industriais (WILLIAM DIAB *et al.*, 2022) se diferencia da IA, segundo Peres *et al*. (PERES *et al.*, 2020) pelas cinco dimensões particulares abaixo:

- **Infraestruturas:** No que diz respeito a hardware e software, há uma grande ênfase nas capacidades de processamento em tempo real, garantindo confiabilidade de nível industrial com altos requisitos de segurança e interconectividade;
- **Dados:** A IA industrial requer dados caracterizados por seu grande volume e alta variedade de velocidade, provenientes de diferentes unidades, produtos, regimes etc;
- **Algoritmos:** Requerem a integração de conhecimentos físicos, digitais e heurísticos. Alta complexidade derivada do gerenciamento, implantação e governança de modelos;
- **Tomada de decisões:** Dado o ambiente industrial, a tolerância ao erro é geralmente muito baixa, sendo extremamente importante o tratamento da incerteza. A eficiência é de especial importância para problemas de otimização em larga escala;
- **Objetivos:** A IA industrial visa, principalmente, a criação de valor concreto por meio de uma combinação de fatores como redução de desperdício, melhoria da qualidade, aumento do desempenho do operador ou dos tempos de aceleração.

O uso de inteligência artificial (IA) tem impulsionado avanços significativos em diversas áreas industriais. Algumas das principais contribuições incluem manufatura inteligente, otimização de processos, controle de qualidade, manutenção preditiva, robótica, diagnóstico de saúde, veículos autônomos e colaboração e ergonomia homem-robô (PERES *et al.*, 2020; WILLIAM DIAB *et al.*, 2022). A facilidade e a velocidade no desenvolvimento de métodos e modelos inteligentes têm o potencial de acelerar a transformação da indústria (WILLIAM DIAB *et al.*, 2022).

No entanto, a adoção dessas tecnologias enfrenta desafios relacionados à necessidade de atender a requisitos de estabilidade, confiabilidade e segurança exigidos por diferentes setores. Gestores e tomadores de decisão precisam exercer cautela para evitar riscos, como

redução de disponibilidade, previsões incorretas, falhas catastróficas ou vulnerabilidades em sistemas de informação.

Um dos principais focos de pesquisa em IA industrial será garantir em três frentes as pesquisas futuras (PERES *et al.*, 2020):

- Dados FAIR (*Findable, Accessible, Interoperable and Reusable* - Localizável, acessível, interoperável e reutilizável) para IA Industrial: garantir que os dados sejam facilmente encontrados, acessíveis, interoperáveis e reutilizáveis é crucial para servir os melhores interesses das comunidades industriais e de investigação, promovendo o avanço da ciência em benefício da sociedade (COMMISSION, [s. d.]) *apud* (PERES *et al.*, 2020);
- Modelos FAIR para IA Industrial: Além dos dados, para permitir a implementação de frameworks FAIR, os ativos digitais, como os modelos, devem ser encontrados de forma confiável e reutilizados quando apropriado, através de identificadores persistentes ligados a metadados ricos (incluindo, por exemplo, a origem) representados através de formatos comuns, padronizados e seguros (PERES *et al.*, 2020);
- Cyber-Infraestruturas: Infraestruturas adequadas são cruciais para garantir o nível de qualidade, segurança e confiabilidade necessários para melhorar a adoção de soluções de IA industrial. Estas incluem operações remotas, segurança cibernética, mecanismos de preservação da privacidade, tecnologia 5G e prognósticos colaborativos, entre outros (PERES *et al.*, 2020).

Sendo assim, o desenvolvimento de aplicações de IAI no mundo real demanda de pesquisadores e do meio produtivo o equilíbrio entre a disponibilização e uso responsável de dados e modelos abertos, desenvolvimento de tecnologia e metodologias que suportem a ciber estruturas e o desenvolvimento gradual de cultura organizacional, investimentos em ciberinfraestrutura e estratégias de segurança para proteção dos negócios.

1.3. Prognóstico de falhas em aerogeradores

Na última década, houve um aumento vertiginoso no número de artigos em conferências e periódicos, abordando diferentes técnicas de prognóstico de falhas em redutores de velocidade de aerogeradores (CARROLL *et al.*, 2019; DAMESHGHI; REFAN, 2019; KOUKOURA, 2018; LEITE; ARAÚJO; ROSAS, 2018; ZHU *et al.*, 2015).

QIAO; LU, (2015a) classificam as falhas dos aerogeradores em falhas causadas pelo desgaste, nas que são ocasionadas após longo tempo de operação e eventos permanentes, e nas falhas aleatórias temporárias que são aquelas que ocorrem em tempos menores causadas por

flutuação da velocidade do vento, questões térmicas, distúrbios na rede, leituras temporárias incorretas em sensores etc.

Diversos autores, (CARROLL *et al.*, 2019; DAMESHGHI; REFAN, 2019; ELASHA *et al.*, 2019; HERP; PEDERSEN; NADIMI, 2019; KOUKOURA, 2018), têm utilizado técnicas de aprendizagem de máquina ou redes neurais artificiais (RNA's) (MURPHY, 2012) para estimar a VUR de engrenagens e rolamentos de redutores de velocidade de aerogeradores e realizar o prognóstico das falhas.

Uma precisão aprimorada na estimativa da vida útil remanescente dos componentes críticos de um aerogerador, como o trem de força, possibilita uma manutenção mais assertiva e economicamente eficiente. Isso se traduz em maior disponibilidade da turbina, uma vez que intervenções de manutenção corretivas são substituídas por ações preventivas programadas. Algoritmos avançados de aprendizado de máquina e inteligência artificial podem explorar grandes volumes de dados históricos coletados por sistemas de condicionamento monitoramento para prognóstico e rastreamento do desgaste dos componentes. Tecnologias como RNA e florestas aleatórias aprendem padrões a partir de variáveis extraídas dos sistemas SCADA e de manutenção baseada na condição, gerando modelos precisos de estimativa de falha. Isso possibilita o prognóstico confiável da vida útil remanescente do conjunto de pás, caixas de engrenagens e outros itens do trem de força com antecedência, apoiando a tomada de decisão sobre a substituição preventiva ideal (CARROLL *et al.*, 2019; YANG; ZHANG, 2020).

No entanto, as técnicas que utilizam inteligência artificial (IA) necessitam que atividades preliminares sejam realizadas nas etapas de processamento de dados e detecção das falhas. Alguns trabalhos (KOUKOURA, 2018; KOUKOURA *et al.*, 2017) destacam a necessidade de realizar a classificação dos dados que possibilite identificar os componentes saudáveis, os componentes com falhas incipientes e os componentes com falha desenvolvida.

Outro ponto importante a ser destacado, é a prática de mercado da contratação do fabricante do aerogerador para realizar a manutenção do equipamento devido à sua expertise e conhecimento. Porém, esta relação pode ser delicada devido a conflitos de interesse que possam surgir, pela possibilidade de o fabricante substituir componentes em intervalos anteriores ao fim da vida útil e ser ele mesmo o fornecedor das peças e subsistemas do aerogerador.

Considerando o aumento consistente da capacidade instalada nos âmbitos regional, nacional e internacional, a partir da instalação de usinas de maior capacidade e utilização de máquinas de maior porte, o desenvolvimento de metodologias de prognóstico de falhas utilizando técnicas de aprendizagem de máquina e inteligência artificial (IA) possibilitarão tomadas de decisão

mais assertivas que gerem mais valor para a operação, a partir do aumento da disponibilidade e manutenção dos aerogeradores.

O uso da inteligência artificial tem sido objeto de estudo em diferentes aplicações e objetivos para a indústria. Destaca-se o interesse de fabricantes em métodos eficazes que possibilitem a detecção precoce e automática de defeitos potenciais com a máquina em operação. Adicionalmente por ser um dos pilares da indústria 4.0, tem grande potencial para ser explorado como catalisador de ganhos operacionais e de manutenção para a área de energia eólica (PERES et al., 2020).

As técnicas de aprendizagem de máquina mais comumente utilizadas para monitoramento da condição em aerogeradores são as RNA`s, as máquinas de vetor de suporte e as árvores de decisão (STETCO *et al.*, 2019)

A falta de grandes bases de dados públicos para desenvolvimento, avaliação e comparação de modelos é um desafio a ser vencido, pois o uso de dados sintéticos, obtidos a partir de testes de bancada, não viabilizam a generalização necessária representar as condições do mundo real (STETCO *et al.*, 2019).

1.4.Objetivos

1.4.1. Objetivo geral

Propor um framework para prognóstico de falhas em aerogeradores que estabeleça critérios para capacidade de generalização dos modelos inteligentes e metodologia para uso eficaz de dados reais de operação, aplicando-o na estimativa da vida útil remanescente do rolamento principal a partir de dados de temperatura, velocidade do vento e de potência ativa da rede do sistema SCADA.

1.4.2. Objetivos específicos

1. Analisar quantitativamente a produção bibliográfica sobre prognóstico de falhas em aerogeradores, identificando lacunas quanto à capacidade de generalização dos modelos e disponibilidade de dados;
2. Propor uma classificação da capacidade de generalização dos modelos inteligentes usados na estimativa de VUR, considerando a quantidade de conjuntos de dados necessários para reduzir o sobreajuste;
3. Desenvolver e validar uma metodologia para uso de subconjuntos de dados reais visando mitigar o problema de sobreajuste devido à escassez de dados de falhas;

4. Elaborar e testar algoritmos para estimativa da VUR utilizando dados sintéticos e de benchmark para avaliação rápida do desempenho dos modelos;
5. Implementar e avaliar diferentes modelos de aprendizagem de máquina (regressão de vetor de suporte, regressão isotônica, árvore de decisão, floresta aleatória, árvores extras e regressão de aumento de gradiente) para prognóstico de falhas do rolamento principal usando dados de temperatura do SCADA;
6. Validar os modelos desenvolvidos com dados reais de três aerogeradores, analisando suas capacidades preditivas através das métricas: erro quadrático médio (MSE), raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e o coeficiente de determinação (R2 Score);
7. Desenvolver um framework que oriente a implantação gradual de métodos de prognóstico de falhas, considerando a complexidade crescente dos dados e modelos;
8. Disseminar o conhecimento gerado através de publicações científicas em periódicos e congressos da área.

1.5.Contribuição científica

A principal contribuição deste trabalho foi o desenvolvimento de um framework para estimativa da Vida Útil Remanescente (VUR) a partir de séries de temperatura do rolamento principal reais provenientes do sistema SCADA a partir de modelos de simples implementação. Adicionalmente, também é proposta uma estratégia de validação cruzada para mitigar o problema dos dados escassos e aumentar a capacidade de generalização dos modelos considerando a natureza da dos dados reais da operação dos aerogeradores.

O trabalho também contribui com discussão sobre a capacidade de generalização de modelos direcionados por dados reais de aerogeradores de usinas eólicas destacando as lacunas da literatura sobre o assunto generalização de modelos, propõe um método de classificação dos modelos de estimativa da VUR a partir da quantidade de conjuntos de dados utilizados.

Também propõe um framework mais abrangente para viabilizar o desenvolvimento e avaliação de modelos a partir de uma abordagem de aumento incremental de complexidade dos dados de entrada com foco na generalização dos resultados.

1.6.Organização da tese

Este trabalho está organizado da seguinte forma:

- No Capítulo 1 é realizada a introdução, com a justificativa e motivação do estudo no âmbito dos aerogeradores destacando sua relevância do ponto de vista da Operação e Manutenção (O&M). Além disso, são apresentados os objetivos, geral e específicos, a contribuição científica e a organização dos capítulos;
- O Capítulo 2 apresenta a revisão bibliográfica sobre o prognóstico de falhas em aerogeradores, contextualizando a relevância do tema e apresentando uma revisão sistemática, baseada no protocolo PRISMA com ênfase na generalização de modelos direcionados por dados para estimativa da VUR dos componentes de aerogeradores;
- No Capítulo 3 são apresentados os conjuntos de dados utilizados no trabalho. Utilizou-se um conjunto de dados sintéticos de temperatura, um conjunto de dados de vibração público e um conjunto de dados do sistema SCADA de três turbinas de duas usinas eólicas. Estes três conjuntos de dados foram utilizados para avaliar o desempenho dos modelos de aprendizagem de máquina para estimar a VUR. Neste capítulo, também são apresentados os modelos de aprendizagem de máquina desenvolvidos para estimativa da VUR com os 3 conjuntos de dados. E é apresentada estratégia do uso de dados escassos para treinamento, validação cruzada e testes dos modelos;
- No Capítulo 4 apresentam-se e discutem-se os resultados dos modelos desenvolvidos a partir de métricas de avaliação e das estimativas da VUR com uso de dados sintéticos, dados públicos e reais do rolamento principal de um aerogerador. Também é proposto um framework para avaliar a viabilidade e estimar a VUR de componentes usando dados reais;
- O Capítulo 5 apresenta as conclusões finais, as sugestões de trabalhos futuros.

2. ESTADO DA ARTE

Este Capítulo tem como objetivo abordar o estado da arte do prognóstico de falhas em aerogeradores. Na primeira e segunda seções serão introduzidos o conceito de prognóstico de falhas e seu uso em máquinas em geral. As Seções 2.3, 2.4, 2.5 e 2.6 apresentam uma revisão sistemática do prognóstico de falhas em aerogeradores utilizando parcialmente o protocolo PRISMA .(GALVÃO; PANSANI; HARRAD, 2015; MOHER et al., 2010). A Seção 2.7 aborda o prognóstico de falhas em rolamentos principais de aerogeradores.

2.1. Prognóstico de falhas

O termo prognóstico está intrinsicamente ligado a quando se deseja referir a saúde de um indivíduo, máquina, equipamento ou sistema. Ele é utilizado em diferentes áreas do conhecimento como a saúde, indústria, engenharia de software dentre outras e tem diferentes definições nas áreas em que é utilizado.

Na área de saúde o termo prognóstico se refere a processos de identificação do estado de saúde, de expectativa de vida de pacientes, das chances de recuperação e do tempo o qual o paciente irá sobreviver (HUI; MO; PAIVA, 2021; KÜHNE *et al.*, 2021; STONE *et al.*, 2021).

Quando se usa o termo prognóstico em aplicação relacionadas a sistemas não humanos, ou não relacionados com a saúde humana, refere-se a processos relacionados à identificação da passagem do estado de ausência para o estado de presença de falha em sistemas físicos ou de software. Também se refere aos processos de monitoramento de grandezas físicas (sensores e sistemas), a coleta, pré-processamento e extração de características estatísticas de dados e a realização de estimativas de quando tempo um determinado componente ou sistema ainda pode operar desempenhando sua função (GUO; LI; LI, 2020a; KIM; CHOI; KIM, 2021b; LYDIA; EDWIN PREM KUMAR, 2023).

Em sistemas complexos, a realização do prognóstico de falhas é importante pois viabiliza um melhor planejamento das intervenções para manutenção, aumenta a segurança, evitando incidentes catastróficos, reduz o *downtime*, aumenta a disponibilidade e a produção e melhora a confiabilidade destes sistemas (ABID; KHAN; IQBAL, 2021; KORDESTANI *et al.*, 2021; WATANABE *et al.*, 2021).

O uso do prognóstico de falhas tem como benefícios

- identificar o estado de saúde de um sistema (GUO; LI; LI, 2020b);
- identificar falhas potenciais (GUO; LI; LI, 2020b);

- detectar e isolar falhas quando do início do seu desenvolvimento (KORDESTANI *et al.*, 2021);
- permite uma operação segura;
- permite o planejamento economicamente mais racional da manutenção (KIM; CHOI; KIM, 2021a);
- realizar estimativa do tempo de vida útil remanescente (BOUGACHA; VARNIER; ZERHOUNI, 2021);
- melhora o processo de decisão pela operação e manutenção (BOUGACHA; VARNIER; ZERHOUNI, 2021).

2.2. Prognóstico de falhas em máquinas em geral

Nesta Seção serão abordados alguns conceitos e questões relacionados ao uso do prognóstico de falhas em equipamentos e máquinas industriais.

Neste sentido, um estudo abrangente realizado por Sikorska, Hodkiewicz, e Ma (SIKORSKA; HODKIEWICZ; MA, 2011) identificou, a partir de propostas de normas e de diferentes autores, que o prognóstico de falhas em aplicações industriais deve reunir o seguinte conjunto de definições:

> a) O prognóstico deve, ou deveria, ser realizado no nível de componente ou subcomponente;
> b) Envolve a predição da progressão do tempo de um modo de falha específico do estado incipiente até a falha do componente;
> c) Uma avaliação da operação futura do componente é necessária, e;
> d) O prognóstico é relativo a, mas não o mesmo que, o diagnóstico.

Sikorska, Hodkiewicz, e Ma (SIKORSKA; HODKIEWICZ; MA, 2011) também destacam que, o prognóstico de falhas é caracterizado por um valor pontual, chamado de estimativa da VUR e um intervalo de confiança, ou incerteza, associado (KRAMTI *et al.*, 2018; SAIDI; HAYDER; SAIDI, 2022; SIKORSKA; HODKIEWICZ; MA, 2011) a esse valor.

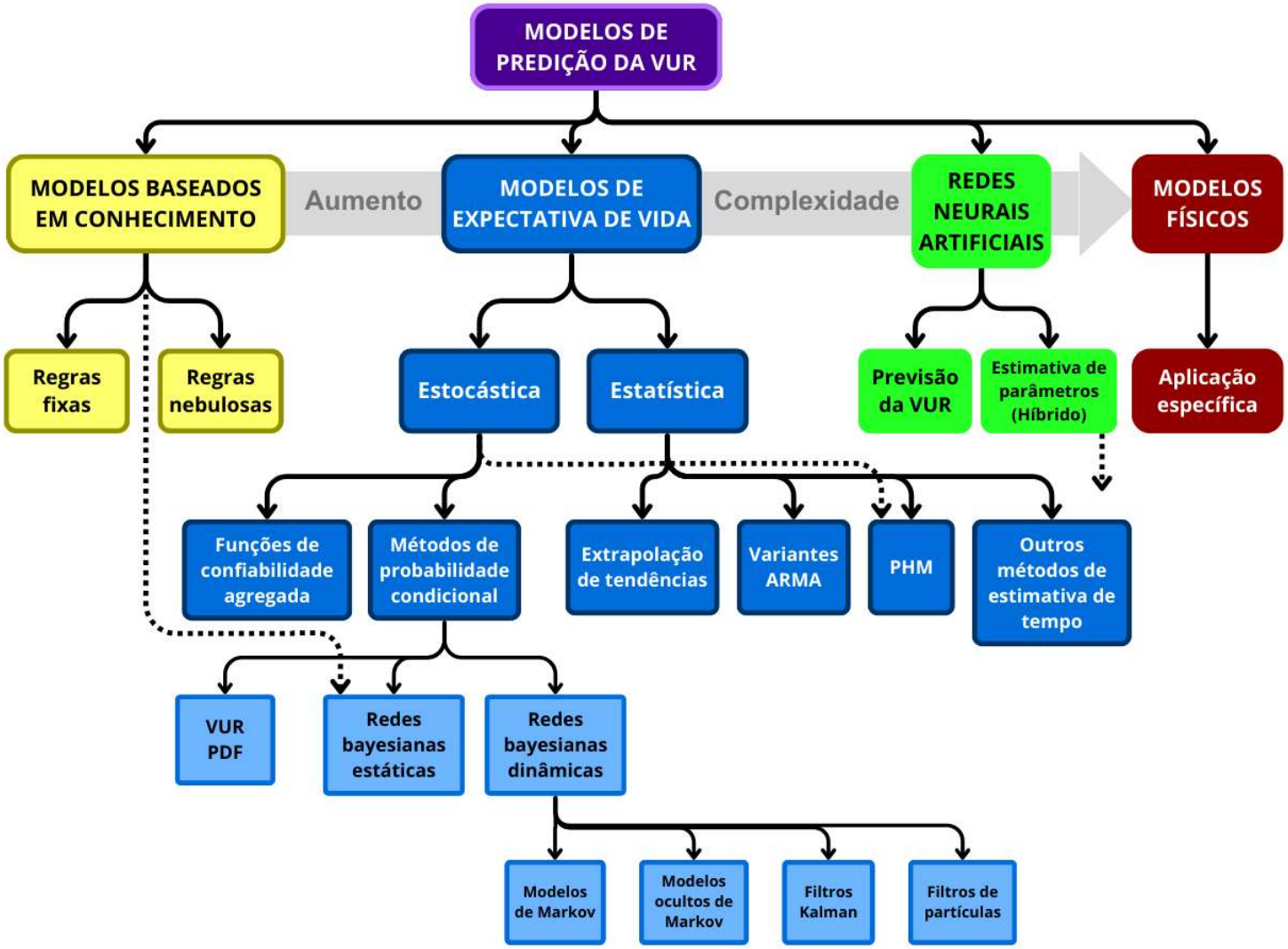
Caso se utilize um sistema de tomada de decisão seja suportada pelo prognóstico de falhas haverá implicações mais abrangentes para o gerenciamento do negócio, sendo necessário considerar fatores não relacionados com a engenharia como questões logísticas, problemas de gerenciamento de estoque/suprimento, opções de planejamento de manutenção (SIKORSKA; HODKIEWICZ; MA, 2011).

Sikorska, Hodkiewicz, e Ma (SIKORSKA; HODKIEWICZ; MA, 2011) propõem uma categorização para diferentes abordagens dos modelos de estimativa da VUR, ver Figura 7,

como modelos de base de conhecimento, modelos de expectativa de vida (estocásticos e estatísticos), redes neurais artificiais (RNA`s) e modelos físicos que aumentam a complexidade nesta ordem.

Ainda segundo Sikorska, Hodkiewicz, e Ma (SIKORSKA; HODKIEWICZ; MA, 2011), os modelos baseados em conhecimento utilizam um mecanismo de comparação para avaliar a semelhança entre um determinado cenário e um repositório de falhas pré-determinadas, permitindo-lhes inferir a VUR prevista com base em ocorrências históricas. Os modelos de expectativa de vida determinam a VUR de componentes individuais da máquina, considerando a probabilidade projetada de degradação, dadas as circunstâncias operacionais conhecidas. As RNA's empregam representações matemáticas derivadas de dados observacionais para calcular direta ou indiretamente uma saída estimada para a VUR de um componente ou máquina. Esta estimativa baseia-se nos dados observados e não numa compreensão abrangente dos mecanismos de falha subjacentes. E os modelos físicos descrevem quantitativamente as características de um modo de falha, utilizando leis físicas derivadas de princípios fundamentais. Esta última abordagem requer uma ampla compreensão do comportamento do sistema.

Figura 7 – Principais categorias de modelos para predição da vida útil remanescente.



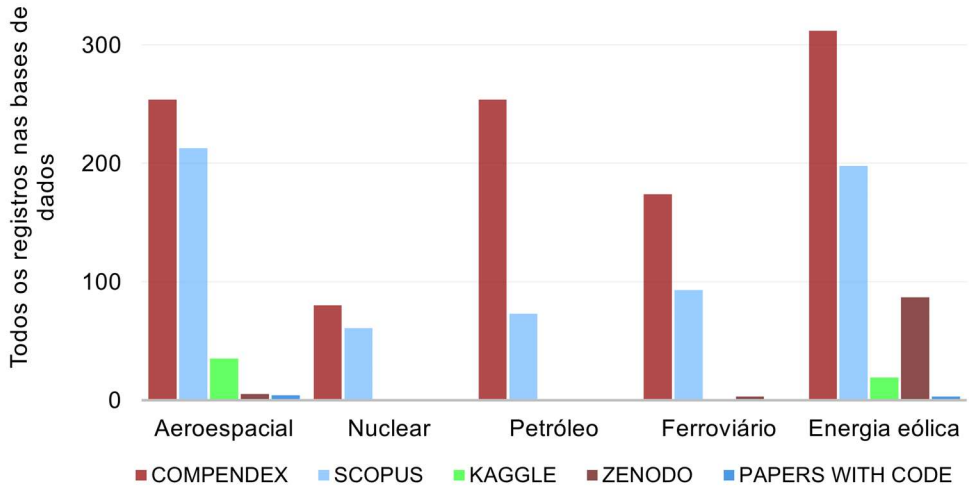
Fonte: Adaptado de Sikorska, Hoedkiewicz & Ma (SIKORSKA; HODKIEWICZ; MA, 2011)

2.3. Prognóstico e Gerenciamento da saúde em diferentes indústrias

O prognóstico de falhas e o gerenciamento de saúde (PHM) é um campo de conhecimento que visa melhorias na confiabilidade de sistemas complexos em diferentes aplicações industriais (TAO *et al.*, 2018). O uso da manutenção preditiva é um requisito da abordagem PHM para alcançar maior desempenho operacional com mitigação de paradas não planejadas e melhoria do conhecimento da dinâmica de degradação das máquinas (KERBOUA; KELAIAIA, 2022). E a métrica do prognóstico mais importante é VUR (SAIDI; HAYDER; SAIDI, 2022) dos componentes de uma determinada máquina ou equipamento.

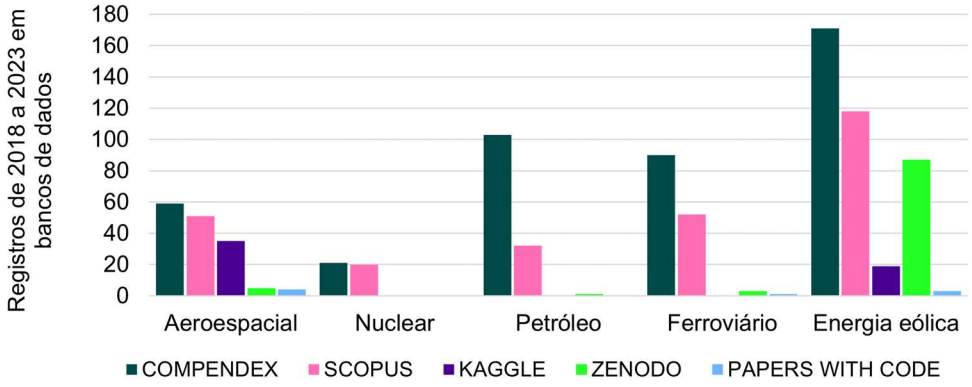
Um olhar sobre as bases de dados científicas permite identificar o interesse dos pesquisadores do PHM em diferentes indústrias A Figura 8 mostra o número de publicações relacionadas a prognósticos de falhas nas indústrias aeroespacial, nuclear, de petróleo e gás, ferroviária e de energia eólica. Na Figura 8 e Figura 9, também é possível reconhecer que a indústria aeroespacial, o petróleo e o gás e a energia eólica têm o maior interesse pela comunidade científica de todos os registros. Não obstante, cabe ressaltar que o prognóstico para aplicações de energia eólica despertou maior interesse por parte dos pesquisadores nos últimos 5 anos, superando em mais de 50% o interesse nas indústrias aeroespacial, de petróleo e gás nas bases Engineering Village e SCOPUS. Da mesma forma, também é observado no Kaggle, Zenodo e Papers with code, conforme mostrado na Figura 9. Destaca-se para as bases de dados Kaggle, Zenodo e Papers with code foram incluídos todos os resultados obtidos, pois não há recurso de filtragem nestas.

Figura 8 – Contagem de registros de publicações de diferentes bases de dados científicas para *strings* de pesquisa com *failure* AND *prognostics* OR *prognosi** e adição do nome da indústria para todos os períodos disponíveis nas bases de dados



Fonte: O autor (2024).

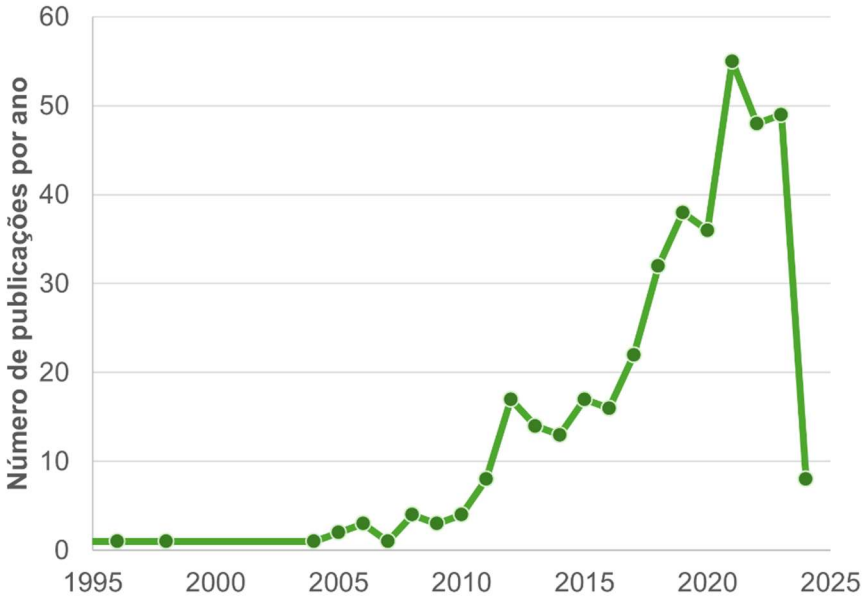
Figura 9 – Contagem de publicações dos últimos 5 anos de diferentes bases de dados científicas para *strings* de pesquisa com *failure* AND *prognostics* OR *prognosi** e adição do nome da indústria para todos os períodos disponíveis nas bases de dados



Fonte: O autor (2024).

Para corroborar isso, houve uma explosão de artigos desde 2014 investigando o uso de métodos e modelos baseados em dados para avaliação da VUR de aerogeradores conforme indicado na Figura 10, abordando diferentes técnicas de prognóstico de falhas no conjunto cinemático, principalmente em redutores de velocidade de aerogeradores.

Figura 10 – Número de registros de publicações relacionadas a prognósticos de falhas de todos os componentes ou subsistemas de aerogeradores na Engineering Village (COMPENDEX).



Fonte: O autor (2024).

Vários autores (CARROLL *et al.*, 2019; DAMESHGHI; REFAN, 2019; ELASHA *et al.*, 2019; HERP; PEDERSEN; NADIMI, 2019; KOUKOURA, 2018) utilizaram técnicas de aprendizado de máquina ou redes neurais artificiais (RNA's) para estimar a VUR de engrenagens e rolamentos em caixas de engrenagens e realizar prognósticos de falhas.

Entretanto, a estimativa da VUR e seu intervalo de confiança (ou incerteza) devem ser convertidos em informações valiosas para o pessoal de manutenção (KERBOUA; KELAIAIA, 2022; KRAMTI *et al.*, 2018; SAIDI; HAYDER; SAIDI, 2022) independentemente do tipo de modelo utilizado para isso.

O cálculo preciso da estimativa da VUR é crucial para uma manutenção econômica e maior disponibilidade de aerogeradores. Algoritmos de aprendizado de máquina e inteligência artificial permitem a utilização de uma enorme quantidade de dados do sistema de manutenção baseado em condições e/ou do sistema de supervisão e aquisição de dados para estimar a VUR dos componentes do trem de força (CARROLL *et al.*, 2019; YANG; ZHANG, 2020).

Alguns métodos e modelos podem alcançar alta precisão na estimativa de VUR e no cálculo do intervalo de confiança, mas, não menos importante, é a capacidade de generalização (GALICKI *et al.*, 2004; VAPNIK, 1999; ZHOU *et al.*, 2023) disso é permitir sua utilização em um grande número de aerogeradores de um mesmo parque eólico, mas esta questão não é discutida com frequência em artigos.

Além disso, também existe potencial para o uso da computação quântica já utilizada para detecção de falhas (CORREA-JULLIAN *et al.*, 2022) em aerogeradores e tem grande potencial para uso em diagnóstico e prognóstico de falhas conforme destacado por *MAIOR et al., (2023)*.

O desafio para pesquisadores e players da indústria eólica é o uso de modelos de aprendizado de máquina na escala dos parques eólicos.

2.4. Importância da avaliação dos modelos e implicações operacionais do prognóstico de falhas em aerogeradores

Avaliar o desempenho de métodos e modelos de prognóstico de falhas é importante para gerar resultados práticos na tomada de decisão das equipes de manutenção (KERBOUA; KELAIAIA, 2022; KRAMTI *et al.*, 2018; SAIDI; HAYDER; SAIDI, 2022). Contudo, é necessário utilizar métricas padronizadas para esta avaliação. Saxena et al., (2008) observam a falta de metodologias padronizadas para medir o desempenho desses modelos.

Para avaliar o desempenho de métodos e modelos prognósticos, devem ser utilizadas métricas padronizadas que estruturem requisitos e especificações de forma quantitativa. Também é essencial realizar muitos experimentos para garantir estimativas representativas do VUR ou um modelo analítico adequado (SAXENA *et al.*, 2008). A escolha das métricas deve ser compatível com o método ou modelo utilizado para evitar inconsistências e garantir a precisão da avaliação (SAXENA *et al.*, 2008).

A realização do prognóstico de falhas também tem implicações operacionais em diferentes indústrias, incluindo a energia eólica, devido à sua capacidade de avaliar o estado operacional dos sistemas (KERBOUA; KELAIAIA, 2022), impactando diretamente nos custos e riscos relacionados à disponibilidade, mão de obra e sistemas/máquinas sobressalentes (SAHU *et al.*, 2023).

Prakash *et al.* (PRAKASH; VUCKOVIC; AMIN, 2023) destacam que as ações de intervenção em sistemas ou máquinas devem ser planejadas quando a incerteza na estimativa de VUR atingir um determinado limite, a fim de permitir a correção dentro do prazo necessário.

Os impactos operacionais do prognóstico também influenciam a disponibilidade e o custo da manutenção, conforme observado por SAHU et al., (2023). Se a VUR for superestimada, poderá ocorrer falha antes da intervenção programada. Se for subestimada, é possível planejar adequadamente os recursos para paradas programadas de manutenção, evitando falhas catastróficas, mas com aumento de custos devido a substituições extras (SAHU *et al.*, 2023).

2.5. Vida útil remanescente e conexões com a manutenção prescritiva de aerogeradores

A manutenção prescritiva é uma estratégia de manutenção proativa que se baseia na análise de medições de parâmetros operacionais da máquina, subsistemas e componentes. É realizada quando os parâmetros escolhidos ultrapassam os limites estabelecidos ou quando são utilizadas análises preditivas e prescritivas para evitar falhas e reduzir custos (FOX *et al.*, 2022). Strack *et al.* (STRACK *et al.*, 2022) ilustram num *framework* como a utilização adequada de dados numa abordagem prescritiva tem um impacto estratégico e tático na realização da manutenção.

Existem poucos estudos sobre a aplicação da manutenção prescritiva em parques eólicos. Alguns trabalhos abordam a utilização de técnicas de monitoramento da condição operacional de sistemas, utilizando dados do SCADA ou do sistema de monitoramento de condição (CMS) (FOX *et al.*, 2022). Esses estudos utilizam diferentes técnicas: baseadas em dados, como inteligência artificial ou modelos estatísticos; baseados em física, como modelos de tempo de

fadiga; estocásticos, como modelos Markovianos; e híbridos, que comparam modelos físicos com dados ambientais e de inspeção (FOX *et al.*, 2022).

Contudo, algumas limitações ao uso da manutenção prescritiva são destacadas por FOX et al., (2022). Dentre elas, destacam-se: dificuldade de acesso a dados de falhas e pequeno número de falhas em bancos de dados, dificultando o treinamento do modelo; a baixa generalização dos modelos, flexibilidade, escalabilidade e adaptabilidade a diferentes turbinas e localizações de usinas; a utilização intensiva de dados históricos em detrimento de dados em tempo real, dificultando a automatização da previsão e do prognóstico de falhas; pouca utilização de dados não estruturados provenientes de relatórios de fiscalização; e a necessidade de atualizar modelos de vida útil remanescente (VUR) para prever a vida útil restante de diferentes componentes e locais (MOMBER *et al.*, 2022).

Além disso, as questões socioculturais também influenciam a adoção bem-sucedida da abordagem prescritiva pelas partes interessadas na manutenção de aerogeradores, conforme identificado STRACK et al., (2022). As fortes capacidades digitais dos operadores, dos fabricantes de equipamento original (OEM) e dos prestadores de serviços independentes (ISP) são destacadas como um fator crucial para uma implementação bem-sucedida. No caso dos operadores e ISPs, alguns aspectos culturais e organizacionais, como a colaboração dinâmica em redes de valor, a colaboração social e a integração de dados e processamento de informação, também mostram relevância para o sucesso da abordagem prescritiva.

Nos últimos 7 anos, vários autores procuraram preencher as lacunas no prognóstico dos aerogeradores e na gestão da saúde. Nesse período, foram publicados 30 artigos de revisão abordando diferentes temas. Uma porcentagem de 36,7% dessas revisões aborda questões relacionadas a algoritmos de inteligência artificial para prognóstico, mostrando o alto interesse dos pesquisadores (BALTAZAR *et al.*, 2019; BERGHOUT; BENBOUZID; AMIRAT, 2023; GAO; LIU, 2021; GARCÍA MÁRQUEZ; PEINADO GONZALO, 2022; KHAN; BYUN, 2024; REZAMAND *et al.*, 2020; SAIDI; BENBOUZID, 2021; SANTIAGO *et al.*, 2024; TENG *et al.*, 2021; YUCESAN; DOURADO; VIANA, 2021; ZHANG; VATN; RASHEED, 2022) por contribuições atuais e potenciais nesta área. Vários estudos (BOYD *et al.*, 2022; BRIJDER *et al.*, 2022; DO; SÖFFKER, 2021; FOX *et al.*, 2022; LEITE; ARAÚJO; ROSAS, 2018; MELLO; FERREIRA DIAS; ROBAINA, 2022; RINALDI; THIES; JOHANNING, 2021; WOO; WHALE, 2022), 26,7% dos artigos, discutem estratégias de manutenção considerando ganhos no prognóstico de falha. Estratégias de monitoramento de condições também despertam interesse de pesquisa em (BADIHI *et al.*, 2022; HELSEN, 2021; LENG *et al.*, 2023; ZIEGLER *et al.*, 2018). Outros artigos de revisão abordam técnicas de monitoramento de condições

(BENBOUZID *et al.*, 2021; MORONEY; VERMA, 2023), uso de algoritmos que não utilizam inteligência artificial (ADEDIPE; SHAFIEE; ZIO, 2020), pré-processamento de sinais (UMA MAHESWARI; UMAMAHESWARI, 2017), confiabilidade (NEJAD *et al.*, 2021) e prognóstico de componentes específicos de aerogeradores (HART *et al.*, 2020; YETER; GARBATOV, 2022). Na literatura existe uma lacuna quanto à investigação da capacidade de generalização de modelos de inteligência artificial para estimar a vida útil remanescente de aerogeradores.

O diagrama da Figura 11 mostra os assuntos abordados pelas revisões, classificados por tema e a lacuna identificada não discutida em artigos de revisão anteriores.

Figura 11 - Artigos de revisão sobre prognóstico de aerogeradores nos últimos 7 anos e lacuna identificada

Fonte: O autor (2024)

2.6. Generalização de métodos para estimativa de vida útil remanescente: Revisão sistemática com metodologia inspirada no protocolo PRISMA

Esta seção apresentará uma revisão sistemática dos métodos de estimativa da VUR para aerogeradores para investigar quais os tipos de conjuntos de dados usados, principais de modelos utilizados, capacidade de prognóstico e desafios de uso no mundo real. Tendo como objetivo identificar a utilização de métodos e modelos capazes de generalizar a estimativa da VUR de componentes de aerogeradores em aplicações do mundo real. As questões de pesquisa, apresentadas na Tabela 1, foram formuladas para orientar a busca em bases de dados e a coleta de dados dos registros de documentos (artigos de periódicos, artigo de conferência, anais de conferência ou capítulo de livro).

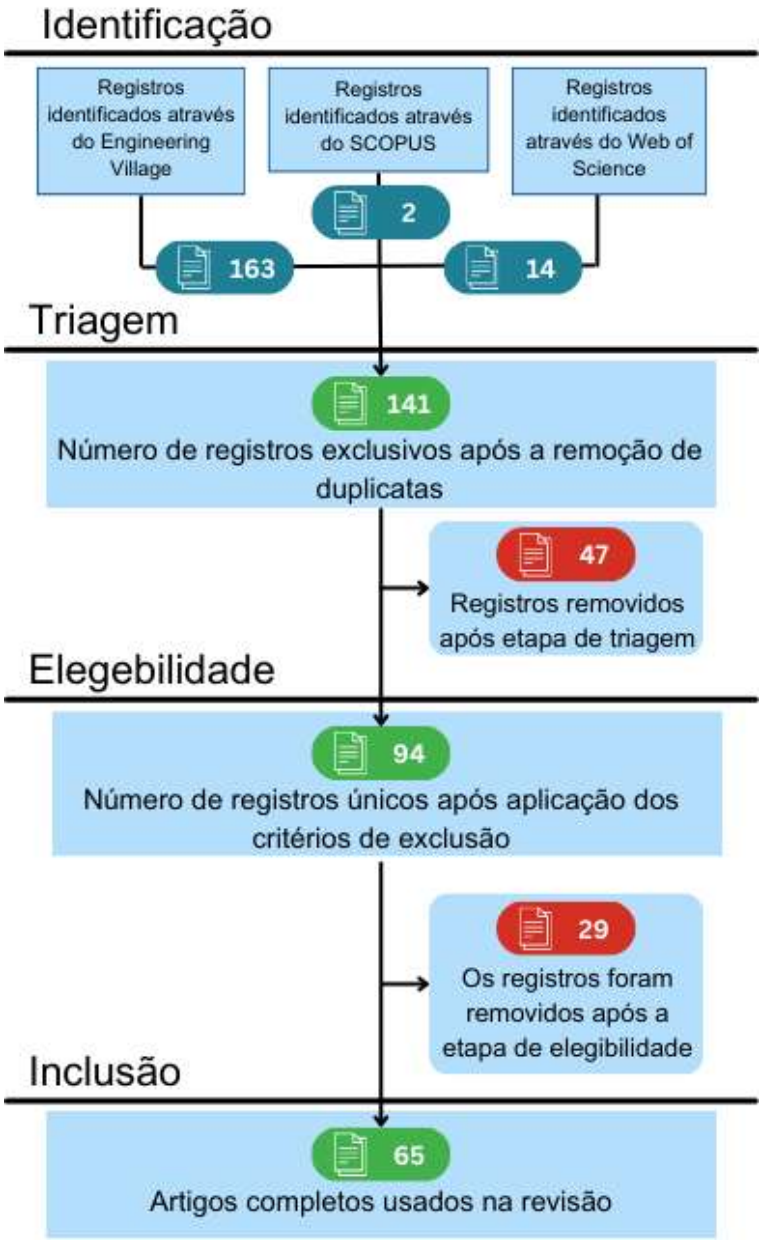
Tabela 1 - Questões de pesquisa

	Questões de Pesquisa
QP1	Quais são os obstáculos para estimar a VUR para componentes de aerogeradores em condições operacionais do mundo real?
QP2	Qual é o nível de generalização alcançado pelos modelos prognósticos para componentes de aerogeradores?
QP3	Quão eficazes são os modelos de estimativa da VUR para componentes de aerogeradores?

Fonte: O autor (2024)

Esta revisão sistemática foi conduzida usando os requisitos do protocolo PRISMA (GALVÃO; PANSANI; HARRAD, 2015; MOHER *et al.*, 2010) para identificar lacunas na literatura sobre a previsão da vida útil restante dos componentes do aerogerador. A Figura 12 mostra um fluxograma com as fases do protocolo PRISMA usado para a revisão sistemática.

Figura 12 – Fluxograma do protocolo PRISMA utilizado para seleção de registros para revisão



Fonte: O autor (2024)

Como desdobramento das questões de pesquisa, foi necessário definir *strings* para uso em mecanismos de busca de bases de dados científicas. Em seguida, foram definidos quatro grupos de palavras-chave para realizar pesquisa bibliográfica (Tabela 2 e Tabela 3). Esses grupos foram combinados para inserir como *strings* de busca em mecanismos de banco de dados selecionados para filtrar registros de documentos relacionados às questões de pesquisa. Inicialmente, para construir a *string* de busca, palavras-chave do Grupo 1 e Grupo 2 foram criadas para construir a *string* de busca. Em seguida, o Grupo 3 e finalmente o Grupo 4 foram sucessivamente incluídos. O número de documentos encontrados foi tabulado e avaliado.

Tabela 2 – Grupos de palavras-chave para encontrar registros relacionados à generalização de métodos ou modelos em bases de dados selecionadas

Grupo 1	Grupo 2	Grupo 3	Grupo 4
Generalization	wind turbine* OR wind farm	-	-
Generalization	wind turbine* OR wind farm	prognostic OR remaining useful life OR RUL or prognosis	-
Generalization	wind turbine* OR wind farm	prognostic OR remaining useful life OR RUL OR prognosis	Reliability-centered maintenance OR RCM

Fonte: O autor (2024)

Tabela 3 - Grupos de palavras-chave para localizar registros relacionados à manutenção prescritiva em bases de dados selecionadas

Grupo 1	Grupo 2	Grupo 3
Prescriptive maintenance	wind turbine* or wind farm	-
Prescriptive maintenance	wind turbine* or wind farm	prognostic or remaining useful life or RUL or prognosis

Fonte: O autor (2024)

Foi decidido utilizar apenas a sequência de busca formada pelo conjunto de palavras-chave dos grupos 1, 2 e 3 juntas para manter uma quantidade razoável para realizar uma revisão.

Primeiramente, na etapa de identificação, os registros de documentos coletados das bases de dados Engineering Village (COMPENDEX), SCOPUS e Web of Science foram quantificados e tabulados. Em seguida, todos os registros de documentos duplicados foram removidos do conjunto de dados. Na terceira etapa, a triagem foi feita considerando os critérios de exclusão indicados na Tabela 4. Após isso, foi realizada a etapa de elegibilidade, na qual os registros que atendem aos requisitos de elegibilidade (Tabela 5) foram lidos na íntegra e alguns dados foram coletados para responder às questões da pesquisa.

Tabela 4 - Critérios de inclusão usados na etapa de triagem

#	Critérios de exclusão
CT1	O registro deve estar em inglês
CT2	O registro deve conter título, ano, fonte, resumo e DOI
CT3	O registro deve ser um artigo de periódico, artigo de conferência, anais de conferência ou capítulo de livro

Fonte: O autor (2024)

Tabela 5 - Critérios de inclusão para etapa de elegibilidade

#	Critérios de elegebilidade
CE1	O registro deve tratar do prognóstico ou estimativa da vida útil remanescente dos aerogeradores
CE2	O texto completo deve estar disponível para os registros encontrados

Fonte: O autor (2024)

A pesquisa bibliográfica foi feita para pesquisar artigos de periódicos, artigos de conferências, anais de conferências e artigos no prelo, livros e capítulos de livros entre novembro de 2022 e janeiro de 2024. Os registros de artigos de revisão, normas e padrões e pré-impressões não foram usados.

Um total de 65 documentos foram lidos completamente após a etapa de elegibilidade. As informações listadas a seguir foram tabuladas para permitir avaliação e representação gráfica para responder às questões de pesquisa.

1) Título
2) Ano
3) Tipo de documento
4) Classificação SCImago
5) O modelo usa conjuntos de dados públicos? (Resposta sim ou não)
6) Dados de um banco de testes, aplicativo de dados do mundo real, sintéticos ou simulados?
7) Pré-processamento (Resposta sim ou não)
8) Qual pré-processamento é feito?
9) Usa conjuntos de dados de operação até a falha? (Resposta sim ou não)
10) Estimativa da VUR com modelos usando dados superajustados? (Resposta sim ou não)
11) Estimativa da VUR por método físico, estatístico ou orientado a dados
12) Qual método/modelo orientado a dados usado
13) Capacidade de generalização (boa, média ou ruim)
14) Com quanto de antecedência o modelo de VUR estima o final de vida?
15) Cálculo do intervalo de confiança (IC) ou a incerteza do cálculo da VUR? (Resposta sim ou não)
16) Objetivo/Meta

2.6.1. Coleta de dados

Para cada um dos artigos que atendiam aos requisitos de elegibilidade, foram incluídos no estudo dois tipos de dados. Primeiramente, foram coletados dados básicos sobre a publicação, incluindo (1) título da publicação, (2) autores, (3) título do periódico, (4) ano da publicação, (5) tipo de documento, (6) DOI, (7) resumo e (8) classificação SCImago (SCIMAGO LAB, 2024) para o periódico, para dar suporte ao processo de elegibilidade.

Após isso, informações detalhadas de cada artigo foram buscadas, tabuladas e analisadas para responder às questões de pesquisa listadas na Seção 2.6.

Para responder à QP1: 'Quais são os obstáculos para estimar a VUR para componentes de aerogeradores em condições operacionais do mundo real?', os dados extraídos de artigos elegíveis foram: (1) o uso de conjuntos de dados públicos de modelos, (2) o uso de bancada de teste, conjunto de dados do mundo real, sintético ou simulado, (3) qualquer pré-processamento de dados realizado e como eles foram pré-processados, (4) tipo (físico, estatístico, orientado a dados ou híbrido) do método usado para cálculo da VUR e (5) qual método/modelo orientado a dados foi usado.

Para responder à QP2: 'Qual é o nível de generalização alcançado pelos modelos prognósticos para componentes de aerogeradores?', foram analisados nos artigos elegíveis os seguintes aspectos: (1) utilização de conjuntos de dados de operação até a falha, (2) estimativa VUR através de modelos, avaliando possível superajuste (*overfit*), e (3) capacidade de generalização do modelo. Esta última foi classificada em três níveis: boa (quando utiliza dados de mais de 3 aerogeradores), média (quando utiliza dados de exatamente 3 aerogeradores) e pequena (quando utiliza dados de menos de 3 aerogeradores).

Para responder à QP3: 'Quão eficazes são os modelos de estimativa de VUR para componentes de aerogeradores?' foram coletados os seguintes dados: (1) com que antecedência da falha o modelo pode estimar a VUR e (2) se o modelo calcula o intervalo de confiança (IC) ou a incerteza para o cálculo da VUR a partir de artigos qualificados. O objetivo/meta dos artigos também foi tabulado para suportar a análise e obter insights para a revisão sistemática.

2.6.2. Resultados da revisão sistemática

Esta Seção apresenta os resultados obtidos a partir da avaliação dos dados coletados, conforme indicado na Seção 2.6.1. Seguindo as diretrizes do protocolo PRISMA, um total de 141 registros únicos foram encontrados em bases de dados científicas e foram submetidos aos

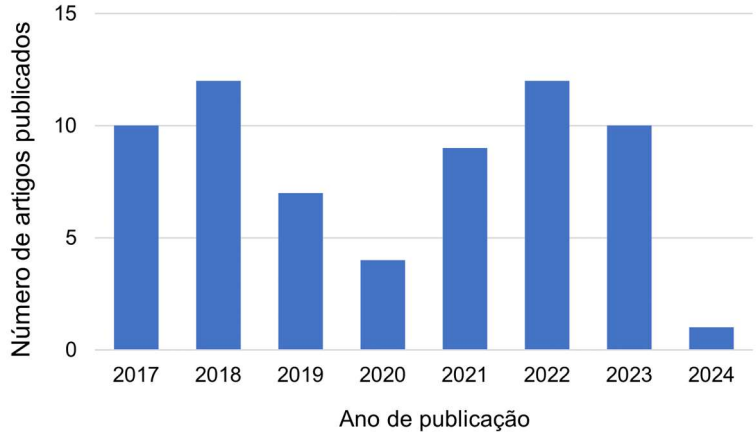
critérios de triagem indicados na Tabela 4. Destes registros, 4 artigos não tinham o texto completo disponível, 2 dos registros eram normas internacionais, 1 era uma dissertação, 1 era um livro, 4 eram artigos em chinês, 7 registros eram artigos de revisão, dos quais 3 foram registrados nas bases de dados como artigos de pesquisa, artigos de conferência ou artigos de periódicos, e 1 artigo era continha plágio e foi publicada uma retratação. Na etapa de elegibilidade, os resumos dos 94 registros foram lidos para identificar quais atendiam aos critérios da Tabela 5. Finalmente, 65 artigos que atendiam aos critérios de elegibilidade foram selecionados e lidos na íntegra para coletar os dados para responder às questões de pesquisa.

Os artigos selecionados foram lidos independentemente da classificação SCImago para identificar a relação entre a qualidade do periódico e a capacidade de generalização dos modelos de artigos.

As publicações que abordaram o tema generalização ou maturidade do modelo tiveram maiores ocorrências nos anos de 2022, 2018, 2023 e 2017 (Figura 13) e os periódicos que mais publicaram artigos abordando generalização estão indicados na Figura 14.

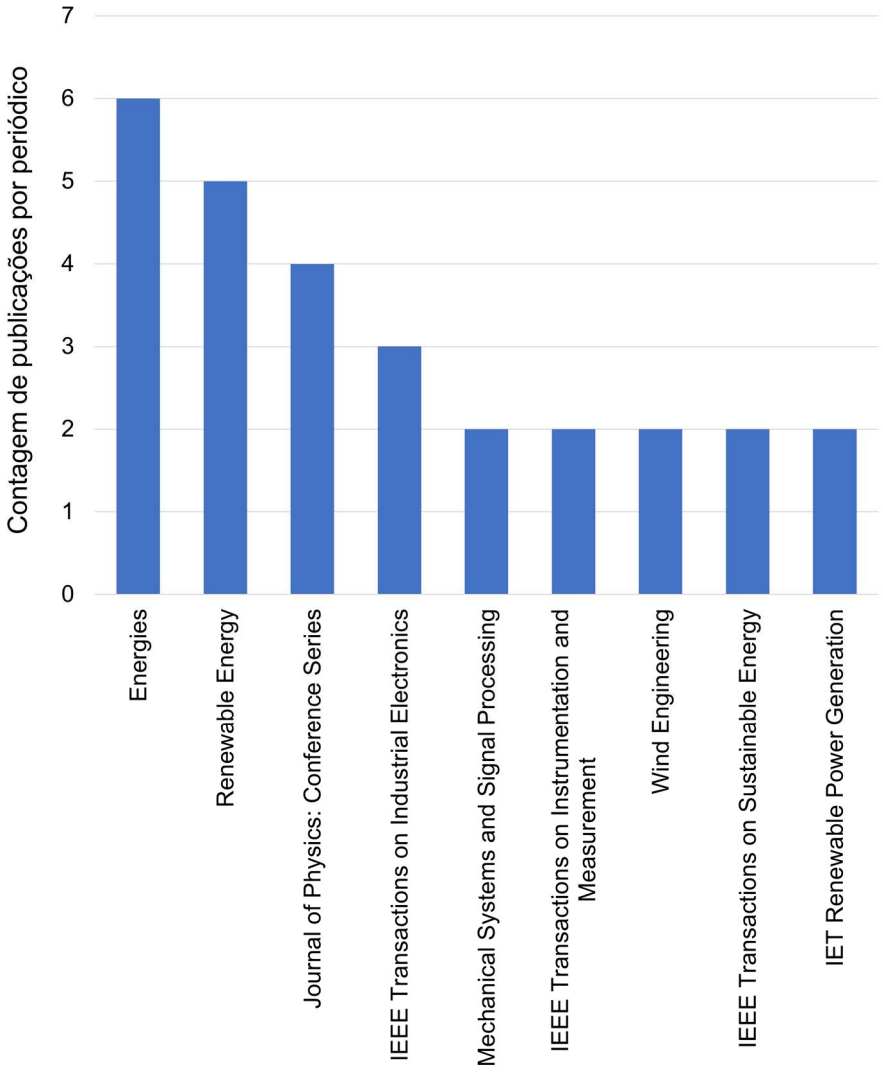
. Verificou-se que 43% - 28 dos 65 registros - foram publicados em 9 periódicos (Figura 14). Ainda 63,1% dos registros, incluindo artigos, artigos de imprensa e artigos de acesso antecipado (*preprint*), eram artigos publicados em periódicos e 36,9% eram artigos publicados como artigos de conferências ou anais de conferências (Figura 15). Isso mostra a relevância do tema para a comunidade científica pelas metodologias já consagradas e pelos novos desenvolvimentos sobre o tema.

Figura 13 – Número de registros de publicações selecionados por ano.



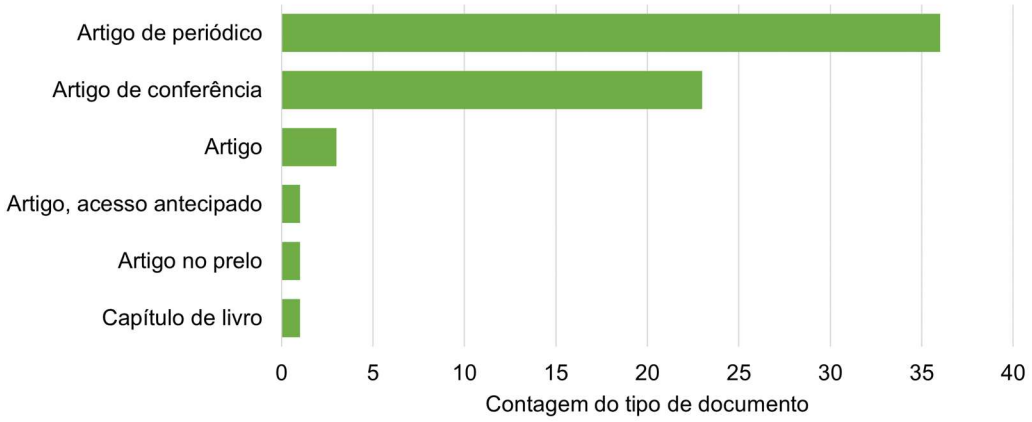
Fonte: O autor (2024)

Figura 14 - Os 9 principais periódicos incluídos na pesquisa em bases de dados de generalização de modelos prognósticos antes da triagem.



Fonte: O autor (2024)

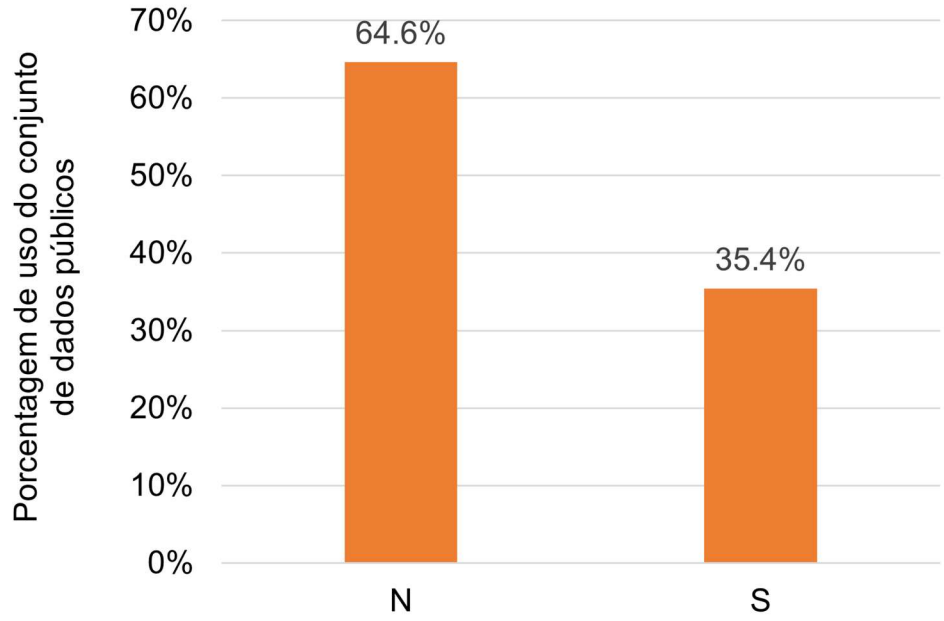
Figura 15 – Classificação de tipos de documentos baseada em bases de dados científicas.



Fonte: O autor (2024)

Identificou-se que a maioria dos artigos utilizou conjuntos de dados privados, não públicos ou abertos, e não estão disponíveis para a comunidade científica e acadêmica, conforme mostra a Figura 16.

Figura 16 – Uso de conjuntos de dados abertos (públicos) utilizados em estudos encontrados em artigos.



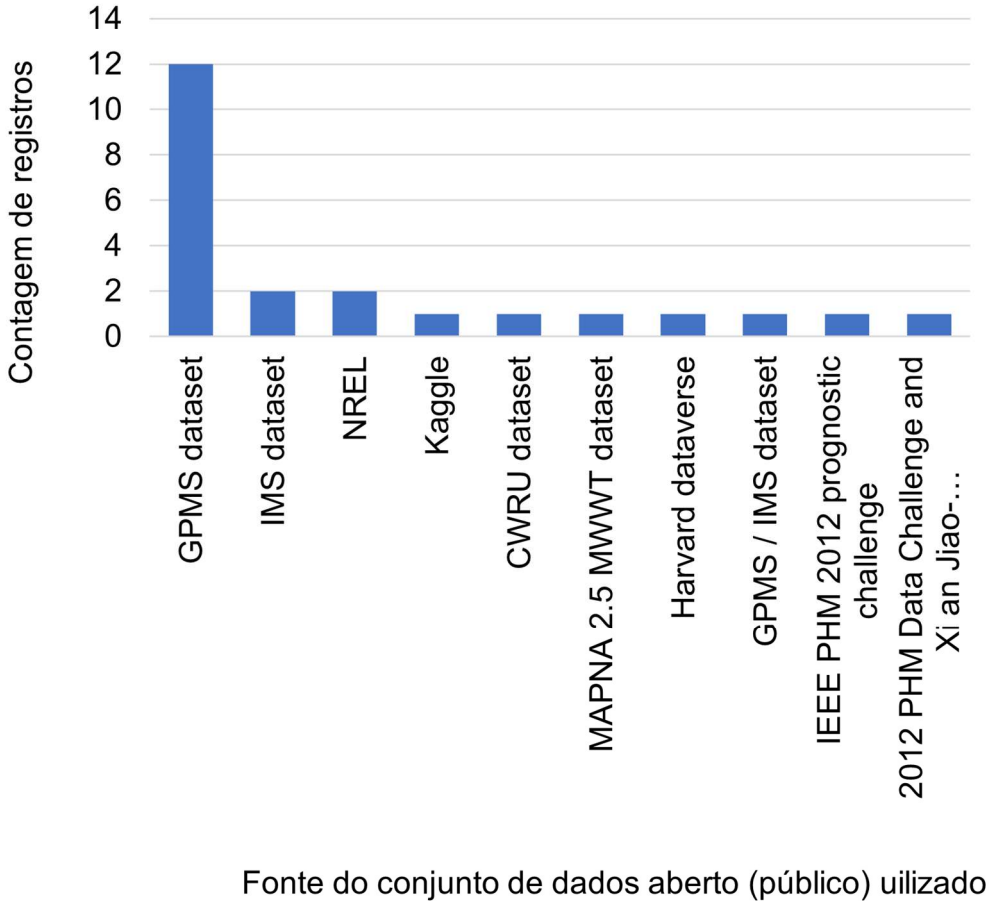
Fonte: O autor (2024)

A Figura 17 mostra as fontes de dados abertos encontradas nos artigos, com destaque para o conjunto de dados Mathworks/GPMS utilizado em 12 artigos.

A indisponibilidade de dados públicos constitui uma barreira a ser superada no estudo prognóstico de determinados componentes do trem de força dos aerogeradores, especialmente aqueles que devem ter uma vida útil longa, como os rolamentos principais.

A utilização de séries temporais com dados até o momento da falha do componente implica num trade-off no prognóstico da falha. A utilização de dados até o fim da vida útil, em condição real de operação do componente, representará a dinâmica da falha com maior fidelidade, mas são de obtenção custosa, pois podem (1) implicar o uso de sensores e equipamentos especiais e (2) na falha catastrófica do aerogerador. Outra forma de obter dados até o fim de sua vida útil é realizar um teste acelerado em bancada com o componente. Esses testes produzem dados com dinâmicas de degradação mais rápidas do que aquelas encontradas na condição operacional dos aerogeradores e, consequentemente, exigem estratégias transformar o sinal de modo que não gerem erros maiores na estimativa do prognóstico.

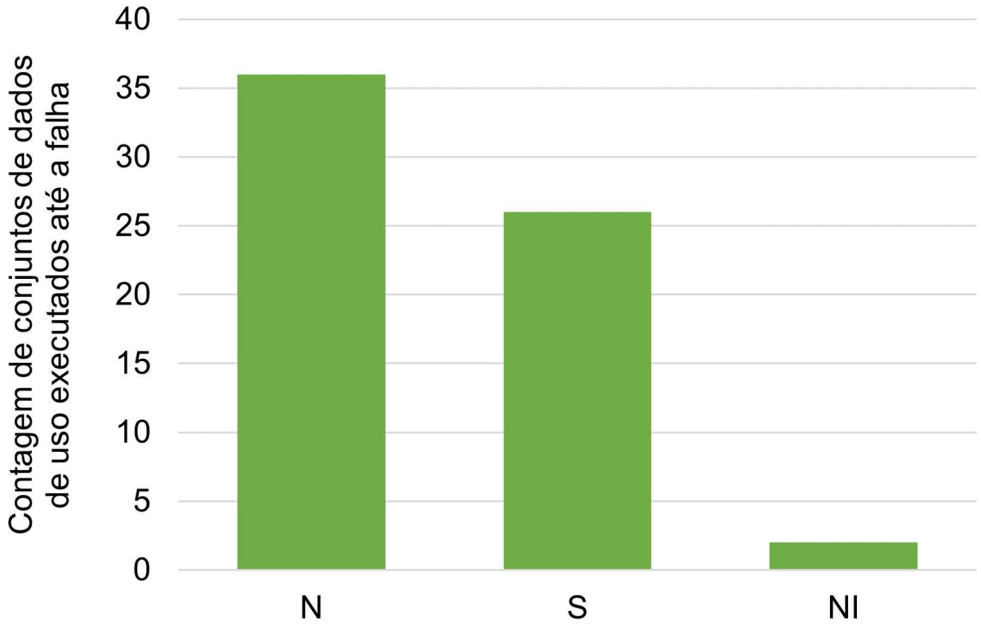
Figura 17 – Fontes de conjuntos de dados abertos encontrados em artigos.



Fonte: O autor (2024)

A Figura 18 mostra que 36 dos 65 (55,4%) estudos escolhidos nesta revisão não utilizam dados (*run-to-failure*) obtidos até que os componentes falhem.

Figura 18 – Uso de conjuntos de dados de execução até falha no estudo.



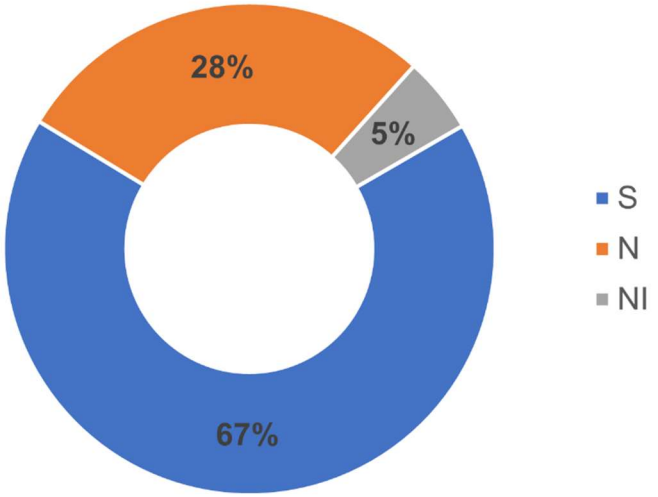
Fonte: O autor (2024)

2.6.2.1. Métodos e modelos para estimar a vida útil remanescente (VUR)

Uma questão fundamental no desenvolvimento de métodos orientados a dados é utilizar um conjunto de dados que represente o comportamento do objeto a ser estudado com a menor influência possível de informações incorretas decorrentes de erros ou mau funcionamento nos sensores, no sistema de aquisição e/ou bancos de dados. Técnicas de pré-processamento tornam o prognóstico de falhas mais confiável (KAMARZARRIN *et al.*, 2022), pois reduzem o número de erros decorrentes de mau funcionamento em sensores e sistemas de aquisição de dados (ZHAO *et al.*, 2017), minimizam ruídos e sobreposição de alarmes (CORREA-JULLIAN *et al.*, 2022; HU; YOUN; WANG, 2018) e reduzem erros na rotulagem de classificação de dados (LEAHY *et al.*, 2018).

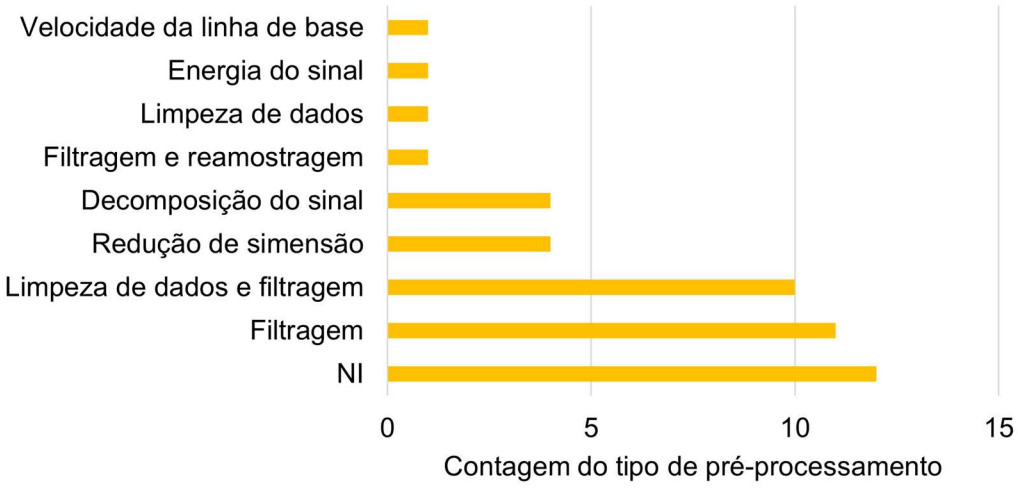
Mais da metade (67%) dos artigos selecionados utilizaram técnicas de pré-processamento de dados, 28% não o fizeram e 5% não informaram (NI) o tipo de pré-processamento, Figura 19. As técnicas mais utilizadas nos artigos foram filtragem de dados (11 artigos - 24,4%) e limpeza e filtragem de dados (10 artigos - 22,2%), técnicas de redução de dimensão variável, por exemplo, PCA (MURPHY, 2012) (4 artigos - 8,9%) e decomposição de sinal - EMD e *Wavelet* (4 artigos – 8,9%), conforme Figura 20. Em 18 artigos, devido aos dados utilizados na modelagem, foram marcados como NA (Não aplicável), são modelos que utilizam resultados de análises de óleo e fadiga ou capítulos de livros. Em 12 artigos (16,7%) os métodos de pré-processamento utilizados não foram indicados.

Figura 19 – Porcentagem de uso de pré-processamento em conjuntos de dados de origem.



Fonte: O autor (2024)

Figura 20 – Grupos de tipos de pré-processamento utilizados.



Fonte: O autor (2024)

Em 10 registros (22,2%) dos artigos foi utilizado um banco de dados público da Green Power Monitoring Systems – GPMS(MATHWORKS INC.; BECHHOEFER, 2017), que possui dados de vibração já processados de um rolamento de eixo de alta velocidade (HSS) de um aerogerador de 2 MW e o pré-processamento não é relatado. Neste caso, a marcação NI foi utilizada para artigos que não informavam o método de pré-processamento.

A falta ou pouco detalhamento do pré-processamento dificulta a compreensão da influência desta etapa na previsão de falhas, conforme destacado por BAPTISTA; GOEBEL; HENRIQUES, (2022).

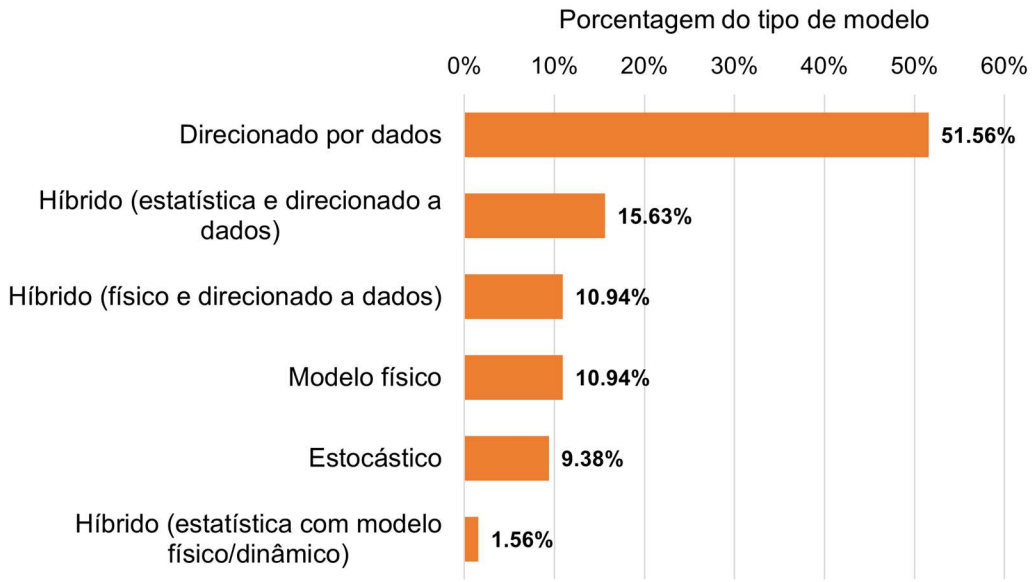
2.6.2.2. Resultados dos modelos e estimativa da vida útil remanescente (VUR)

A escolha do modelo (físico, baseado em dados, estatísticos ou híbrido) para estimar a vida útil remanescente depende de (1) da quantidade, qualidade e tipos de dados disponíveis e (2) da presença de ruído nos dados causado pelo processamento das medições, seja processo ou sistema que executam a coleta, (3) do tipo de reparo (recuperação perfeita ou imperfeita), (4) como modelar o número de modos de falha e (5) como se gerenciam os novos tipos de falha. Porém, o conhecimento técnico do mecanismo de falha e do processo de coleta de dados é essencial para a utilização de modelos prognósticos. (SIKORSKA; HODKIEWICZ; MA, 2011).

Nesta revisão, foi identificado o uso majoritário de modelos baseados em dados (51,56%), híbridos (baseados em dados e estatísticos) (15,63%), híbridos (baseados em dados e físicos) (10,94%) e físicos (10,94%), conforme mostrado na

Figura 21, indicando que a disponibilidade de linguagens de programação de código aberto (Python e R), com uma grande variedade de pacotes e frameworks facilita a implementação de modelos estatísticos e baseados em dados. Os modelos físicos, por outro lado, têm como principal desafio uma alta complexidade matemática e aplicação para resolver um mecanismo de falha individual. (SIKORSKA; HODKIEWICZ; MA, 2011).

Figura 21 – Tipos de modelos para prognóstico de falhas de aerogeradores identificados nos artigos selecionados

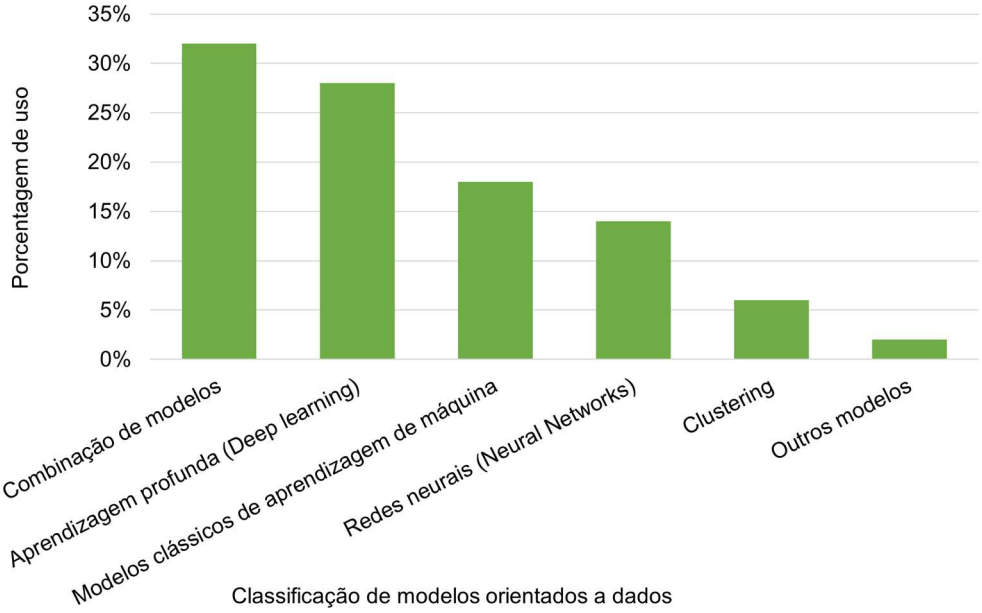


Fonte: O autor (2024)

Para modelos orientados a dados ou híbridos, identificou-se que os métodos mais utilizados para estimativa de VUR são combinações de modelos (GÉRON, 2019; MURPHY, 2012), 32%, desenvolvidos com mais de uma abordagem, abordagem de aprendizagem

profunda (GOODFELLOW; BENGIO; COURVILLE, 2016) 28%, modelos clássicos de aprendizado de máquina (Support Vector Machines (SVM), k-Nearest Neighbours (kNN), Naive Bayes, Logistic Regression e Random Forest) (MURPHY, 2012), correspondem a 18%, os modelos de redes neurais artificiais (RNA's) são 14% conforme mostrado na Figura 22. Clustering (MURPHY, 2012) e outros modelos são os menos utilizados, respectivamente, com percentuais de 6% e 2%.

Figura 22 – Modelos baseados em dados mais comuns para estimativa de VUR para prognósticos de aerogeradores.



Fonte: O autor (2024)

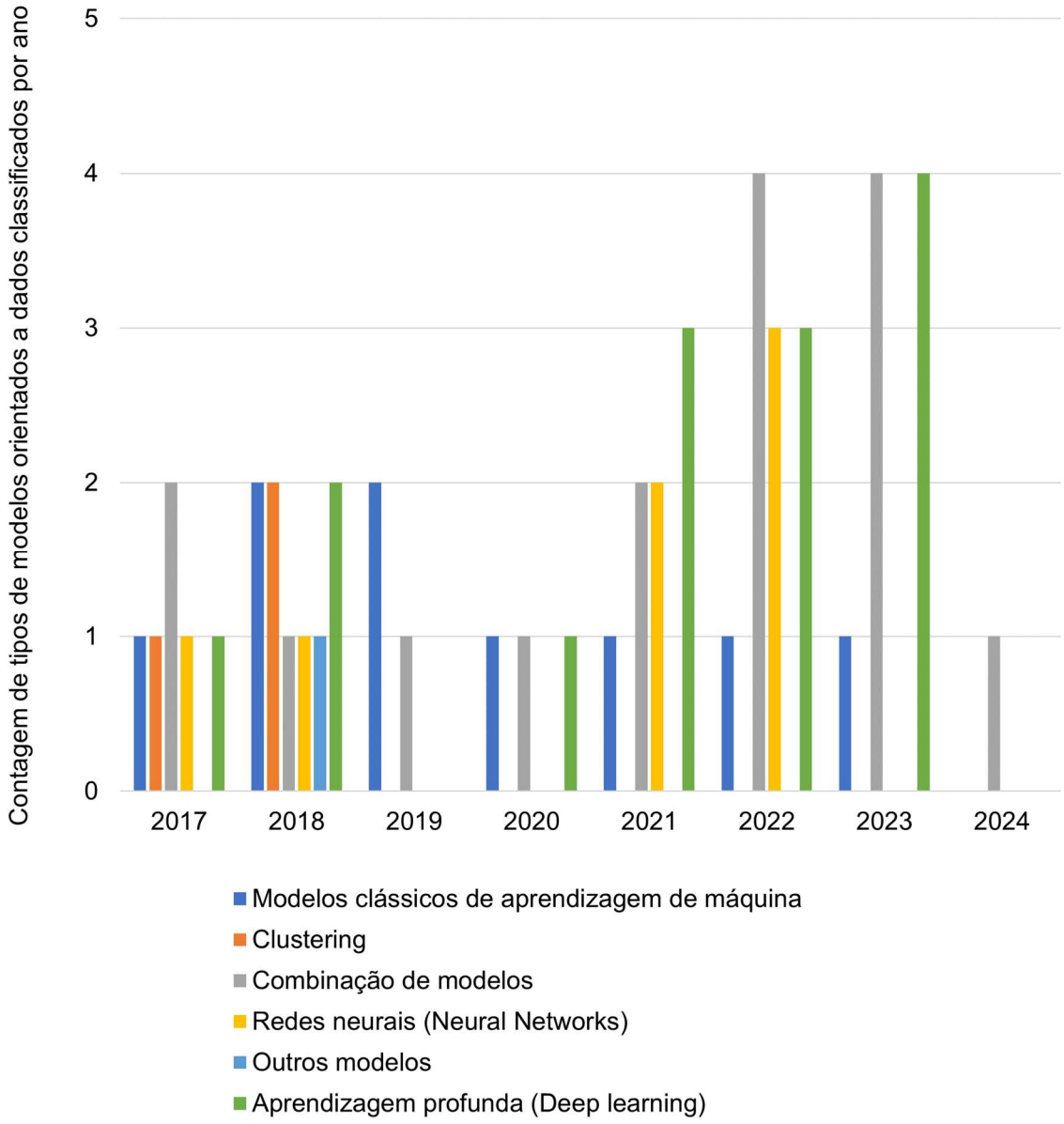
A Figura 23 apresenta os modelos de estimativa da VUR organizados por ano. E deixa claro que a combinação de modelos é a mais utilizada, durante o período de avaliação e, que a utilização de modelos de aprendizagem profunda aumentou desde 2021.

Um dos desafios comuns em modelos baseados em dados é o sobreajuste (*overfitting*) que pode ocorrer ao usar amostras de um único conjunto de dados para treinamento e teste. A utilização de diferentes conjuntos de dados implica um compromisso: por um lado, há uma redução no *overfitting*, mas, por outro, há uma diminuição potencial no desempenho das métricas do modelo baseado em dados.

Neste estudo, os métodos baseados em dados que usaram apenas um único conjunto de dados de aerogeradores foram considerados sobreajustados (*overfitted*). Nos casos em que um

conjunto de dados foi utilizado para treinamento e outro para teste, os métodos foram considerados como não sobreajustados.

Figura 23 – Modelos baseados em dados para estimativa de VUR para prognósticos de aerogeradores organizados por ano.



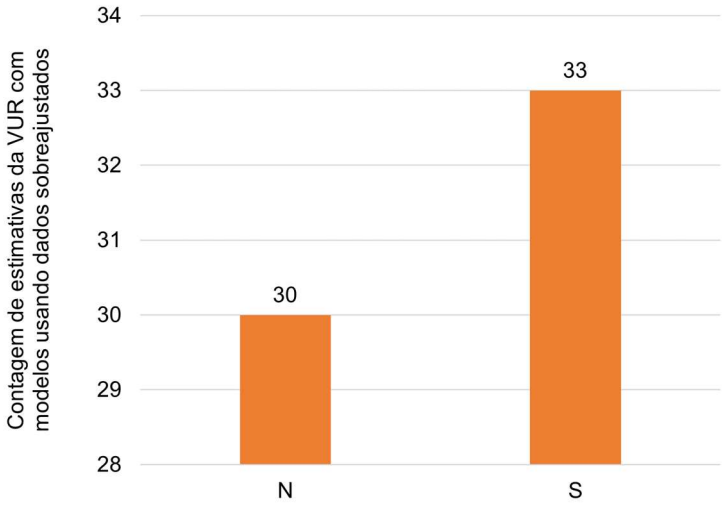
Fonte: O autor (2024)

Identificou-se que 52,4% (33 artigos) utilizaram modelos com dados sobreajustados (*overfitted*), utilizando mesmo conjunto de dados para treinamento e teste, e que 47,6% (30 artigos) não apresentaram sobreajuste, usando conjuntos de treinamento e teste diferentes (Figura 24). A Figura 25 mostra que os artigos de periódicos classificados como SCImago Q2 e Q4 apresentam mais modelos sem dados com *overfitting* do que com overfitting, porém os

periódicos classificados com SCImago Q1 e Q3 apresentaram mais artigos que utilizavam dados com overfitting.

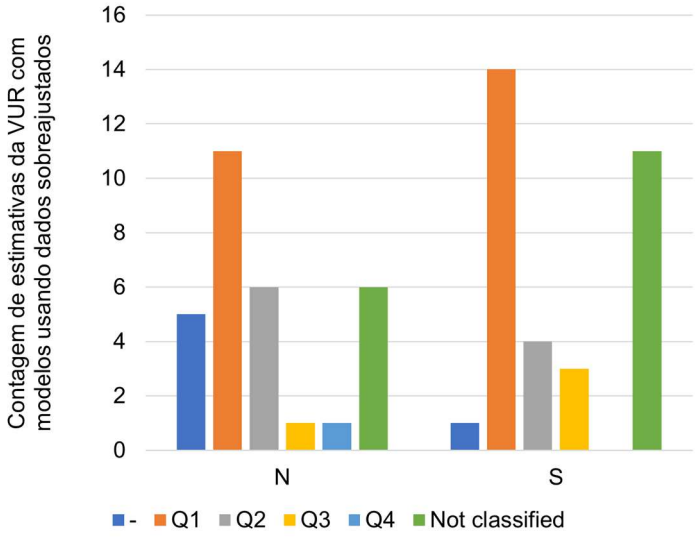
Uma forma de minimizar ou mitigar o problema de sobreajuste (*overfitting*) é maior disponibilidade de dados para uso nas etapas de treinamento, validação e teste. A existência de repositórios de dados abertos ajudaria nesta questão.

Figura 24 – Estimativa da VUR usando dados sobreajustados



Fonte: O autor (2024)

Figura 25 – Estimativa da VUR usando dados superajustados com classificação SCImago.



Fonte: O autor (2024)

As estimativas da VUR são expressas em horas, dias e meses, dependendo da escala de tempo escolhida pelos autores. Apenas 32 dos 65 artigos apresentaram valores numéricos. Observou-se que apesar da dispersão, as estimativas entre 20 e 35 dias são as mais frequentes,

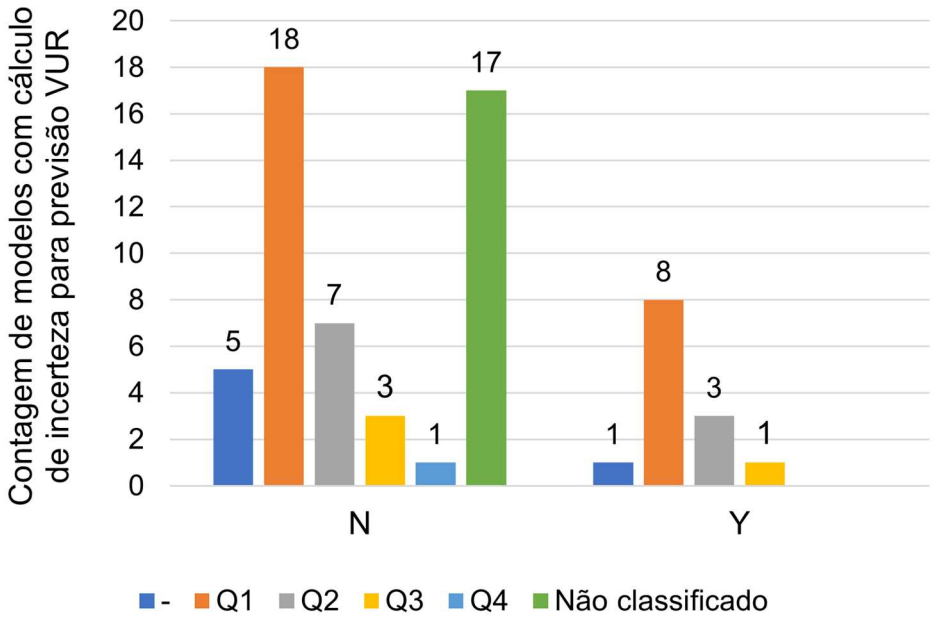
seguidas das estimativas de 65 a 195 dias. Em 30 estudos o valor da VUR não foi estimado ou indicado.

2.6.2.3. Cálculo da incerteza do modelo (ou nível de confiança)

Conforme indicado nas normas (1856-2017 - IEEE Standard Framework for Prognostics and Health Management of Electronic Systems, 2017; VOGL; WEISS; DONMEZ, 2014) e por alguns autores, o cálculo da incerteza ou intervalo de confiança da estimativa da VUR é importante porque se traduz em consistência no desempenho do modelo (SAHU *et al.*, 2023) e tem implicações operacionais, indicando a urgência de intervenção para reparo deve ser realizada (PRAKASH; VUCKOVIC; AMIN, 2023).

Contudo, observou-se que em apenas 13 dos 64 (20,3%) artigos avaliados é computada a incerteza na estimativa da VUR (CHENG et al., 2019; CHENG; QU; QIAO, 2018; DU et al., 2023; HU et al., 2018; KAMARZARRIN et al., 2022; LI et al., 2022; MEHLAN; KELLER; NEJAD, 2023; REZAMAND et al., 2019; SAIDI et al., 2018, [s. d.]; WIESE et al., 2020; XIANG et al., 2023; XU et al., 2024). Observou-se que, nos 26 artigos publicados nas revistas SCImago Q1, 18 não calculam o intervalo de confiança e 8 calculam o intervalo de confiança, conforme mostrado na Figura 26.

Figura 26 – Intervalo de confiança ou cálculo de incerteza da estimativa da VUR com classificação SCImago



Fonte: O autor (2024)

No entanto, o cálculo de um intervalo de confiança para estimativas da VUR requer dados abundantes, conforme observado por PRAKASH; VUCKOVIC; AMIN (2023) e CHENG et al. (2019). A falta de um repositório aberto de dados sobre aerogeradores representa um obstáculo ao desenvolvimento de modelos que prevejam a VUR com intervalos de confiança. Essa escassez ajuda a explicar por que os estudos usam, mais comumente, dados proprietários para prever falhas em aerogeradores.

Também observou-se que a maioria das estimativas da VUR não levou em conta a incerteza, colunas indicadas em cinza. As colunas em azul indicam os estudos que calcularam a incerteza no cálculo da VUR.

2.6.2.4. Capacidade de generalização baseada na minimização de overfitting

ZHOU et al., (2023) destacam que a generalização de domínio (DG) visa alcançar a generalização fora de distribuição usando apenas dados de origem para aprendizado de modelo, melhorando as capacidades de aprendizado de máquina em diversas aplicações.

SUN; SUN, (2018) destacam que um único modelo aplicado sob diferentes condições e dados utilizados para treinamento pode ter desempenho diferente devido a volatilidades ambientais, diferentes modos de degradação e discrepâncias operacionais. E propõe análise de variância para diferenciar as tendências subjacentes dos dados e seleção adequada de amostras de treinamento com níveis flutuantes semelhantes.

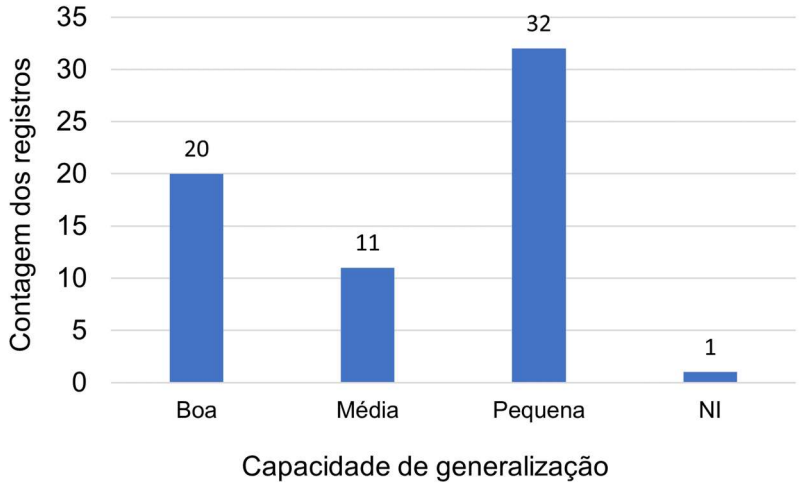
A capacidade de generalização permite que modelos clássicos de aprendizado de máquina (*machine learning*), redes neurais artificiais (RNA's), redes neurais recorrentes (RNN's) e aprendizagem profunda (*deep learning*) trabalhem com vasta infraestrutura de dados com tratamentos matemáticos e estatísticos intensificados por diferenças nos dados permitindo que os modelos façam previsões precisas (MACKENZIE, 2015), melhorem o aprendizado. processo e capacidade em diversas aplicações (GALICKI *et al.*, 2004; ZHOU *et al.*, 2023).

Nesta revisão, os métodos e modelos utilizados para estimativa de VUR de aerogeradores, foram classificados com base em suas capacidades de generalização. Isso foi feito, considerando que 76,9% dos métodos revisados eram baseados em dados ou híbridos e que as boas práticas recomendam separar os dados em conjuntos distintos para treinamento, validação e teste e garantir quantidade e diversidade suficiente para evitar *overfitting*. Sendo assim, foi definido que os métodos desenvolvidos usando dados de mais de 3 aerogeradores foram classificados como tendo boa capacidade de generalização. Os métodos que utilizaram dados de 3 ou menos turbinas foram classificados como capacidade de generalização média. Os

métodos que utilizam dados de menos de 3 conjuntos foram classificados com capacidade de generalização fraca ou ruim.

Como ilustrado na Figura 27, 20 métodos foram classificados como tendo boa generalização, 9 como médios e a maioria (32) como fraca ou ruim. Um dos estudos não indicou quantas turbinas foram usadas no conjunto de dados. O particionamento adequado dos dados pode ajudar a avaliar até que ponto os métodos se estendem a novas situações, em vez de meramente memorizar amostras existentes.

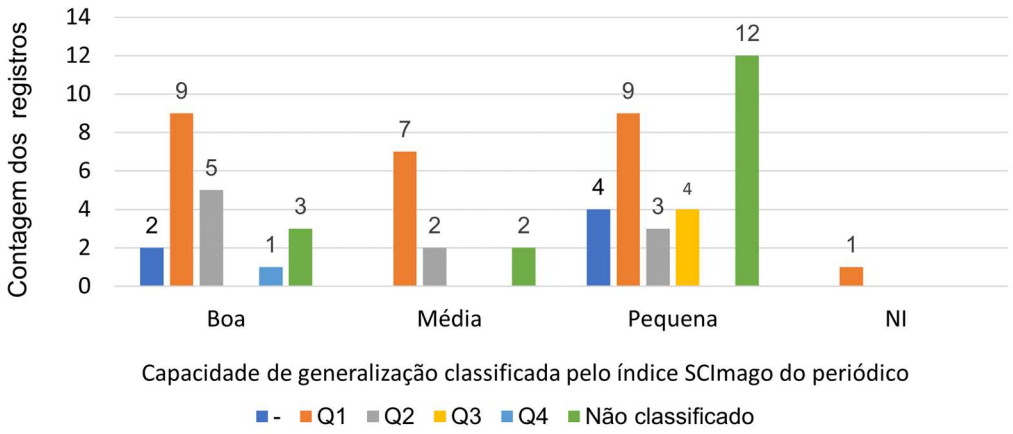
Figura 27 – Capacidade de generalização de modelos baseados em critérios desenvolvidos.



Fonte: O autor (2024)

Observou-se também que no extrato SCImago Q1, 9 métodos apresentam baixa capacidade de generalização, 7 métodos média e 9 boa capacidade de generalização, indicando uma distribuição e aceitação uniforme pelos periódicos sem uma avaliação mais rigorosa do *overfitting* dos dados (Figura 28).

Figura 28 – Capacidade de generalização de modelos baseados em critérios de desenvolvimento versus classificação de periódicos SCImago.



Fonte: O autor (2024)

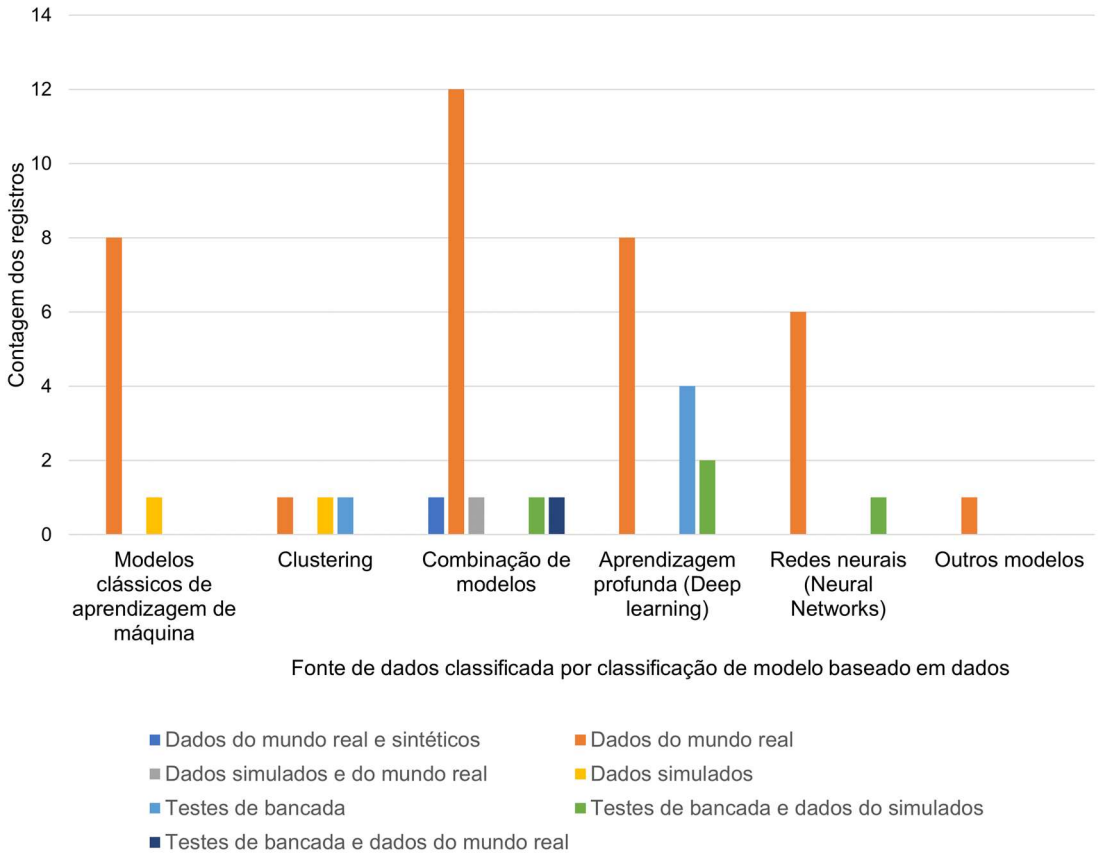
A Tabela 6 apresenta as vantagens e desvantagens dos métodos mais utilizados na literatura visando generalizar as estimativas da VUR.

Observou-se que os métodos utilizam principalmente dados reais para estimar a VUR (Figura 29), destacando as combinações de modelos, os modelos de aprendizagem profunda e os modelos clássicos de ML. Os métodos que utilizam combinações de modelos e os de deep learning se destacam por serem os que mais apresentam o cálculo da incerteza da estimativa VUR (Figura 30).

Observando a Tabela 6, é possível verificar que os métodos que utilizam os modelos de aprendizagem profunda, ML clássico e modelos de *clustering* podem indicar o fim da vida (EOL) com maior antecipação do que outros.

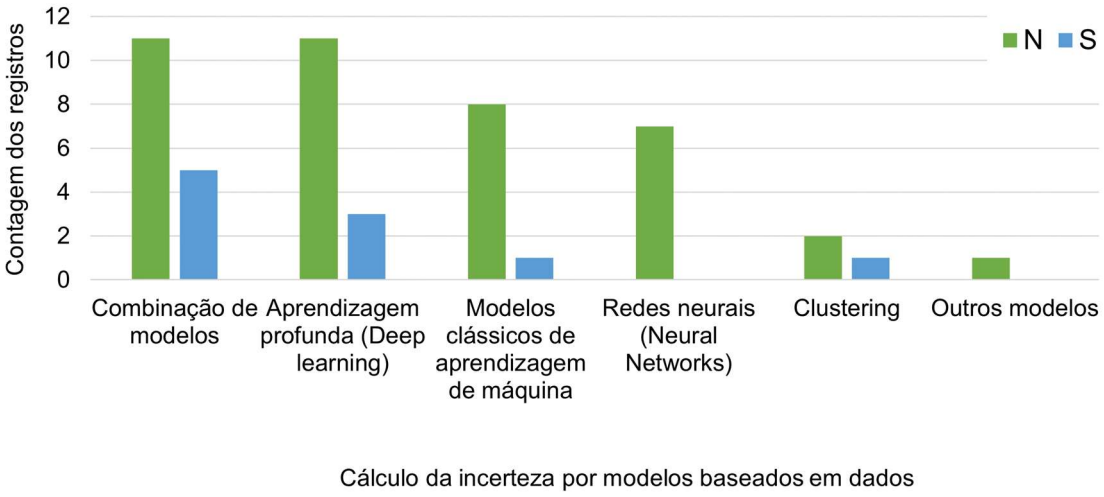
Além disso, devido à utilização de conjuntos de dados com maior número de turbinas, os modelos clássicos de ML, as combinações de modelos e os modelos de RNA's apresentam, respectivamente, a 1ª, 2ª e 3ª melhores capacidades de generalização das estimativas de VUR.

Figura 29 – Fonte de conjuntos de dados usados para cálculo da VUR por diferentes modelos baseados em dados.



Fonte: O autor (2024)

Figura 30 – Cálculo da incerteza considerando a classificação do modelo.



Fonte: O autor (2024)

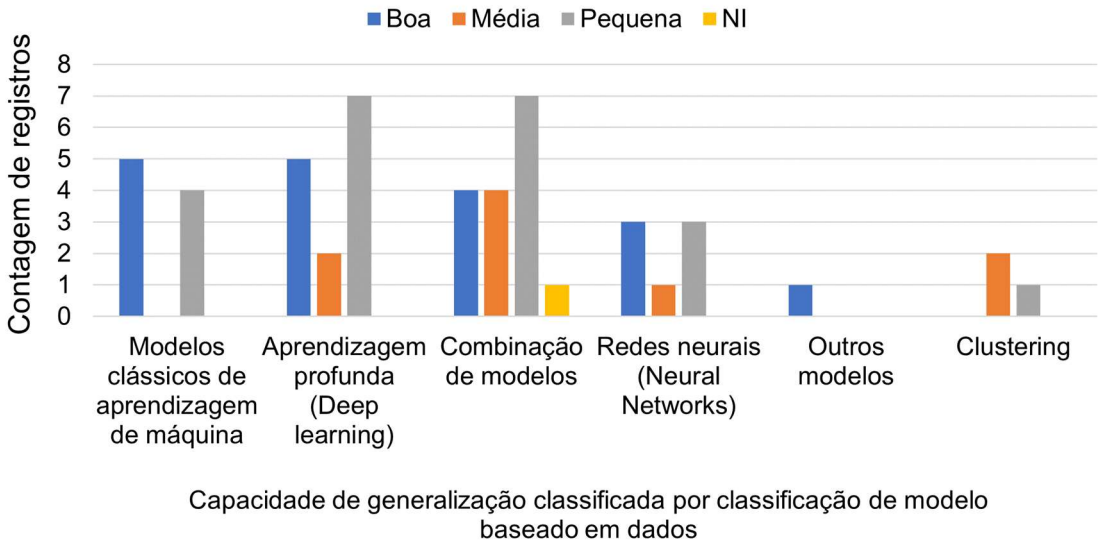
Tabela 6 - As vantagens e desvantagens dos métodos mais utilizados conforme revisão da literatura.

Tipo do Modelo	Vantagens	Desvantagens
Combinação de modelos	2ª melhor capacidade de generalização	Na maioria dos modelos, não computa incerteza.
Aprendizagem profunda (*Deep learning*)	Bom desempenho de métricas. Bom custo computacional	A maioria dos modelos desenvolvidos não calcula a incerteza. Apresenta capacidade de generalização altamente variável, de média a pequena
Modelos clássicos de aprendizagem de máquina (*Machine learning*)	1ª melhor capacidade de generalização Baixos custos computacionais, para modelos com pequeno número de variáveis. A maioria dos modelos desenvolvidos apresenta boa capacidade de generalização	Não calcula a incerteza.
Redes neurais artificiais (RNA's)	3ª melhor capacidade de generalização. Todos os modelos usam conjuntos de dados do mundo real.	Não calcula a incerteza. Não apresenta a estimativa da VUR. Apresenta habilidades de generalização altamente variáveis
Agrupamento (*Clustering*)	-	50% dos métodos calculam a incerteza

Fonte: O autor (2024)

A Figura 29 mostra a relação entre os diferentes métodos e fontes de dados. A Figura 30 mostra os tipos de modelos que consideram o cálculo da incerteza (em verde) e não (em azul) e a Figura 31 mostra a capacidade de generalização separada por tipo de modelo.

Figura 31 – Capacidade de generalização de diferentes modelos baseados em dados para computação da VUR.



Fonte: O autor (2024)

2.6.3. Descobertas acerca dos prognósticos de falhas e os ganhos potenciais

Nesta Seção, apresentamos uma discussão dos achados durante a leitura, coleta de dados e avaliação crítica dos artigos selecionados na revisão sistemática para responder às questões de pesquisa elencadas na Seção 2.6.1.

Primeiramente, observou-se que os modos de falha identificados nas estimativas de prognóstico de falha foram desgaste dos rolamentos HSS (23,4%), falhas do gerador (9,52%) e falhas nos rolamentos principais (7,94%). Falhas na caixa de engrenagens e falhas no sistema de passo são responsáveis, cada uma, por 6,3% dos modelos de falha. As falhas nos rolamentos do gerador, na caixa de engrenagens, no gerador e falhas no enrolamento do gerador são o quinto conjunto de modos de falha mais frequentes, representando 3,1% cada, conforme mostrado na Figura 32. Destaca-se o uso intensivo nos artigos selecionados de um conjunto de dados público com dados do HSS (MATHWORKS INC.; BECHHOEFER, 2017).

Figura 32 – Os modos de falha foram identificados nos artigos selecionados

Fonte: O autor (2024)

No contexto da real condição de operação dos aerogeradores, buscando responder ao QP1, destacamos a seguir os obstáculos e oportunidades de melhorias que impactam a realização das estimativas de RUL, com base nas avaliações dos artigos selecionados nesta revisão:

Conforme identificado nesta revisão sistemática, ver Figura 16, um pequeno número de métodos (AFRIDI et al., 2023; BINSBERGEN et al., 2022; DU et al., 2023; HUANG; TANG; VANZWIETEN, 2022, [s. d.]; JELLALI; MAATALLAH; OUNI, 2022; KAMARZARRIN et al., 2022; LAN et al., 2023; LI et al., 2021, 2017; MEHLAN; KELLER; NEJAD, 2023; XU et

al., 2024; ZHAO et al., 2023) faz uso de bases de dados públicas (HAGMEYER; MAUTHE; ZEILER, 2021) relevantes para o setor de energia eólica.

A disponibilidade de repositórios abertos, com dados operacionais reais, pode beneficiar a academia e o meio produtivo no desenvolvimento de abordagens mais robustas e com maior capacidade de generalização.

Usar dados reais no desenvolvimento, validação e teste de modelos baseados em dados pode ser um desafio devido à complexidade dos dados e ao pré-processamento intensivo.

A utilização de conjuntos de dados sintéticos, conjuntos de dados obtidos por simulações ou testes de bancada facilitam o desenvolvimento de métodos baseados em dados, pois são mais simples de tratar, identificar e analisar o comportamento de degradação de um determinado componente, conforme observado em alguns estudos (AFRIDI *et al.*, 2023; CHENG; QU; QIAO, 2018; DING *et al.*, 2018; DING; TIAN, 2021; DJEZIRI; BENMOUSSA; SANCHEZ, 2018; DU *et al.*, 2023; HONG *et al.*, 2017; KAMARZARRIN *et al.*, 2022; KONG *et al.*, 2023; KÖNIG *et al.*, 2021; LAN *et al.*, 2023; LI *et al.*, 2017; PENG *et al.*, 2017; ZHANG *et al.*, 2021). A utilização inicial de dados sintéticos, simulados ou de bancada possibilita usar uma estratégia de aumento gradual da complexidade das informações de entrada, permitindo a utilização de diferentes pré-processamentos e simplificando a modelagem e interpretação dos resultados obtidos.

Desta forma, ao compreender o comportamento de um determinado modelo, é possível realizar as etapas de treinamento, validação e teste com dados reais de sistemas de aerogeradores SCADA e CMS e observar o desempenho com essas entradas mais complexas. Entendemos também que esta abordagem reduz potencialmente o *overfitting*, conforme identificado em artigos nos quais foram utilizados dados sintéticos e testes de bancada (DAMESHGHI; REFAN, 2019), dados simulados e dados reais (LI *et al.*, 2022) ou conjuntos de dados de teste usando dados de bancada e reais (KRAMTI *et al.*, 2021; LI *et al.*, 2021).

No caso de aerogeradores, é necessário pré-processar os dados do sistema SCADA (TUTIV'EN *et al.*, 2021; ZHAO *et al.*, 2017), com foco na limpeza e balanceamento dos dados, e do CMS (MAZIDI *et al.*, 2017), executando a limpeza, aumentando a relação sinal-ruído (SNR) e escolha adequada das características estatísticas. E estes dados devem possuir comportamento monotônico, com tendência e capacidade prognóstica.

Os métodos que apresentaram capacidade de generalização boa ou intermediária utilizaram principalmente as técnicas de pré-processamento como limpeza de dados (AFRIDI *et al.*, 2023; CHEN *et al.*, 2017; LEAHY *et al.*, 2017, 2018; MAMMADOV; FARROKHABADI; CANIZARES, 2021; MAZIDI *et al.*, 2017; PETER *et al.*, 2022;

REZAMAND *et al.*, 2019; ZHANG; LANG, 2020) e redução de dimensionalidade (DJEZIRI; BENMOUSSA; SANCHEZ, 2018; ZHAO *et al.*, 2017), filtragem de dados (CHEN *et al.*, 2017; CHENG; QU; QIAO, 2018; HENKEL; WEIJTJENS; DEVRIENDT, 2019; LEAHY *et al.*, 2017, 2018; LEE; KANG; LEE, 2023; MAMMADOV; FARROKHABADI; CANIZARES, 2021; MAZIDI *et al.*, 2017; MEHLAN; KELLER; NEJAD, 2023; PETER *et al.*, 2022; REZAMAND *et al.*, 2019; SAIDI *et al.*, 2018; SANTOS *et al.*, 2022; XIANG *et al.*, 2023; XU *et al.*, 2024; ZHANG; LANG, 2020), decomposição de sinal (CAO *et al.*, 2018; DU *et al.*, 2023) e velocidade da turbina (velocidade basal) (LI *et al.*, 2022).

Consideramos que 68,5% dos artigos utilizaram métodos direcionados por dados devido à sua maior facilidade de implementação. Também foi observada a oportunidade de uma maior utilização de modelos CNN e *autoencoders* (CORREA-JULLIAN *et al.*, 2022; HUANG; TANG; VANZWIETEN, 2022b; LI *et al.*, 2021; TUTIV'EN *et al.*, 2021), identificados em combinações de modelos.

Esta revisão também observou que certos artigos (CAO; QIAN; PEI, 2019; DAMESHGHI; REFAN, 2019; KÖNIG *et al.*, 2021), descrevendo os métodos e modelos baseados em dados com mais profundidade, permitiriam que outros pesquisadores replicassem a análise ou adaptassem as técnicas para seus propósitos de modelagem. A comunicação clara das especificidades técnicas é importante para o avanço no campo através da reimplementação e melhoria dos esforços anteriores.

Portanto, consideramos que os obstáculos para a realização de estimativas da VUR em condições reais de operação estão relacionados: (1) ao baixo volume de dados, de diferentes aerogeradores, disponíveis para aplicação dos métodos, (2) à ausência de uma estratégia de utilização de dados sintéticos, simulados ou testes de bancada que possibilitariam a compreensão do desempenho dos modelos para depois aplicá-los a dados reais e (3) a necessidade de utilização de técnicas intensivas de pré-processamento dos dados, principalmente os de vibração, a serem utilizados pelos métodos.

Para responder à QP2, detalhamos a seguir questões e oportunidades de melhorias relacionadas ao nível de generalização obtido pelos métodos e modelos de prognóstico de aerogeradores, utilizando as avaliações dos artigos selecionados.

A representação do comportamento de degradação de um determinado componente, obtido por métodos baseados em dados, é melhorada com a disponibilidade de medições de grandezas operacionais até o momento da falha. No entanto, a sua aquisição envolve riscos que podem resultar em falhas catastróficas do aerogerador. Ademais, estudos que utilizaram esse tipo de dado demonstraram pequena capacidade de generalização (ABID; SAYED-

MOUCHAWEH; LAURENCE, 2019; CAO; QIAN; PEI, 2019; DAMESHGHI; REFAN, 2019; HAYDER; SAIDI, 2021; HU et al., 2018; HUANG; TANG; VANZWIETEN, 2022b; KAMARZARRIN et al., 2022; KERBOUA; KELAIAIA, 2022; LI et al., 2017; PENG et al., 2017; SAIDI et al., [s. d.], [s. d.]; SAIDI; HAYDER; SAIDI, 2022) ou média capacidade de generalização (CAO *et al.*, 2018; CHENG; QU; QIAO, 2018; HUANG; TANG; VANZWIETEN, 2022b; KRAMTI *et al.*, 2018; SAIDI *et al.*, 2018; ZHANG; LANG, 2020) de acordo com os critérios apresentados na Seção 2.6.2.4, não justificando sua utilização devido aos custos e riscos para se dispor destes dados.

Para obter um método capaz de generalizar os resultados, é essencial eliminar ou reduzir ao máximo o *overfitting*. Apenas 43,9% dos métodos identificados nesta pesquisa utilizaram estratégias para reduzir o *overfitting*.

Conforme destacado Seção 2.6.2.4, a capacidade de generalização dos métodos baseados em dados dependerá da utilização de uma metodologia que garanta um bom desempenho em condições diferentes dos dados de treinamento. Usar uma métrica da quantidade de dados de diferentes turbinas pode ajudar a obter métodos mais robustos e com desempenho mais estável. Entende-se que a combinação do uso de um volume de dados de 3 ou mais aerogeradores e a seleção apropriada de dados para os subconjuntos de treinamento, validação e teste resultarão em maior generalização dos métodos.

Para responder à QP3, discutiremos questões e oportunidades relativas à eficácia de métodos de estimativa da VUR visando a área de manutenção de aerogeradores. As estimativas da VUR mais frequentes foram entre 20 e 35 dias antes da falha, mas apenas 49% dos métodos dos artigos selecionados apresentaram o valor desta estimativa. Os demais métodos apresentam apenas resultados de indicadores de saúde ou métricas de desempenho (precisão, recall, f1 score, por exemplo).

Os valores das estimativas de VUR encontrados nos artigos podem contribuir para o planejamento da manutenção. No entanto, métodos que consigam fazer estimativas confiáveis num horizonte de 1 ano têm o potencial de contribuir estrategicamente para a manutenção, considerando que alguns parques eólicos planejam paradas gerais em períodos de ventos mais fracos.

Consideramos essa desconexão em alguns estudos entre o prognóstico de falha e a estimativa da VUR como uma lacuna, o que inviabiliza o uso eficaz dos métodos para aplicações no mundo real.

Outra lacuna e oportunidade de melhoria no prognóstico de falhas de aerogeradores é a incorporação do cálculo de incerteza nas estimativas de VUR, pois apenas 16,8% dos estudos selecionados consideraram incerteza em seus resultados.

A Tabela 7 sintetiza os achados da revisão sistemática realizada de modo a identificar as lacunas em relação à estimativa da VUR e a capacidade de generalização dos métodos e modelos.

Tabela 7 - Considerações sobre problemas e oportunidades de melhoria para prognósticos de aerogeradores

	Questão de pesquisa	Problemas associados	Oportunidades de melhoria
QP1	Quais são os obstáculos para estimar a VUR para componentes de aerogeradores em condições operacionais do mundo real?	Baixa disponibilidade de conjuntos de dados públicos; Ausência de conjunto de dados sintéticos, simulados ou experimentais para avaliação de métodos anteriores; Necessidade de técnicas intensivas de pré-processamento.	Criação de um repositório público de conjuntos de dados para detecção, diagnóstico e prognóstico de AE.
QP2	Qual é o nível de generalização alcançado pelos modelos prognósticos para componentes de aerogeradores?	Indisponibilidade de conjuntos de dados de 3 ou mais AE; Seleção adequada de subconjuntos de treinamento, validação e teste.	Disponibilidade de conjunto de dados com medições de pelo menos 4 AE; Estratégias para construir subconjuntos para lidar com a escassez de dados especifica para AE.
QP3	Quão eficazes são os modelos de estimativa da VUR para componentes de aerogeradores?	Prognósticos de falha sem cálculo estimado VUR; Baixo número de métodos que calculam a incerteza.	Estimativas de longo prazo da VUR; Estimativa da VUR com cálculo de incerteza.

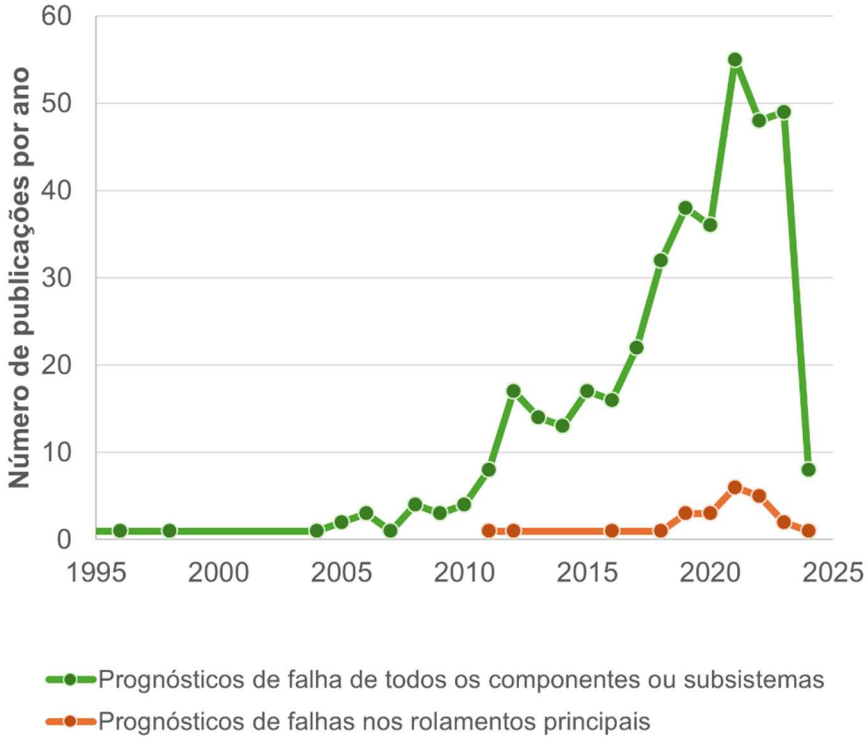
Fonte: O autor (2024)

2.7. Prognóstico de falhas em rolamentos principais de aerogeradores

A próxima fronteira que a indústria de energia eólica busca alcançar é a realização de prognósticos de falhas com base em dados coletados por técnicas preditivas, já utilizadas em diversas plantas. Essas técnicas são particularmente importantes para componentes do aerogerador que envolvem altos custos de aquisição e substituição, além de tempos de inatividade (*downtime*) que reduzem a disponibilidade da máquina e a produtividade da usina. Conforme destacado na Seção 2.3, apesar da proliferação de artigos desde 2014 sobre métodos

e modelos baseados em dados para avaliação da VUR e do conjunto cinemático, especialmente em redutores de velocidade de aerogeradores, há um número significativamente menor de estudos focados no prognóstico de rolamentos principais (HART *et al.*, 2020; LEITE; ARAÚJO; ROSAS, 2018), conforme comparação mostrada na Figura 33.

Figura 33 - Número de registros de publicações relacionadas a prognósticos de falhas de todos os componentes ou subsistemas e prognósticos de falhas do rolamento principal dos aerogeradores da Engineering Village (COMPENDEX).



Fonte: O autor (2024).

Alguns estudos (YUCESAN; VIANA, 2019, 2020), empregando modelos híbridos com redes neurais profundas, concentram-se nas principais estimativas de fadiga de rolamentos. Outros estudos avançam na estimativa da VUR do rolamento principal utilizando modelos de redes neurais naturais (do tipo Long-Short Term Memory - LSTM) (HERP; PEDERSEN; NADIMI, 2019). Os autores enfatizam a necessidade de manipulação de dados, aumento de dados (YUCESAN; VIANA, 2019) e reamostragem (HERP; PEDERSEN; NADIMI, 2019) para estimar fadiga ou VUR.

Os desafios no prognóstico de falhas dos mancais principais de aerogeradores decorrem da falta de compreensão da relação entre carga e dano, bem como da frequência e modos de falha em aplicações reais (HART *et al.*, 2020).

O cálculo preciso da estimativa VUR é crucial para uma manutenção economicamente viável e para maior disponibilidade de aerogeradores. Algoritmos de aprendizado de máquina e inteligência artificial permitem a utilização de uma enorme quantidade de dados do sistema de manutenção baseado em condições e/ou do sistema de supervisão e aquisição de dados para estimar o VUR dos componentes do trem de força (CARROLL *et al.*, 2019; YANG; ZHANG, 2020).

A utilização de variáveis como temperatura é proposta como alternativa para identificação de falhas e realização de prognósticos. Os sistemas de controle supervisório de aerogeradores (SCADA) monitoram as temperaturas de vários componentes do conjunto cinemático, desde o mancal principal até o mancal do lado oposto do acoplamento do gerador, e da Nacele e da temperatura ambiente. É possível definir limites de temperatura com base em informações técnicas sobre os componentes e lubrificantes utilizados.

2.8.Conclusões do Capítulo estado da arte

Neste Capítulo, identificamos as principais fontes de dados, técnicas de pré-processamento, tipos de métodos de prognóstico, capacidade de estimar a vida útil remanescente e modos de falha abordados na literatura. Também foi proposta uma classificação dos métodos de acordo com o volume de dados de aerogeradores utilizado para evitar *overfitting*, pois observou-se que diversos métodos de prognóstico da área de energia eólica não utilizam as boas práticas para garantir a capacidade de generalização. Além disso, se discutiu a importância e necessidade de conjuntos de dados públicos maiores e sintéticos para avaliação dos métodos, bem como a seleção cuidadosa de subconjuntos de treinamento, validação e teste que avalize a capacidade de generalização dos modelos de aprendizagem de máquina aplicados no prognóstico de falha da área de energia eólica. Concluiu-se que o aumento da disponibilidade de dados abertos e o uso de conjuntos de dados de múltiplas turbinas podem contribuir para o desenvolvimento de métodos ainda mais eficazes de prognóstico de falhas em aerogeradores, em especial para o rolamento principal.

3. FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica apresenta os conceitos e técnicas que embasam o desenvolvimento deste trabalho, com foco na aplicação de modelos de aprendizagem de máquina para prognósticos de falhas. Este capítulo explora as bases teóricas e metodológicas necessárias para compreender os métodos utilizados, contextualizando-os dentro do campo de prognóstico de falhas e sua aplicação no setor eólico. A adoção de técnicas avançadas, como os modelos regressivos, busca superar os desafios relacionados à previsão da vida útil remanescente e à generalização em diferentes contextos operacionais.

Dentre os modelos discutidos, destacam-se aqueles voltados para tarefas de regressão, como o *Support Vector Machine* (SVM), modelos isotônicos e diversas abordagens baseadas em árvores de decisão e *ensembles*. Esses modelos são particularmente relevantes devido à sua capacidade de capturar padrões não lineares, lidar com dados de alta dimensionalidade e oferecer interpretabilidade em cenários complexos. Ao longo deste capítulo, destacaremos suas vantagens, limitações e aplicações específicas ao domínio do estudo.

3.1. Modelos de aprendizagem de máquina

Apesar do termo aprendizagem de máquina recebido atenção mais recentemente pelas demandas da indústria 4.0 e aplicativos de redes sociais, anteriormente já se utilizava os modelos de aprendizagem de máquina para criação de filtros automáticos de spam em correios eletrônicos, no reconhecimento ótico de caracteres dentre outros.

A aprendizagem de máquina é uma ciência, ou subárea da ciência da computação, que permite programar computadores para aprenderem a partir de dados (GÉRON, 2019) e utilizando algoritmos que “dependem de coleções de exemplos de determinado fenômeno” (BURKOV, 2019).

Também podemos entender a aprendizagem de máquina como um processo de (1) “acessar um conjunto de dados e (2) construir algoritmicamente modelos estatísticos com base nestes dados” de modo a resolver um problema prático.

Os modelos de aprendizagem podem ser classificados como supervisionados, semi-supervisionados, não-supervisionados e por reforço.

3.2. Modelos regressivos de aprendizagem de máquina

Os modelos regressivos de aprendizagem de máquina são amplamente utilizados em tarefas de previsão, especialmente no contexto de prognósticos. Esses modelos buscam mapear

relações entre variáveis preditoras e a variável de interesse (alvo), permitindo estimar valores futuros ou compreender padrões complexos nos dados. Nesta seção, exploraremos alguns dos modelos mais relevantes.

3.2.1. Modelo de vetor de suporte (Support vector machine (SVM))

O modelo de SVM é uma técnica robusta baseada na maximização da margem entre os dados de treinamento e a função de decisão. Originalmente concebido para classificação, o SVM pode ser adaptado para problemas de regressão por meio do método de "*Support Vector Regression*" (SVR). Essa abordagem utiliza uma função de perda insensível a desvios menores que um limiar, promovendo uma solução com boa generalização e robustez a *outliers* (MURPHY, 2012).

A formulação matemática do SVR se baseia em encontrar uma função $f(x)$ que minimize o erro de previsão, ajustando o equilíbrio entre erro empírico e regularização. Essa característica torna o SVM particularmente eficaz em cenários com conjuntos de dados pequenos e de alta dimensionalidade (BURKOV, 2019).

3.2.2. Modelo isotônico (Isotonic regression)

O modelo isotônico é uma técnica de regressão não paramétrica que assume uma relação monotônica entre as variáveis preditoras e a variável de saída. É ideal para problemas onde há conhecimento prévio de que essa relação monotônica existe. O ajuste é realizado minimizando a soma dos erros quadráticos sob a restrição de monotonicidade, resultando em uma função de predição que respeita essa estrutura (GÉRON, 2019).

Embora menos flexível do que modelos paramétricos, o modelo isotônico é valioso em casos em que há necessidade de interpretar a relação direta entre as variáveis ou onde as suposições de linearidade não se aplicam.

3.2.3. Modelos de árvores e conjuntos (ensembles)

Os modelos de árvores são amplamente utilizados em tarefas de regressão devido à sua simplicidade, interpretabilidade e capacidade de modelar relações não lineares. Além disso, a combinação de várias árvores em conjuntos (*ensembles*) pode melhorar significativamente o desempenho geral.

3.2.3.1. Árvore de decisão (Decision tree)

As árvores de decisão particionam o espaço de características em regiões homogêneas, onde cada divisão é feita com base na maximização da redução de variância ou do erro residual. Embora sejam intuitivas e fáceis de interpretar, sua tendência ao sobreajuste (*overfitting*) em conjuntos de dados pequenos é um desafio significativo (GOODFELLOW; BENGIO; COURVILLE, 2016).

3.2.3.2. Aumento de gradiente (Gradient boosting)

O aumento de gradiente (*Gradient Boosting*) é um método de *ensemble* que constrói modelos em sequência, onde cada modelo subsequente tenta corrigir os erros do anterior. Esse método otimiza uma função de perda específica, como o erro absoluto ou quadrático, utilizando gradientes para ajustar os parâmetros. Métodos populares como XGBoost e LightGBM são implementações práticas desse conceito (BURKOV, 2019).

3.2.3.3. Floresta aleatória (Random forest)

A floresta aleatória combina várias árvores de decisão, cada uma construída a partir de uma amostra aleatória do conjunto de dados. Este método incorpora aleatoriedade tanto na seleção das amostras quanto das características, reduzindo significativamente o risco de sobreajuste (*overfitting*). Além disso, as florestas aleatórias oferecem uma estimativa robusta da importância das características no modelo (GÉRON, 2019).

3.2.3.4. Árvores extras (Extra trees)

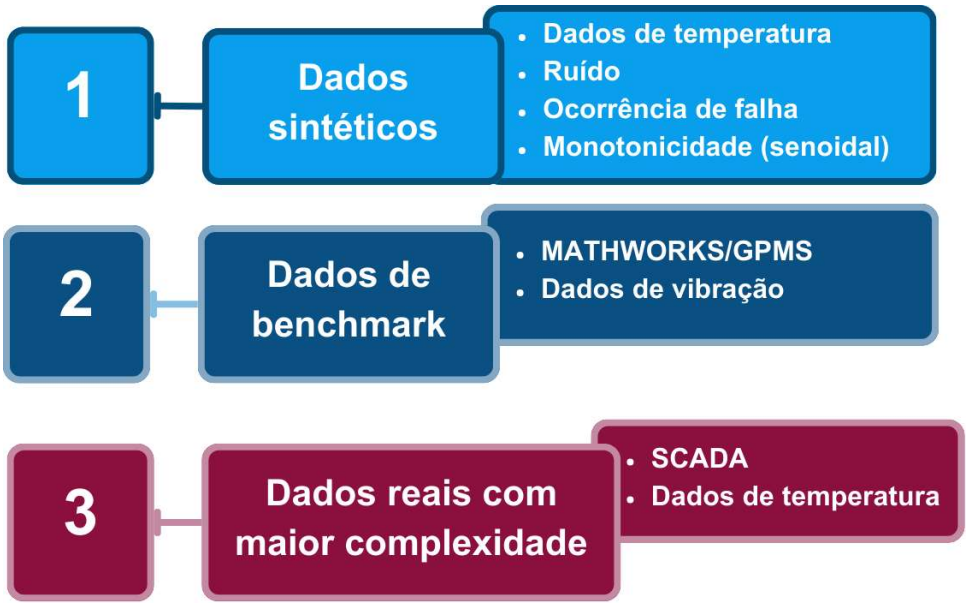
As Árvores Extras diferem das florestas aleatórias principalmente na forma como realizam as divisões nos nós. Em vez de selecionar a melhor divisão com base em um critério de impureza, escolhem divisões aleatórias. Essa abordagem acelera o treinamento e aumenta a diversificação entre as árvores do ensemble, promovendo maior robustez e melhor desempenho em certos cenários (GOODFELLOW; BENGIO; COURVILLE, 2016).

4. DESENVOLVIMENTO DE MODELO COMPUTACIONAL PARA ESTIMATIVA DA VIDA ÚTIL REMANESCENTE

Neste Capítulo serão apresentados os modelos computacionais desenvolvidos para estimar a VUR de aerogeradores. Os modelos foram elaborados utilizando uma abordagem com aumento gradual de complexidade, partindo-se de séries temporais com comportamento

monotônico e mais simples de se interpretar até séries temporais mais complexas provenientes do sistema SCADA, em condições reais de operação, conforme apresentado na Figura 34.

Figura 34 - Diagrama do uso de diferentes bases de dados com aumento gradual da complexidade



Fonte: O autor (2024).

Foram utilizadas uma série temporal sintética, outra série de medições intensivas coletadas em condição real de operação, que chamaremos de dados de benchmark, e por último dados reais com maior complexidade extraídos diretamente do sistema SCADA, de diferentes aerogeradores provenientes de duas usinas eólicas em operação. Considerou-se o 3º tipo de dado como os com maior complexidade pois não receberam pré-processamentos e com períodos sem informação do sinal de temperatura.

A série temporal sintética randômica foi criada com valores de 30°C (operação normal), 42°C (início da degradação) a 80°C (fim de vida) para emular uma variação de temperatura que poderia ser identificada no rolamento principal de um aerogerador com potência de aproximadamente 1,5 MW. Esses dados de temperatura foram retirados uma tabela de alarmes do aerogerador o qual se utilizou os dados do sistema SCADA.

A série temporal de medições intensivas é uma base de dados pública disponibilizada no Github pela Mathworks em conjunto com a empresa Green Power Monitoring Systems (MATHWORKS INC.; BECHHOEFER, 2017).

A série temporal mais complexa é uma série temporal proveniente de medições de temperatura do rolamento principal coletada pelo sistema SCADA de duas usinas eólicas do

nordeste brasileiro com mais de 10 anos de início de operação, na qual operam aerogeradores de 1,65 MW.

O detalhamento da construção da série temporal sintética e das séries temporais em condições reais de operação serão descritas nas seções a seguir.

As estimativas da VUR foram realizadas utilizando modelos de aprendizagem de máquina (AM) com técnicas de regressão. Optou-se por utilizar-se esses modelos pois 18% dos modelos para estimativa da VUR são os modelos clássicos de AM e em alguns casos são utilizados em combinações com outros modelos e por serem modelos de simples implementação. Isto permitiu investir mais tempo na seleção, tratamento e análise crítica dos dados operacionais mais complexos adquiridos pelo sistema SCADA.

A seguir serão apresentados: (1) as bases de dados utilizadas, os pré-processamentos realizados e uma estratégia para lidar com os dados escassos – no caso do rolamento principal, (2) os modelos de AM regressivos desenvolvidos e (3) os parâmetros utilizados no processo de validação cruzada.

Os resultados com os subconjuntos de testes, métricas dos modelos de AM e os resultados das estimativas da VUR serão apresentados no capítulo 5.

Nas seções e subseções a seguir utilizaremos o termo modelos de estimativa da VUR para se referir aos modelos de AM desenvolvidos para o cálculo da VUR.

4.1. Dados sintéticos

Nesta subseção será descrita a metodologia para elaboração do conjunto de dados sintéticos utilizado para avaliação inicial dos modelos de estimativa da VUR.

Para criação do conjunto de dados sintéticos foi usado como fonte de consulta a lista de alarmes definidos pelo fabricante no projeto do aerogerador que se utilizou os dados do sistema SCADA os modelos com dados reais.

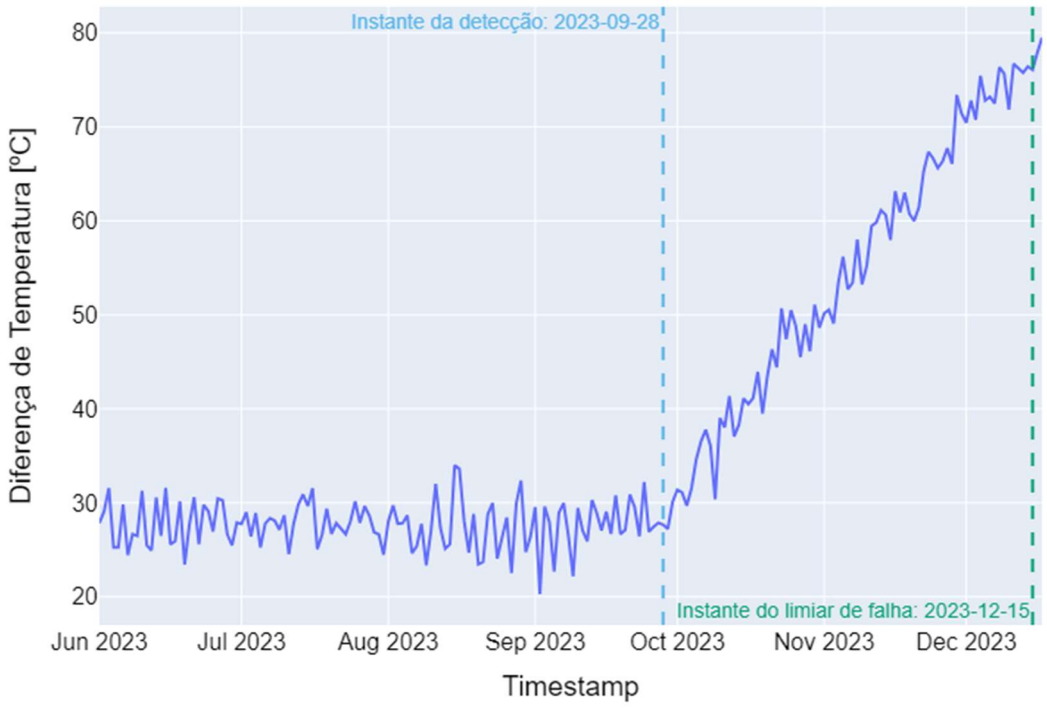
Como o instante da falha é caracterizado pelo alcance de limites superiores ou inferiores para uma determinada grandeza medida, o conjunto de dados sintéticos foi criado com as seguintes premissas:

1) Utilizou-se uma variação aleatória, entre 25.0 e 30.0, correspondente aos valores considerados a temperatura de operação normal do rolamento principal do aerogerador;
2) Foram definidos (1) o instante de detecção da falha (*first predition time* - FPT), e (2) o instante do limiar de falha (*failure threshold time* - FTT);

3) Sinal da série temporal com comportamento monotônico;
4) Inserção de ruído aleatória no sinal.

A Figura 35 apresenta a série temporal de temperatura com as indicações do instante de detecção e o do instante em que se atinge o limiar da falha

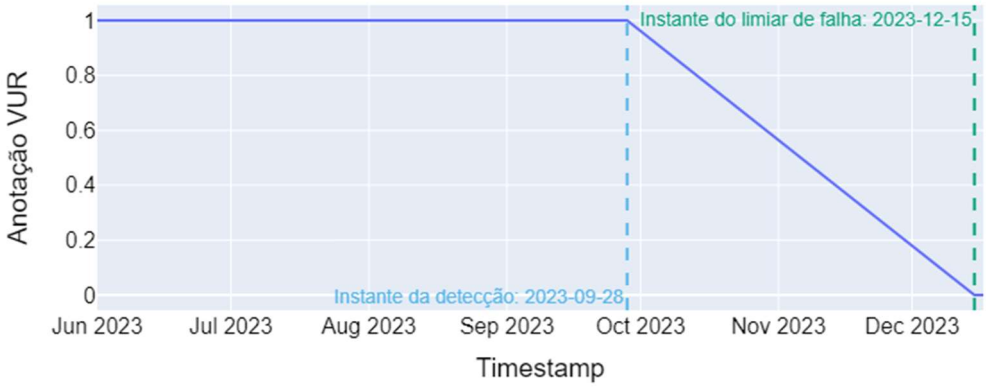
Figura 35 – Série temporal do conjunto de dados sintéticos



Fonte: O autor (2024)

Em seguida foi atribuída uma anotação do conjunto de dados sintéticos para cálculo da estimativa da VUR atribuindo valores entre 1(um) e 0(zero), sendo 1(um) para a condição de operação normal e 0 (zero) no instante do limiar da falha. Os demais dados com valores de temperatura acima dos definidos como operação normal foram interpolados linearmente entre 1 e 0, conforme mostrado na Figura 36.

Figura 36 – Anotação dos dados entre o intervalo 1 (estado saudável) até zero (falha do componente) para estimativa da VUR.



Fonte: O autor (2024)

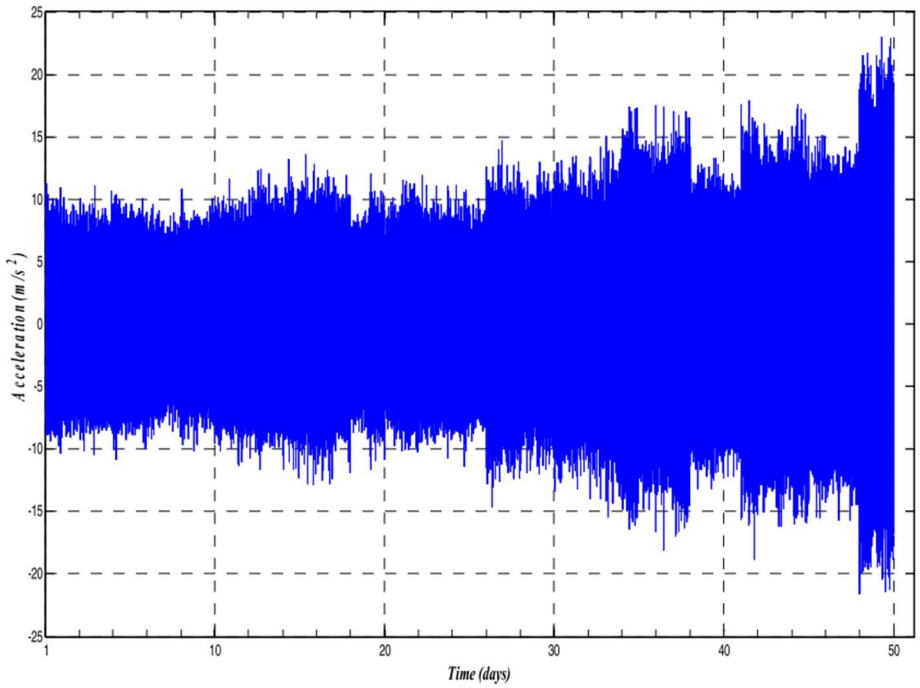
4.2. Dados GPMS – Vibração do rolamento de alta velocidade

Nesta subseção será descrita a leitura e pré-processamento de um conjunto de dados públicos utilizado para avaliação dos modelos de estimativa da VUR. Foi escolhida a base de dados Mathworks/GPMS (MATHWORKS INC.; BECHHOEFER, 2017), utilizada em 12 artigos científicos que abordam o tema da generalização de modelos de estimativa da VUR.

A base de dados Mathworks/GPMS é uma base dados pública com dados de vibração do rolamento de rolos cônicos (SKF 32222 J2) do eixo de alta velocidade (HSS) de um aerogerador Suzlon de 2,2 MW, coletados com uma taxa de amostragem em torno de 100 kHz durante 50 dias, com um sensor especial desenvolvido pela Green Power Monitoring Systems. Essa base de dados está disponível em (MATHWORKS INC.; BECHHOEFER, 2017). Esse conjunto é disponibilizado com dados já pré-processados o que facilita bastante a sua utilização.

Os dados, disponibilizados em 50 arquivos em formato .mat, foram lidos e agrupados em um único dataframe. A Figura 37 mostra a série temporal com os dados de vibração constantes na base de dados.

Figura 37 - Sinais brutos de vibração de rolamento até a falha na pista interna



Fonte: (KRAMTI *et al.*, 2018)

Em seguida foram extraídas as features de média, desvio padrão, RMS, pico-a-pico, assimetria, curtose e fator de crista da série temporal para identificar visualmente, qual característica estatística no domínio do tempo mostra mais claramente o instante de detecção da falha (*first predition time* - FPT). O instante do limiar de falha (*failure threshold time* - FTT) foi considerado como o final da série temporal.

As características estatísticas no domínio do tempo foram calculadas de acordo com as equações de (1) a (7) a seguir:

Média do sinal (μ):

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

Onde

x – o sinal de tempo de população

N – o número de amostras colhidas

i – o índice de amostra

Desvio padrão (σ):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (2)$$

RMS

$$RMS = \sqrt{\frac{1}{N}\left[\sum_{i=1}^{N}(x_i)^2\right]} \quad (3)$$

Pico-a-pico

$$Pico - a - pico = \max(x_i) - \min(x_i) \quad (4)$$

Assimetria (*Skewness*)

$$Assimetria = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3 \quad (5)$$

Onde,

x – o sinal de tempo de amostra original

N – o número de amostras colhidas

i – o índice de amostra

Curtose (*Kurtosis*)

$$Curtose = \frac{N\sum_{i=1}^{N}(x_i - \mu)^4}{[\sum_{i=1}^{N}(x_i - \mu)^2]^2} \quad (6)$$

Fator de crista

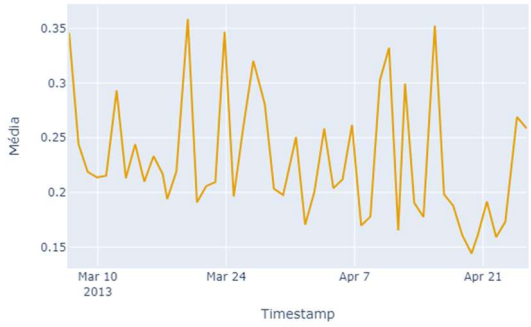
$$Fator\ de\ crista = \frac{\max(|x_i|)}{RMS} \quad (7)$$

A Figura 38 mostra os valores das características estatísticas no período de medição.

A curtose foi a característica estatística que apresentou mais claramente a evolução do estado saudável para o estado com falha. Desta forma ela foi utilizada para a inclusão de uma anotação dos dados.

Em seguida, utilizando a curtose foi atribuída uma anotação do conjunto de dados Mathworks/GPMS para cálculo da estimativa da VUR atribuindo valores entre 1 (um) e 0 (zero), sendo 1 (um) para a condição de operação normal, até a data de 01/04/2013, e 0 (zero) no instante do limiar da falha. Os demais dados com valores acima dos definidos como operação normal foram interpolados linearmente entre 1 e 0.

Figura 38 -Características estatísticas do sinal adquirido por medições intensivas Mathworks/GPMS. (a) Média, (b) Desvio-padrão, (c) RMS, (d) Pico-a-Pico, (e) Assimetria (Skewness), (f) Curtose (Kurtosis) e (g) Fator de crista

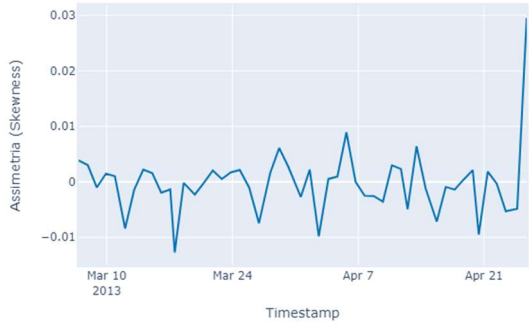


(a)

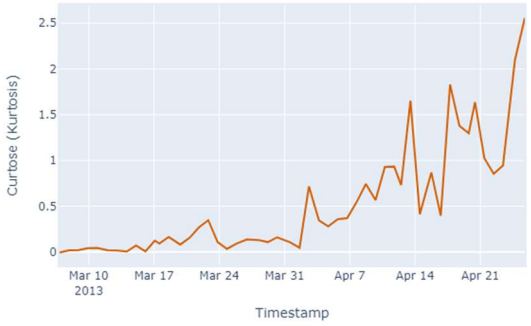
(b)

(c)

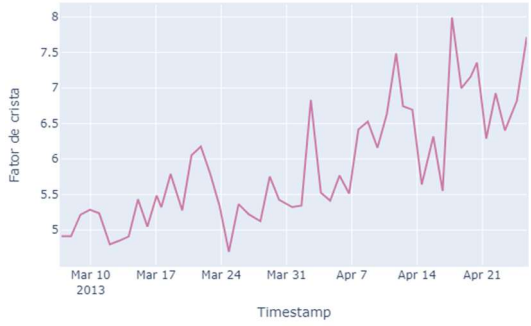
(d)



(e)



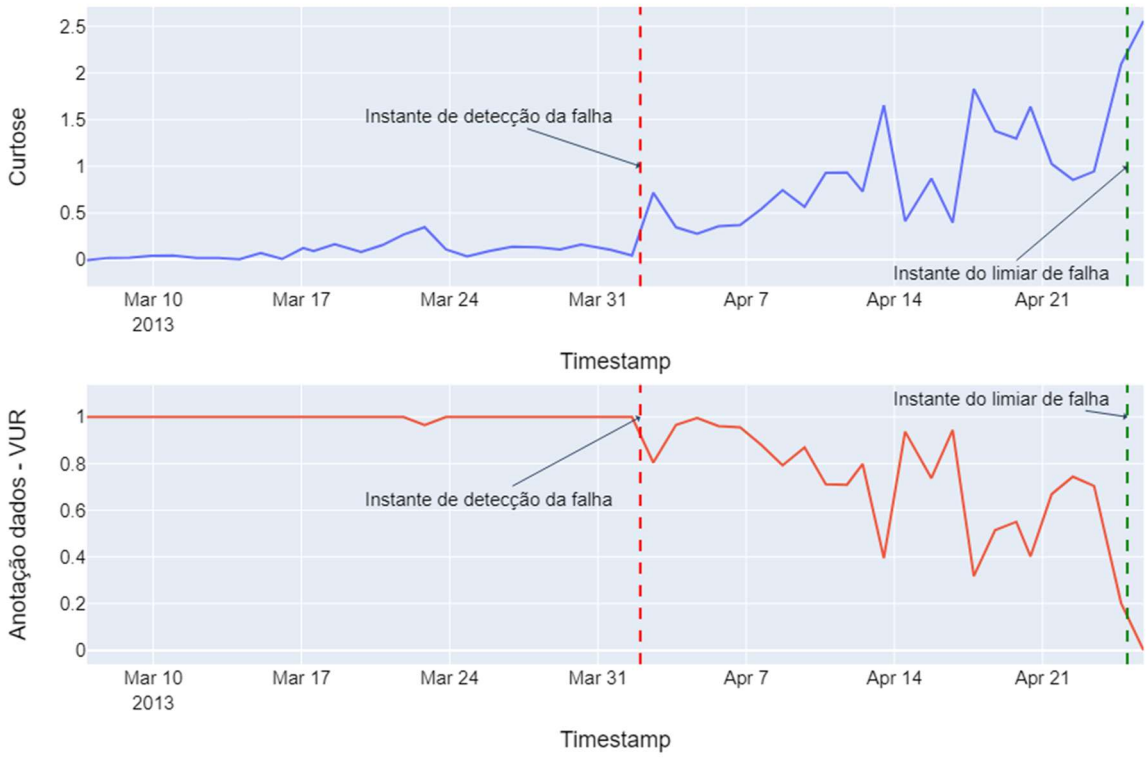
(f)



(g)

Fonte: O autor (2024)

Figura 39 - Série temporal do conjunto de dados Mathworks/GPMS e anotação do conjunto de dados para estimativa da VUR



Fonte: O autor (2024)

4.3. Dados SCADA – Conjunto de dados de temperatura do rolamento principal

Nesta Seção serão apresentados os dados de temperatura do rolamento principal, as etapas do pré-processamento, criação de uma anotação nos dados e definição do perfil de degradação e a metodologia para definição dos subconjuntos de dados para treinamento, validação e testes.

4.3.1. Leitura e pré-processamento de dados do sistema SCADA

O sistema SCADA tem uma taxa de aquisição de dados menor do que os sistemas de manutenção baseados em condições (HU; YOUN; WANG, 2018) (pp.303-342). Os sinais são registrados em amostras de 10 minutos em um determinado intervalo de tempo, mas ainda podem identificar falhas geradas por componentes e subsistemas que envolvem alterações nas grandezas medidas. Os dados brutos das turbinas dos dois parques eólicos foram fornecidos em arquivos no formato .csv e foram lidos e condensados em um arquivo *parquet*. Foi criado um único *dataframe* contendo todas as variáveis SCADA com os dados do arquivo *parquet*, que

chamaremos de dataset de dados brutos. A partir do conjunto de dados brutos, foi criado um novo *dataframe* contendo apenas as séries temporais com as variáveis de interesse, de 34 aerogeradores (WT1 – WT34) dos dois parques eólicos.

Para pré-processar o sinal, foram realizadas as etapas mostradas na Figura 40, os dados foram filtrados para remover medições onde a velocidade do vento era maior ou igual a 3,5 m/s, a velocidade do gerador estava abaixo de 1198 RPM ou acima de 1200 RPM e a temperatura ambiente era menor ou igual a 50 graus Celsius, o que estava fora das condições de operação da turbina. A série de temperaturas foi então reamostrada em intervalos de 1 hora para filtrar variações intra-hora. Um recurso foi criado subtraindo a temperatura ambiente do valor principal da temperatura do rolamento, conforme descrito por WIESE et al., (2020), que chamaremos de variação de temperatura, para retirar a influência da temperatura ambiente no comportamento do mancal principal.

Figura 40 – Pré-processamento do sinal

Fonte: O autor (2024).

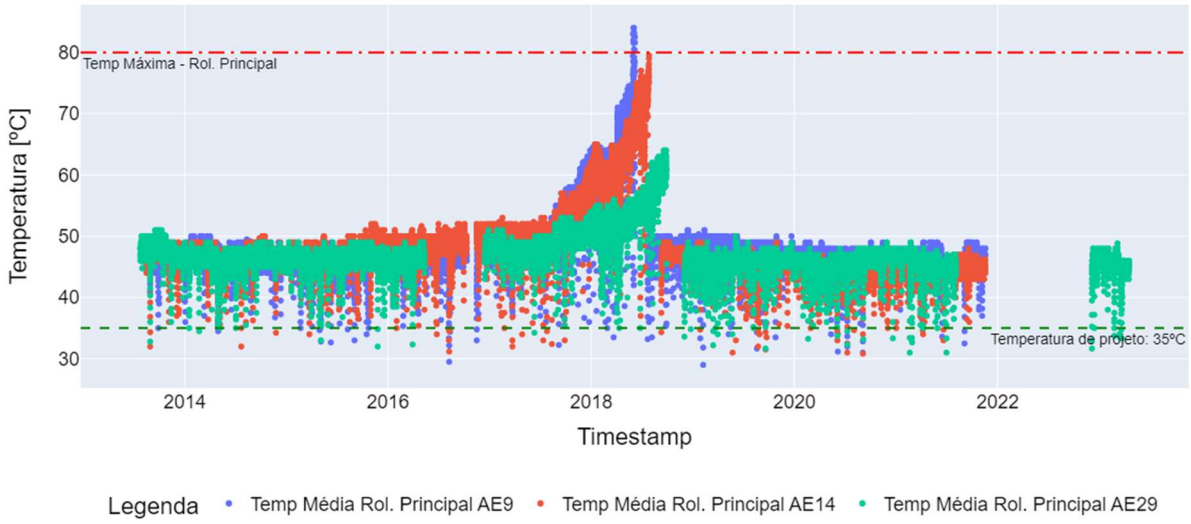
Na terceira etapa do pré-processamento, os valores discrepantes de variação de temperatura foram removidos. Isso envolveu a eliminação de valores de variação abaixo de 10 graus Celsius, que indicam a operação do aerogerador em baixa velocidade, e os valores acima de 98% do percentil da série temporal, considerando a faixa de temperatura de operação (-30ºC a +110ºC) da graxa SKF LGWM 1 usada para lubrificar o rolamento principal. A Figura 41 mostra a série temporal das turbinas com falha (WT9, WT14 e WT29) com os valores discrepantes, enquanto a Figura 42 mostra a série sem os valores discrepantes e com a reamostragem de 1 dia necessária para a decomposição sazonal. Na sequência, foi realizada remoção das medições que apresentavam valores ausentes (*Not a Numbers – Nans)* e que não podem ser convertidos em outro tipo de valor diferente do tipo *float* (GEEKS FOR GEEKS, 2023; PANDAS, 2024), da variação da temperatura do rolamento principal, para obter o resultado desejado na última etapa do pré-processamento.

A etapa final de pré-processamento envolveu a aplicação de uma decomposição sazonal aditiva à série temporal. Este método estima a tendência da série temporal aplicando um filtro de convolução aos dados, remove a tendência calculada da série e, em seguida, calcula a média

da série sem tendência (PERKTOLD; SEABOLD; TAYLOR, 2023). Neste trabalho foi utilizada a decomposição sazonal aditiva para períodos de 7 dias, e o resultado é mostrado na Figura 43.

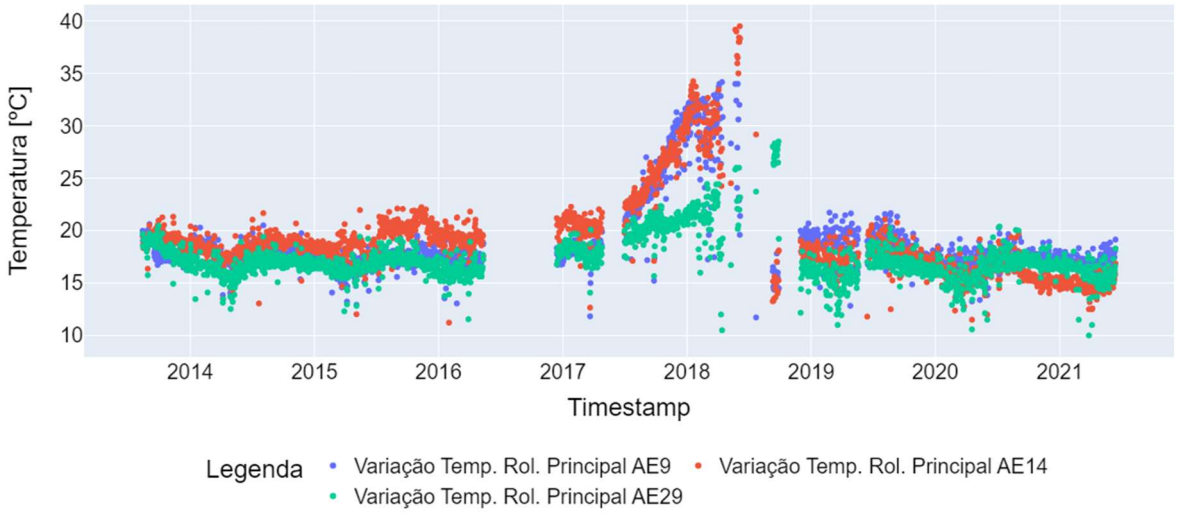
Após o pré-processamento as séries temporais apresentaram intervalos de tempo sem dados. Estes vazios estão relacionados com períodos em que os aerogeradores estavam parados em manutenção ou com algum problema no processo de aquisição dos sinais. Por se tratar de uma situação que pode ocorrer durante a operação dos aerogeradores, estes intervalos não foram preenchidos para que os modelos pudessem lidar e reproduzir esse fenômeno.

Figura 41 – Série de temperatura média dos rolamentos principais com valores discrepantes (*outliers*) (WT9, WT14 e WT29)



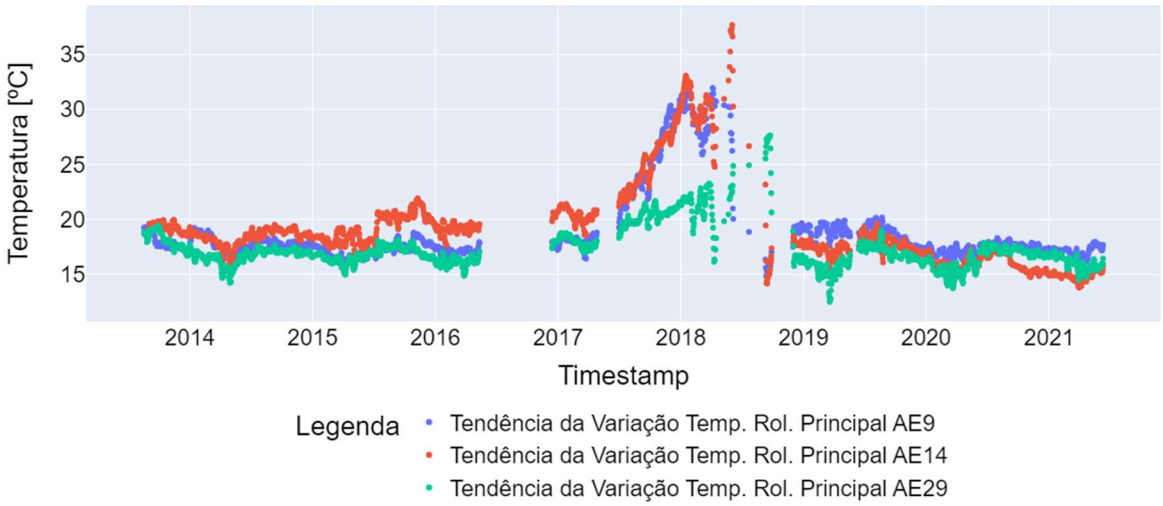
Fonte: O autor (2024).

Figura 42 – Variação de temperatura do rolamento principal após remoção dos valores discrepantes (*outliers*) (WT9, WT14 e WT29)



Fonte: O autor (2024).

Figura 43 – Variação de temperatura para o rolamento principal após decomposição sazonal (WT9, WT14 e WT29)



Fonte: O autor (2024).

4.3.2. Criação de anotação dos dados de variação de temperatura

Nesta seção apresentaremos o processo de anotação dos dados para o uso nos modelos regressivos. As etapas desse processo (1) identificação do instante de detecção da falha (*first predition time* - FPT) e do instante do limiar de falha (*failure threshold time* - FTT) analisando graficamente as séries temporais, (2) definição do valor mínimo da anotação com base na variação da temperatura e (3) uso de interpolação linear para classificar os dados entre os instantes FPT e FTT.

Para realizar as estimativas da VUR foi necessário classificar os dados de variação de temperatura das séries temporais das turbinas que apresentaram falha. Através de análise gráfica das séries, identificou-se (1) o instante de detecção da falha (*first predition time* - FPT), depois identificou-se (2) o instante do limiar de falha (*failure threshold time* - FTT), momento em que ocorre a falha do componente. A partir do FPT que se considerou o início do processo de degradação do componente e ao atingir o limiar de falha considera-se que o componente está completamente degradado. A vida útil remanescente é esse intervalo de tempo entre o FTT e o FPT.

Após o pré-processamento dos dados, Figura 40, foi necessário classificar os dados de variação de temperatura criando assim a variável dependente utilizada nos modelos de regressão. Essa anotação foi definida manualmente, observando o comportamento da série temporal da variação de temperatura do rolamento principal, após a etapa de decomposição sazonal.

A partir da avaliação gráfica das séries temporais da variação de temperatura, foram identificados os instantes de tempo a partir dos quais houve aumento monotônico. Este instante foi considerado como o FPT. E todos os dados anteriores a este instante foram classificados como saudáveis e foram classificados com o valor 1. O instante de tempo que se registrou o maior valor da variação de temperatura foi considerado como o FTT. A partir deste instante os dados foram classificados com o valor zero. Quando a variação de temperatura voltou a apresentar valores não nulos, se classificou os dados como saudáveis novamente, correspondente ao completo restabelecimento da vida útil a partir deste instante, que chamaremos de instante de restabelecimento (RT).

Identificou-se que o valor máximo da variação de temperatura de cada uma das turbinas era diferente. Os valores máximos registrados foram 34,17, 39,5 e 28,5 em graus Celsius, para as turbinas WT9, WT14 e WT29. Esses valores foram normalizados e utilizados para definir o valor mínimo da anotação dos dados, conforme equação (8).

$$Anot_{min} = 1 - Temperatura_{norm} \tag{8}$$

Onde:

$Temperatura_{norm} = \frac{MaxTemp_{AE(i)}}{MaxTemp_{AEs}}$

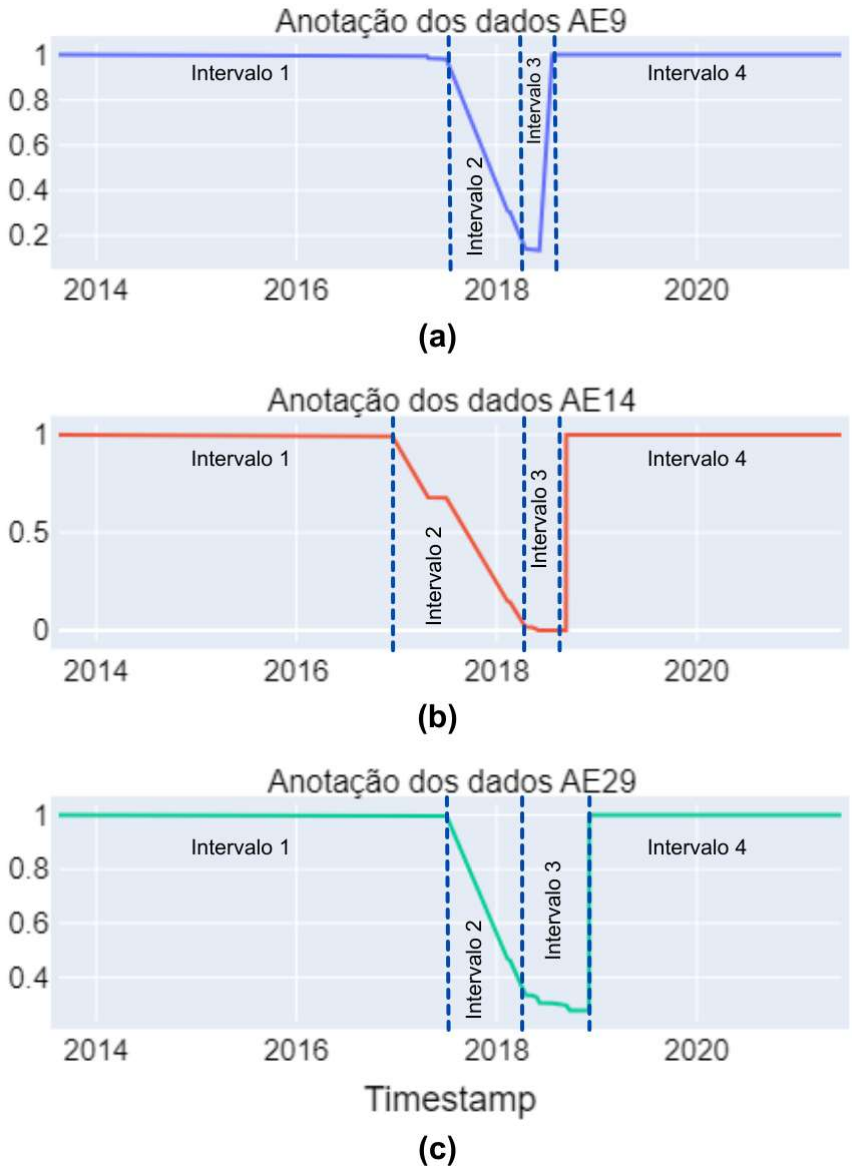
$MaxTemp_{AE(i)}$ – A maior temperatura atingida por um dado aerogerador

$MaxTemp_{AEs}$ – A maior temperatura identificada num conjunto de aerogeradores

Com os instantes FPT, FTT e RT e os respectivos valores da anotação dos dados utilizou-se uma interpolação linear para classificar os dados de variação de temperatura entre os instantes FPT e FTT.

Para o período do início da série temporal até antes do início do aumento observável da temperatura do rolamento principal (Intervalo 1) os dados foram classificados com o valor de 1, do início do aumento da temperatura do rolamento principal até a primeira ocorrência da temperatura máxima (Intervalo 2), valores decrescentes interpolados linearmente entre 1 e valor mínimo da anotação ($Anot_{min}$), do intervalo da primeira ocorrência da temperatura máxima até a parada da turbina, na qual a se atribui valor zero a anotação dos dados. Quando do retorno do aerogerador a operação considerou-se a anotação igual a 1 novamente. A Figura 44 apresenta a anotação dos dados ao longo de toda as séries temporais para as 3 turbinas selecionadas, onde se observa diferentes dinâmicas de degradação dos rolamentos principais, ora mais rápida, para a WT9, ora mais lenta, para a WT29.

Figura 44 - Anotação dos dados estimados a partir das principais séries temporais de temperatura dos mancais para os aerogeradores: (a) WT9, (b) WT14 e (c) WT29



Fonte: O autor (2024).

4.3.3. Criação de subconjuntos de dados

Um problema comum com modelos de aprendizado de máquina é o *overfitting*, que pode ocorrer ao usar amostras do mesmo conjunto de dados para treinamento e teste. No entanto, o uso de diferentes conjuntos de dados envolve uma compensação: reduzir o *overfitting* em detrimento das métricas do modelo de regressão.

Neste estudo, empregamos uma estratégia para lidar com a escassez de dados de aerogeradores, conforme sugerido por (ABID; SAYED-MOUCHAWEH; LAURENCE, 2019; REZAMAND *et al.*, 2019), e minimizar o *overfitting*. Isso envolveu o uso de subconjuntos de

dados, definidos por meio de uma combinação com permutação, das três séries temporais das turbinas WT9, WT14 e WT29, que exibiram crescimento monotônico da temperatura do mancal principal (ver Tabela 8).

Tabela 8 – Subconjuntos usados para lidar com escassez dados de aerogeradores

ID	Subconjunto de Treinamento	Subconjunto de Validação	Subconjunto de Teste
MB-SC01	WT9	WT14	WT29
MB-SC02	WT9	WT29	WT14
MB-SC03	WT14	WT9	WT29
MB-SC04	WT14	WT29	WT9
MB-SC05	WT29	WT9	WT14
MB-SC06	WT29	WT14	WT9

Fonte: O autor (2024).

Os IDs, indicados na Tabela 8, representam uma combinação de dados de uma turbina usada para treinar o modelo (subconjunto de treinamento), dados de outra turbina usada para validar o modelo (subconjunto de validação) e dados de um terceiro turbina usada para testar o modelo após validação (subconjunto de teste). Esses IDs serão utilizados ao longo do texto para se referir aos casos de simulação realizados com os modelos regressivos para estimativa da VUR do rolamento principal.

4.4.Modelos de regressão para estimativa da vida útil remanescente

Nesta Seção serão apresentados os modelos de aprendizagem de máquina, o treinamento e a validação cruzada, quando realizada, com estes. Foram escolhidos modelos supervisionados de regressão e todos eles foram utilizados com todas as bases de dados. Os modelos foram desenvolvidos usando a biblioteca scikit-learn (SCIKIT-LEARN DEVELOPERS, 2024) do Python.

Foram selecionados os seguintes modelos de aprendizagem de máquina com uso de regressão:

Regressão por Vetor de suporte [*support vector regression* (SVR)]. Este modelo por ser versátil com regressão de dados linear e não linear. Inicialmente, foram feitos testes com kernel utilizando RBF (*Radial Basis Function*) e polinômio do 3º grau, e os resultados do R2 score foram melhores com o polinômio do 3º grau. Então, o modelo foi construído utilizando um polinômio do 3º grau, como kernel, devido a anotação dos dados realizada, que tem características de uma função linear por partes (*piecewise*). Este modelo possui uma boa capacidade de generalização, reduzindo o risco de sobreajuste (*overfitting*). Este tipo de

regressão tem como desvantagem a ser mais adequado para conjuntos de dados de pequeno ou médio porte (GÉRON, 2019; MURPHY, 2012), tem baixa escalabilidade e precisa de uma escolha adequada do kernel da função de regressão.

Para o modelo de regressão por vetor de suporte utilizaram-se os seguintes parâmetros:

- kernel="poly" – kernel polinomial escolhido devido à natureza e à anotação dos dados utilizada;
- C=100 – O parâmetro de regularização
- gamma="auto" – O parâmetro de coeficiente do kernel. Neste caso uses auto é igual a 1 / número de features
- degree=3 – O parâmetro que define a ordem ou grau do polinômio usado no kernel;
- epsilon=0.1 – O parâmetro de valor não negativo. "Define uma margem de tolerância na qual nenhuma penalidade será aplicada aos erros. Deve ser menor para conjuntos de dados maiores"(HARRISON, 2020).
- coef0=1 – O parâmetro do termo independente na função do kernel

Regressão isotônica [*isotonic regression* (ISOR)]. Este modelo foi escolhido pela característica de ajustar os valores previsto dados crescentes ou decrescentes utilizando uma função linear por partes (*piecewise*) e pela facilidade de interpretar relações monotônicas. Possui menor flexibilidade que os modelos de árvores ou conjuntos (*ensembles*) e é limitada a uso de dados monotônicos.

Na regressão isotônica o ajuste dos dados unidimensionais é realizado utilizando uma função real não decrescente, que resolva o problema:

$$min \sum_{i} \omega_i (y_i - \hat{y}_i)^2 \tag{9}$$

De modo que $\hat{y}_i \leq \hat{y}_j$ sempre que $X_i \leq X_j$ e o pesos ω_i são necessariamente positivos, e x e y são quantidades reais arbitrárias.

Para o modelo de regressão isotônica utilizaram-se os seguintes parâmetros:

- y_min=0 – O limite inferior do valor previsto;
- y_max=1 – O limite superior do valor previsto, e
- increasing='auto '– O parâmetro que "determina se as previsões devem ser restringidas para aumentar ou diminuir com X. Quando definido como 'auto', o

> algoritmo decidirá com base no sinal da estimativa de correlação de Spearman.”(SCIKIT-LEARN DEVELOPERS, 2024).

Regressão por Árvore de decisão [Decision tree regression (DTR)]. Este método não paramétrico de aprendizagem supervisionada foi escolhido por fazer aproximação constante por partes (SCIKIT-LEARN DEVELOPERS, 2024) e prediz o valor de cada nó da árvore (GÉRON, 2019). São os componentes fundamentais das árvores aleatórias. Mas, “são propensas a sobrejuste (*ovefitting*) em tarefas de regressão”, sensíveis a pequenas variações dos dados (GÉRON, 2019). Também possuem menor precisão que os modelos de conjuntos (*ensembles*).

Para o modelo de regressão por árvore de decisão foram utilizados
os seguintes parâmetros:

- criterion='squared_error';
- splitter='best'

Regressão de aumento de gradiente [Gradient boosting regression (GBR)]. Este tipo de modelo regressivo trata-se de um conjunto (ensemble) de modelos. Ele é caracterizado ser um conjunto de modelos de árvore de decisão que utiliza o resíduo da árvore anterior para realizar a regressão (GÉRON, 2019) e foi escolhido pela sua alta precisão através da combinação de modelos e tem menor tendência a sobrejuste (*ovefitting*). Demandam um tempo de treinamento maior, em especial para conjuntos de dados maiores e sensibilidade ao ajuste dos hiperparâmetros, precisando de um ajuste cuidadoso.

Para o modelo de regressão de aumento de gradiente utilizaram-se os seguintes parâmetros:

- loss = "squared_error" – O parâmetro do termo;
- learning_rate = 0,01 – O parâmetro do termo;
- n_estimators = 100 – O parâmetro do termo

Regressão por floresta aleatória [Random forest regression (RFR)]

O algoritmo de floresta aleatória é um modelo de conjunto (ensemble) de árvores de decisão que “introduz aleatoriedade extra ao “cultivar” árvores; em vez de procurar a melhor característica ao dividir um nó”, resultando num modelo melhor por “trocar um viés maior por uma variância menor”(GÉRON, 2019). Tem como vantagens a menor propensão a sobrejuste (*ovefitting*), pode lidar com dados faltantes e, geralmente, apresenta alta precisão. No entanto,

trata-se de um modelo mais complexo, com tempos de treinamento e predição maiores e apresenta um maior consumo de memória para armazenar múltiplas árvores.

Para o modelo de regressão por floresta aleatória utilizaram-se os seguintes parâmetros:

- max_depth=2;
- random_state=0

Regressão por Árvores extras [Extra trees (ETR)]

Este algoritmo também é um modelo de conjunto (*ensemble*) como os modelos de regressão por floresta aleatória e regressão de aumento de gradiente, e troca mais viés por uma variância menor, o que a torna mais rápida que a floresta aleatória (GÉRON, 2019). Tem a menor propensão a sobrejuste (*ovefitting*), é um modelo mais rápido de treinar e possui boa capacidade de generalização. Porém, depende dos dados usados na inicialização das "sementes" e é mais difícil de interpretar que outros métodos de *ensemble*.

Para o modelo de regressão por floresta aleatória utilizaram-se os seguintes parâmetros:

- n_estimators=100;
- random_state=0

Nas subseções a seguir serão detalhados as etapas de treinamento e validação cruzada, quando realizada, dos modelos para cada conjunto de dados elencado nas Seções 4.1, 4.2 e 4.3.

4.4.1. Estudo de caso – Uso de dados sintéticos

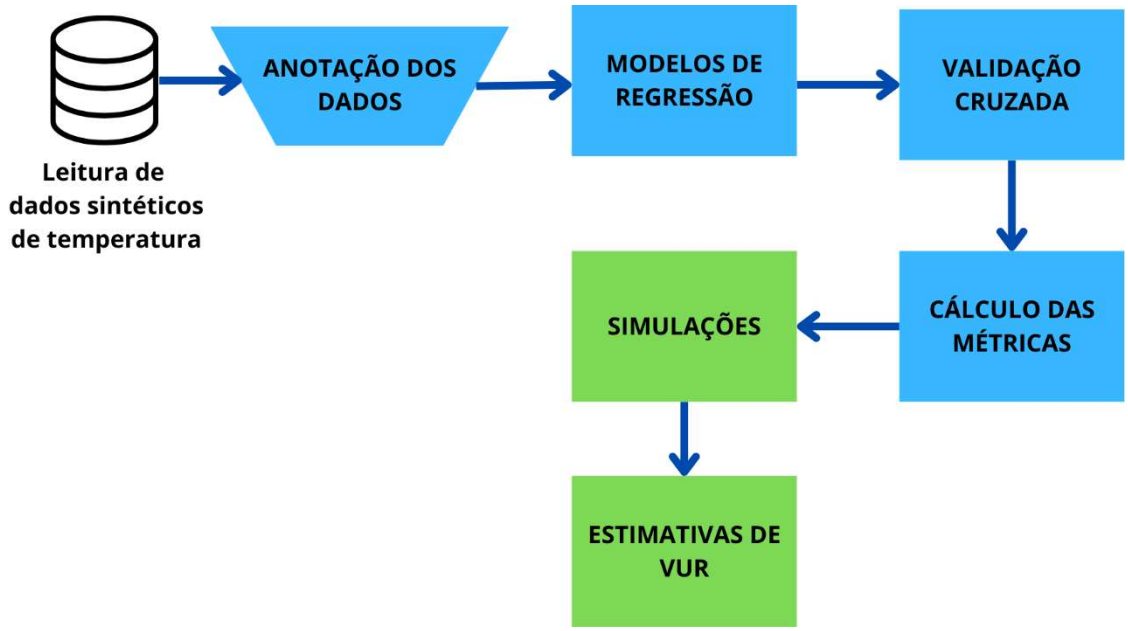
Nesta Seção serão apresentados os modelos de regressão para estimativa da VUR utilizando os dados sintéticos apresentados na Seção 4.1. Dados sintéticos.

Utilizou-se um único conjunto de dados para treinamento e para a validação cruzada. Após a construção da série temporal, os dados de temperatura receberam uma anotação decrescente da VUR, de 1 a zero, de acordo com valores de temperatura: (1) de 30°C (operação normal), (2) 42°C (início da degradação) a (3) 80°C (fim de vida) para representar o processo de degradação de componente.

Nesta etapa utilizou-se o fluxograma indicado na Figura 45 para: (1) criar anotação nos dados de temperatura, (2) treinar os modelos de regressão desenvolvidos e (3) o cálculo de métricas para avaliação dos modelos.

Com as informações de temperatura e anotação, os dados foram normalizados para o treinamento, de acordo com os respectivos intervalos usando MinMax Scaler, do Scikit learn.

Figura 45 - Fluxograma de anotação, treinamento, validação, verificação do desempenho com uso de dados sintéticos e testes para estimativa da VUR.

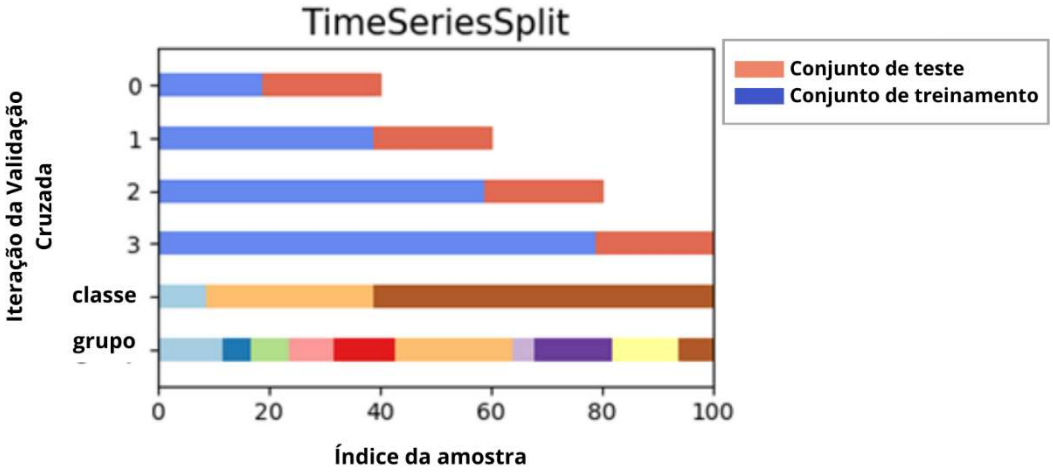


Fonte: O autor (2024).

As variáveis independentes utilizadas nos subconjuntos de treinamento e teste dos modelos incluíram apenas as variações de temperatura do conjunto de dados sintéticos. A variável dependente foi o dado de anotação da VUR definidos na construção do conjunto de dados.

Nesta etapa foi utilizada a divisão de conjuntos de dados – treinamento e teste – com a classe TimeSeriesSplit, para validação cruzada. Essa classe realiza a separação dos dados de criando conjunto de treinamento e teste sequenciados até atingir a totalidade dos dados (Figura 46). Para o conjunto de dados sintéticos utilizaram-se 5, 6, 7 e 8 divisões.

Figura 46 – Exemplo de 4 divisões (splits) para validação cruzada usando a classe TimeSeriesSplit



Fonte: Adaptado de Scikit-Learn Developers (SCIKIT-LEARN DEVELOPERS, 2024)

Para avaliar e comparar os resultados dos modelos de regressão, foram calculadas as seguintes métricas: erro quadrático médio (MSE), raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e o coeficiente de determinação (R2 Score).

4.4.2. Estudo de caso – Uso de dados GPMS

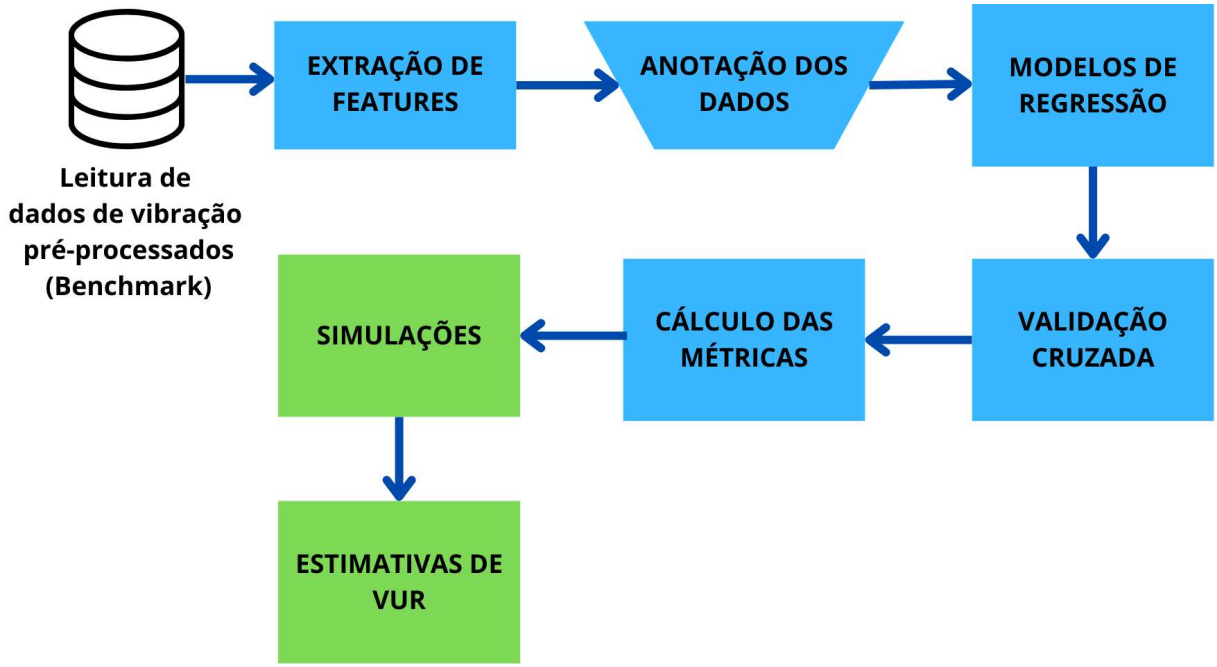
Nesta Seção serão apresentados os modelos de regressão para estimativa da VUR utilizando os dados de vibração do eixo de alta velocidade, do conjunto de dados da Mathworks/GPMS considerado como benchmark pela comunidade científica, apresentados na Seção 4.2. Dados GPMS – Vibração do rolamento de alta velocidade.

Nesta etapa utilizou-se o fluxograma indicado na Figura 45 para: (1) extrair as características estatísticas (features), (2) aplicar anotação nos dados de vibração de acordo com uma feature escolhida, (3) treinar os modelos de regressão desenvolvidos, (4) realizar a validação cruzada e (5) o cálculo de métricas para avalição dos modelos.

A feature selecionada para classificar os dados foi a curtose devido à característica monotônica e que destaca claramente o início do processo de degradação do rolamento do HSS (Figura 38(f)).

Com as informações de curtose e anotação, os dados foram normalizados para o treinamento e validação, de acordo com os respectivos intervalos usando MinMax Scaler, do Scikit learn.

Figura 47 - Fluxograma de anotação, treinamento, validação, verificação do desempenho com uso de dados de benchmark (Mathworks/GPMS) e testes para estimativa da VUR.



Fonte: O autor (2024).

A variável independente utilizada nos subconjuntos de treinamento, validação e teste dos modelos foi a curtose do sinal de vibração. A variável dependente foi o dado de anotação da VUR.

Nesta etapa também foi utilizada a divisão de conjuntos de dados – treinamento e teste – com a classe TimeSeriesSplit, para validação cruzada e também se utilizou 5, 6, 7 e 8 divisões para o conjunto de dados Mathworks/GPMS.

E se utilizou MSE, RMSE, MAE e o R2 Score para avaliar e comparar os resultados dos modelos de regressão.

4.4.3. Estudo de caso – Uso de dados SCADA – Temperatura rolamento principal

Esta Seção apresenta uma metodologia para estimar a VUR de rolamentos principais a partir de modelos de regressão utilizando dados reais de aerogeradores (Seção 4.3. Dados SCADA – Conjunto de dados de temperatura do rolamento principal). Os modelos foram construídos com a definição a priori de um perfil VUR através da análise de séries temporais de temperatura local, anotação dos dados, criação de conjuntos de treinamento, validação e testes com séries temporais de diferentes aerogeradores, avaliação de métricas como MSE, RMSE, MAE e R2 Score e estimativa da VUR usando o conjunto de dados de teste.

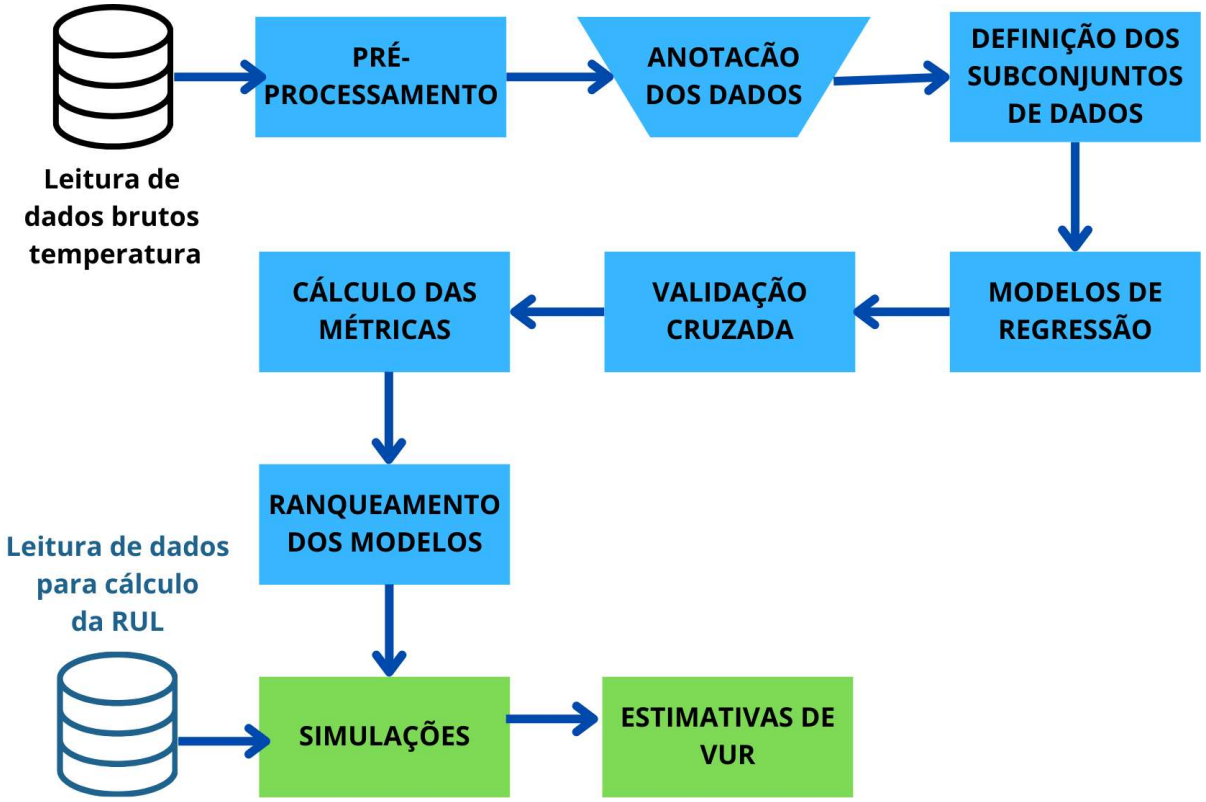
4.4.3.1. Estimativa da VUR de aerogeradores usando dados de temperatura do rolamento principal

Esta Seção apresenta um framework para prever falhas em aerogeradores, estimando a VUR do rolamento principal, utilizando dados de temperatura monitorados pelo sistema SCADA. A estrutura é baseada no uso de dados brutos de temperatura dos rolamentos principais e modelos de regressão para calcular a VUR de turbinas de dois parques eólicos instalados na costa do Nordeste do Brasil. As turbinas são todas do mesmo modelo e possuem potência nominal acima de 1 MW. Informações detalhadas não podem ser disponibilizadas por motivos de confidencialidade.

O procedimento consiste na leitura e pré-processamento dos dados brutos do sistema SCADA, seguido da criação de um perfil de degradação (anotação dos dados) baseado no comportamento da temperatura e com os limites estabelecidos no projeto da turbina. A temperatura média do rolamento principal, a temperatura média do ambiente, a velocidade do vento e os dados de potência ativa da rede dentro da faixa operacional dos aerogeradores foram considerados como variáveis de interesse. As condições ambientais e as propriedades mecânicas do lubrificante do rolamento principal de três aerogeradores com diferentes dinâmicas de degradação também foram contabilizadas. Os casos de simulação foram definidos a partir dos conjuntos de dados dessas três turbinas, respectivamente para treinamento, validação e teste de modelos de aprendizado de máquina usando modelos de regressão.

A Figura 48 mostra o fluxograma para estimar a VUR do rolamento principal a partir dos dados de temperatura.

Figura 48 – Metodologia - Estimativa da VUR por modelos de aprendizado de máquina de regressão



Fonte: O autor (2024).

O método consiste na (1) leitura dos dados brutos do sistema SCADA, numa (2) etapa de pré-processamento dos dados com técnicas de limpeza, filtragem e reamostragem, em seguida (3) a etapa de anotação dos dados das séries temporais, realizada manualmente. Na etapa (4) é feita a criação de subconjuntos de dados para treinamento, validação e teste são separados os dados disponíveis em subconjuntos. A próxima etapa (5) é a criação e treinamento dos modelos de regressão. Depois realiza-se a (6) validação cruzada para identificação dos melhores hiper parâmetros dos modelos e em seguida (7) do cálculo das métricas. Os modelos foram (8) ranqueados e selecionou-se os que apresentam melhores resultados. Em seguida realizaram-se as simulações para estimar a vida útil remanescente a partir (9) da leitura dos dados dos subconjuntos de teste, (10) da simulação com os melhores modelos regressivos e (11) cálculo da VUR.

As etapas (1), (2), (3) e (4) deste método foram descritas na Seção 4.3. As demais etapas serão descritas a seguir.

Conforme citado anteriormente, seis modelos regressivos de aprendizado de máquina: *support vector regression* (SVR), *isotonic regression* (ISOR), *gradient boosting regression*

(GBR), *decision tree* (DTR), *extra trees* (ETR) e *random forest regression* (RFR), que foram desenvolvidos para estimar a VUR usando a biblioteca sci-kit-learn (SCIKIT-LEARN DEVELOPERS, 2024) do Python.

Os dados de treinamento, validação e teste foram normalizados de acordo com os respectivos intervalos dos dados (MinMax Scaler) de modo a padronizar a o intervalo das variáveis independentes (TUTIV'EN *et al.*, 2021).

As variáveis independentes utilizadas nos subconjuntos de treinamento, validação e teste dos modelos incluíram as variações de temperatura do mancal principal, a velocidade do vento e a potência ativa líquida dos aerogeradores nos subconjuntos. As variáveis dependentes foram os dados anotados, apresentados na Figura 44, estimados a partir da série temporal de variação de temperatura. Para garantir o uso de diferentes subconjuntos para treinamento, validação e teste, os subconjuntos de dados foram empregados conforme indicado na Seção 4.3.3.

A validação cruzada realizada neste conjunto de dados usou a classe GridSearchCV

Para este caso, a validação cruzada é um processo crítico, pois serve para melhorar o desempenho das estimativas do modelo, reduzir o *overfitting* e auxiliar na seleção do melhor modelo e seus hiperparâmetros, melhorando assim a robustez dos testes.

A escolha de conjuntos de dados de diferentes turbinas e diferentes usinas teve como objetivo evitar a ocorrência de pontuações de modelos inflacionadas com hiperparâmetros ajustados por validação cruzada, conforme observado por LEAHY et al., (2018).

Após o treinamento do modelo, a validação cruzada foi empregada para identificar o modelo e os hiperparâmetros que minimizariam a perda de regressão do erro quadrático médio. Os intervalos de parâmetros para cada modelo de regressão são detalhados na Tabela 9.

Para avaliar e comparar os resultados dos modelos de regressão, foram calculadas as seguintes métricas para cada caso e são mostradas na Tabela 8: MSE, RMSE, MAE e R2 score. Também foram calculadas as médias e desvios padrão das métricas de treinamento, validação e teste para cada subconjunto de casos simulados.

As métricas obtidas dos 6 modelos desenvolvidos e dos 6 casos apresentaram pequenas variações. Consequentemente, selecionamos os 2 modelos com os menores valores de MAE e os valores do R2 Score mais próximos de 1. A VUR foi estimada com cada um dos modelos. Em seguida foi feita uma comparação entre os valores estimados e os valores observados, considerando as datas de parada registradas para os aerogeradores, identificadas através das séries temporais de temperatura dos rolamentos principais.

O cálculo da VUR depende da disponibilidade de informações sobre o primeiro instante de previsão; neste estudo esse tempo foi identificado por meio da série temporal avaliada.

As estimativas da VUR foram feitas usando dados das turbinas WT9, WT14 e WT29. Foi então feita uma comparação com a data da falha também identificada na série histórica.

Tabela 9 – Lista de parâmetros e intervalos de parâmetros usados para etapa de validação cruzada

Parâmetros	**Support vector regression (SVR)**	**Decision tree regression (DTR)**	**Isotonic regression (ISOR)**	**Gradient boosting regression (GBR)**	**Random forest (RFR)**	**Extra trees (ETR)**
Parâmetro 1	Regularization parameter - C	Criterion: function to measure the quality of a split.	Lower bound on the lowest predicted value (y_min)	Loss	Criterion: function to measure the quality of a split.	Criterion: function to measure the quality of a split.
Intervalo	0,1, 1,0, 10,0	Squared error, absolute error, and Friedman MSE	0, 0,1, 0,25, 0,5	Squared error, absolute error, Huber, quantile	Squared error, absolute error, and Friedman MSE	Squared error, absolute error, and Friedman MSE
Parâmetro 2	Polynomial degree	The maximum depth of the tree.	Upper bound on the highest predicted value (y_max)	Learning rate	The maximum depth of the tree.	The maximum depth of the tree.
Intervalo	5, 6, 7, 8, 9	None, 5, 10	0,5, 0.75, 1	0,01, 0,05 and 0,1	None, 5, 10	None, 5, 10
Parâmetro 3	-	Minimum number of samples required to split an internal node	Whether computing data is increasing or decreasing (increasing)	Number of estimators	Random_state	Random_state
Intervalo	-	2, 5, 10	True, False, 'auto'	100, 250, 500	None, 10, 100	None, 10, 100
Parâmetro 4	-	-	Handles how X values outside of the training domain are handled during prediction. (out_of_bounds)	Criterion	-	-
Intervalo	-	-	clip	Friedman MSE, squared error	-	-

4.5. Conclusões do Capítulo desenvolvimento de modelo computacional para estimativa da vida útil remanescente

Neste Capítulo, três conjuntos de dados distintos foram utilizados para o desenvolvimento e validação do modelo computacional de prognóstico de falhas em aerogeradores: um conjunto sintético, um conjunto de benchmark e um conjunto com dados reais de uma usina eólica. Os conjuntos sintéticos e de benchmark facilitaram o pré-processamento devido ao seu tamanho e ausência de valores ausentes, incorretos ou excesso de outliers. Em contraste, o conjunto de dados reais demandou um pré-processamento mais completo, incluindo filtragem de acordo com condições operacionais, criação de variáveis, remoção de valores discrepantes e NaN's, e aplicação de um filtro de convolução para representar a sazonalidade dos dados.

Além disso, a estratégia de dividir os conjuntos de dados em treinamento, validação e teste, utilizada com o conjunto de dados reais, mostrou-se efetiva para minimizar o *overfitting* nos modelos de prognóstico de aerogeradores. Os modelos regressivos escolhidos, devido às características dos dados e do fenômeno de degradação, puderam ser facilmente implementados utilizando a biblioteca Scikit-learn. Da mesma forma, a validação cruzada implementada com a classe GridSearchCV ocorreu sem problemas, utilizando os intervalos de parâmetros indicados na Tabela 9.

Em suma, este Capítulo demonstrou a viabilidade do uso de diferentes conjuntos de dados e modelos regressivos para o desenvolvimento de um modelo computacional confiável de prognóstico de falhas em aerogeradores.

5. RESULTADOS E DISCUSSÃO

Neste Capítulo serão apresentados e discutidos os resultados dos modelos desenvolvidos a partir de métricas de avaliação e das estimativas da VUR realizados com os dados sintéticos, com a base pública com dados de vibração do eixo de alta velocidade (HSS) e com os dados de rolamento principal de uma usina eólica em operação.

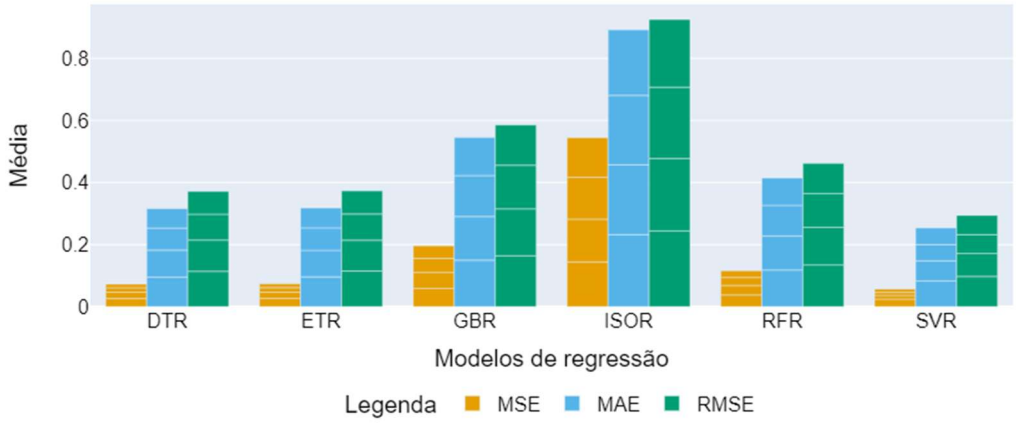
A partir da metodologia utilizada para estimar a VUR de componentes usando os três conjuntos de dados, é proposto um framework com abordagem de complexidade crescente para realizar estimativas da VUR com dados reais.

5.1. Resultados do Estudo de caso – Uso de dados sintéticos

Nesta Seção serão apresentados os resultados – métricas de desempenho e estimativas da VUR – dos modelos regressivos de aprendizagem de máquina, considerando as divisões dos conjuntos de dados (splits) criadas pela classe TimeSeriesSplit na preparação dos dados, conforme mostrado nas etapas simulações e estimativas da VUR na Figura 45.

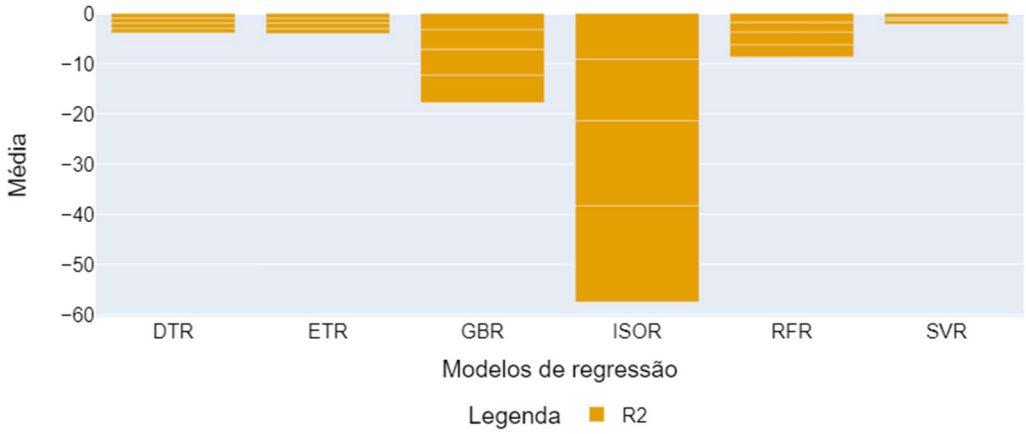
A Figura 49 e a Figura 50 apresentam os resultados gerais das métricas MAE, MSE e RMSE e R2 Score para os modelos utilizados. Da Figura 51 à Figura 56, são apresentadas estimativas da VUR do rolamento do HSS utilizando-se os modelos DTR, ETR, GBR, ISOR, RFR e SVR, respectivamente.

Figura 49 - Métricas MSE, MAE e RMSE na etapa de teste dos modelos de regressão usando o conjunto de dados sintéticos elaborados



Fonte: O autor (2024)

Figura 50 - Métrica R2 Score na etapa de teste dos modelos de regressão usando o conjunto de dados sintéticos



Fonte: O autor (2024)

Avaliando os valores de MAE, MSE e RMSE na etapa de teste, os modelos SVR, ETR, DTR respectivamente, apresentam os melhores resultados nos quais as médias destas métricas não são superiores a 0,096 (MAE), 0,026 (MSE) e 0,115 (RMSE) e as médias do coeficiente de determinação variando de -0,936 a -0,158.

Os modelos RFR e GBR apresentam resultados intermediários, com bons valores para a média das métricas MAE, MSE e RMSE, porém com valores mais negativos, entre -1,786 e -3,254 para o coeficiente de determinação (R2 Score).

O pior resultado foi o do modelo ISOR que mostrou valores mais altos para as métricas de MAE, MSE e RMSE e valor mais negativo para o coeficiente de determinação (-9,128).

Tabela 10 - Valores máximos para as métricas MAE, MSE e RMSE, na etapa de teste, para o conjunto de dados sintéticos de temperatura

Modelo	**MAE**	**MSE**	**RMSE**
	Média	**Média**	**Média**
DTR	0,095	0,026	0,114
ETR	0,096	0,026	0,115
GBR	0,149	0,059	0,164
ISOR	0,232	0,144	0,244
RFR	0,118	0,037	0,134
SVR	0,083	0,024	0,097

Fonte: O autor (2024)

Tabela 11 - Valores máximos para a métrica do coeficiente de determinação (R2 Score), na etapa de teste, para o conjunto de dados sintéticos de temperatura

Modelo	R2
	Valor
DTR	-0,872
ETR	-0,936
GBR	-3,254
ISOR	-9,128
RFR	-1,786
SVR	-0,158

Fonte: O autor (2024)

Os valores negativos para o coeficiente de determinação (R2 Score), mostrado na Figura 50 e na Tabela 11, demonstram falta de capacidade dos modelos em captar a relação real entre as variáveis que podem estar relacionadas com o tamanho da amostra, ausência de mais variáveis preditoras ou nível muito alto de ruído.

Os gráficos com os resultados da métricas de cada modelo para cada um dos splits podem ser visualizados no Apêndice.

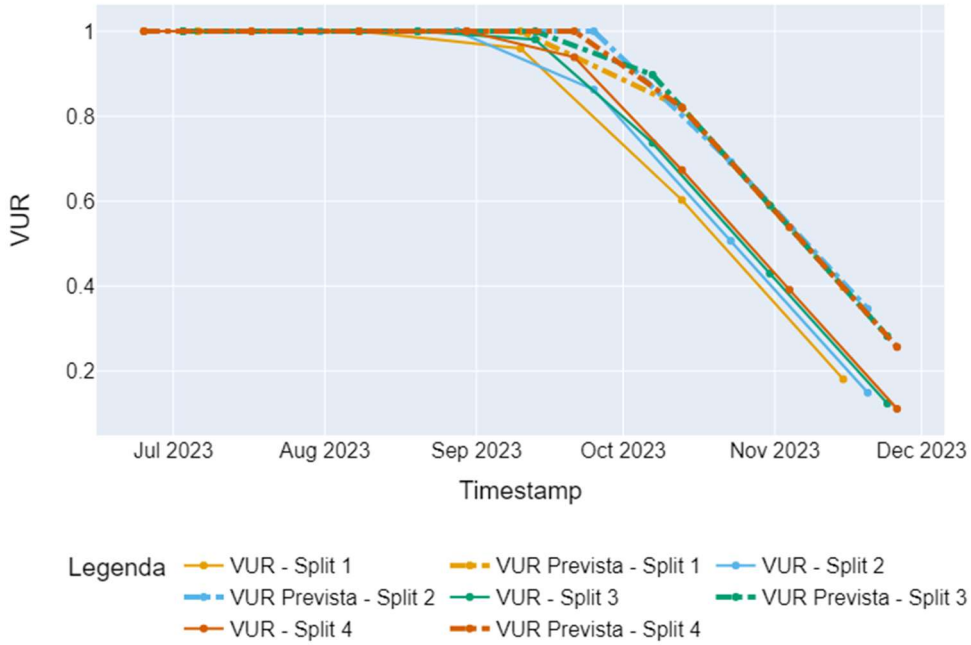
Em seguida, foi feita a análise dos resultados das estimativas de VUR de cada modelo para comparação com as métricas de regressão obtidas nas simulações dos modelos com o conjunto de dados sintéticos.

Observou-se que as previsões realizadas pelos modelos DTR, ETR, GBR, ISOR e RFR mostraram um offset em relação ao instante de ocorrência da falha.

Destacamos que o modelo ISOR apresentou maior dificuldade de realizar a previsão, o que se justifica também pelo valor mais negativo do R2 Score obtido por esse modelo. De acordo com o gráfico da VUR prevista deste modelo, (ver Figura 62), observa-se há dificuldade de realizar a regressão após o início da degradação (instante de detecção da falha).

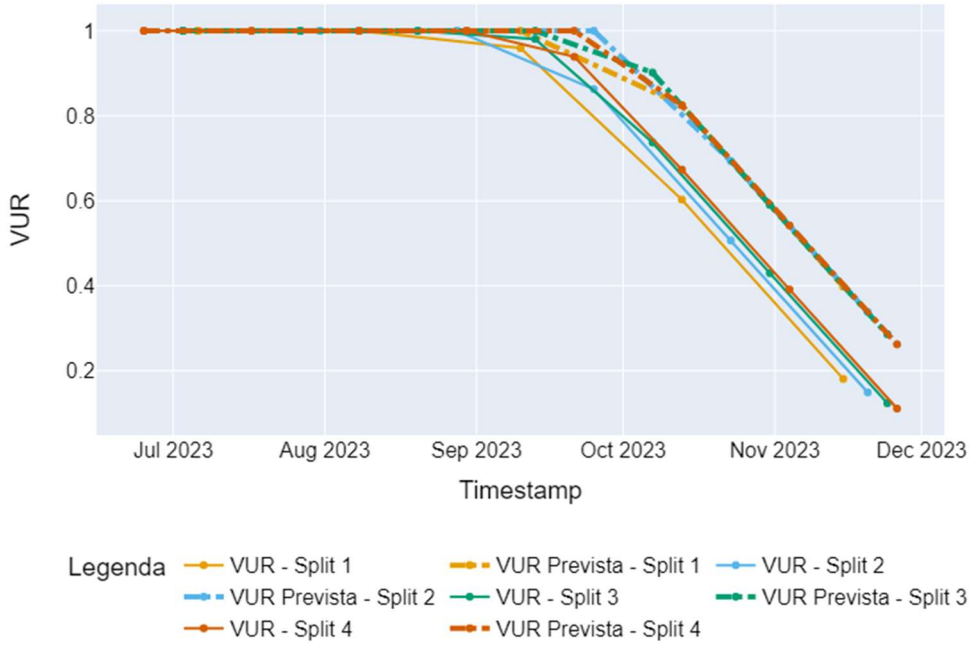
A precisão da estimativa da VUR do SVR (Figura 56) foi a maior, sobretudo para os modelos com 6 e 8 splits.

Figura 51 - Estimativa e VUR do modelo DTR utilizando conjunto de dados sintéticos



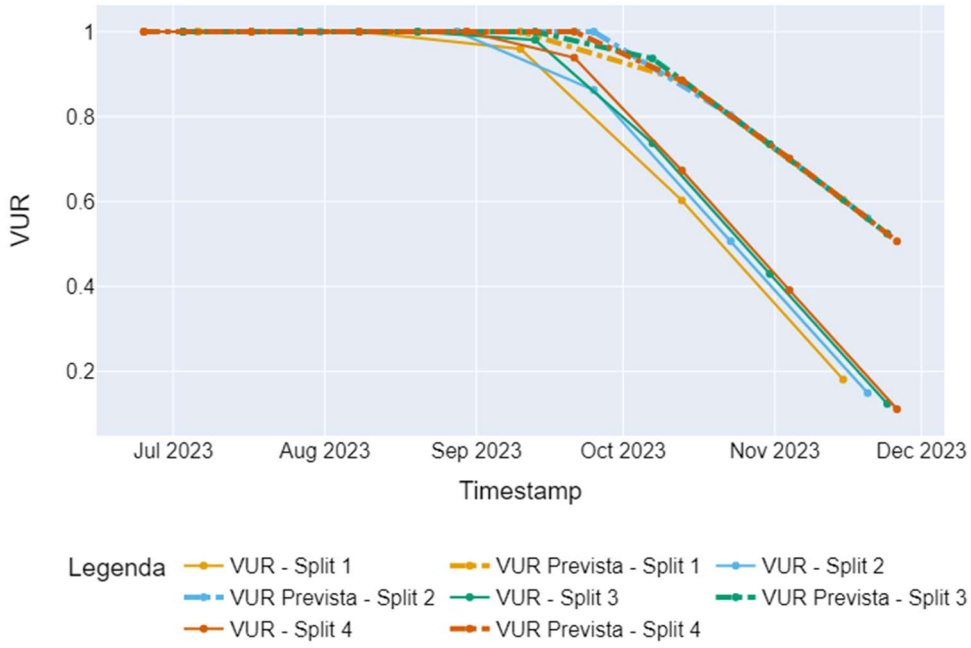
Fonte: O autor (2024)

Figura 52 - Estimativa e VUR do modelo ETR utilizando conjunto de dados sintéticos



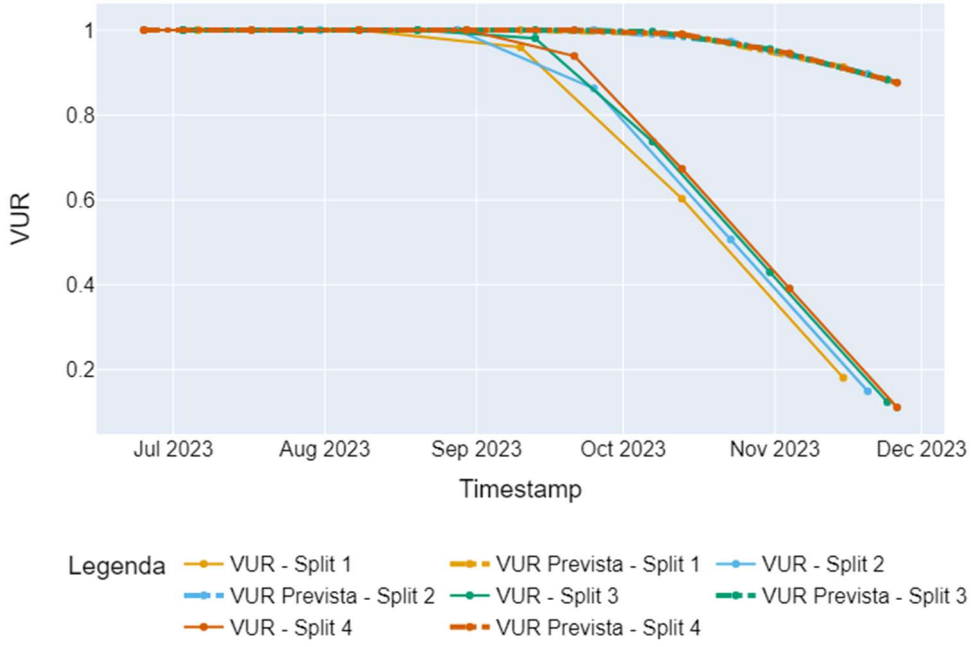
Fonte: O autor (2024)

Figura 53 - Estimativa e VUR do modelo GBR utilizando conjunto de dados sintéticos



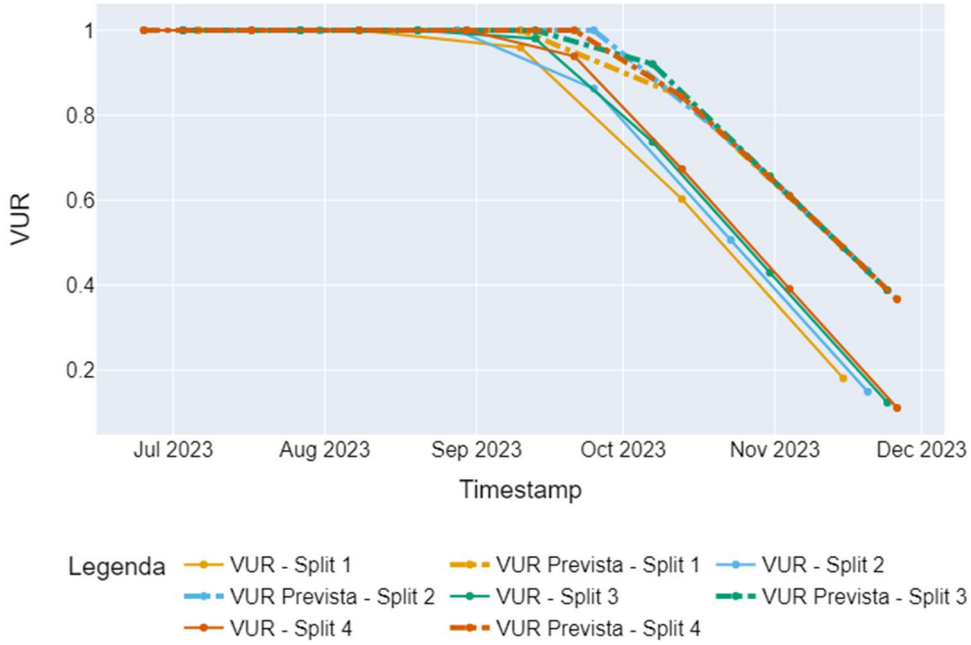
Fonte: O autor (2024)

Figura 54 - Estimativa e VUR do modelo ISOR utilizando conjunto de dados sintéticos



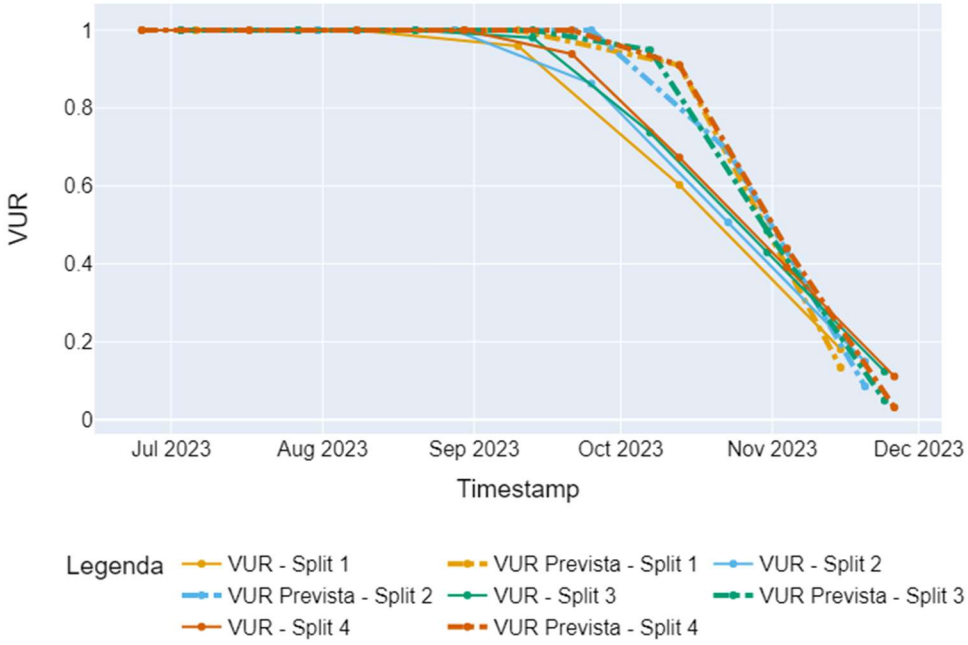
Fonte: O autor (2024)

Figura 55 - Estimativa e VUR do modelo RFR utilizando conjunto de dados sintéticos



Fonte: O autor (2024)

Figura 56 - Estimativa e VUR do modelo SVR utilizando conjunto de dados sintéticos



Fonte: O autor (2024)

5.2.Resultados do Estudo de caso – Uso de dados Mathworks/GPMS

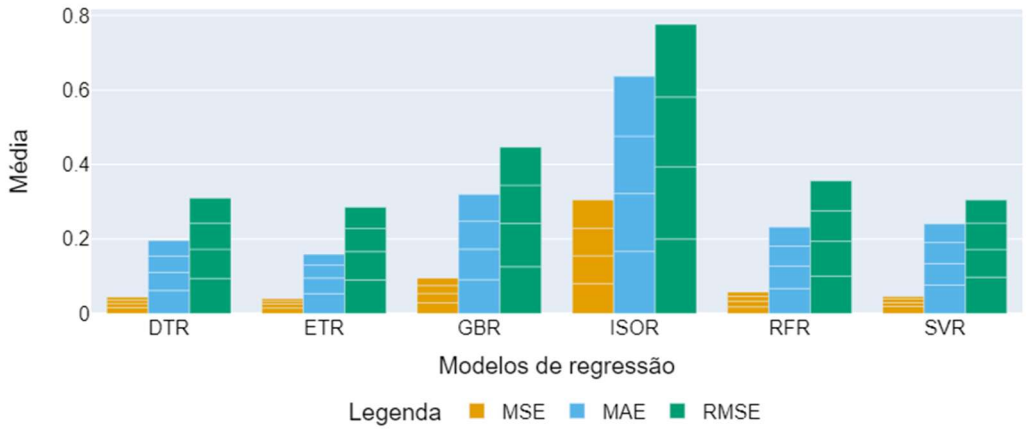
Nesta Seção serão apresentados os resultados com as simulações realizadas com os modelos de regressão desenvolvidos e utilizando diferentes quantidade de divisões da série

temporal usando a classe do Scikit-learn TimeSeriesSplit, conforme mostrado nas etapas simulações e estimativas da VUR na Figura 47.

A Figura 57 e a Figura 58 apresentam os resultados gerais das métricas MAE, MSE e RMSE e R2 Score para os modelos utilizados na etapa de teste. E da Figura 59 à Figura 64 são apresentadas estimativas da VUR do rolamento do HSS utilizando-se o modelo SVR, ISOR, DTR, GBR, ETR e RFR, respectivamente.

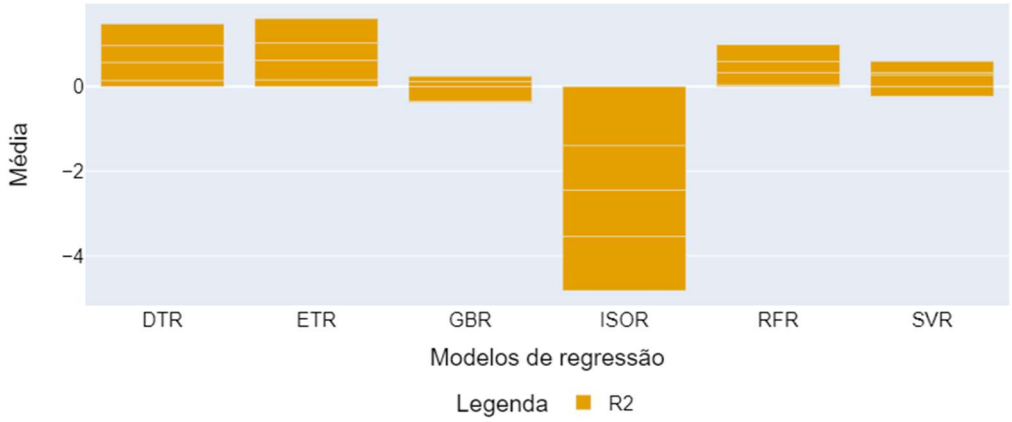
Analisando os valores das métricas MAE, MSE e RMSE, os modelos baseados em árvores decisão, ETR, DTR e RFR respectivamente, apresentam os melhores resultados com valores máximos para a média não superiores a 0,067 (MAE), 0,018 (MSE) e 0,1 (RMSE) e valores variando de 0,399 a 0,575 para a média do coeficiente de determinação na etapa de teste.

Figura 57 - Métricas MSE, MAE e RMSE, na etapa de teste, dos modelos de regressão usando o conjunto de dados Mathworks/GPMS



Fonte: O autor (2024)

Figura 58 - Métrica R2 Score, na etapa de teste, dos modelos de regressão usando o conjunto de dados Mathworks/GPMS



Fonte: O autor (2024)

Os modelos SVR e GBR apresentam resultados intermediários, com bons valores para a média das métricas MAE, MSE e RMSE, porém com valores pequenos, entre 0,129 e 0,282 para o coeficiente de determinação (R2 Score).

O pior resultado foi do modelo ISOR que mostrou valores mais altos para as métricas de MAE, MSE e RMSE e valor negativo para o coeficiente de determinação. De acordo com o gráfico da VUR prevista deste modelo, (ver Figura 62), observa-se há dificuldade de realizar a regressão após o início da degradação (instante de detecção da falha).

Tabela 12 - Valores máximos para as métricas MAE, MSE e RMSE, na etapa de teste, para o conjunto de dados GPMS

Modelo	**MAE**	**MSE**	**RMSE**
	Média	**Média**	**Média**
DTR	0,062	0,016	0,094
ETR	0,053	0,015	0,090
GBR	0,091	0,028	0,125
ISOR	0,167	0,080	0,200
RFR	0,067	0,018	0,100
SVR	0,076	0,019	0,097

Fonte: O autor (2024)

Tabela 13 - Valores máximos para a métrica do coeficiente de determinação (R2 Score), na etapa de teste, para o conjunto de dados GPMS

Modelo	R2
	Valor
DTR	0,517
ETR	0,575
GBR	0,129
ISOR	-1,050
RFR	0,399
SVR	0,282

Fonte: O autor (2024)

Os valores negativos para o coeficiente de determinação (R2 Score), mostrados na Figura 58 e Tabela 13, demonstram falta de capacidade de alguns modelos (GBR, ISOR e SVR) em captar a relação real entre as variáveis podendo estar vinculadas ao tamanho da amostra, a ausência de mais variáveis preditoras, ao nível muito alto de ruído ou uso de modelos lineares que não representem o comportamento não linear (caso do modelo ISOR).

Os gráficos com os resultados da métricas de cada modelo para cada um dos splits podem ser visualizados no Apêndice.

Posteriormente, foi feita a análise dos resultados das estimativas de VUR de cada modelo para comparação com as métricas de regressão obtidas nas simulações dos modelos com o conjunto de dados de benchmark (Mathworks/GPMS).

Os modelos que apresentaram melhor desempenho foram os modelos SVR, ETR e RFR que apresentaram estimativas mais próximas da VUR observada durante as medições do rolamento do HSS.

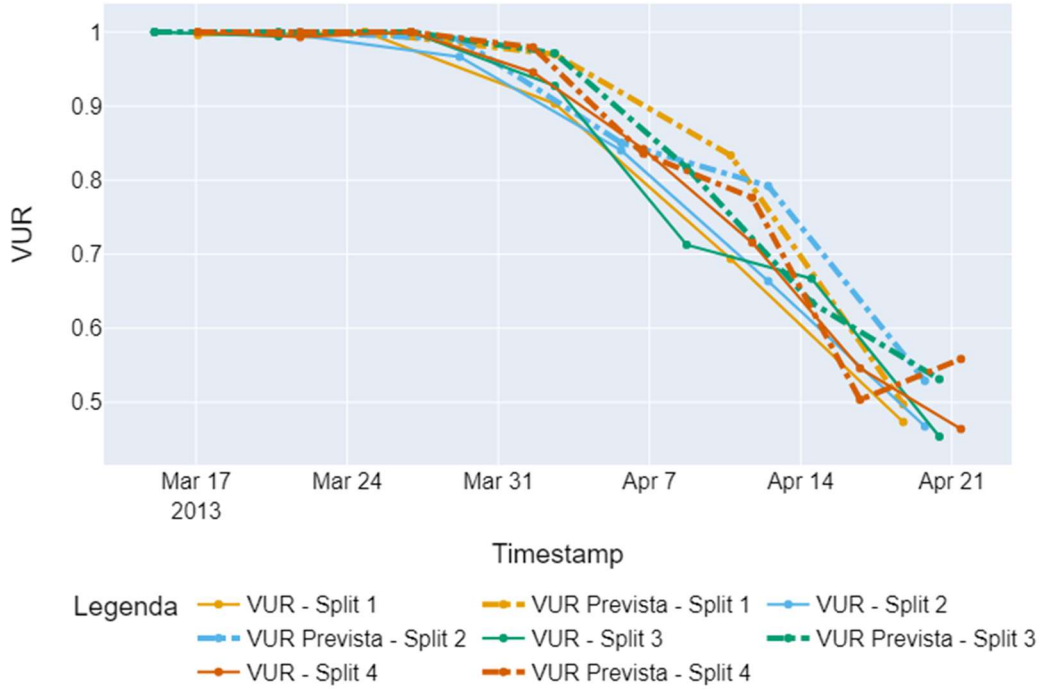
As piores estimativas foram, respectivamente, do modelo ISOR, GBR e DTR, pois as estimativas indicavam a VUR maior em relação ao medido em campo no rolamento.

Similarmente, com o que aconteceu com o conjunto de dados sintéticos, o modelo ISOR foi o que apresentou maior dificuldade de realizar a previsão, o que justifica também pelo valor mais negativo do R2 Score obtido por esse modelo. De acordo com o gráfico da VUR prevista deste modelo, (ver Figura 62), observa-se há dificuldade de realizar a regressão após o início da degradação (instante de detecção da falha).

Destaca-se ainda que quanto maior a quantidade de dados antes da falha melhor deverão ser os resultados, conforme destacado por SAXENA et al., (2008) por representar melhor a dinâmica de degradação do componente.

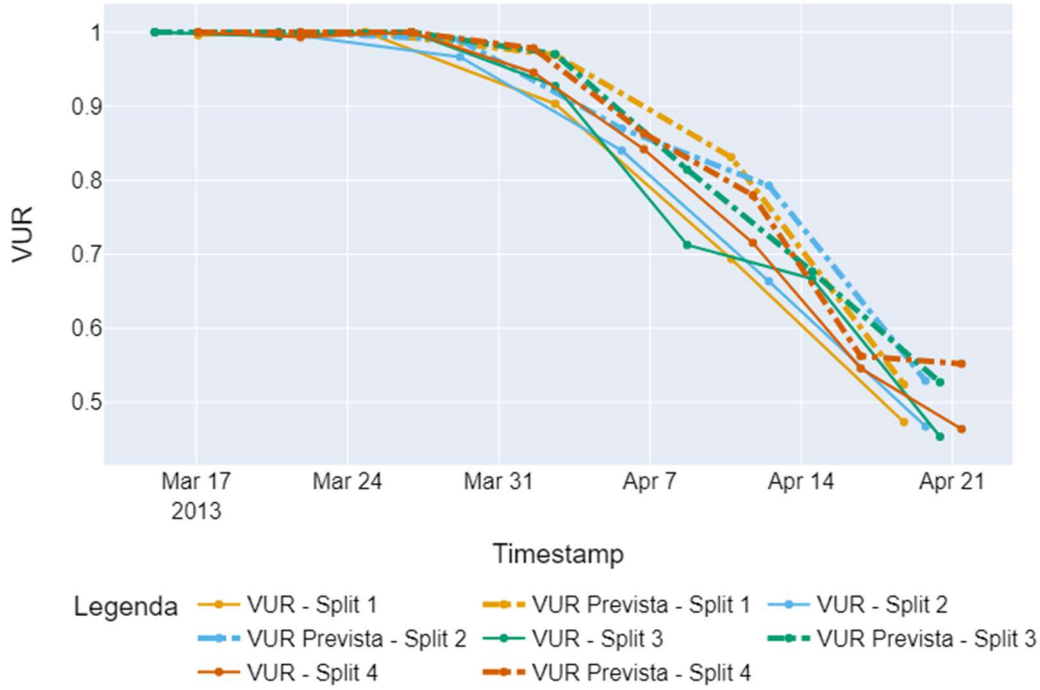
Também neste caso, o SVR (Figura 64) demonstrou maior precisão que os outros modelos, seguido pelo modelo ETR e RFR.

Figura 59 – Estimativa e VUR do rolamento do HSS utilizando modelo DTR



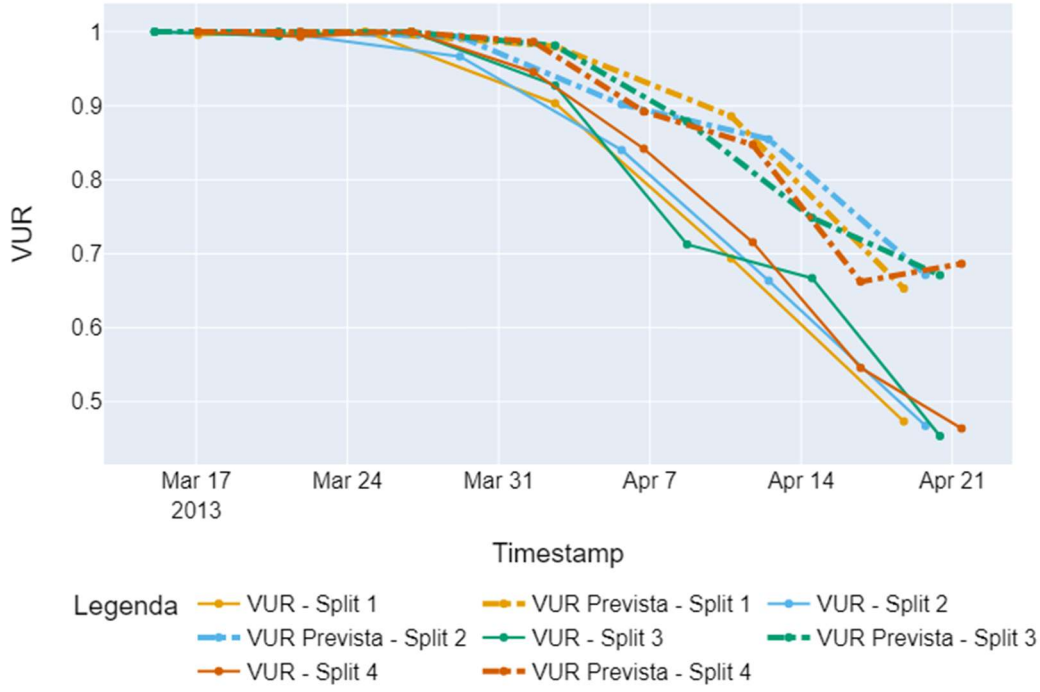
Fonte: O autor (2024)

Figura 60 - Estimativa e VUR do rolamento do HSS utilizando modelo ETR



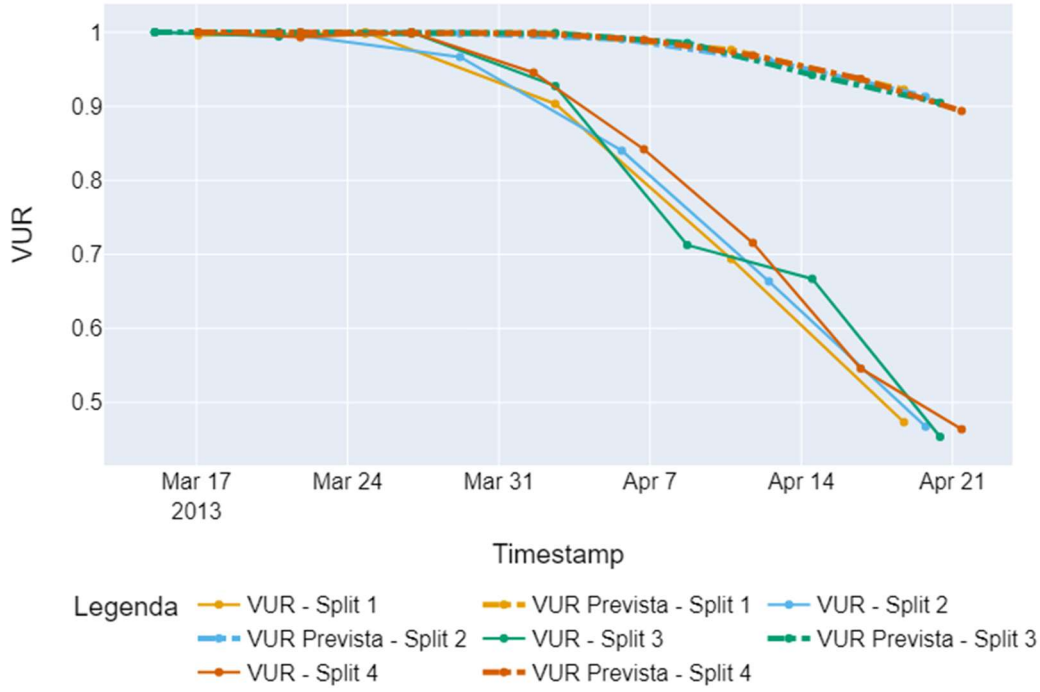
Fonte: O autor (2024)

Figura 61 - Estimativa e VUR do rolamento do HSS utilizando modelo GBR



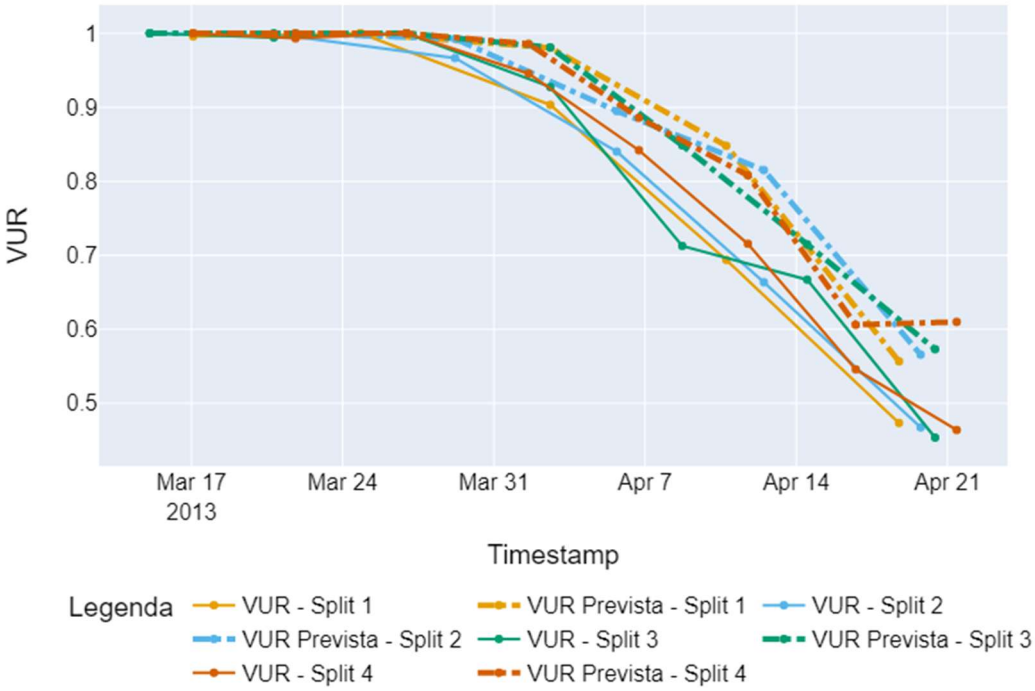
Fonte: O autor (2024)

Figura 62 - Estimativa e VUR do rolamento do HSS utilizando modelo ISOR



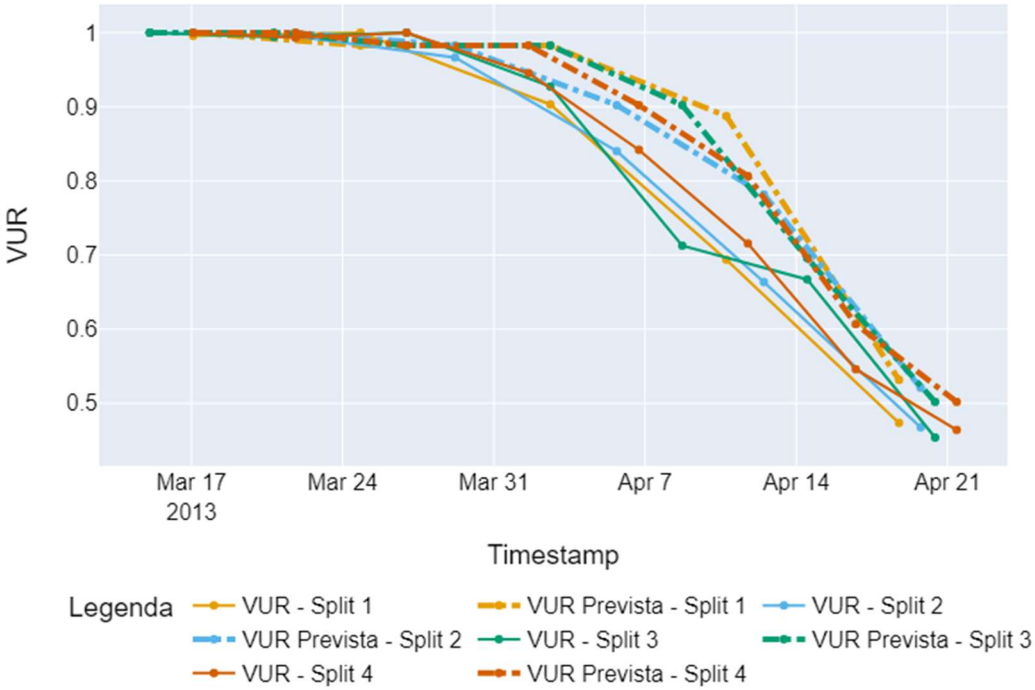
Fonte: O autor (2024)

Figura 63 - Estimativa e VUR do rolamento do HSS utilizando modelo RFR



Fonte: O autor (2024)

Figura 64 - Estimativa e VUR do rolamento do HSS utilizando modelo SVR



Fonte: O autor (2024)

5.3. Resultados do Estudo de caso – Uso de dados SCADA – Temperatura rolamento principal

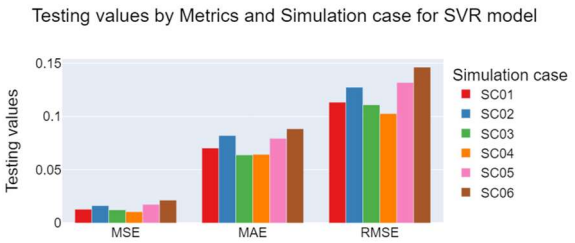
Esta Seção apresenta os resultados obtidos com os modelos nas etapas de treinamento, validação e testes. Primeiramente são apresentados os resultados das métricas do modelo de regressão e dos modelos selecionados e, em segundo lugar, são apresentados os resultados das simulações para estimativa da VUR. A seguir, são apresentadas as observações sobre os resultados da estimativa da VUR e uma discussão sobre os resultados da Seção.

Para avaliação dos modelos de regressão foram escolhidas as métricas que permitissem uma fácil interpretação (MAE, MSE, RMSE), que tivessem uma menor sensibilidade aos outliers (MAE) e explicassem a influência das variáveis independentes na variável dependente – VUR – (R2 Score). Quanto menores os valores de MSE e RMSE melhores os resultados dos modelos. O MAE deveria variar de 0 a 1, devido ao intervalo da VUR sendo desejáveis valores mais próximos de zero. E valores mais próximos de 1 são desejáveis para o R2 score.

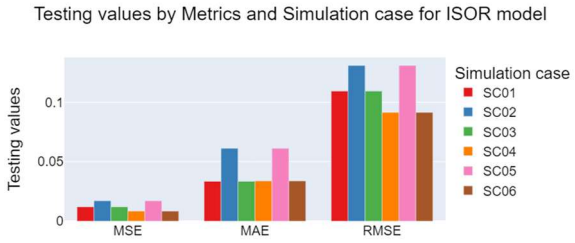
A Figura 65 ilustra os valores médios das métricas MSE, RMSE e MAE durante a fase de teste do modelo. Pequenas variações foram observadas no MAE e no MSE.

Apesar de a média do MAE ter variado de 0,02 a 0,07 e o desvio padrão médio, obtido a partir dos seis casos simulados, apresentou variação de 0,002 e 0,029 indicando que os modelos devem apresentar diferenças nas estimativas da VUR. Os modelos com as menores médias do MAE foram o RFR e DTR. Os menores valores de desvio padrão foram de observados para os modelos GBR, ETR e DTR Consequentemente, estes modelos deverão fazer estimativas com menores erros, o que se comprovou nos modelos GBR e RFR, como será mostrado na Seção 5.3.1.

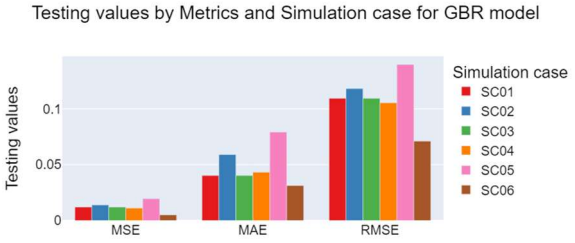
Figura 65 - Resultados de MAE, MSE e RMSE para cada modelo de regressão e cada caso de simulação: (a) Vetor de suporte, (b) Isotônica, (c) Aumento de gradiente, (d) Árvore de decisão (e) Árvores extras e (f) Floresta aleatória



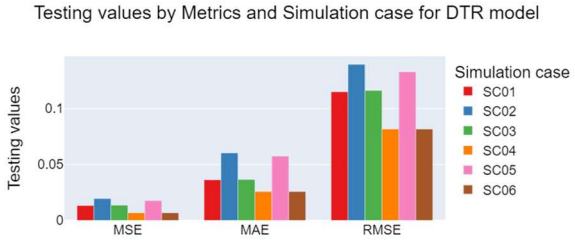
(a) – Modelo Support vector regression



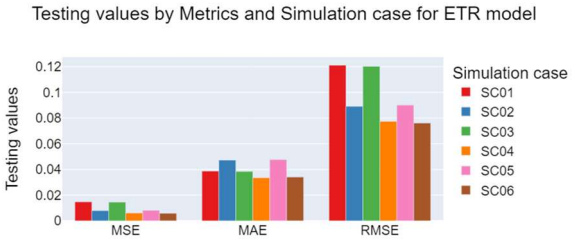
(b) – Modelo Isotonic regression



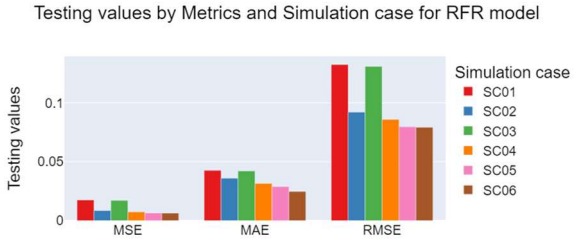
(c) – Modelo Gradient Boosting regression



(d) – Modelo Decision Tree regression



(e) – Modelo Extra trees regression

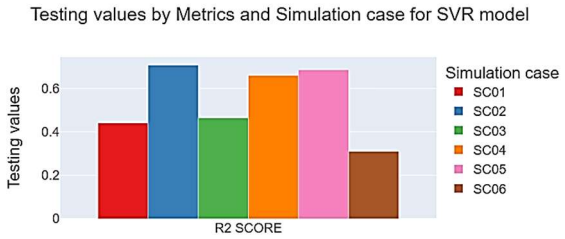


(f) – Modelo Random forest regression

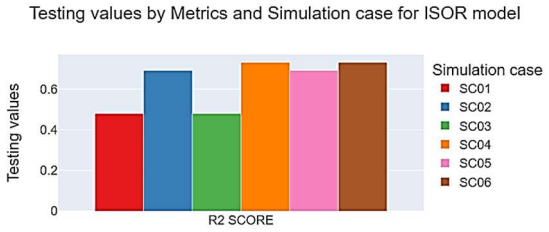
Fonte: O autor (2024)

No caso da métrica R2 Score, os valores médios dos subconjuntos de treinamento, validação e teste variaram de 0,894 a 0,639, com desvio padrão variando de 0,073 a 0,402 (ver Figura 66). Os modelos RFR, ETR e GBR representaram a melhor influência da variação de temperatura do rolamento principal na VUR, maiores valores de R2 Score. Os menores valores de desvio padrão foram observados para os modelos ISOR, ETR e GBR. Além disso, os resultados do R2 Score indicaram que os subconjuntos de dados 01 e 03 exibiram os valores mais baixos.

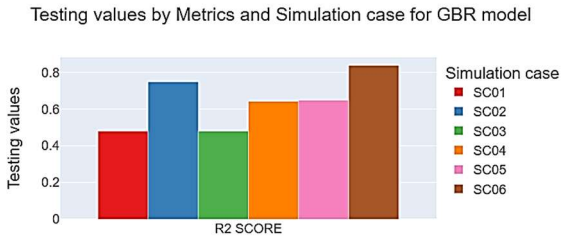
Figura 66 - Resultados do R2 Score para cada modelo de regressão e cada caso de simulação: (a) Vetor de suporte, (b) Isotônica, (c) Aumento de gradiente, (d) Árvore de decisão (e) Árvores extras e (f) Floresta aleatória



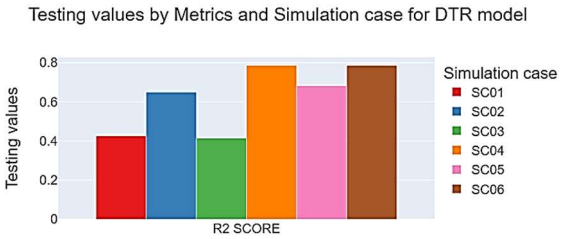
(a) – Modelo Support vector regression



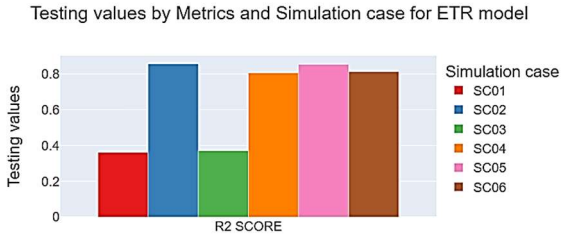
(b) – Modelo Isotonic regression



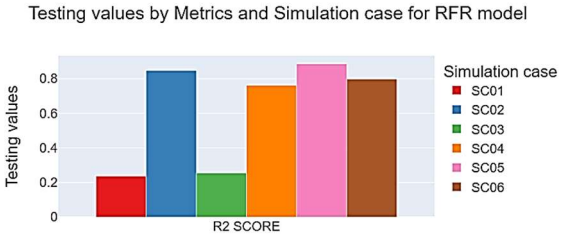
(c) – Modelo Gradient Boosting regression



(d) – Modelo Decision Tree regression



(e) – Modelo Extra trees regression



(f) – Modelo Random forest regression

Fonte: O autor (2024)

Dentre as métricas, houve pouca variação entre as etapas de treinamento e validação:

- Para o MAE, as médias variaram de 0,27 no treinamento a 0,25 na validação, Tabela 14;
- Para o MSE, as médias variaram de 0,005 no treinamento a 0,004 na validação, Tabela 15;
- Para o RMSE, as médias variaram de 0,066 no treinamento a 0,063 na validação, Tabela 16;
- Para o R2 Score, as médias variaram de 0,839 no treinamento a 0,86 na validação, Tabela 17.

As pequenas diferenças entre as etapas dificultaram a anotação precisa dos modelos testados.

Tabela 14 - Resultados do erro absoluto médio (MAE) para todos os casos de simulação, para cada modelo, para etapas de treinamento, validação e teste

Modelos	Média dos valores de treinamento	Média dos valores de validação	Média dos valores de teste
DTR	0,014	0,014	0,040
ETR	0,023	0,023	0,040
GBR	0,026	0,026	0,049
ISOR	0,025	0,025	0,043
RFR	0,018	0,014	0,034
SVR	0,057	0,051	0,075
Média geral do MAE	0,027	0,025	0,047

Fonte: O autor (2024)

Tabela 15 - Resultados do erro quadrático médio (MSE) para todos os casos de simulação, para cada modelo, para as etapas de treinamento, validação e teste

Modelos	Média dos valores de treinamento	Média dos valores de validação	Média dos valores de teste
DTR	0,004	0,004	0,013
ETR	0,004	0,004	0,010
GBR	0,004	0,004	0,012
ISOR	0,005	0,005	0,013
RFR	0,005	0,003	0,011
SVR	0,007	0,006	0,015
Média geral do MSE	0,005	0,004	0,012

Fonte: O autor (2024)

Tabela 16 - Resultados do erro quadrático médio (RMSE) para todos os casos de simulação, para cada modelo, para etapas de treinamento, validação e teste

Modelos	Média dos valores de treinamento	Média dos valores de validação	Média dos valores de teste
DTR	0,059	0,058	0,111
ETR	0,062	0,062	0,096
GBR	0,060	0,059	0,109
ISOR	0,071	0,071	0,111
RFR	0,063	0,054	0,100
SVR	0,082	0,077	0,122
Média geral do RMSE	0,066	0,063	0,108

Fonte: O autor (2024)

Tabela 17 - Resultados do coeficiente de determinação (R2 Score) para todos os casos de simulação, para cada modelo, para as etapas de treinamento, validação e teste

Modelos	Média dos valores de treinamento	Média dos valores de validação	Média dos valores de teste
DTR	0,877	0,880	0,623
ETR	0,864	0,865	0,677
GBR	0,861	0,885	0,639
ISOR	0,836	0,836	0,633
RFR	0,807	0,901	0,630
SVR	0,790	0,798	0,545
Média geral do R^2 Score	0,839	0,861	0,625

Fonte: O autor (2024)

Devido à pequena variação na média do MAE e à variação na média e desvio padrão do R2 Score, decidiu-se realizar simulações com todos os modelos e avaliar os resultados avaliando o desempenho dos modelos para cada caso simulado.

5.3.1. Estimativa da vida útil remanescente do conjunto de amostras de turbinas com falha

Nesta Seção, serão apresentadas as simulações realizadas com os modelos utilizando os dados de turbina definidos no subconjunto de teste para os casos apresentados na Seção 4.3.3. A Tabela 18 apresenta os melhores resultados obtidos nas simulações com os subconjuntos de teste.

A coluna **Estimativa da VUR [dias]** indica o valor calculado da vida útil remanescente obtida pelo respectivo modelo e com o respectivo conjunto de dados e a na coluna **Erro de cálculo da estimativa VUR [dias]** são indicadas as diferenças entre os resultados das colunas data prevista e data registrada da falha.

Valores negativos na coluna de erro de estimativa da VUR indicam que a data estimada foi menor que a data registrada. Valores positivos nesta mesma coluna indicam que a data estimada foi maior que a data registrada.

Os resultados com os modelos SVR e ETR foram obtidos sem as informações dos dados de detecção de falhas. Os resultados para os modelos GBR e RFR foram obtidos com informações sobre a data de detecção de falhas.

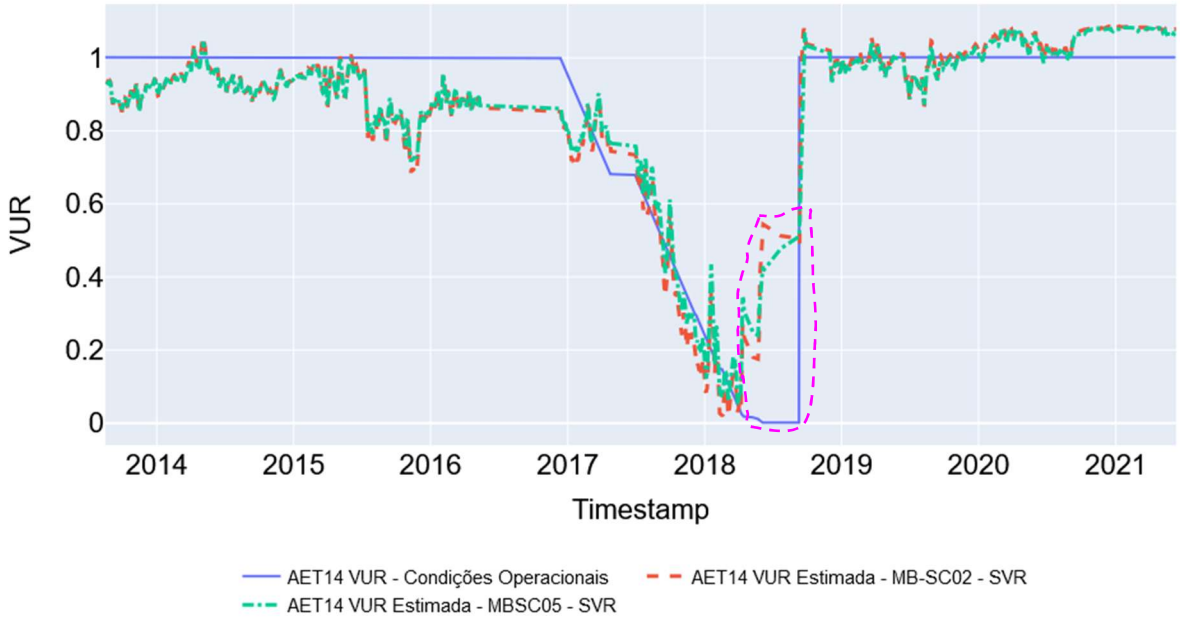
Tabela 18 - Resultados de simulação com subconjuntos de teste

Casos de simulação	Modelo	Estimativa da VUR [dias]	Erro de cálculo da estimativa VUR [dias]
MB-SC02	SVR	1638.0	-111
MB-SC05	SVR	1685.0	-64
MB-SC04	GBR	393.0	0
MB-SC04	RFR	461.0	68
MB-SC02	ETR	1856.0	107
MB-SC06	ETR	1855.0	118

Fonte: O autor (2024)

A Figura 67 mostra os resultados dos modelos SVR para os dois casos de simulação (SC02 e SC05). Observa-se que os modelos, treinados e validados com dados de outras turbinas, conseguem representar o comportamento da WT14, usada nos testes até próximo da data da falha. O modelo apresenta erros maiores para representar o intervalo entre a falha (parada da máquina) e o retorno a operação (vide região destacada em magenta).

Figura 67 - Resultados de simulações com modelos SVR (MB-SC02 e MB-SC05)



Fonte: O autor (2024)

5.3.2. Observações sobre os resultados da estimativa da VUR

Os modelos desenvolvidos demonstraram desempenho satisfatório para as métricas comumente usadas para avaliar modelos de regressão. No entanto, as pequenas diferenças no desempenho tornaram desafiador estabelecer uma estratégia de ranqueamento dos modelos.

Nas simulações para estimar a VUR, os modelos SVR e ETR puderam reproduzir o comportamento de degradação com base nos dados da série temporal. O modelo SVR exibiu resultados mais conservadores, prevendo a data da falha antes da registrada.

Por outro lado, os modelos GBR e RFR exigiam informações sobre a data de detecção para fazer uma estimativa mais precisa da VUR.

5.3.3. Estimativa da vida útil restante do conjunto de amostras de turbinas com falha

Os resultados mostraram que é possível realizar estimativas da VUR com uso de poucos dados e sem sobreajuste (*overfitting*) com o uso de dados de três aerogeradores. Isso o torna uma opção complementar viável para a tomada de decisão por gerentes de manutenção de parques eólicos. No entanto, conforme destacado por vários autores (CORREA-JULLIAN *et al.*, 2022; TUTIV'EN *et al.*, 2021; ZHAO *et al.*, 2017), técnicas intensivas de pré-processamento de dados reais foram necessárias para obter séries temporais que permitiriam resultados minimamente razoáveis a partir de modelos de aprendizado de máquina regressivos. Esta etapa requer a análise do comportamento mecânico de componentes, subsistemas e sistemas em estudo, combinados com técnicas de processamento de dados para gerar efetivamente informações úteis para a operação e manutenção de aerogeradores.

Os resultados das estimativas de vida útil remanescente (VUR) dos aerogeradores permitem interpretação distintas do ponto de vista da manutenção. Pode-se considerá-las em assertiva, como o modelo GBR que previu a data com erro de zero dias; conservadora, nos casos dos modelos SVR que subestimaram a vida útil; e não conservadora, como os modelos ETR e RFR que superestimaram os valores encontrados nas séries temporais (ver Tabela 18). Portanto, estimativas conservadoras, como as do SVR, e assertivas, como as do GBR, podem subsidiar um planejamento adequado da manutenção dos recursos necessários para a realização de paradas programadas e intervenções nas turbinas, evitando falhas catastróficas que poderiam reduzir a disponibilidade do parque eólico. Dessa forma, as estimativas subsidiam a prevenção de falhas graves e garantem a operação das turbinas (SAHU *et al.*, 2023).

A estratégia de criar casos de simulação provou ser eficaz para lidar com a questão da escassez de dados específica dos aerogeradores, sem a necessidade de usar preenchimento de dados ou dados sintéticos.

No entanto, para a aplicação bem-sucedida do método são necessários dados das variáveis de interesse, indicadas na Seção 4.3.1 de pelo menos 3 aerogeradores que tenham apresentado a falha do rolamento principal, de modo a minimizar o sobreajuste (*overfitting*). Outra limitação do método é a necessidade de dados temperatura do rolamento principal com

crescimento monotônico e se possível com registros próximos da data de parada da turbina para substituições. O uso de valores baixos de temperatura para treinamento e/ou validação do modelo podem enviesar os resultados e o método poderá indicar uma substituição antes do tempo real do final da vida deste componente.

5.4.Conclusões do Capítulo resultados e discussões

Este Capítulo apresentou os resultados das simulações realizadas com modelos regressivos aplicados a três tipos de conjuntos de dados: sintéticos, de benchmark e reais de uma usina eólica.

A utilização de conjuntos de dados sintéticos e de benchmark permitiu uma avaliação rápida e eficiente do desempenho dos modelos, evidenciando suas capacidades de estimativa e métricas associadas. As simulações com dados sintéticos revelaram a necessidade de aprimorar a geração de dados, incorporando informações mais complexas, devido aos valores obtidos do coeficiente de determinação para todos os modelos, independentemente de sua propensão ao *overfitting*. Por outro lado, os dados de benchmark apresentaram uma evolução de falha mais próxima das aplicações reais, destacando os modelos com melhores métricas (ETR, DTR e RFR) e previsões da VUR (SVR, ETR, DTR e RFR).

Os resultados com o conjunto de dados reais demonstraram a eficácia da estratégia de utilização de subconjuntos de dados, conforme detalhado na Seção 4.3.3, para redução do *overfitting* e obtenção de boas métricas regressivas com o conjunto de testes. As métricas médias alcançadas (MAE: 0,047, MSE: 0,012, RMSE: 0,108 e R2 SCORE: 0,625) evidenciam a capacidade de generalização dos modelos. Além disso, as estimativas da VUR apresentaram resultados similares ou superiores aos encontrados na literatura, com uma média de erro de 20 dias para a previsão do final da vida útil.

Estes resultados demonstram o potencial significativo do uso dessas previsões como ferramenta de apoio à tomada de decisões pelas equipes responsáveis pela operação e manutenção de usinas eólicas, contribuindo para uma gestão mais eficiente e proativa dos ativos.

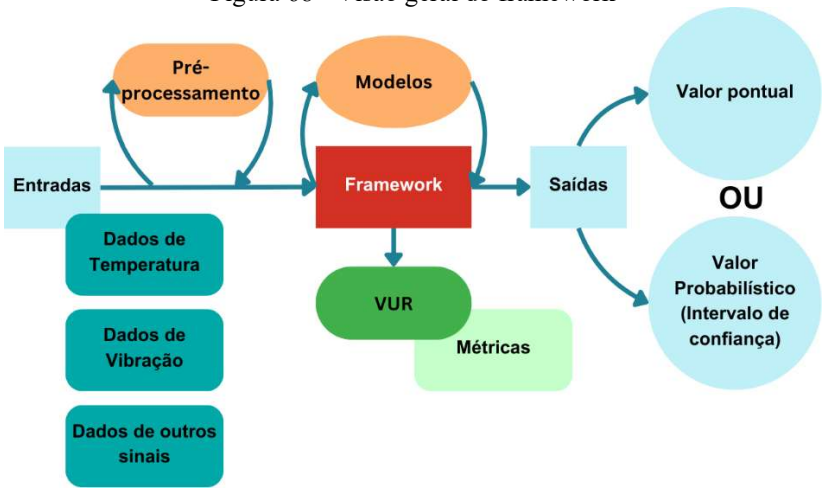
5.4.1. Proposta de framework para estimativa da VUR de aerogeradores com vistas a generalização

Com base nas conclusões da Seção 2.6 do Capítulo 2, da revisão sistemática e dos resultados apresentados, propomos um framework como uma forma para estimar a VUR usando dados reais de operação de aerogeradores.

O framework proposto do fluxograma, mostrado na Figura 68, é composto por 5 etapas que não são necessariamente obrigatórias, mas complementares entre si, detalhada na Figura 69. A principal característica é a utilização de dados de entrada com aumento gradativo de complexidade, permitindo ajustes nos modelos que possibilitem a generalização dos resultados e a manutenção do bom desempenho das métricas de desempenho.

Os dados de entrada dos modelos podem ser temperatura, vibração e outro sinal específico de um determinado componente ou subsistema, ver Figura 68, adquirido por SCADA ou CMS. Ao utilizar dados SCADA, o conjunto de dados também deve conter pelo menos dados sobre velocidade do vento, potência ativa gerada e rotação – mínima e máxima – do gerador para permitir a filtragem dos dados com base nas condições de operação do aerogerador. Em todas as fases da estrutura, as métricas do modelo devem ser calculadas e avaliadas quanto ao desempenho e ao *overfitting*.

Figura 68 - Visão geral do framework

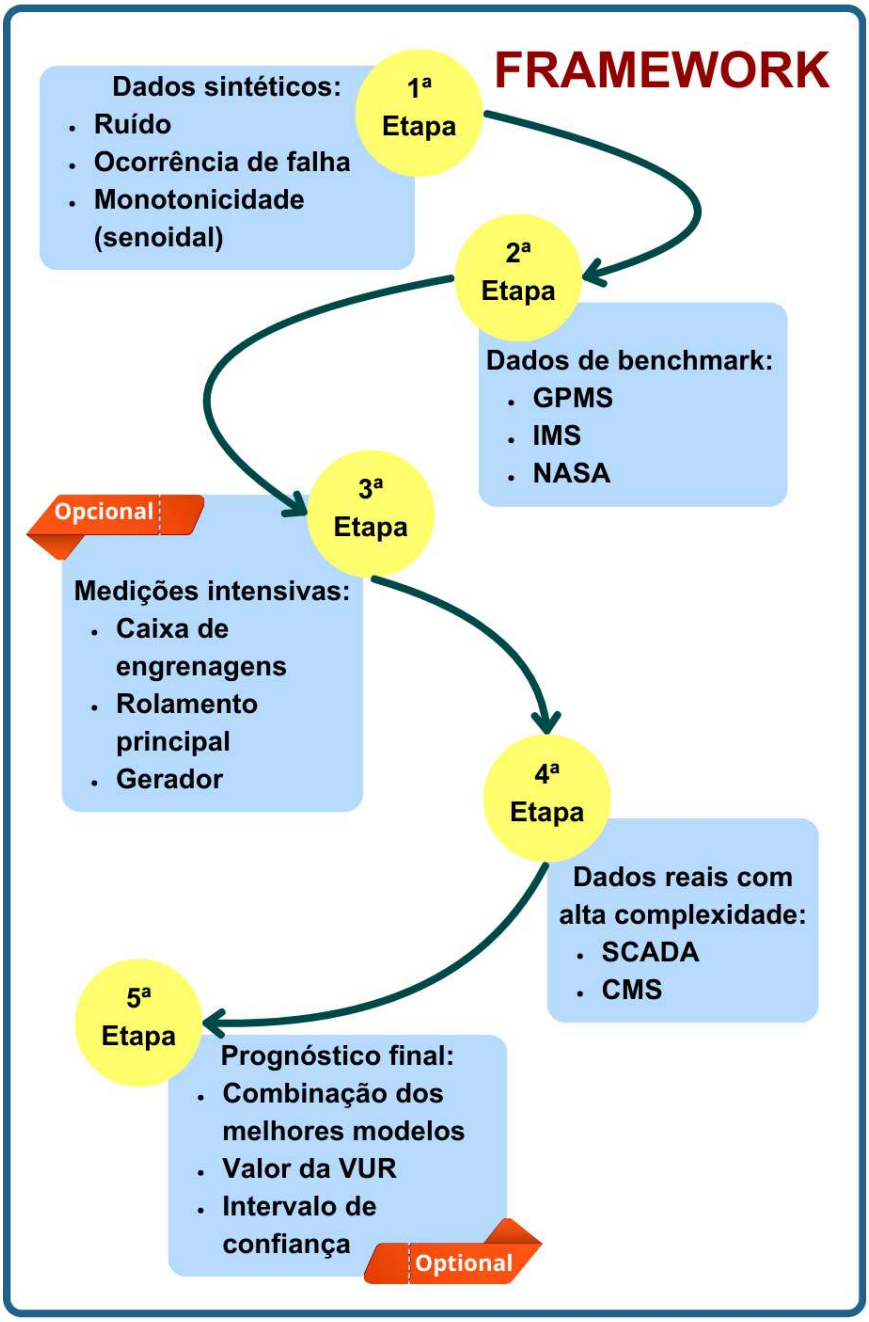


Fonte: O autor: (2024)

Na primeira etapa propõe-se a realização de modelagem e testes com dados sintéticos, com característica monotônica, com oscilação senoidal e ruído aleatório presente no sinal e pelo menos um período de falha incluído. O período de falha é caracterizado pelo alcance de limites superiores ou inferiores para uma determinada grandeza medida. As estimativas da VUR, feitas

pelos modelos em teste, deverão ser calculadas com valores pontuais, desconsiderando a incerteza.

Figura 69 - Etapas detalhadas do framework.



Fonte: O autor: (2024)

Na segunda etapa, a modelagem e os testes devem ser realizados utilizando medições de grandezas em bancada (testes de falha acelerados) ou em condições reais de operação até o final da vida útil de um determinado componente utilizando conjuntos de dados públicos, por exemplo, Dados de rolamentos definido em IMS (LEE *et al.*, 2007), NASA, Mathworks/GPMS

(MATHWORKS INC.; BECHHOEFER, 2017). Nesta fase, apenas devem ser calculados também os valores pontuais da VUR, sem estimar a incerteza para verificar a capacidade do modelo realizar a estimativa.

A terceira etapa é opcional e depende da possibilidade de realizar medições intensivas de quantidades de dados de entrada de determinados componentes com taxa de aquisição superior à comumente encontrada nos sistemas SCADA e CMS com sensores e sistemas de aquisição dedicados. O processamento dos dados deverá ser realizado em sistema separado do SCADA e do CMS. Nesta fase, uma etapa de validação cruzada deve ser incluída antes dos testes para ajustar os hiperparâmetros dos modelos. As estimativas da VUR, feitas pelos modelos no teste, deverão ser calculadas com valores pontuais e computando a incerteza.

A quarta etapa do framework deve utilizar conjuntos de dados de maior complexidade, preferencialmente de pelo menos 4 aerogeradores em condições reais de operação, também recomendado por BARBER et al., (2023). Modelagem, validação cruzada e testes com os modelos devem ser realizados. Novamente, as estimativas da VUR deverão conter o cálculo da incerteza.

Na quinta etapa, que é opcional, propõe-se que seja feita uma combinação dos melhores modelos e utilizados conjuntos de dados de maior complexidade, preferencialmente de pelo menos 4 aerogeradores em condições reais de operação. As estimativas da VUR deverão novamente apresentar o cálculo da incerteza.

A abordagem de aumento gradual da complexidade dos dados para estimativa de VUR busca permitir a compreensão do comportamento dos modelos e métricas nas estimativas da VUR, incluindo a incerteza com foco no desenvolvimento de modelos capazes de generalizar a dinâmica de degradação dos componentes do aerogerador.

6. CONCLUSÕES E TRABALHOS FUTUROS

Nesta seção, apresentamos as principais conclusões das investigações realizadas, a partir do estudo do estado da arte e dos resultados obtidos com os modelos de aprendizagem de máquina desenvolvidos. Também são propostas algumas direções para trabalhos futuros na área de estimativa da VUR de aerogeradores.

A revisão da literatura, apresentada no Capítulo 2, identificou as principais fontes de dados, as técnicas de pré-processamento, os tipos e métodos utilizados para a estimativa da VUR, os horizontes temporais dessas estimativas, os métodos cálculo incerteza da VUR e os principais modos de falha abordados.

Na revisão da literatura foram identificadas as lacunas e as propostas para preencher elas e contribuir cientificamente apresentadas na Tabela 19.

Tabela 19 - Lacunas identificadas e propostas deste trabalho

Lacuna	Propostas
Carência de discussão sobre a quantidade de dados e a capacidade de generalização, com vistas ao sobreajuste de métodos de estimativa da VUR.	Discussão do tema no âmbito dos modelos de estimativa de VUR para aplicação na indústria eólica e proposição de um método de criação de subconjuntos de dados para as etapas de treinamento, validação e teste.
Ausência de uma classificação da capacidade de generalização dos modelos de estimativa da VUR aplicados em aerogeradores.	Proposta de utilização de uma classificação ca capacidade de generalização com base na quantidade de dados de diferentes aerogeradores para construção dos modelos de AM.
Baixa disponibilidade de dados públicos (brutos ou tratados) adequados para desenvolvimentos de métodos de estimativa de VUR de aerogeradores.	Criação de banco de dados públicos e abertos de séries temporais de diferentes sistemas de monitoramento de aerogeradores para o desenvolvimento de modelos de VUR
Inexistência de um framework para o desenvolvimento de métodos para estimativa da VUR usando dados com aumento gradual de complexidade.	Proposição de um framework que use diferente fontes de dados para criação, treinamento, validação e testes de modelos de AM para estimava da VUR de aerogeradores.

Uma contribuição significativa deste trabalho, específica para área de aerogeradores, foi a proposição de uma classificação da capacidade de generalização dos modelos inteligentes usados na estimativa de VUR, considerando a quantidade de conjuntos de dados utilizados para reduzir o sobreajuste (*overfitting*). Esta proposta preenche uma lacuna na literatura, que ainda carece de discussão sobre o volume mínimo de dados, provenientes de diferentes aerogeradores e de períodos mais longos de operação, conforme as boas práticas de generalização, para obter modelos com capacidade de generalização adequada para aplicação em aerogeradores.

Destacou-se a escassez de conjuntos de dados públicos e a falta de dados sintéticos ou experimentais, especificamente de aerogeradores, para avaliação de métodos. E concluiu-se que a disponibilização de dados abertos pode impulsionar o desenvolvimento de modelos mais eficazes e promover sinergia entre academia e indústria, estimulando inovações no prognóstico de falhas e no gerenciamento da saúde de aerogeradores (BARBER *et al.*, 2023).

No que se refere aos modelos desenvolvidos e simulações realizadas, concluiu-se que o uso de conjuntos de dados sintéticos e de benchmark permitiu uma avaliação rápida e eficiente do desempenho dos modelos, evidenciando suas capacidades de estimativa e métricas associadas.

Também se desenvolveu e testou uma abordagem para realizar o prognóstico de falhas de rolamentos principais de aerogeradores, utilizando dados de temperatura do sistema SCADA. Os dados de temperatura de 3 aerogeradores e diferentes modelos de regressão de aprendizado de máquina foram usados para estimar a VUR do mancal principal.

Os modelos propostos, testados em dados reais de três aerogeradores localizadas no nordeste do Brasil, apresentaram métricas MAE, MSE, RMSE e R2 Score satisfatórias nos estágios de treinamento, validação e testes. Os resultados da simulação demonstraram que os modelos (SVR, ETR, GBR e RFR) superaram a previsão da data da falha do equipamento, com uma diferença média de 20 dias e um máximo de 100 dias. O modelo ISOR se mostrou inadequado para previsão da VUR devido a sua capacidade de lidar com dados monotônicos, não sendo adequado para dados que contém informações sobre operação, parada e retorno a operação. O método de uso de subconjuntos de dados reais foi eficaz na mitigação do sobreajuste (*overfitting*) devido os dados escassos dos aerogeradores. Outra contribuição expressiva deste trabalho devido à disponibilidade limitada de dados de séries temporais sobre falhas de mancais principais.

Concluiu-se que com a metodologia proposta que as estimativas conservadoras (SVR), e assertivas (GBR) podem subsidiar um planejamento adequado da manutenção, evitando falhas catastróficas e redução da disponibilidade de usinas eólicas.

Outra conclusão deste trabalho, foi a proposição de um framework para estimar a VUR usando dados reais de operação de aerogeradores, que pode servir para o processo de implantação e amadurecimento de métodos de prognóstico de falha, com aumento gradual da complexidade dos dados e modelos, com potencial impacto na cultura organizacional em relação aos cuidados para coleta e análise de dados e do uso de tecnologias de IAI.

Enfatiza-se a importância da colaboração entre indústria e academia para aumentar a disponibilidade de dados para prognóstico de falhas de aerogeradores em repositórios públicos, permitindo o desenvolvimento de modelos mais robustos e eficazes.

6.1. TRABALHOS FUTUROS

Propõem-se como sugestões de trabalhos futuros:

- O desenvolvimento de uma estrutura que use dados sintéticos, dados de bancada e dados reais para construção modelos mais complexos usando abordagens, de aprendizagem profunda (*deep learning*), de combinação de modelos, de *autoencoders* (GOODFELLOW; BENGIO; COURVILLE, 2016) ou de modelos de linguagem de grande escala (*large language models*);
- O desenvolvimento de modelos que consideram as incertezas das estimativas da VUR para diferentes componentes dos aerogeradores;
- O desenvolvimento de metodologias que considerem os fatores logísticos e seus impactos nas estimativas da VUR para prognósticos de falhas mais complexos e eficazes do ponto de vista das empresas interessadas;
- O desenvolvimento de metodologias que consideram a influência e a relevância das estimativas da VUR de componentes de aerogeradores para a gestão da manutenção e planejamento de paradas gerais de usinas eólicas e/ou manutenções oportunísticas de componentes com vida útil próxima do fim
- O desenvolvimento de uma metodologia para determinação dos requisitos desejáveis (necessários) – qualitativos e quantitativos – para os conjuntos de dados de entrada (CMS, SCADA ou outros) para realização do prognóstico de falha de componentes de aerogeradores, que considere a estimativa da VUR e incerteza.
- O desenvolvimento de modelos de inteligência artificial que consigam classificar e indicar a possibilidade de utilização de conjuntos de dados de entrada (CMS, SCADA ou outros) para realização das etapas de detecção, diagnóstico e prognóstico de falhas em aerogeradores

6.2. ARTIGOS PUBLICADOS

Apresenta-se abaixo a lista de artigos, publicados ou submetidos, em periódicos ao longo do desenvolvimento deste trabalho.

Periódicos

VIEIRA, J. L. de M.; FARIAS, F. C.; OCHOA, A. A. V.; DE MENEZES, F. D.; COSTA, A. C. A. da; DA COSTA, J. Â. P.; DE NOVAES PIRES LEITE, G.; VILELA, O. de C.; DE SOUZA, M. G. G.; MICHIMA, P. S. A. Remaining Useful Life Estimation Framework for the Main Bearing of Wind Turbines Operating in Real Time. **Energies**, v. 17, n. 6, p. 1430, 2024. Disponível em: https://doi.org/10.3390/en17061430

Submetidos (periódicos)

VIEIRA, J. L. de M.; LEITE, G. D. N. P.; OCHOA, A. A. V.; COSTA, A. C. A. da; DA COSTA, J. Â. P.; VILELA, O. de C. Wind turbines prognostic review: success, perspectives, and obstacles. **Renewable and Sustainable Energy Reviews**. 2024.

Submetidos (eventos)

NETO, SAMUEL, B. D G.; LEITE, G. D. N. P.; VIEIRA, J. L. de M.; OCHOA, A. A. V.; DROGUETT, E. L.; COSTA, A. C.A D.; SOUZA, M. G. G. D. Main bearing degradation prediction in wind turbines via quantum support vector machine regression. **LVI Simpósio Brasileiro de Pesquisa Operacional.** Fortaleza. 2024,

REFERÊNCIAS BIBLIOGRÁFICAS

1856-2017 - IEEE Standard Framework for Prognostics and Health Management of Electronic Systems. *[S. l.]*: IEEE, 2017.

ABID, A.; KHAN, M. T.; IQBAL, J. A review on fault detection and diagnosis techniques: basics and beyond. **Artificial Intelligence Review**, v. 54, n. 5, p. 3639–3664, 2021. Disponível em: https://doi.org/10.1007/s10462-020-09934-2

ABID, K.; SAYED-MOUCHAWEH, M.; LAURENCE, C. Adaptive Machine Learning Approach for Fault Prognostics based on Normal Conditions - Application to Shaft Bearings of Wind Turbine. **Annual Conference of the PHM Society**, Hammamet, Tunisia, v. 11, n. 1, p. 46–50, 2019. Disponível em: https://doi.org/10.36001/phmconf.2019.v11i1.838

ADEDIPE, T.; SHAFIEE, M.; ZIO, E. Bayesian Network Modelling for the Wind Energy Industry: An Overview. **Reliability Engineering and System Safety**, v. 202, n. May, p. 107053, 2020. Disponível em: https://doi.org/10.1016/j.ress.2020.107053

AFRIDI, Y. S.; HASAN, L.; ULLAH, R.; AHMAD, Z.; KIM, J. M. LSTM-Based Condition Monitoring and Fault Prognostics of Rolling Element Bearings Using Raw Vibrational Data. **Machines**, v. 11, n. 5, 2023. Disponível em: https://doi.org/10.3390/machines11050531

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Brasil termina 2021 com maior acréscimo em potência instalada desde 2016**. *[s. l.]*, 2021. Disponível em: https://www.gov.br/aneel/pt-br/assuntos/noticias/2022/brasil-termina-2021-com-maior-acrescimo-em-potencia-instalada-desde-2016. Acesso em: 7 fev. 2022.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Matriz elétrica teve ampliação de 1,5 GW em abril, com 47 novas usinas**. *[s. l.]*, 2024. Disponível em: https://www.gov.br/aneel/pt-br/assuntos/noticias/2024/matriz-eletrica-teve-ampliacao-de-1-5-gw-em-abril-com-47-novas-usinas. Acesso em: 16 maio. 2024.

AMELANG, SÖREN; WEHRMANN, B. **German onshore wind power – output, business and perspectives**. *[s. l.]*, 2020. Disponível em: https://www.cleanenergywire.org/factsheets/german-onshore-wind-power-output-business-and-perspectives. Acesso em: 9 dez. 2021.

BADIHI, H.; ZHANG, Y.; JIANG, B.; PILLAY, P.; RAKHEJA, S. A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis. **Proceedings of the IEEE**, v. 110, n. 6, p. 754–806, 2022. Disponível em: https://doi.org/10.1109/JPROC.2022.3171691

BALTAZAR, S.; LI, C.; DANIEL, H.; VALENTE DE OLIVEIRA, J. A Review on Neurocomputing Based wind Turbines Fault Diagnosis and Prognosis. **Proceedings - 2018 Prognostics and System Health Management Conference, PHM-Chongqing 2018**, p. 437–443, 2019. Disponível em: https://doi.org/10.1109/PHM-Chongqing.2018.00081

BAPTISTA, M. L.; GOEBEL, K.; HENRIQUES, E. M. P. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. **Artificial Intelligence**, v. 306, 2022. Disponível em: https://doi.org/10.1016/j.artint.2022.103667

BARBER, S.; IZAGIRRE, U.; SERRADILLA, O.; OLAIZOLA, J.; ZUGASTI, E.; AIZPURUA, J. I.; MILANI, A. E.; SEHNKE, F.; SAKAGAMI, Y.; HENDERSON, C. Best Practice Data Sharing Guidelines for Wind Turbine

Fault Detection Model Evaluation. **Energies**, v. 16, n. 8, 2023. Disponível em: https://doi.org/10.3390/en16083567

BENBOUZID, M.; BERGHOUT, T.; SARMA, N.; DJUROVIĆ, S.; WU, Y.; MA, X. Intelligent Condition Monitoring of Wind Power Systems: State of the Art Review. **Energies**, v. 14, n. 18, p. 5967, 2021. Disponível em: https://doi.org/10.3390/en14185967

BERGHOUT, T.; BENBOUZID, M.; AMIRAT, Y. Interaction between Machine Learning and Maintenance Planning for Wind Energy Systems: A Theoretical Review and Analysis. *In*: 2023, **2023 IEEE 2nd Industrial Electronics Society Annual On-Line Conference (ONCON)**. : IEEE, 2023. p. 1–6.Disponível em: https://doi.org/10.1109/ONCON60463.2023.10431363

BINSBERGEN, D. van; SOARES, M. N.; PEDERSEN, E.; NEJAD, A. R. A physics-, SCADA-based remaining useful life calculation approach for wind turbine drivetrains. **Journal of Physics: Conference Series**, v. 2265, n. 3, p. 032079, 2022. Disponível em: https://doi.org/10.1088/1742-6596/2265/3/032079

BOUGACHA, O.; VARNIER, C.; ZERHOUNI, N. Review of Post-Prognostics Decision-Making in Prognostics and Health Management. **International Journal of Prognostics and Health Management**, v. 11, n. 2, 2021. Disponível em: https://doi.org/10.36001/ijphm.2020.v11i2.2928

BOYD, D.; KARIMIRAD, M.; SIVAKUMAR, V.; JALILVAND, S.; DESMOND, C. A REVIEW OF END-OF-LIFE DECISION MAKING FOR OFFSHORE WIND TURBINES. *In*: 2022, Hamburg. **Proceedings of the ASME 2022 41st Internationa**. Hamburg: ASME, 2022. p. 1–10.

BRIJDER, R.; HAGEN, C. H. M.; CORTÉS, A.; IRIZAR, A.; THIBBOTUWA, U. C.; HELSEN, S.; VÁSQUEZ, S.; OMPUSUNGGU, A. P. Review of corrosion monitoring and prognostics in offshore wind turbine structures: Current status and feasible approaches. **Frontiers in Energy Research**, v. 10, 2022. Disponível em: https://doi.org/10.3389/fenrg.2022.991343

BURKOV, A. **The Hundred-Page Machine Learning Book**. *[S. l.]*: Andriy Burkov, 2019.

CAMELO, H. do N.; CARVALHO, P. C. M.; LEAL JUNIOR, J. B. V.; ACCIOLY FILHO, J. B. P. Análise estatística da velocidade de vento do estado do Ceará. **Revista Tecnologia**, v. 29, n. 2, p. 211–223, 2008. Disponível em: https://ojs.unifor.br/tec/article/view/38. Acesso em: 3 jun. 2024.

CAO, L.; QIAN, Z.; PEI, Y. Remaining Useful Life Prediction of Wind Turbine Generator Bearing Based on EMD with an Indicator. *In*: 2019, **Proceedings - 2018 Prognostics and System Health Management Conference, PHM-Chongqing 2018**. : Institute of Electrical and Electronics Engineers Inc., 2019. p. 375–379.Disponível em: https://doi.org/10.1109/PHM-Chongqing.2018.00070

CAO, L.; QIAN, Z.; ZAREIPOUR, H.; WOOD, D.; MOLLASALEHI, E.; TIAN, S.; PEI, Y. Prediction of remaining useful life of wind turbine bearings under non-stationary operating conditions. **Energies**, v. 11, n. 12, 2018. Disponível em: https://doi.org/10.3390/en11123318. Acesso em: 21 jun. 2023.

CARROLL, J.; KOUKOURA, S.; MCDONALD, A.; CHARALAMBOUS, A.; WEISS, S.; MCARTHUR, S. Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. **Wind Energy**, v. 22, n. 3, p. 360–375, 2019. Disponível em: https://doi.org/10.1002/we.2290

CHEN, N.; YU, R.; CHEN, Y.; XIE, H. Hierarchical method for wind turbine prognosis using SCADA data. **IET Renewable Power Generation**, v. 11, n. 4, p. 403–410, 2017. Disponível em: https://doi.org/10.1049/iet-rpg.2016.0247

CHENG, F.; QU, L.; QIAO, W. Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis. **IEEE Transactions on Sustainable Energy**, v. 9, n. 1, p. 157–167, 2018. Disponível em: https://doi.org/10.1109/TSTE.2017.2719626

CHENG, F.; QU, L.; QIAO, W.; HAO, L. Enhanced Particle Filtering for Bearing Remaining Useful Life Prediction of Wind Turbine Drivetrain Gearboxes. **IEEE Transactions on Industrial Electronics**, v. 66, n. 6, p. 4738–4748, 2019. Disponível em: https://doi.org/10.1109/TIE.2018.2866057

COMMISSION, E. Final Report and Action Plan from the European Commission Expert Group on FAIR Data TURNING FAIR INTO REALITY 2018 Research and Innovation. *[s. d.]*. Disponível em: https://doi.org/10.2777/54599

CORREA-JULLIAN, C.; COFRE-MARTEL, S.; MARTIN, G. S.; DROGUETT, E. L.; LEITE, G. de N. P.; COSTA, A. Exploring Quantum Machine Learning and Feature Reduction Techniques for Wind Turbine Pitch Fault Detection. **Energies**, v. 15, n. 8, 2022. Disponível em: https://doi.org/10.3390/en15082792

DAMESHGHI, A.; REFAN, M. H. Combination of condition monitoring and prognosis systems based on current measurement and PSO-LS-SVM method for wind turbine DFIGs with rotor electrical asymmetry. **Energy Systems**, n. 0123456789, 2019. Disponível em: https://doi.org/10.1007/s12667-019-00357-9

DANTAS BARROS, J.; LUIZA SILVA FURTADO, M.; SANTOS MARINHO, G.; MOREIRA DA SILVA, F. Sazonalidade do vento na cidade de Natal/RN pela distribuição de Weibull. **Sociedade e Território** , v. 25, n. 2º, p. 78–92, 2013.

DING, F.; TIAN, Z. Integrated Prognosis for Wind Turbine Gearbox Condition-Based Maintenance Considering Time-Varying Load and Crack Initiation Time Uncertainty. **International Journal of Reliability, Quality and Safety Engineering**, v. 28, n. 4, 2021. Disponível em: https://doi.org/10.1142/S0218539321500248

DING, F.; TIAN, Z.; ZHAO, F.; XU, H. An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions. **Renewable Energy**, v. 129, p. 260–270, 2018. Disponível em: https://doi.org/10.1016/j.renene.2018.05.074

DJEZIRI, M. A.; BENMOUSSA, S.; SANCHEZ, R. Hybrid method for remaining useful life prediction in wind turbine systems. **Renewable Energy**, v. 116, p. 173–187, 2018. Disponível em: https://doi.org/10.1016/j.renene.2017.05.020

DO, M. H.; SÖFFKER, D. State-of-the-art in integrated prognostics and health management control for utility-scale wind turbines. **Renewable and Sustainable Energy Reviews**, v. 145, 2021. Disponível em: https://doi.org/10.1016/J.RSER.2021.111102. Acesso em: 30 out. 2022.

DU, X.; JIA, W.; YU, P.; SHI, Y.; GONG, B. RUL prediction based on GAM–CNN for rotating machinery. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 45, n. 3, 2023. Disponível em: https://doi.org/10.1007/s40430-023-04062-8

DURAKOVIC, A. **World's Largest, Most Powerful Wind Turbine Stands Complete**. *[s. l.]*, 2021. Disponível em: https://www.offshorewind.biz/2021/11/12/worlds-largest-most-powerful-wind-turbine-stands-complete/. Acesso em: 19 dez. 2021.

ELASHA, F.; SHANBR, S.; LI, X.; MBA, D. Prognosis of a wind turbine gearbox bearing using supervised machine learning. **Sensors (Switzerland)**, v. 19, n. 14, p. 1–17, 2019. Disponível em: https://doi.org/10.3390/s19143092

FOX, H.; PILLAI, A. C.; FRIEDRICH, D.; COLLU, M.; DAWOOD, T.; JOHANNING, L. A Review of Predictive and Prescriptive Offshore Wind Farm Operation and Maintenance. 2022. Disponível em: https://doi.org/10.3390/en15020504

GALICKI, M.; LEISTRITZ, L.; ZWICK, E. B.; WITTE, H. Improving Generalization Capabilities of Dynamic Neural Networks. **Neural Computation**, v. 16, n. 6, p. 1253–1282, 2004. Disponível em: https://doi.org/10.1162/089976604773717603

GALVÃO, T. F.; PANSANI, T. de S. A.; HARRAD, D. Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. **Epidemiologia e Serviços de Saúde**, v. 24, n. 2, p. 335–342, 2015. Disponível em: https://doi.org/10.5123/S1679-49742015000200017

GAO, Z.; LIU, X. An Overview on Fault Diagnosis, Prognosis and Resilient Control for Wind Turbine Systems. **Processes**, v. 9, n. 2, p. 300, 2021. Disponível em: https://doi.org/10.3390/pr9020300

GARCÍA MÁRQUEZ, F. P.; PEINADO GONZALO, A. A Comprehensive Review of Artificial Intelligence and Wind Energy. **Archives of Computational Methods in Engineering**, v. 29, n. 5, p. 2935–2958, 2022. Disponível em: https://doi.org/10.1007/s11831-021-09678-4

GEEKS FOR GEEKS. **Check for NaN in Pandas DataFrame**. *[s. l.]*, 2023. Disponível em: https://www.geeksforgeeks.org/check-for-nan-in-pandas-dataframe/. Acesso em: 1 mar. 2024.

GENERAL ELETRIC. **An industry first Haliade-X offshore wind turbine**. *[s. l.]*, 2021. Disponível em: https://www.ge.com/renewableenergy/wind-energy/offshore-wind/haliade-x-offshore-turbine.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2nd edition ed. Sebastopol: O'Reilly Media, Inc., 2019.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. *[S. l.]*: MIT Press, 2016. *E-book*. Disponível em: http://www.deeplearningbook.org. Acesso em: 21 ago. 2024.

GORDON, C. A. K.; BURNAK, B.; ONEL, M.; PISTIKOPOULOS, E. N. Data-Driven Prescriptive Maintenance: Failure Prediction Using Ensemble Support Vector Classification for Optimal Process and Maintenance Scheduling. **Industrial & Engineering Chemistry Research**, v. 59, n. 44, p. 19607–19622, 2020. Disponível em: https://doi.org/10.1021/acs.iecr.0c03241

GUO, J.; LI, Z.; LI, M. A Review on Prognostics Methods for Engineering Systems. **IEEE Transactions on Reliability**, v. 69, n. 3, p. 1110–1129, 2020 a. Disponível em: https://doi.org/10.1109/TR.2019.2957965

GUO, J.; LI, Z.; LI, M. A Review on Prognostics Methods for Engineering Systems. **IEEE Transactions on Reliability**, v. 69, n. 3, p. 1110–1129, 2020 b. Disponível em: https://doi.org/10.1109/TR.2019.2957965

HAGMEYER, S.; MAUTHE, F.; ZEILER, P. Creation of Publicly Available Data Sets for Prognostics and Diagnostics Addressing Data Scenarios Relevant to Industrial Applications. **International Journal of Prognostics and Health Management**, v. 12, n. 2, 2021. Disponível em: https://doi.org/10.36001/ijphm.2021.v12i2.3087

HARRISON, M. **Machine Learning – Guia de Referência Rápida**. São Paulo: Novatec Editora Ltda, 2020.

HART, E.; CLARKE, B.; NICHOLAS, G.; KAZEMI AMIRI, A.; STIRLING, J.; CARROLL, J.; DWYER-JOYCE, R.; MCDONALD, A.; LONG, H. A review of wind turbine main bearings: Design, operation, modelling,

damage mechanisms and fault detection. **Wind Energy Science**, v. 5, n. 1, p. 105–124, 2020. Disponível em: https://doi.org/10.5194/wes-5-105-2020

HAYDER, A. N.; SAIDI, L. Applications of Artificial Neural Networks With Input and output Degradation data for Renewable Energy Systems Fault Prognosis. *In*: 2021, **2021 12th International Renewable Energy Congress, IREC 2021**. : Institute of Electrical and Electronics Engineers Inc., 2021.Disponível em: https://doi.org/10.1109/IREC52758.2021.9624894

HELSEN, J. Review of Research on Condition Monitoring for Improved O&M of Offshore Wind Turbine Drivetrains. **Acoustics Australia**, v. 49, n. 2, p. 251–258, 2021. Disponível em: https://doi.org/10.1007/s40857-021-00237-2

HENKEL, M.; WEIJTJENS, W.; DEVRIENDT, C. Validation of virtual sensing on subsoil strain data of an offshore wind turbine. *In*: 2019, **8th IOMAC - International Operational Modal Analysis Conference, Proceedings**. *[S. l.: s. n.]* p. 765–774.

HERP, J.; PEDERSEN, N. L.; NADIMI, E. S. Assessment of early stopping through statistical health prognostic models for empirical rul estimation in wind turbine main bearing failure monitoring. **Energies**, v. 13, n. 1, p. 1–18, 2019. Disponível em: https://doi.org/10.3390/en13010083

HONG, W.; WANG, S.; TOMOVIC, M. M.; LIU, H.; SHI, J.; WANG, X. A novel indicator for mechanical failure and life prediction based on debris monitoring. **IEEE Transactions on Reliability**, v. 66, n. 1, p. 161–169, 2017. Disponível em: https://doi.org/10.1109/TR.2016.2628412

HU, C.; YOUN, B. D.; WANG, P. **Engineering Design under Uncertainty and Health Prognostics**. 1. ed. Cham: Springer International Publishing AG, 2018. (Springer Series in Reliability Engineering). *E-book*. Disponível em: https://doi.org/10.1007/978-3-319-92574-5. Acesso em: 17 out. 2024.

HU, Y.; LI, H.; SHI, P.; CHAI, Z.; WANG, K.; XIE, X.; CHEN, Z. A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process. **Renewable Energy**, v. 127, p. 452–460, 2018. Disponível em: https://doi.org/10.1016/j.renene.2018.04.033

HUANG, Y.; TANG, Y.; VANZWIETEN, J. Prognostics with Variational Auto-encoder by Generative Adversarial Learning Adversarial Learning. **IEEE Transactions on Industrial Electronics**, v. 69, n. PREPRINT, p. 856–867, 2022 a. Disponível em: https://doi.org/10.36227/techrxiv.12751172.v1

HUANG, Y.; TANG, Y.; VANZWIETEN, J. Prognostics With Variational Autoencoder by Generative Adversarial Learning. **IEEE Transactions on Industrial Electronics**, v. 69, n. 1, p. 856–867, 2022 b. Disponível em: https://doi.org/10.1109/TIE.2021.3053882

HUI, D.; MO, L.; PAIVA, C. E. The Importance of Prognostication: Impact of Prognostic Predictions, Disclosures, Awareness, and Acceptance on Patient Outcomes. **Current Treatment Options in Oncology**, v. 22, n. 2, p. 12, 2021. Disponível em: https://doi.org/10.1007/s11864-020-00810-3

HUTCHINSON, M.; ZHAO, F. **GLOBAL WIND REPORT 2023**. Brussels: *[s. n.]*, 2023.

JELLALI, A.; MAATALLAH, H.; OUNI, K. Predicting Remaining Useful Life of Wind Turbine Bearing using Linear Regression. *In*: 2022, **2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)**. : IEEE, 2022. p. 357–362.Disponível em: https://doi.org/10.1109/IC_ASET53395.2022.9765939

KAMARZARRIN, M.; REFAN, M. H.; AMIRI, P.; DAMESHGHI, A. A new intelligent fault diagnosis and prognosis method for wind turbine doubly-fed induction generator. **Wind Engineering**, v. 46, n. 1, p. 308–340, 2022. Disponível em: https://doi.org/10.1177/0309524X211027808

KERBOUA, A.; KELAIAIA, R. RECURRENT NEURAL NETWORK OPTIMIZATION FOR WIND TURBINE CONDITION PROGNOSIS. **Diagnostyka**, v. 23, n. 3, 2022. Disponível em: https://doi.org/10.29354/diag/151608

KHAN, P. W.; BYUN, Y.-C. A Review of machine learning techniques for wind turbine's fault detection, diagnosis, and prognosis. **International Journal of Green Energy**, v. 21, n. 4, p. 771–786, 2024. Disponível em: https://doi.org/10.1080/15435075.2023.2217901

KIM, S.; CHOI, J. H.; KIM, N. H. **Challenges and opportunities of system-level prognostics**. *[S. l.: s. n.]* Disponível em: https://doi.org/10.3390/s21227655

KIM, S.; CHOI, J.-H.; KIM, N. H. Challenges and Opportunities of System-Level Prognostics. **Sensors**, v. 21, n. 22, p. 7655, 2021 b. Disponível em: https://doi.org/10.3390/s21227655

KONG, Y.; HAN, Q.; CHU, F.; QIN, Y.; DONG, M. Spectral ensemble sparse representation classification approach for super-robust health diagnostics of wind turbine planetary gearbox. **Renewable Energy**, v. 219, 2023. Disponível em: https://doi.org/10.1016/j.renene.2023.119373

KÖNIG, F.; MARHEINEKE, J.; JACOBS, G.; SOUS, C.; ZUO, M. J.; TIAN, Z. Data-driven wear monitoring for sliding bearings using acoustic emission signals and long short-term memory neural networks. **Wear**, v. 476, 2021. Disponível em: https://doi.org/10.1016/j.wear.2021.203616

KORDESTANI, M.; SAIF, M.; ORCHARD, M. E.; RAZAVI-FAR, R.; KHORASANI, K. Failure Prognosis and Applications—A Survey of Recent Literature. **IEEE Transactions on Reliability**, v. 70, n. 2, p. 728–748, 2021. Disponível em: https://doi.org/10.1109/TR.2019.2930195

KOUKOURA, S. Failure and Remaining Useful Life Prediction of Wind Turbine Gearboxes. **Annual Conference of the PHM Society**, v. 10, n. 1, 2018. Disponível em: https://doi.org/10.36001/phmconf.2018.v10i1.712

KOUKOURA, S.; CARROLL, J.; WEISS, S.; MCDONALD, A. Wind turbine gearbox vibration signal signature and fault development through time. **25th European Signal Processing Conference, EUSIPCO 2017**, v. 2017- Janua, p. 1380–1384, 2017. Disponível em: https://doi.org/10.23919/EUSIPCO.2017.8081435

KRAMTI, S. E.; BEN ALI, J.; SAIDI, L.; SAYADI, M.; BECHHOEFER, E. Direct Wind Turbine Drivetrain Prognosis Approach Using Elman Neural Network. *In*: 2018, **2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)**. : IEEE, 2018. p. 859–864.Disponível em: https://doi.org/10.1109/CoDIT.2018.8394926

KRAMTI, S. E.; BEN ALI, J.; SAIDI, L.; SAYADI, M.; BOUCHOUICHA, M.; BECHHOEFER, E. A neural network approach for improved bearing prognostics of wind turbine generators. **EPJ Applied Physics**, v. 93, n. 2, 2021. Disponível em: https://doi.org/10.1051/epjap/2021200259

KÜHNE, F.; HERMANN, M.; PREISLER, M.; ROHRMOSER, A.; LETSCH, A.; GOERLING, U. Prognostic Awareness in Advanced Disease: A Review Update and Concept Analysis. **Frontiers in Psychology**, v. 12, 2021. Disponível em: https://doi.org/10.3389/fpsyg.2021.629050

LAN, X.; LIANG, D.; LIU, W.; WANG, Y.; HAN, Y.; YU, Z. Remaining Useful Life Prediction of Wind Turbine Generator Bearings Based on Nonlinear Wiener Process. *In*: 2023, **Proceedings - 2023 3rd Power System and Green Energy Conference, PSGEC 2023**. : Institute of Electrical and Electronics Engineers Inc., 2023. p. 1124–1129.Disponível em: https://doi.org/10.1109/PSGEC58411.2023.10255862

LEAHY, K.; GALLAGHER, C.; BRUTON, K.; O'DONOVAN, P.; O'SULLIVAN, D. T. J. Automatically Identifying and Predicting Unplanned Wind Turbine Stoppages Using SCADA and Alarms System Data: Case Study and Results. **Journal of Physics: Conference Series**, v. 926, n. 1, p. 012011, 2017. Disponível em: https://doi.org/10.1088/1742-6596/926/1/012011

LEAHY, K.; GALLAGHER, C.; O'DONOVAN, P.; BRUTON, K.; O'SULLIVAN, D. T. J. A robust prescriptive framework and performance metric for diagnosing and predicting wind turbine faults based on SCADA and alarms data with case study. **Energies**, v. 11, n. 7, 2018. Disponível em: https://doi.org/10.3390/en11071738

LEE, J.; QIU, H.; YU, G.; LIN, J.; REXNORD TECHNICAL SERVICES. **Bearing Data Set**. Moffett Field: NASA Ames Research Center, 2007. Disponível em: https://phm-datasets.s3.amazonaws.com/NASA/4.+Bearings.zip. Acesso em: 1 abr. 2024.

LEE, J.; ZHAO, F. **GLOBAL WIND REPORT 2024**. Brussels: *[s. n.]*, 2024. Disponível em: www.gwec.net. Acesso em: 23 maio. 2024.

LEE, S.; KANG, S.; LEE, G.-S. Predictions for Bending Strain at the Tower Bottom of Offshore Wind Turbine Based on the LSTM Model. **Energies**, v. 16, n. 13, p. 4922, 2023. Disponível em: https://doi.org/10.3390/en16134922

LEITE, G. de N. P.; ARAÚJO, A. M.; ROSAS, P. A. C. Prognostic techniques applied to maintenance of wind turbines: a concise and specific review. **Renewable and Sustainable Energy Reviews**, v. 81, n. May 2016, p. 1917–1925, 2018. Disponível em: https://doi.org/10.1016/j.rser.2017.06.002

LENG, J.; GARDONI, P.; WANG, M.; LI, Z.; KRÓLCZYK, G.; FENG, S.; INCECIK, A.; LI, W. Condition-based structural health monitoring of offshore wind jacket structures: Opportunities, challenges, and perspectives. **Structural Health Monitoring**, v. 22, n. 5, p. 3558–3575, 2023. Disponível em: https://doi.org/10.1177/14759217221148688

LI, B.; TANG, B.; DENG, L.; ZHAO, M. Self-Attention ConvLSTM and Its Application in RUL Prediction of Rolling Bearings. **IEEE Transactions on Instrumentation and Measurement**, v. 70, 2021. Disponível em: https://doi.org/10.1109/TIM.2021.3086906

LI, C.; CERRADA, M.; CABRERA, D.; SANCHEZ, R. V.; PACHECO, F.; ULUTAGAY, G.; OLIVEIRA, J. V. De. Some preliminary results on the comparison of fcm, gk, fcmfp and fn-dbscan for bearing fault diagnosis. *In*: 2017, **Proceedings - 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC 2017**. : Institute of Electrical and Electronics Engineers Inc., 2017. p. 41–46.Disponível em: https://doi.org/10.1109/SDPC.2017.18

LI, N.; XU, P.; LEI, Y.; CAI, X.; KONG, D. A self-data-driven method for remaining useful life prediction of wind turbines considering continuously varying speeds. **Mechanical Systems and Signal Processing**, v. 165, p. 108315, 2022. Disponível em: https://doi.org/10.1016/j.ymssp.2021.108315. Acesso em: 31 ago. 2022.

LIU, Z.; ZHANG, L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. **Measurement: Journal of the International Measurement Confederation**, v. 149, p. 107002, 2020. Disponível em: https://doi.org/10.1016/j.measurement.2019.107002

LYDIA, M.; EDWIN PREM KUMAR, G. Condition monitoring in wind turbines. *In*: **Non-Destructive Testing and Condition Monitoring Techniques in Wind Energy**. *[S. l.]*: Elsevier, 2023. p. 229–247. Disponível em: https://doi.org/10.1016/B978-0-323-99666-2.15003-8

MACKENZIE, A. The production of prediction: What does machine learning want? **European Journal of Cultural Studies**, v. 18, n. 4–5, p. 429–445, 2015. Disponível em: https://doi.org/10.1177/1367549415577384

MAIOR, C. B. S.; ARAUJO, L. M. M.; LINS, I. D.; MOURA, M. D. C.; DROGUETT LOPEZ, E. L. Prognostics and Health Management of Rotating Machinery via Quantum Machine Learning. **IEEE Access**, v. 11, p. 25132–25151, 2023. Disponível em: https://doi.org/10.1109/ACCESS.2023.3255417

MAMMADOV, E.; FARROKHABADI, M.; CANIZARES, C. A. AI-enabled Predictive Maintenance of Wind Generators. *In*: 2021, **2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)**. : IEEE, 2021. p. 1–5.Disponível em: https://doi.org/10.1109/ISGTEurope52324.2021.9640162

MATHWORKS INC.; BECHHOEFER, E. **Wind Turbine High-Speed Bearing Prognosis**. *[S. l.: s. n.]* Disponível em: https://github.com/mathworks/WindTurbineHighSpeedBearingPrognosis-Data. Acesso em: 19 jun. 2024.

MAZIDI, P.; DU, M.; BERTLING TJERNBERG, L.; SANZ BOBI, M. A. A health condition model for wind turbine monitoring through neural networks and proportional hazard models. **Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability**, v. 231, n. 5, p. 481–494, 2017. Disponível em: https://doi.org/10.1177/1748006X17707902

MEHLAN, F. C.; KELLER, J.; NEJAD, A. R. Virtual sensing of wind turbine hub loads and drivetrain fatigue damage. **Forschung im Ingenieurwesen**, v. 87, n. 1, p. 207–218, 2023. Disponível em: https://doi.org/10.1007/s10010-023-00627-0

MELLO, G.; FERREIRA DIAS, M.; ROBAINA, M. Evaluation of the environmental impacts related to the wind farms end-of-life. **Energy Reports**, v. 8, p. 35–40, 2022. Disponível em: https://doi.org/10.1016/j.egyr.2022.01.024

MOHER, D.; LIBERATI, A.; TETZLAFF, J.; ALTMAN, D. G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. **International Journal of Surgery**, v. 8, n. 5, p. 336–341, 2010. Disponível em: https://doi.org/10.1016/j.ijsu.2010.02.007

MOMBER, A. W.; MÖLLER, T.; LANGENKÄMPER, D.; NATTKEMPER, T. W.; BRÜN, D. A Digital Twin concept for the prescriptive maintenance of protective coating systems on wind turbine structures. **Wind Engineering**, v. 46, n. 3, p. 949–971, 2022. Disponível em: https://doi.org/10.1177/0309524X211060550

MORONEY, P. D.; VERMA, A. S. Durability and Damage Tolerance Analysis Approaches for Wind Turbine Blade Trailing Edge Life Prediction: A Technical Review. **Energies**, v. 16, n. 24, p. 7934, 2023. Disponível em: https://doi.org/10.3390/en16247934

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**. *[S. l.]*: The MIT Press, 2012.

NEJAD, A. *et al.* Wind turbine drivetrains: state-of-the-art technologies and future development trends. **Wind Energy Science Discussions**, v. 7, n. 1, p. 1–35, 2021. Disponível em: https://doi.org/10.5194/wes-2021-63

PANDAS. **Pandas User guide**. *[s. l.]*, 2024. Disponível em: https://pandas.pydata.org/pandas-docs/stable/user_guide/missing_data.html. Acesso em: 1 mar. 2024.

PENG, H.; ZHANG, H.; FAN, Y.; SHANGGUAN, L.; YANG, Y. A Review of Research on Wind Turbine Bearings' Failure Analysis and Fault Diagnosis. **Lubricants**, v. 11, n. 1, p. 14, 2022. Disponível em: https://doi.org/10.3390/lubricants11010014

PENG, Y.; CHENG, F.; QIAO, W.; QU, L. Fault prognosis of drivetrain gearbox based on a recurrent neural network. **IEEE International Conference on Electro Information Technology**, p. 593–599, 2017. Disponível em: https://doi.org/10.1109/EIT.2017.8053433

PEREIRA, E. B.; MARTINS, F. R.; GONÇALVES, A. R.; COSTA, R. S.; LIMA, F. J. L.; RÜTHER, R.; ABREU, S. L.; TIEPOLO, G. M.; PEREIRA, S. V.; SOUZA, J. G. **Atlas brasileiro de energia solar (Vol. 1).** 2ª ed. São José dos Campos: *[s. n.]*, 2017.

PERES, R. S.; JIA, X.; LEE, J. A. Y.; SUN, K.; COLOMBO, A. W.; BARATA, J. Industrial Artificial Intelligence in Industry 4.0 -Systematic Review, Challenges and Outlook. **IEEE Access**, n. December, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2020.3042874

PERKTOLD, J.; SEABOLD, S.; TAYLOR, J. **Statsmodels Documentation**. *[s. l.]*, 2023. Disponível em: https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html#statsmodels.tsa.seasonal.seasonal_decompose. Acesso em: 17 jan. 2024.

PETER, R.; ZAPPALÁ, D.; SCHAMBOECK, V.; WATSON, S. J. Wind turbine generator prognostics using field SCADA data. *In*: 2022, **Journal of Physics: Conference Series**. : Institute of Physics, 2022.Disponível em: https://doi.org/10.1088/1742-6596/2265/3/032111

PRAKASH, S.; VUCKOVIC, K.; AMIN, S. Prognostic Model Evaluation Metrics. *In*: 2023, **2023 IEEE Aerospace Conference**. : IEEE, 2023. p. 1–11.Disponível em: https://doi.org/10.1109/AERO55745.2023.10115952

QIAO, W.; LU, D. A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis - Part II: Signals and Signal Processing Methods. **IEEE Transactions on Industrial Electronics**, v. 62, n. 10, p. 6546–6557, 2015 a. Disponível em: https://doi.org/10.1109/TIE.2015.2422394

QIAO, W.; LU, D. A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part I: Components and Subsystems. **IEEE Transactions on Industrial Electronics**, v. 62, n. 10, p. 6536–6545, 2015 b. Disponível em: https://doi.org/10.1109/TIE.2015.2422112

REZAMAND, M.; CARRIVEAU, R.; TING, D. S. K.; DAVISON, M.; DAVIS, J. J. Aggregate reliability analysis of wind turbine generators. **IET Renewable Power Generation**, v. 13, n. 11, p. 1902–1910, 2019. Disponível em: https://doi.org/10.1049/iet-rpg.2018.5909

REZAMAND, M.; KORDESTANI, M.; CARRIVEAU, R.; TING, D. S.-K.; ORCHARD, M. E.; SAIF, M. Critical Wind Turbine Components Prognostics: A Comprehensive Review. **IEEE Transactions on Instrumentation and Measurement**, v. 69, n. 12, p. 9306–9328, 2020. Disponível em: https://doi.org/10.1109/TIM.2020.3030165

RICHARD, C. **Vestas launches new 15MW offshore wind turbine with 236-metre rotor**. *[s. l.]*, 2021. Disponível em: https://www.windpowermonthly.com/article/1706915/vestas-launches-new-15mw-offshore-wind-turbine-236-metre-rotor. Acesso em: 1 jan. 2022.

RINALDI, G.; THIES, P. R.; JOHANNING, L. Current Status and Future Trends in the Operation and Maintenance of Offshore Wind Turbines: A Review. **Energies**, v. 14, n. 9, p. 2484, 2021. Disponível em: https://doi.org/10.3390/en14092484

SAHU, A.; JAMBHALE, R.; ADIGA, D. T.; POWAR, N.; MCKINLEY, T. Formulation of Model Stability Metrics for Remaining Useful Life Models of Engine Components. *In*: 2023, **2023 IEEE Aerospace Conference**. : IEEE, 2023. p. 1–11.Disponível em: https://doi.org/10.1109/AERO55745.2023.10115610

SAIDI, L.; ALI, B.; BECHHOEFER, E.; BENBOUZID, M. **Particle filter-based prognostic approach for High-Speed Shaft Bearing Wind Turbine Progressive Degradations**. *[S. l.: s. n.]*.

SAIDI, L.; ALI, B.; BENBOUZID, M.; BECHHOEFER, E. **Wind Turbine Drivetrain Prognosis Approach Based on Kalman Smoother with Confidence Bounds**. *[S. l.: s. n.]*.

SAIDI, L.; BEN ALI, J.; BENBOUZID, M.; BECHHOFER, E. An integrated wind turbine failures prognostic approach implementing Kalman smoother with confidence bounds. **Applied Acoustics**, v. 138, p. 199–208, 2018. Disponível em: https://doi.org/10.1016/j.apacoust.2018.04.005

SAIDI, L.; BENBOUZID, M. Renewable Energy Systems Prognostics and Health Management: A Review of Recent Advances. *In*: 2021, **IECON Proceedings (Industrial Electronics Conference)**. : IEEE Computer Society, 2021.Disponível em: https://doi.org/10.1109/IECON48115.2021.9589308

SAIDI, L.; HAYDER, A. N.; SAIDI, M. Remaining useful life prognosis for wind turbine using a neural network with a long-term prediction. **Wind Engineering**, p. 0309524X2210851, 2022. Disponível em: https://doi.org/10.1177/0309524X221085133

SANTIAGO, R. A. da F.; BARBOSA, N. B.; MERGULHÃO, H. G.; CARVALHO, T. F. de; SANTOS, A. A. B.; MEDRADO, R. C.; FILHO, J. B. de M.; PINHEIRO, O. R.; NASCIMENTO, E. G. S. Data-Driven Models Applied to Predictive and Prescriptive Maintenance of Wind Turbine: A Systematic Review of Approaches Based on Failure Detection, Diagnosis, and Prognosis. **Energies**, v. 17, n. 5, p. 1010, 2024. Disponível em: https://doi.org/10.3390/en17051010

SANTOS, F. D. N.; NOPPE, N.; WEIJTJENS, W.; DEVRIENDT, C. Data-driven farm-wide fatigue estimation on jacket-foundation OWTs for multiple SHM setups. **Wind Energy Science**, v. 7, n. 1, p. 299–321, 2022. Disponível em: https://doi.org/10.5194/wes-7-299-2022

SAXENA, A.; CELAYA, J.; BALABAN, E.; GOEBEL, K.; SAHA, B.; SAHA, S.; SCHWABACHER, M. Metrics for evaluating performance of prognostic techniques. *In*: 2008, **2008 International Conference on Prognostics and Health Management**. : IEEE, 2008. p. 1–17.Disponível em: https://doi.org/10.1109/PHM.2008.4711436

SCIKIT-LEARN DEVELOPERS. **scikit learn - User Manual**. *[s. l.]*, 2024. Disponível em: https://scikit-learn.org/stable/supervised_learning.html. Acesso em: 25 jun. 2024.

SCIMAGO LAB. **Scimago Journal & Country Rank**. *[s. l.]*, 2024.

SIKORSKA, J. Z.; HODKIEWICZ, M.; MA, L. Prognostic modelling options for remaining useful life estimation by industry. **Mechanical Systems and Signal Processing**, v. 25, n. 5, p. 1803–1836, 2011. Disponível em: https://doi.org/10.1016/j.ymssp.2010.11.018

STETCO, A.; DINMOHAMMADI, F.; ZHAO, X.; ROBU, V.; FLYNN, D.; BARNES, M.; KEANE, J.; NENADIC, G. Machine learning methods for wind turbine condition monitoring: A review. **Renewable Energy**, v. 133, p. 620–635, 2019. Disponível em: https://doi.org/10.1016/j.renene.2018.10.047. Acesso em: 30 jun. 2021.

STONE, P.; WHITE, N.; OOSTENDORP, L. J. M.; LLEWELLYN, H.; VICKERSTAFF, V. Comparing the performance of the palliative prognostic (PaP) score with clinical predictions of survival: A systematic review. **European Journal of Cancer**, v. 158, p. 27–35, 2021. Disponível em: https://doi.org/10.1016/j.ejca.2021.08.049

STRACK, B.; FRANK, J.; STICH, V.; LENART, M.; PFAU, F. Sociotechnical Implementation of Prescriptive Maintenance for Onshore Wind Turbines. *In*: 2022, **Proceedings of the 55th Hawaii International**

Conference on System Sciences. *[S. l.: s. n.]* p. 1288–1297.Disponível em: https://hdl.handle.net/10125/79490

STRIELKOWSKI, W.; CIVÍN, L.; TARKHANOVA, E.; TVARONAVIČIENĖ, M.; PETRENKO, Y. **Renewable energy in the sustainable development of electrical power sector: A review**. *[S. l.: s. n.]* Disponível em: https://doi.org/10.3390/en14248240

SUN, Z.; SUN, H. Health Status Assessment for Wind Turbine with Recurrent Neural Networks. **Mathematical Problems in Engineering**, v. 2018, 2018. Disponível em: https://doi.org/10.1155/2018/6972481

TAO, F.; ZHANG, M.; LIU, Y.; NEE, A. Y. C. Digital twin driven prognostics and health management for complex equipment. **CIRP Annals**, v. 67, n. 1, p. 169–172, 2018. Disponível em: https://doi.org/10.1016/j.cirp.2018.04.055

TENG, W.; DING, X.; TANG, S.; XU, J.; SHI, B.; LIU, Y. **Vibration analysis for fault detection of wind turbine drivetrains—a comprehensive investigation**. *[S. l.: s. n.]* Disponível em: https://doi.org/10.3390/s21051686. Acesso em: 24 nov. 2021.

TUTIV'EN, C.; BENALCAZAR–PARRA, C.; ESCUELA, A. E.; VIDAL, Y.; PURUNCAIAS, B.; FAJARDO, M. Wind Turbine Main Bearing Condition Monitoring via Convolutional Autoencoder Neural Networks. *In*: 2021, **2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)**. : IEEE, 2021. p. 01–06.Disponível em: https://doi.org/10.1109/ICECCME52200.2021.9590937

UMA MAHESWARI, R.; UMAMAHESWARI, R. Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train - A contemporary survey. **Mechanical Systems and Signal Processing**, v. 85, p. 296–311, 2017. Disponível em: https://doi.org/10.1016/j.ymssp.2016.07.046

VAPNIK, V. N. An overview of statistical learning theory. **IEEE Transactions on Neural Networks**, v. 10, n. 5, p. 988–999, 1999. Disponível em: https://doi.org/10.1109/72.788640

VOGL, G. W.; WEISS, B. A.; DONMEZ, M. A. **Standards Related to Prognostics and Health Management (PHM) for Manufacturing**. Gaithersburg, MD: *[s. n.]*, 2014. Disponível em: https://doi.org/10.6028/NIST.IR.8012.

WATANABE, A. T. Y.; SEBEM, R.; LEAL, A. B.; HOUNSELL, M. da S. Fault prognosis of discrete event systems: An overview. **Annual Reviews in Control**, v. 51, p. 100–110, 2021. Disponível em: https://doi.org/10.1016/j.arcontrol.2021.03.005

WIESE, B.; PEDERSEN, N. L.; NADIMI, E. S.; HERP, J. Estimating the remaining power generation of wind turbines-an exploratory study for main bearing failures. **Energies**, v. 13, n. 13, 2020. Disponível em: https://doi.org/10.3390/en13133406

WILLIAM DIAB, W.; FERRARO, A.; KLENZ, B.; LIN, S.-W.; LIONGOSARI, E.; ELIE TANNOUS, W.; ZARKOUT, B. **Industrial IoT Artificial Intelligence Framework**. Boston: *[s. n.]*, 2022.

WOO, S. M.; WHALE, J. A mini-review of end-of-life management of wind turbines: Current practices and closing the circular economy gap. **Waste Management & Research: The Journal for a Sustainable Circular Economy**, v. 40, n. 12, p. 1730–1744, 2022. Disponível em: https://doi.org/10.1177/0734242X221105434

WORLD WIND ENERGY ASSOCIATION. **Worldwide Wind Capacity Reaches 744 Gigawatts – An Unprecedented 93 Gigawatts added in 2020**. *[s. l.]*, 2021. Disponível em: https://wwindea.org/worldwide-wind-capacity-reaches-744-gigawatts/. Acesso em: 13 dez. 2021.

WORLD WIND ENERGY ASSOCIATION. **WWEA Annual Report 2023**. Bonn: *[s. n.]*, 2024. Disponível em: www.wwindea.org.

XIANG, S.; QIN, Y.; LUO, J.; WU, F.; GRYLLIAS, K. A concise self-adapting deep learning network for machine remaining useful life prediction. **Mechanical Systems and Signal Processing**, v. 191, p. 110187, 2023. Disponível em: https://doi.org/10.1016/j.ymssp.2023.110187

XU, J.; JIANG, X.; LIAO, S.; KE, D.; SUN, Y.; YAO, L.; MAO, B. Probabilistic Prognosis of Wind Turbine Faults with Feature Selection and Confidence Calibration. **IEEE Transactions on Sustainable Energy**, v. 15, n. 1, p. 52–67, 2024. Disponível em: https://doi.org/10.1109/TSTE.2023.3272317

YANG, L.; ZHANG, Z. Wind Turbine Gearbox Failure Detection Based on SCADA Data: A Deep Learning Based Approach. **IEEE Transactions on Instrumentation and Measurement**, p. 1–1, 2020. Disponível em: https://doi.org/10.1109/tim.2020.3045800

YETER, B.; GARBATOV, Y. Structural integrity assessment of fixed support structures for offshore wind turbines: A review. **Ocean Engineering**, v. 244, p. 110271, 2022. Disponível em: https://doi.org/10.1016/j.oceaneng.2021.110271

YUCESAN, Y. A.; DOURADO, A.; VIANA, F. A. C. A survey of modeling for prognosis and health management of industrial equipment. **Advanced Engineering Informatics**, v. 50, p. 101404, 2021. Disponível em: https://doi.org/10.1016/j.aei.2021.101404

YUCESAN, Y. A.; VIANA, F. A. C. **Wind Turbine Main Bearing Fatigue Life Estimation with Physics-informed Neural Networks**. Orlando, FL: *[s. n.]*, 2019.

YUCESAN, Y. A.; VIANA, F. A. C. A hybrid model for main bearing fatigue prognosis based on physics and machine learning. *In*: 2020, **AIAA Scitech 2020 Forum**. : American Institute of Aeronautics and Astronautics Inc, AIAA, 2020.Disponível em: https://doi.org/10.2514/6.2020-1412

ZHANG, S.; BASU, M.; ROBINSON, E.; FITZGERALD, B.; BASU, B. Fault Prediction and Classification for a Doubly-Fed Induction Generator based Wind Turbine by using Random Forest Classifier. *In*: 2021, **The 9th Renewable Power Generation Conference (RPG Dublin Online 2021)**. : Institution of Engineering and Technology, 2021. p. 149–154.Disponível em: https://doi.org/10.1049/icp.2021.1353. Acesso em: 7 set. 2023.

ZHANG, S.; LANG, Z. Q. SCADA-data-based wind turbine fault detection: A dynamic model sensor method. **Control Engineering Practice**, v. 102, 2020. Disponível em: https://doi.org/10.1016/j.conengprac.2020.104546

ZHANG, W.; VATN, J.; RASHEED, A. A review of failure prognostics for predictive maintenance of offshore wind turbines. *In*: 2022, **Journal of Physics: Conference Series**. : Institute of Physics, 2022.Disponível em: https://doi.org/10.1088/1742-6596/2362/1/012043

ZHAO, W.; ZHANG, C.; WANG, J.; PEYRANO, O. G.; GU, F.; WANG, S.; LV, D. Research on main bearing life prediction of direct-drive wind turbine based on digital twin technology. **Measurement Science and Technology**, v. 34, n. 2, 2023. Disponível em: https://doi.org/10.1088/1361-6501/ac99f4

ZHAO, Y.; LI, D.; DONG, A.; KANG, D.; LV, Q.; SHANG, L. Fault prediction and diagnosis of wind turbine generators using SCADA data. **Energies**, v. 10, n. 8, 2017. Disponível em: https://doi.org/10.3390/en10081210

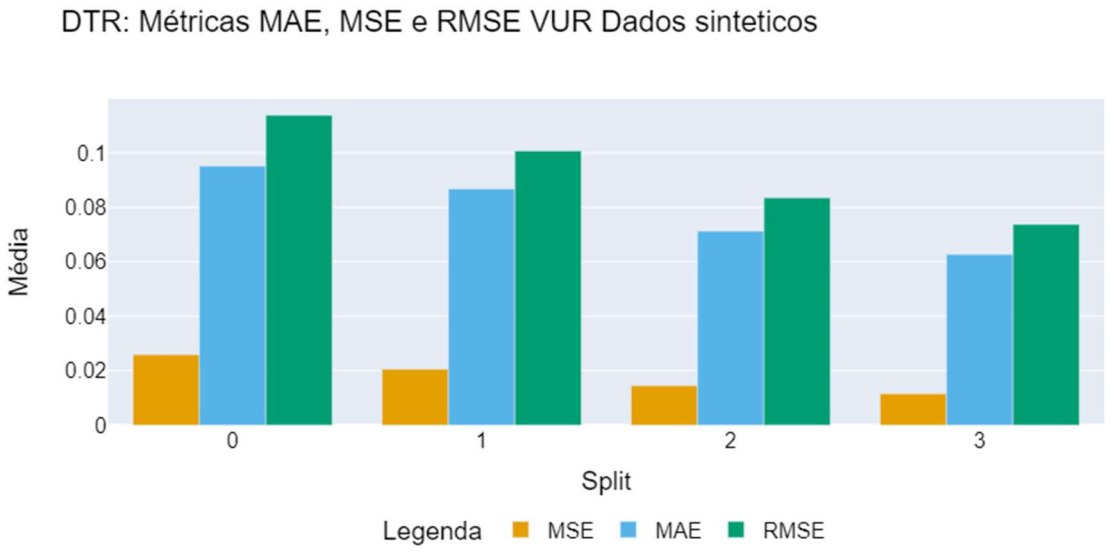
ZHOU, K.; LIU, Z.; QIAO, Y.; XIANG, T.; LOY, C. C. Domain Generalization: A Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 45, n. 4, p. 4396–4415, 2023. Disponível em: https://doi.org/10.1109/TPAMI.2022.3195549

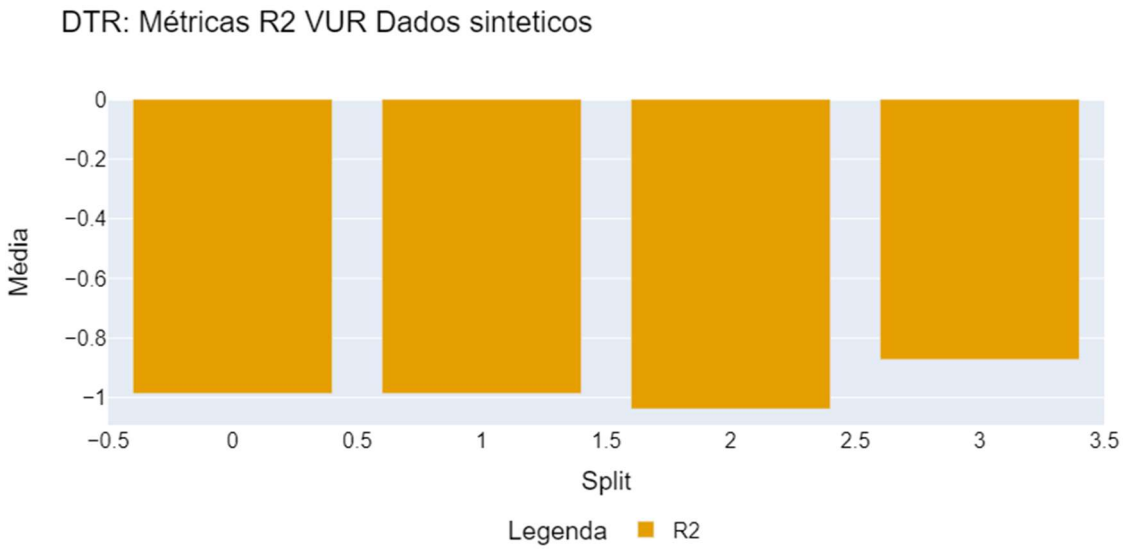
ZHU, J.; YOON, J. M.; HE, D.; BECHHOEFER, E. Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines. **Wind Energy**, v. 18, n. 6, p. 1131–1149, 2015. Disponível em: https://doi.org/10.1002/we.1746

ZIEGLER, L.; GONZALEZ, E.; RUBERT, T.; SMOLKA, U.; MELERO, J. J. Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK. **Renewable and Sustainable Energy Reviews**, v. 82, p. 1261–1271, 2018. Disponível em: https://doi.org/10.1016/j.rser.2017.09.100

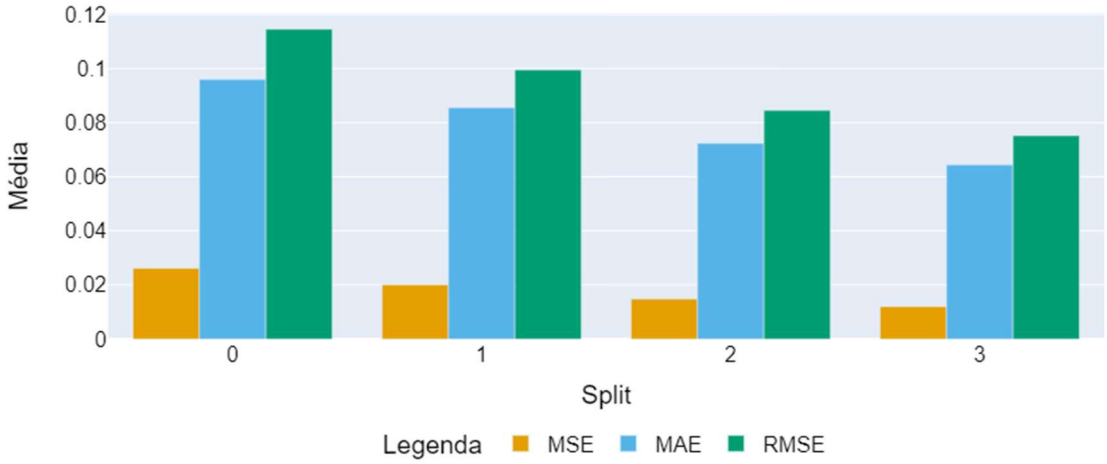
APÊNDICES

Métricas dos modelos para cada split – Dados sintéticos de temperatura

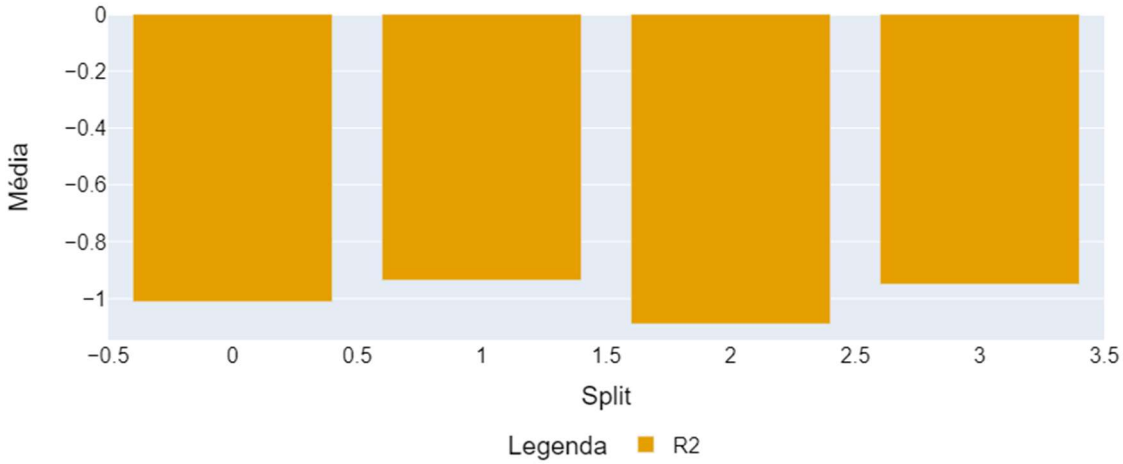




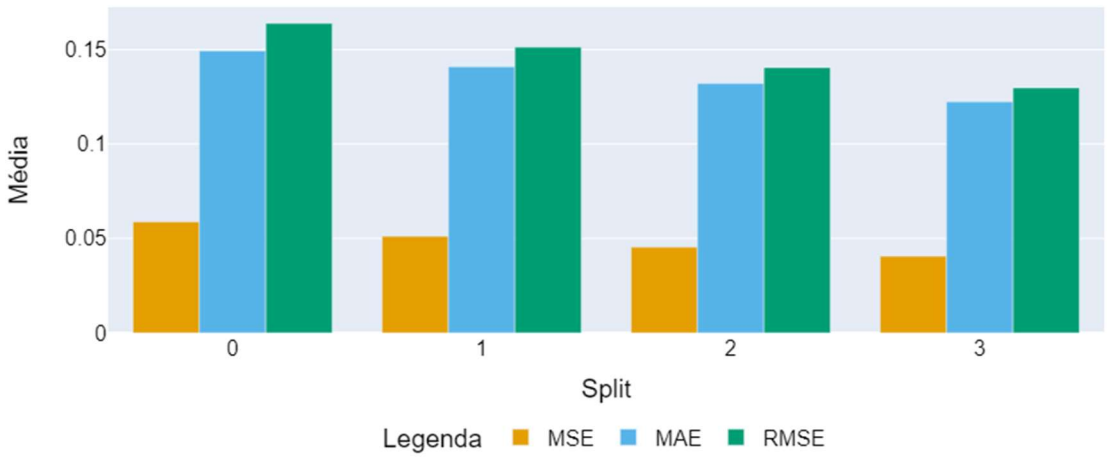
ETR: Métricas MAE, MSE e RMSE VUR Dados sinteticos



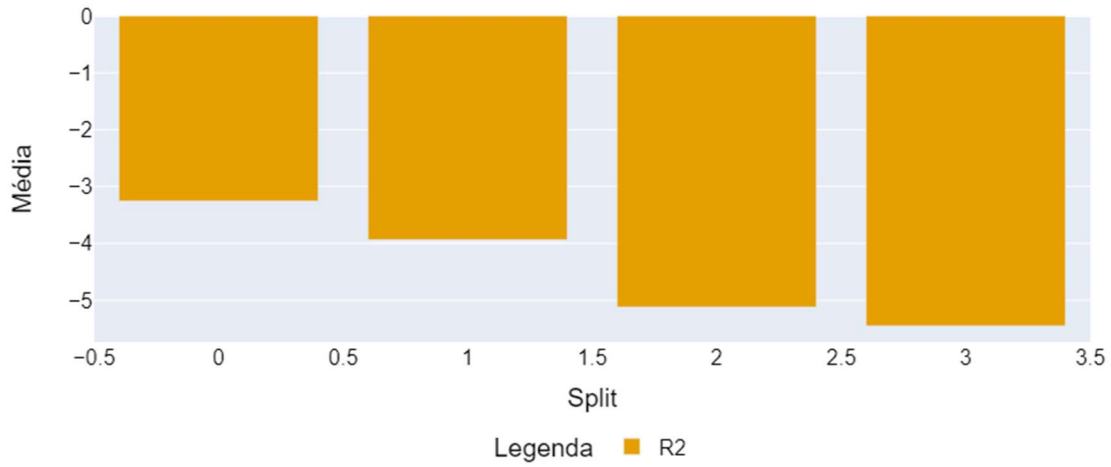
ETR: Métricas R2 VUR Dados sinteticos



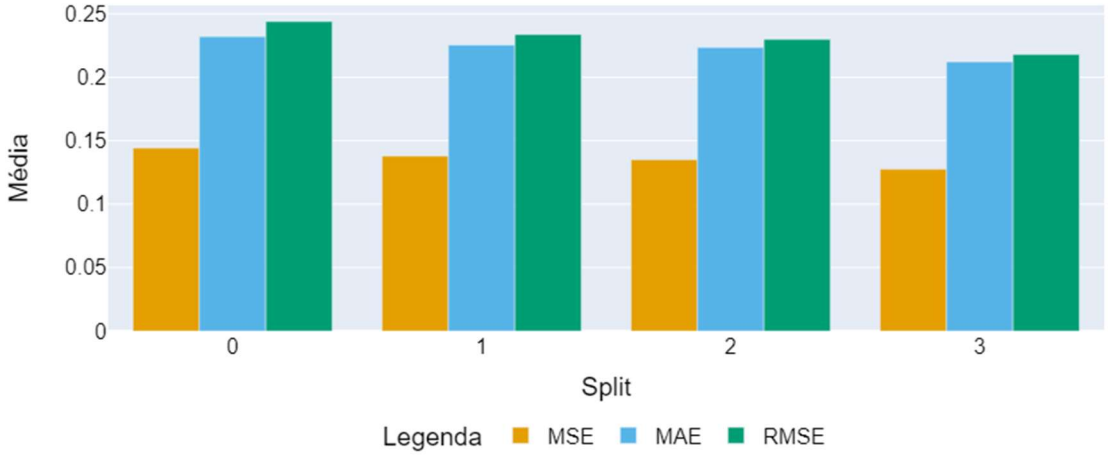
GBR: Métricas MAE, MSE e RMSE VUR Dados sinteticos



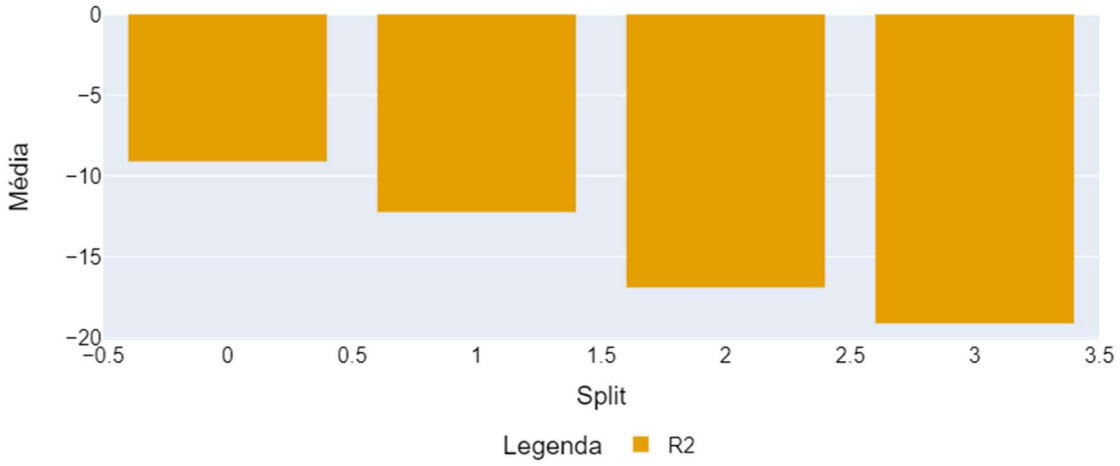
GBR: Métricas R2 VUR Dados sinteticos

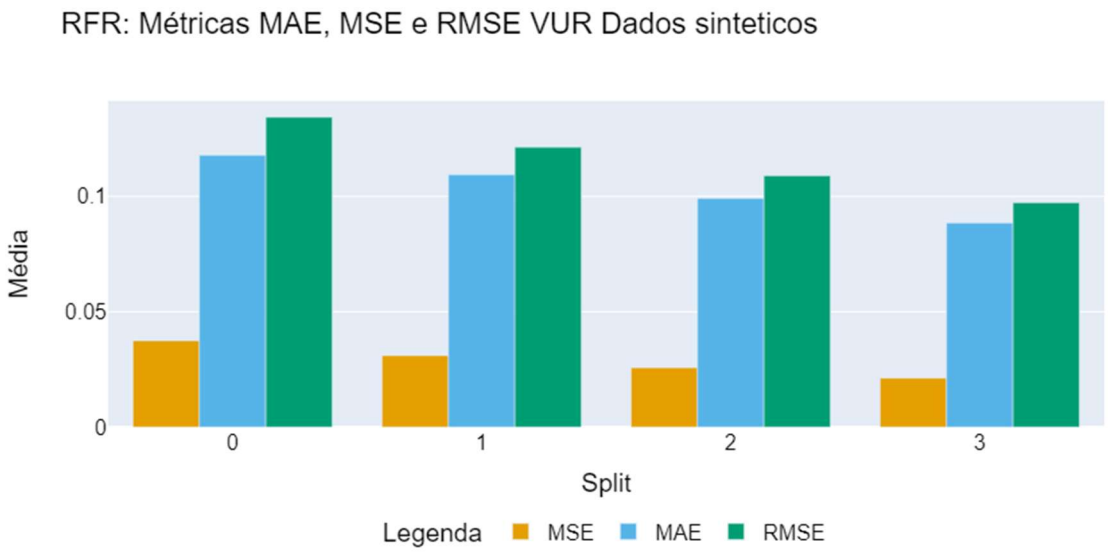


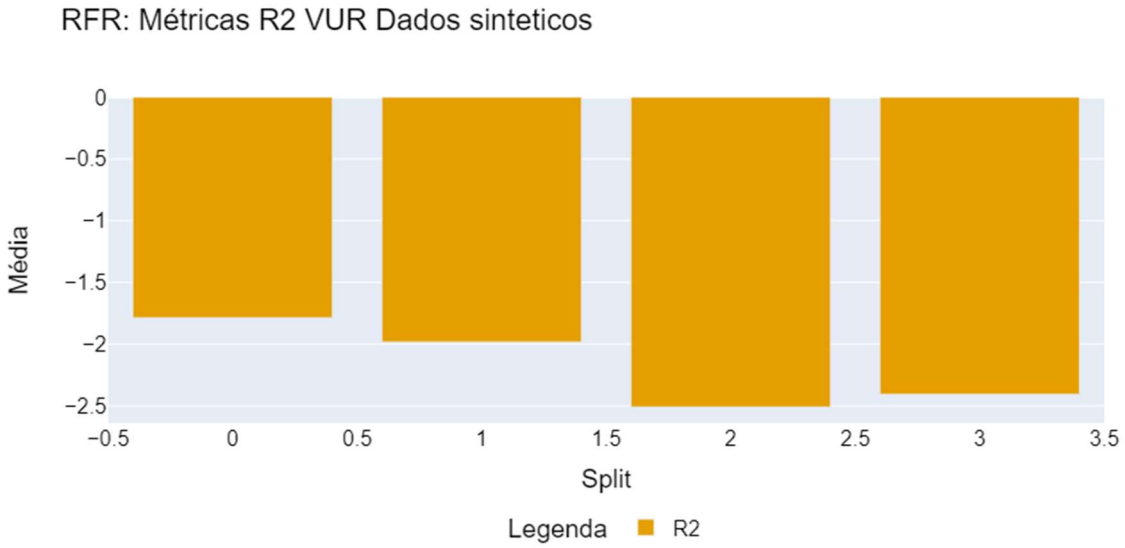
ISOR: Métricas MAE, MSE e RMSE VUR Dados sinteticos

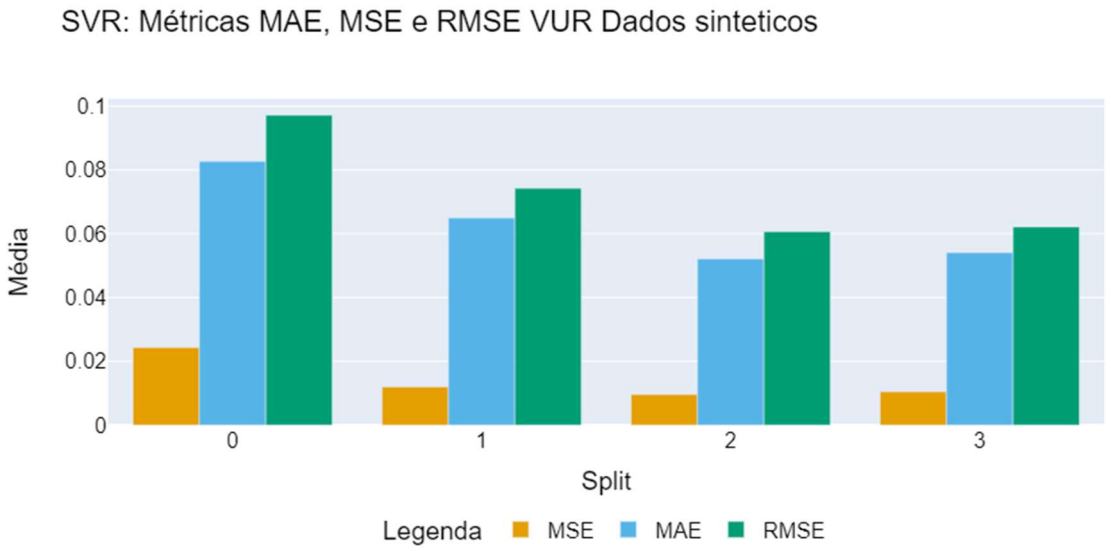


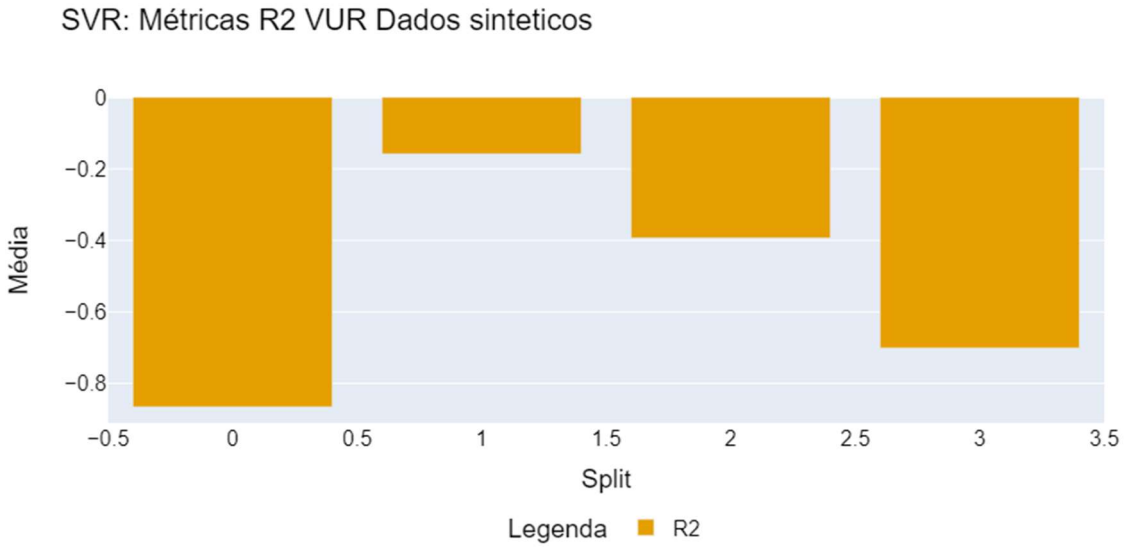
ISOR: Métricas R2 VUR Dados sinteticos



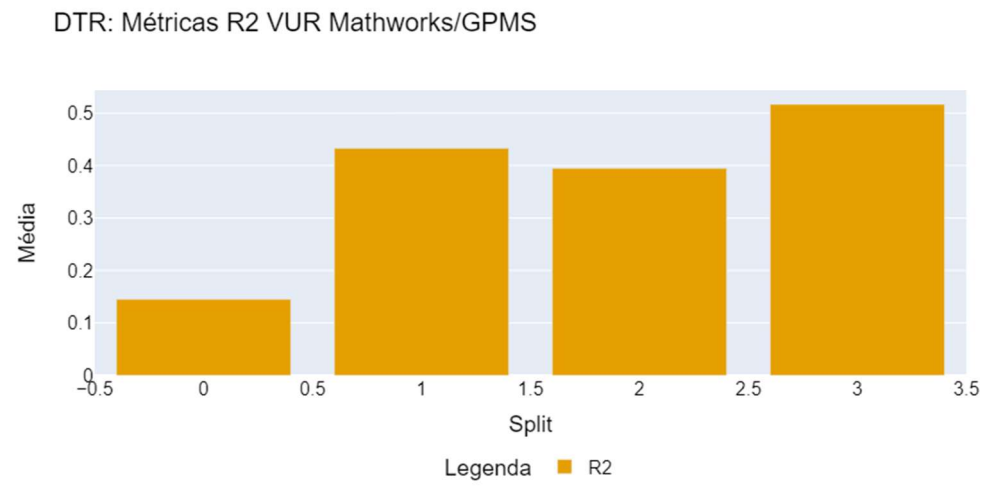

RFR: Métricas MAE, MSE e RMSE VUR Dados sinteticos
Média
0.1
0.05
0
0
1
2
3
Split
Legenda
MSE
MAE
RMSE

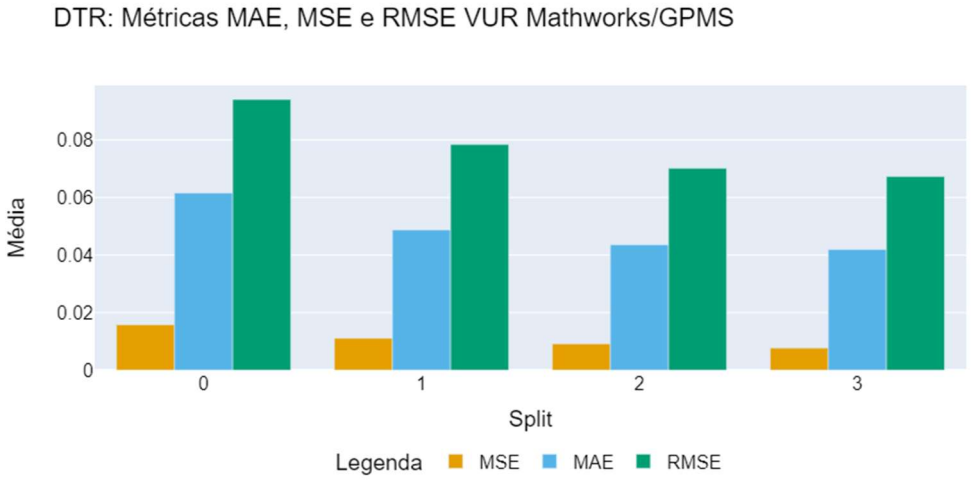

RFR: Métricas R2 VUR Dados sinteticos
Média
0
−0.5
−1
−1.5
−2
−2.5
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
Split
Legenda
R2


SVR: Métricas MAE, MSE e RMSE VUR Dados sinteticos
Média
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
Split
Legenda
MSE
MAE
RMSE

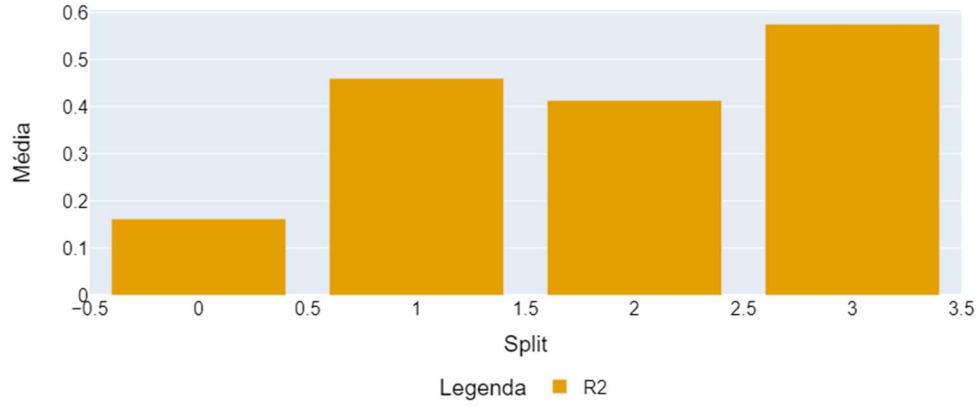

SVR: Métricas R2 VUR Dados sinteticos
Média
0
−0.2
−0.4
−0.6
−0.8
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
Split
Legenda
R2

Métricas dos modelos para cada split – Dados Mathworks/GPMS

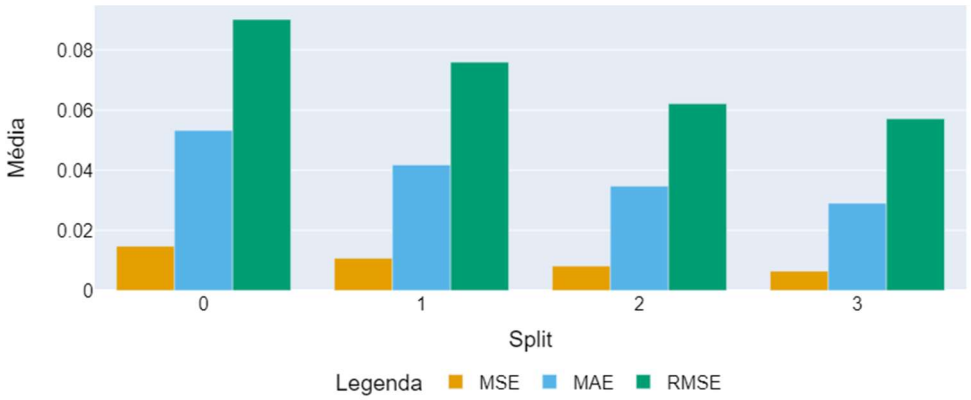




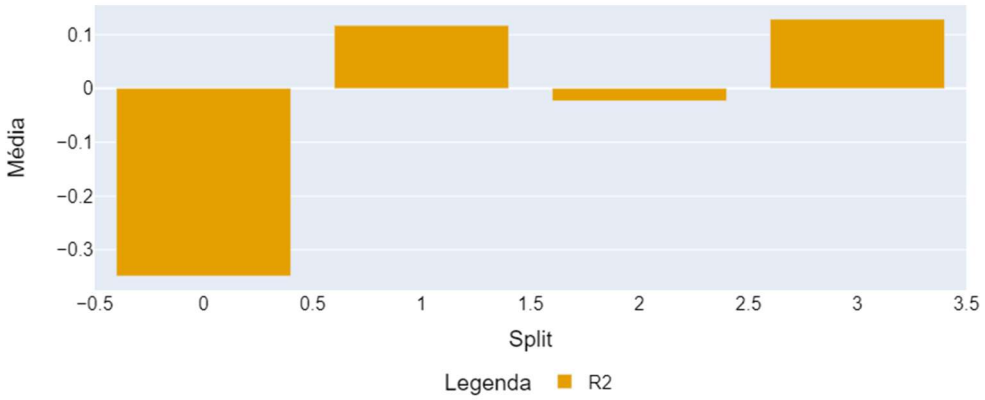
ETR: Métricas R2 VUR Mathworks/GPMS



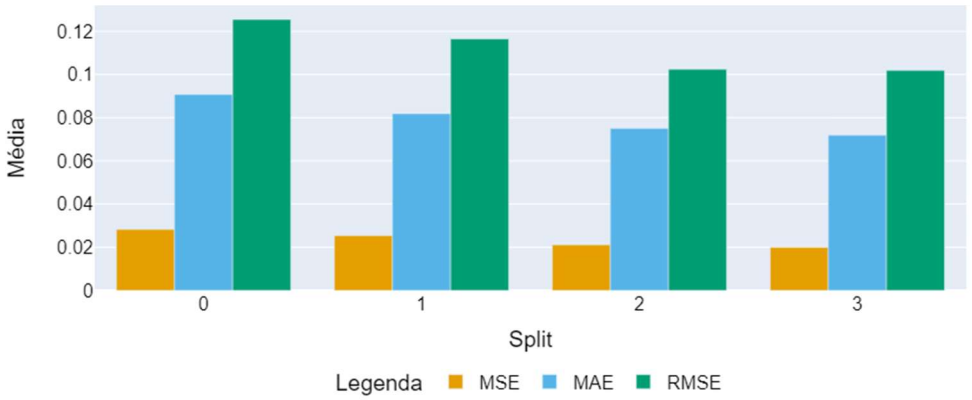
ETR: Métricas MAE, MSE e RMSE VUR Mathworks/GPMS

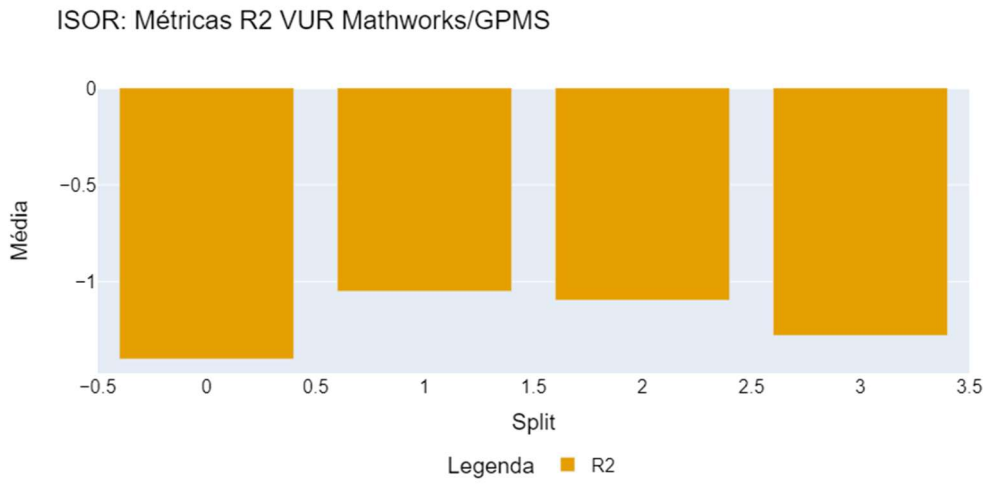


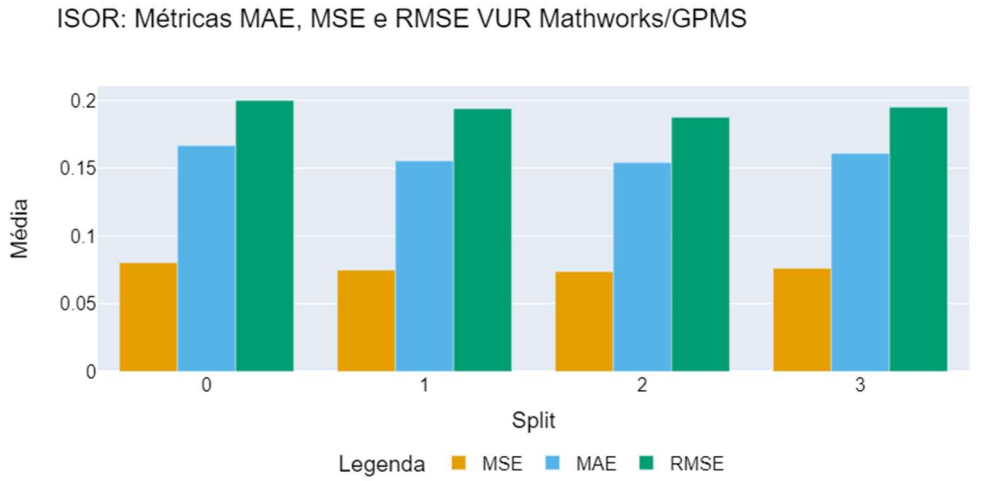
GBR: Métricas R2 VUR Mathworks/GPMS



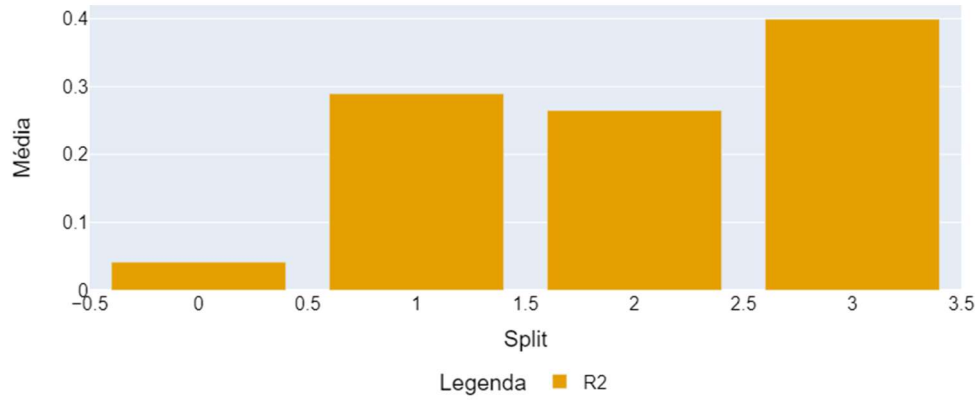
GBR: Métricas MAE, MSE e RMSE VUR Mathworks/GPMS



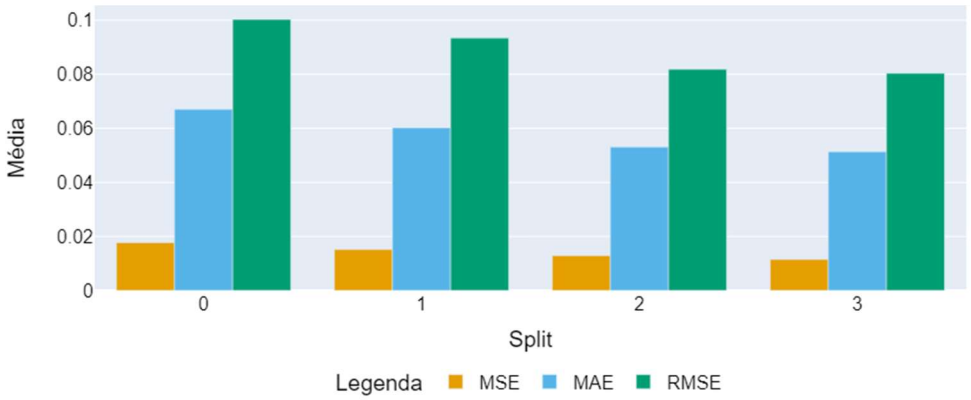

ISOR: Métricas R2 VUR Mathworks/GPMS
Média
0
−0.5
−1
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
Split
Legenda
R2

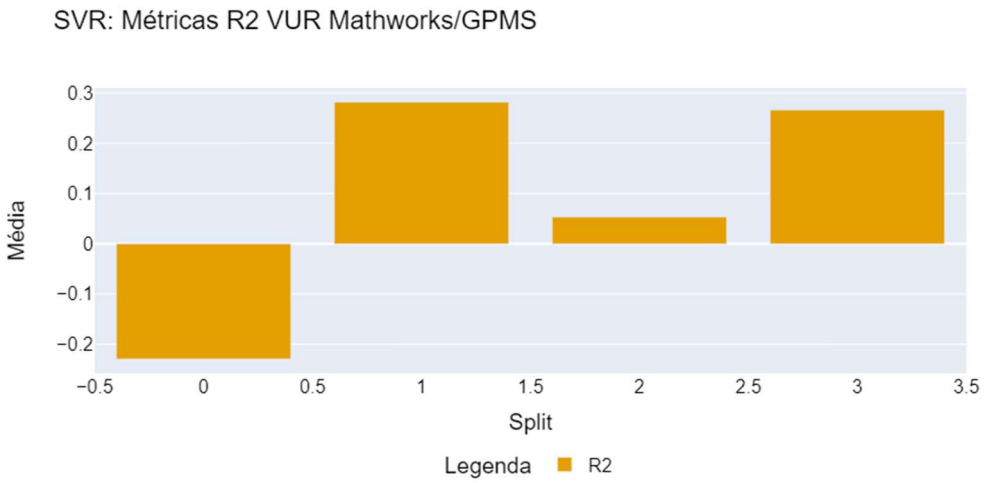

ISOR: Métricas MAE, MSE e RMSE VUR Mathworks/GPMS
Média
0.2
0.15
0.1
0.05
0
0
1
2
3
Split
Legenda
MSE
MAE
RMSE

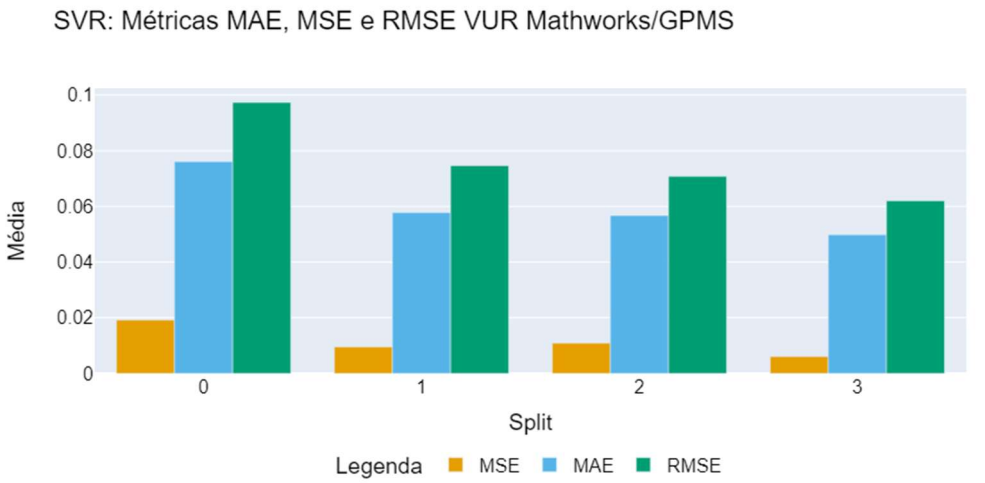
RFR: Métricas R2 VUR Mathworks/GPMS



RFR: Métricas MAE, MSE e RMSE VUR Mathworks/GPMS




SVR: Métricas R2 VUR Mathworks/GPMS
Média
0.3
0.2
0.1
0
−0.1
−0.2
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
Split
Legenda
R2


SVR: Métricas MAE, MSE e RMSE VUR Mathworks/GPMS
Média
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
Split
Legenda
MSE
MAE
RMSE