

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGEEL

MATEUS NICOLADELLI DE OLIVEIRA

**ESTUDO DE CASO SOBRE O USO DE ALGORITMO COM APRENDIZADO DE
MÁQUINA PARA DETECÇÃO DE DESVIO OPERACIONAL EM ATIVOS DE
GERAÇÃO ELÉTRICA E SOLUÇÕES IIOT**

JOINVILLE

2020

MATEUS NICOLADELLI DE OLIVEIRA

**ESTUDO DE CASO SOBRE O USO DE ALGORITMO COM APRENDIZADO DE
MÁQUINA PARA DETECÇÃO DE DESVIO OPERACIONAL EM ATIVOS DE
GERAÇÃO ELÉTRICA E SOLUÇÕES IIOT**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Ademir Nied

Coorientador: Rafael Stubs Parpinelli

JOINVILLE

2020

Oliveira, Mateus Nicoladelli de
Estudo de Caso sobre o uso de Algoritmo com
Aprendizado de Máquina para Detecção de Desvio
Operacional em Ativos de Geração Elétrica e Soluções
IIoT / Mateus Nicoladelli de Oliveira. - Joinville,
2020.

108 p.

Orientador: Ademir Nied.

Coorientador: Rafael Stubs Parpinelli.

Dissertação (Mestrado) - Universidade do Estado
de Santa Catarina, Centro de Ciências Tecnológicas,
Programa de Pós-Graduação em Engenharia Elétrica,
Joinville, 2020.

1. Redes neurais. 2. Aprendizado de máquina. 3.
Internet das coisas. 4. Detecção de desvio operacional.
I. Nied, Ademir . II. Parpinelli, Rafael Stubs . III.
Universidade do Estado de Santa Catarina, Centro de
Ciências Tecnológicas, Programa de Pós-Graduação em
Engenharia Elétrica. IV. Título.

MATEUS NICOLADELLI DE OLIVEIRA

**ESTUDO DE CASO SOBRE O USO DE ALGORITMO COM APRENDIZADO DE
MÁQUINA PARA DETECÇÃO DE DESVIO OPERACIONAL EM ATIVOS DE
GERAÇÃO ELÉTRICA E SOLUÇÕES IIOT**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Ademir Nied

Coorientador: Rafael Stubs Parpinelli

BANCA EXAMINADORA

Prof. Dr. Ademir Nied
CCT UDESC
(presidente)

Membros:

Dr. Walter Evaldo Kuchenbecker
WEG

Prof. Dr. Douglas Wildgrube Bertol
CCT UDESC

Joinville, 30 de novembro de 2020

Dedico este trabalho à minha família, minha esposa Maiara, meus pais Sebastião e Vânia e meus irmãos Henrique e Sebastião.

AGRADECIMENTOS

Agradeço a minha esposa Maiara pelo amor, carinho, suporte e por proporcionar um ambiente familiar muito agradável, tanto para o convívio quanto para o estudo.

Aos meus pais, Sebastião e Vânia, pelo amor incondicional, carinho e ensinamentos de vida. Aos meus irmãos, Henrique e Sebastião, pelo companheirismo e imensurável contribuição nos momentos de descontração e lazer.

Ao orientador professor Ademir Nied, pela sabedoria compartilhada nos momentos decisivos e pela valiosa orientação despendida nesta etapa tão importante da minha vida.

Ao coorientador professor Rafael Parpinelli, que aceitou participar do projeto e contribuiu com importantes direcionamentos.

A UDESC, professores e servidores que permitiram esta jornada.

Aos colegas de mestrado, Rodrigo Piontkewicz, Roberto Schmitt e Ederson Manarin.

A WEG e seus gestores, Elissa S. de Carvalho, Carlos Ogawa e João C. Tabalipa, que incentivaram a realização do mestrado e a realização deste estudo.

Ao Gregory Marchesan, colega de trabalho na WEG, pelos momentos de discussão e análise sobre os resultados deste estudo.

Não posso deixar de citar o João Vitor, Heron Pereira e Ana Carolina, também colegas de trabalho na WEG, que me ensinaram muito, tanto no aspecto pessoal como no profissional.

“All models are wrong, but some are useful.”
(George E.P. Box)

RESUMO

Conhecer o real status operacional dos ativos de uma frota é desafiador e altamente desejável, isso pode maximizar o tempo entre paradas inesperadas e melhorar o desempenho da frota. Atualmente, a maioria dos ativos de elevado valor agregado, como é o caso de aerogeradores e hidrogeradores, possuem dispositivos para retirar o equipamento de serviço em caso de problemas severos como temperatura elevada e vibração excessiva. Em geral, estes dispositivos são incapazes de identificar um desvio operacional incipiente, que no médio-longo prazo pode causar um desligamento repentino ou período de indisponibilidade. Com objetivo de propor uma solução conceitual com algoritmo de aprendizado de máquina e identificar os principais *players* neste segmento, uma breve pesquisa bibliografia e de patentes é apresentada. Neste estudo, é proposto um *workflow* para orientar o processo de criação de tal algoritmo, com foco na detecção de desvio operacional em uma aplicação industrial, a partir dos dados do sistema SCADA(*Supervisory Control and Data Acquisition*). Para avaliar o *workflow* proposto, é construído um modelo de classificação e outro de regressão, o primeiro aplica um *Random Forest* para identificar o padrão operacional de um hidrogerador. O segundo, uma rede neural artificial configurada como regressor, para emular a temperatura do estator de um aerogerador, o que é explorado como indicador de anormalidade. A rede utilizada foi a TLFN focada(*Focused Time Lagged Feedforward Network*). O *workflow* proposto neste estudo usa a metodologia CRISP-DM como referência, indicando um fluxo de etapas bem definido para organizar os dados e a elaboração do algoritmo. Adicionalmente, são mostradas as áreas de competência e interface entre elas, algo observado como relevante durante o desenvolvimento deste trabalho. Os resultados obtidos mostram que a abordagem proposta e aplicada pode ser uma boa ferramenta para detecção de desvio operacional em séries temporais multivariadas, o que é comum em ativos industriais.

Palavras-chave: Redes neurais. Aprendizado de máquina. Internet das coisas. Detecção de desvio operacional.

ABSTRACT

Understanding the real operational status of a fleet's assets is challenging and highly desirable, it can maximize the time between unexpected shutdowns and improve fleet performance. Currently, most high value-added assets, such as wind turbines and hydrogenerators, they are equipped with monitoring devices which can remove the asset from service in case of severe problems such as high temperature and excessive vibration. In general, these devices are unable to identify an incipient operational deviation, which in the medium to long term may cause a sudden shutdown or unavailability period. In order to proposing a conceptual solution with machine learning algorithm and to identify the main players in this segment, a brief bibliographic and patent research is presented. In this study, a workflow is proposed to guide the algorithm creation process, focusing on the operational deviation detection of an industrial application, based on SCADA system data. To evaluate the proposed workflow, a classification model and a regression model are built, the first applies a Random Forest to identify the operational pattern in data from a hydrogenerator. The second, an artificial neural network (with TLFN focused) configured as a regressor is applied to emulate the wind turbine stator temperature, which is used as abnormality indicator. The proposed workflow uses the CRISP-DM methodology as a reference, which indicates a well-defined step flow to organize the data and the algorithm creation. Additionally, the areas of competence and interface between them are shown at proposed workflow. It is a relevant point observed during the study development. The initial results shows that the proposed and applied approach can be a good tool for detecting operational deviation in multivariate time series, which is common in industrial assets.

Keywords: Neural networks. Machine learning. Internet of things. Detection of operational pattern deviation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Matriz energética brasileira.	21
Figura 2 – Geração de energia (MWmed) hidrelétricas em 2019.	22
Figura 3 – Geração de energia (MWmed) eólica em 2019.	22
Figura 4 – Hidrogerador em operação.	23
Figura 5 – Aplicação de turbinas hidráulicas até 30MW.	24
Figura 6 – Curva de rendimento para turbinas hidráulicas em função da potência.	24
Figura 7 – Turbinas Pelton (A), Francis (B) e Kaplan (C).	25
Figura 8 – Gerador de polos salientes com excitação brushless.	25
Figura 9 – Diagrama elétrico do gerador com excitação brushless.	26
Figura 10 – Curva de capacidade do gerador e limites de operação.	27
Figura 11 – Diagrama simplificado do sistema de automação em uma hidroelétrica.	28
Figura 12 – Sistema SCADA para planta de hidrogeração.	29
Figura 13 – Evolução da comunicação industrial e tecnologias relacionadas.	29
Figura 14 – Esquemático planta de hidrogeração (a) layout geral e (b) diagrama de blocos.	30
Figura 15 – Diagrama de blocos do sistema de excitação do gerador.	31
Figura 16 – Ponto de instalação do sistema de vedação do eixo.	32
Figura 17 – Detalhes do sistema de vedação do eixo com anéis de gaxeta	32
Figura 18 – Aplicação dos sensores de vazão.	34
Figura 19 – Parque de geração eólico com AGW 110/2.1.	35
Figura 20 – Visão interna da nacelle do aerogerador.	35
Figura 21 – Esquemático simplificado do aerogerador direct-drive.	36
Figura 22 – Regiões da curva de potência do aerogerador.	37
Figura 23 – Curva de potência, vento e velocidade do aerogerador.	38
Figura 24 – Modelo térmico aerogerador proposto por Nienhaus.	39
Figura 25 – Resultado do modelo térmico de Nienhaus.	40
Figura 26 – Ações habilitadas pelas soluções digitais no segmento O&M.	41
Figura 27 – Relação entre eventos de manutenção, complexidade e abordagem.	42
Figura 28 – Estratégias de manutenção em função da disponibilidade e custo.	43
Figura 29 – Estágios no desenvolvimento das soluções na Indústria 4.0.	45
Figura 30 – Níveis básicos para uma solução IIoT.	45
Figura 31 – Arquitetura simplificada de uma solução IIoT.	46
Figura 32 – Recursos necessários em uma solução IIoT.	48
Figura 33 – Quadrantes mágicos Gartner para Plataformas IIoT.	52
Figura 34 – CRISP-DM.	54
Figura 35 – Exemplo de <i>workflow</i> para limpeza e preparação dos dados.	57
Figura 36 – Exemplo de identificação e adequação dos <i>outliers</i> com segundo método.	59
Figura 37 – Dados brutos de fevereiro a abril de 2019 - caso 1 hidrogerador.	60

Figura 38 – Informação estatística e histograma da potência ativa.	60
Figura 39 – Informação estatística e histograma da vibração do mancal.	61
Figura 40 – Informação estatística e histograma da temperatura mancal.	61
Figura 41 – Informação estatística e histograma da temperatura da gaxeta.	62
Figura 42 – Mapa de correlação entre um grupo de variáveis do hidrogerador.	62
Figura 43 – Pairplot das grandezas.	63
Figura 44 – Dados do sensor de vibração e o respectivo <i>boxplot</i>	64
Figura 45 – Visualização de série temporal multivariada com <i>boxplot</i> (valores em pu). . .	64
Figura 46 – Decomposição da série temporal sensor de vibração.	65
Figura 47 – Tensão e corrente terminal do hidrogerador.	66
Figura 48 – Gráfico Volts/Hertz com limite operacional IEC60034-3.	66
Figura 49 – Curva de capacidade do gerador com dados reais.	67
Figura 50 – Histograma de potência do gerador.	68
Figura 51 – Vibração em mm/s RMS dos mancais no tempo.	68
Figura 52 – Histograma de vibração para os mancais.	69
Figura 53 – Boxplot da vibração para os mancais.	69
Figura 54 – Gráfico de dispersão entre potência ativa, vazão e abertura distribuidor. . . .	70
Figura 55 – Gráfico de dispersão potência ativa e queda líquida.	71
Figura 56 – <i>Boxplot</i> queda líquida para condição de turbinas ligada e desligada.	71
Figura 57 – Técnicas de aprendizado de máquina.	72
Figura 58 – <i>Workflow</i> para construção dos algoritmos com aprendizado de máquina. . .	73
Figura 59 – Representação do aprendizado supervisionado.	75
Figura 60 – Processo de treino e teste de um modelo de classificação.	76
Figura 61 – Arquitetura simplificada de uma rede neural.	78
Figura 62 – Processo <i>Feedforward</i> e <i>Backpropagation</i> de uma rede neural.	79
Figura 63 – Representação de um neurônio artificial.	79
Figura 64 – Funções de ativação.	81
Figura 65 – Gradiente descendente e minimização do erro.	82
Figura 66 – Rede neural TLFN focada.	82
Figura 67 – Modelo físico térmico simplificado do sistema de arrefecimento.	83
Figura 68 – Composição do vetor de entrada e saída do modelo.	84
Figura 69 – Proposta de workflow para criação de algoritmo com aprendizado <i>Data-Informed</i> e <i>Data-Driven</i>	87
Figura 70 – Importância e necessidades para o time de dados.	89
Figura 71 – Dados preparados de fevereiro a abril de 2019 - hidro.	92
Figura 72 – Saída do modelo de classificação com 3 padrões.	93
Figura 73 – Saída do modelo de classificação com 4 padrões.	93
Figura 74 – Curva de aprendizado do modelo com a rede 1 e 2.	96
Figura 75 – Saída modelo para rede 1 e rede 2, indicador desvio operacional.	96

Figura 76 – Carta de autorização WEG para utilização dos dados.	106
Figura 77 – Sistema de arrefecimento do aerogerador.	108
Figura 78 – Dados brutos de janeiro de 2019 a agosto de 2020 - aerogerador.	108
Figura 79 – Dados brutos de janeiro de 2020 a agosto de 2020 - aerogerador.	108
Figura 80 – Dados preparados para utilização no modelo - Caso 2.	108
Figura 81 – <i>Players</i> por status legal.	108
Figura 82 – Publicações por ano.	108
Figura 83 – Tecnologias e aplicações.	108
Figura 84 – Dependências entre os <i>players</i> por citações.	108

LISTA DE TABELAS

Tabela 1 – Rotação gerador em função da frequência e número de polos.	27
Tabela 2 – Variáveis presentes em um hidrogerador.	33
Tabela 3 – Variáveis de um aerogerador utilizadas no estudo.	38
Tabela 4 – Aspectos relevantes para O&M.	41
Tabela 5 – Níveis de aquisição de dados para ativos industriais.	47
Tabela 6 – Comparação geral dos recursos nas soluções de mercado.	49
Tabela 7 – Etapas e duração estimada para implantação da solução <i>Data-Driven</i>	50
Tabela 8 – Patentes na área de solução digital e análise avançada.	51
Tabela 9 – Principais códigos IPC da pesquisa realizada.	52
Tabela 10 – Comparativo entre o modelo físico e modelo baseado em dados.	74
Tabela 11 – Matriz de confusão.	85
Tabela 12 – Principais etapas adotadas na criação do algoritmo.	89
Tabela 13 – Aprimoramentos de <i>software</i> e aquisição de dados necessários para tornar o estudo uma solução.	91
Tabela 14 – Configuração da Rede 1 e Rede 2.	95
Tabela 15 – Valores nominais para conversão do <i>dataset</i> para pu.	108
Tabela 16 – Períodos de treino e teste do modelo - Caso 2.	108
Tabela 17 – Períodos de treino e teste do modelo, status do equipamento e métricas para avaliação da rede 1 e 2.	108

LISTA DE ABREVIATURAS E SIGLAS

IoT	<i>Internet of Things</i> (Internet das Coisas)
IIoT	<i>Industrial Internet Of Things</i> (Internet das Coisas Industriais)
O&M	<i>Operation and Maintenance</i> (Operação e Manutenção)
OEM	<i>Original Equipment Manufacture</i> (Fabricante Original do Equipamento)
TLFN	<i>Time Lagged Feedforward Network</i> (Rede neural alimentada adiante focada atrasada no tempo)
EDA	<i>Exploratory Data Analysis</i> (Análise Exploratória dos Dados)
API	<i>Application Programming Interface</i> (Interface de Programação de Aplicações)
MTBF	<i>Mean Time Between Failures</i> (Tempo Médio Entre Falhas)
MTBI	<i>Mean Time Between Interruptions</i> (Tempo Médio Entre Interrupções)
SCADA	<i>Supervisory Control and Data Acquisition</i> (Sistema Supervisório e Aquisição dos Dados)

SUMÁRIO

1	INTRODUÇÃO	16
1.1	DEFINIÇÃO AO PROBLEMA	17
1.2	LIMITAÇÕES E ASPECTOS RELEVANTES	18
1.3	MOTIVAÇÃO E OBJETIVO	19
1.4	PESQUISA BIBLIOGRÁFICA E CONTRIBUIÇÃO	19
2	GERAÇÃO DE ENERGIA ELÉTRICA E IIOT	21
2.1	HIDROELÉTRICAS	22
2.1.1	Aspectos Construtivos	23
2.1.2	Principais Variáveis de Operação e Aquisição	33
2.2	AEROGERADORES	34
2.2.1	Aspectos construtivos	34
2.2.2	Principais variáveis de operação e aquisição	38
2.2.3	Estudo de caso sobre a estimativa de temperatura	38
2.3	ASPECTOS RELEVANTES PARA O&M	40
2.3.1	Indicadores e estatísticas para O&M	43
2.4	IIOT E INDÚSTRIA 4.0	44
2.5	EXPLORAÇÃO COMERCIAL E PATENTES	47
2.6	CONSIDERAÇÕES	53
3	CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA	54
3.1	SÉRIE TEMPORAL	55
3.1.1	Univariado e multivariado	55
3.2	CORRELAÇÃO E CAUSALIDADE	56
3.3	PREPARAÇÃO DOS DADOS	56
3.4	ORIGEM E REMOÇÃO DE <i>OUTLIERS</i>	58
3.5	ANÁLISE EXPLORATÓRIA	59
3.5.1	Contexto estatístico dos dados	59
3.5.2	Contexto operacional dos dados	65
3.6	APRENDIZADO DE MÁQUINA	72
3.6.1	Modelo físico versus modelo baseado em dados	73
3.6.2	Aprendizado supervisionado e não-supervisionado	74
3.6.3	Classificadores	75
3.6.4	Regressores	77
3.6.5	Redes neurais artificiais e TLFN	77
3.6.6	Criação do vetor de entrada do modelo	82
3.6.7	Períodos de treinamento e teste do modelo	84

3.6.8	Métricas para avaliação	85
3.7	CONSIDERAÇÕES	86
4	WORKFLOW PROPOSTO PARA CONSTRUÇÃO DO ALGORITMO .	87
4.1	CONSTRUÇÃO DO ALGORITMO	88
4.2	RECOMENDAÇÕES PARA IMPLEMENTAÇÃO DO ALGORITMO EM PRODUÇÃO	90
4.3	CONSIDERAÇÕES	91
5	RESULTADOS E DISCUSSÕES	92
5.1	CASO 1 – HIDROGERADOR, TEMPERATURA DA GAXETA	92
5.2	CASO 2 – EÓLICO, SISTEMA DE ARREFECIMENTO	94
5.3	CONSIDERAÇÕES	97
6	CONCLUSÃO	98
	REFERÊNCIAS	100
	GLOSSÁRIO	104
	 ANEXOS	 105
	ANEXO A – CARTA DE AUTORIZAÇÃO E AVISO LEGAL	106
	ANEXO B – INFORMAÇÕES CONFIDENCIAIS	107

1 INTRODUÇÃO

A quarta revolução industrial está em curso, o termo “Indústria 4.0” foi introduzido em 2011 pelo *Communication Promoters Group* no *Industry-Science Research Alliance*, na Alemanha. Segundo a academia alemã Acatech (SCHUH et al., 2017), a Indústria 4.0 ou IIoT (*Industrial Internet Of Things*) busca aplicar tecnologias como computação de borda (*edge computing*), computação na nuvem (*cloud computing*), *big data*, sensores sofisticados e algoritmos avançados com aprendizado de máquina (*machine-learning*) para elevar os níveis de disponibilidade e eficiência dos ativos.

Dados do Departamento de Energia dos EUA (EUA, 2020, *online*) indicam que a IIoT pode reduzir de 25% a 35% os custos de manutenção, reduzir de 70% a 75% as paralisações de produção e gerar um aumento de 35% a 45% no tempo de atividade da produção.

Na gestão de ativos industriais, a manutenção preditiva desempenha um papel crucial, busca elevar os níveis de eficiência e disponibilidade da planta. Para isso, atualmente é adotado o conhecimento especialista humano, presente na equipe interna ou em agentes externos como consultores, empresas de serviços ou fabricantes de ativos (OEM, *original equipment manufacturer*). Na indústria do futuro, parte do conhecimento especialista estará embarcado em sistemas “inteligentes” e autônomos, por meio dos algoritmos.

Durante a operação anormal do ativo ou após uma falha, é comum a equipe de manutenção realizar uma análise sobre os dados de operação e uma investigação física sobre o(s) componente(s) da ocorrência. Este tipo de abordagem é conhecido como *troubleshooting*. A efetividade deste processo depende basicamente de três aspectos: da experiência da equipe, da expertise da equipe e da qualidade dos dados e registros. Nota-se que, parte do valor está no conhecimento e parte nos dados. Por este motivo é importante que uma solução digital combine estes dois elementos da melhor maneira possível.

Este estudo possui basicamente seis tópicos. No Capítulo 1 é apresentado a Introdução sobre o tema, definição do problema, limitações, motivação e objetivo. Além disso, é apresentado uma breve pesquisa de patentes, pesquisa bibliográfica e a contribuição deste estudo.

O Capítulo 2 aborda a geração de energia elétrica com hidrogeradores e aerogeradores. Uma explanação básica sobre os aspectos construtivos, operacionais e relevantes para O&M (Operação & Manutenção) são apresentados. São também apresentadas as principais variáveis de cada sistema e a maneira como os dados são coletados.

O Capítulo 3 apresenta a fundamentação teórica básica utilizada neste estudo. Alguns conceitos sobre série temporal univariada e multivariada. A diferença entre correlação e causalidade é discutida. Sobre *outliers*, são apresentadas as possíveis origens e como removê-los do *dataset*. A importância da análise exploratória e preparação dos dados é comentada. Neste capítulo, ainda é comentado sobre o aprendizado supervisionado e não supervisionado, a diferença entre os algoritmos de classificação e regressão. A rede neural utilizada TLFN focada (*Focused Time Lagged Feedforward Network*) e o modo como o vetor de entrada e saída foram criados são

também apresentados. Por fim, é demonstrado como é selecionado o período de treinamento/teste e quais métricas foram utilizadas para avaliação do algoritmo criado.

O Capítulo 4 mostra o *workflow* proposto para análise exploratória do *dataset* e criação do algoritmo com aprendizado de máquina. O algoritmo criado neste estudo não ficou apto para experimentação em ambiente de produção por algumas limitações técnicas e de tempo. As principais melhorias e ações necessárias são listadas na Tabela 13.

O Capítulo 5 apresenta os resultados em caráter conceitual e discussões sobre os dois casos abordados. O primeiro de um hidrogerador, no sensor de temperatura da gaxeta¹ e outro, de um aerogerador, no sistema de arrefecimento² do gerador elétrico. A implementação do algoritmo em um parque eólico foi impossibilitada, mas testes no banco de dados foram feitos e conclusões relevantes são apresentadas. Por fim, o Capítulo 6 apresenta a conclusão do estudo e comentários finais.

1.1 DEFINIÇÃO AO PROBLEMA

Durante a operação de um ativo industrial há falhas ou anormalidades que se manifestam de modo evolutivo e gradual. Em geral, os sistemas de automação/proteção disponíveis na indústria atuam de modo instantâneo ou temporizado (na casa de segundos) para níveis de alarme ou trip (*threshold*) sobre uma determinada variável. Em alguns casos, o sistema permite a alteração do nível de *threshold* conforme lógica condicional, sobre uma segunda variável.

Esta abordagem atende muitos casos da indústria, principalmente os casos com variável univariada e de contexto operacional simples. Para ocorrências anormais com dados multivariados e de contexto operacional pouco trivial, as técnicas tradicionais apresentam eficácia limitada. Além disso, poucos sistemas de automação tradicionais possuem recurso para implementação de abordagens sofisticadas, limitando o uso de técnicas com aprendizado de máquina.

O que se busca desenvolver neste estudo é a criação de um *workflow* de trabalho para construção de um algoritmo com aprendizado de máquina, para detecção do desvio operacional incipiente. Para isso serão utilizados os dados do sistema SCADA que, em geral, já estão disponíveis digitalmente nas aplicações industriais. É importante reforçar que tal abordagem busca a detecção e diagnóstico incipiente de falhas ou desvio operacional e não o prognóstico.

Este estudo aborda duas aplicações reais de geração de energia elétrica, uma com hidrogerador e outra com aerogerador. Para o hidrogerador são utilizados os dados de operação com desvio operacional na temperatura da gaxeta, com evolução gradual em um período de uma semana. No caso do aerogerador os dados utilizados possuem desvio operacional na velocidade da bomba do fluido de arrefecimento e na abertura da válvula proporcional do sistema de arrefecimento.

Em ambos os casos, seria desejável que um sistema identificasse o desvio operacional

¹ Elemento do sistema de vedação do eixo (Figura 17).

² Circuito com fluido específico para troca térmica entre os elementos de potência e os radiadores (Figura 77).

com a máxima antecedência possível, minimizando o risco do componente sofrer um dano severo, especialmente no caso do aerogerador onde os custos de manutenção corretiva são elevados.

1.2 LIMITAÇÕES E ASPECTOS RELEVANTES

Durante o desenvolvimento deste estudo foram observadas algumas limitações e aspectos relevantes que influenciam no desenvolvimento do estudo:

- Necessidade de conhecimento sobre a construção do ativo, em especial sobre o subsistema sob investigação (sistema gaxeta e sistema de arrefecimento);
- A geração eólica possui grande variação da condição operacional (dinâmica), característica natural da aplicação. Tal condição insere um grande número de cenários de aprendizado;
- Na indústria e na manutenção preditiva, as ocorrências de falso positivo e falso negativo são pouco toleradas. Por este motivo, deve-se buscar abordagens confiáveis e de fácil comprovação;
- A interação com a equipe de O&M e engenharia do produto é fundamental para o entendimento do contexto no *dataset*;
- Alguns sistemas industriais possuem limitação para armazenamento dos dados e ausência de sinais relevantes, por motivos diversos. Isso pode limitar o alcance de uma solução por algoritmo.

Outras limitações são apresentadas no Anexo B. Para atender algumas exigências de sigilo e anonimato impostas pela empresa WEG, algumas informações confidenciais e sensíveis foram colocadas no Anexo B e serão suprimidas na versão pública.

Os dados de operação do hidrogerador foram obtidos em um projeto IIoT realizado pela empresa WEG. Tal projeto foi realizado em caráter de prova de conceito, com objetivo de buscar aprendizados sobre o monitoramento remoto de uma pequena central hidrelétrica (PCH). Por motivos técnicos, o período de dados disponível é de apenas 60 dias. Os dados e informações sobre o hidrogerador possui nível de sigilo intermediário e serão apresentadas neste estudo.

Os dados de operação do aerogerador foram obtidos do ambiente de produção, proveniente de um dos parques que a empresa WEG realiza a operação e manutenção (O&M). O período de dados disponível é de 2 anos ³. Os dados e informações sobre os aerogeradores possuem nível de sigilo elevado. Por este motivo, a análise exploratória dos dados apresentada no Capítulo 2 é feita com algumas variáveis de operação do hidrogerador.

³ Ver limitações no Anexo B.

1.3 MOTIVAÇÃO E OBJETIVO

O tema "inteligência artificial" está cada vez mais em destaque, tanto na academia como na indústria. A capacidade de um algoritmo com aprendizado de máquina emular o poder cognitivo humano é fascinante. Compreender como esta tecnologia pode ser aplicada na indústria, para resolver problemas que os métodos tradicionais não conseguem, é um dos principais motivadores deste estudo.

O principal objetivo deste estudo é explorar as atividades que precedem a criação de um algoritmo com aprendizado de máquina, com foco na aplicação industrial. Como resultado deste processo, é proposto o *workflow* da Figura 69.

O objetivo secundário está no processo exploratório e de criação de um algoritmo com aprendizado de máquina, para consumir os dados do sistema SCADA (*Supervisory Control and Data Acquisition*) e identificar um desvio operacional incipiente de ativos, de modo autônomo. Tal abordagem é discutida neste estudo e dois casos reais são apresentados, um na Seção 5.1 e outro na Seção 5.2.

1.4 PESQUISA BIBLIOGRÁFICA E CONTRIBUIÇÃO

Para o desenvolvimento do primeiro objetivo, a principal referência utilizada é a metodologia CRISP-DM⁴ (WIRTH; HIPPE, 2000). Adicionalmente, consultas sobre outras metodologias e sobre o processo de construção dos algoritmos de aprendizado foram realizadas na documentação do *Sklearn*, *Keras* e *Tensorflow* e em sites como *towardsdatascience.com*, *medium.com* e *machinelearningmastery.com*. A troca de experiências com colegas de trabalho que atuam nesta área também é uma fonte importante de informação.

Para o segundo objetivo, a pesquisa bibliográfica é feita em duas etapas, a primeira focada no entendimento das técnicas de aprendizado de máquina, mais especificadamente na rede neural aplicada às séries temporais, com as referências (MAFRA, 2019), (GUPTA et al., 2014), (NIED, 2007) e (HAYKIN, 2007). A segunda etapa trata do uso das técnicas de aprendizado de máquina em aplicações reais, para compreender os detalhes envolvidos no desafio e as principais diferenças entre os modelos físicos e modelos baseados em dados. Neste caso as referências mais relevantes observadas na pesquisa são (FAWAZ et al., 2019), (ZHOU; HU; TOMLIN, 2016), (JI et al., 2012) e (NIENHAUS; HILBERT, 2012). A contribuição destas referências é apresentada na Seção 2.2.3 e Seção 3.6.5.

A principal contribuição deste estudo é a criação de um *workflow* para orientar o processo de criação de algoritmo com aprendizado de máquina para aplicação industrial. Não é escopo deste estudo explorar de modo aprofundado as técnicas de aprendizado de máquina ou criar uma técnica inovadora.

⁴ Cross Industry Standard Process for Data Mining.

O *workflow* proposto (Figura 69) é aplicado em dois estudos de caso, com dados reais de um hidrogerador e um aerogerador, que são apresentados no Capítulo 5.

A contribuição secundária é a exploração, em caráter conceitual, de algoritmos de aprendizado de máquina (*Random Forest* e rede neural TLFN focada) com dados de série temporal multivariada, proveniente de aplicação real. Sendo o objetivo, a detecção de desvio operacional em sistema ou subsistema industrial. Espera-se que partes do código e algumas descobertas realizadas neste estudo possam auxiliar o time de O&M WEG na construção de uma solução robusta, aplicável para o caso investigado ou em outros casos que tal tecnologia possa ser empregada.

2 GERAÇÃO DE ENERGIA ELÉTRICA E IIOT

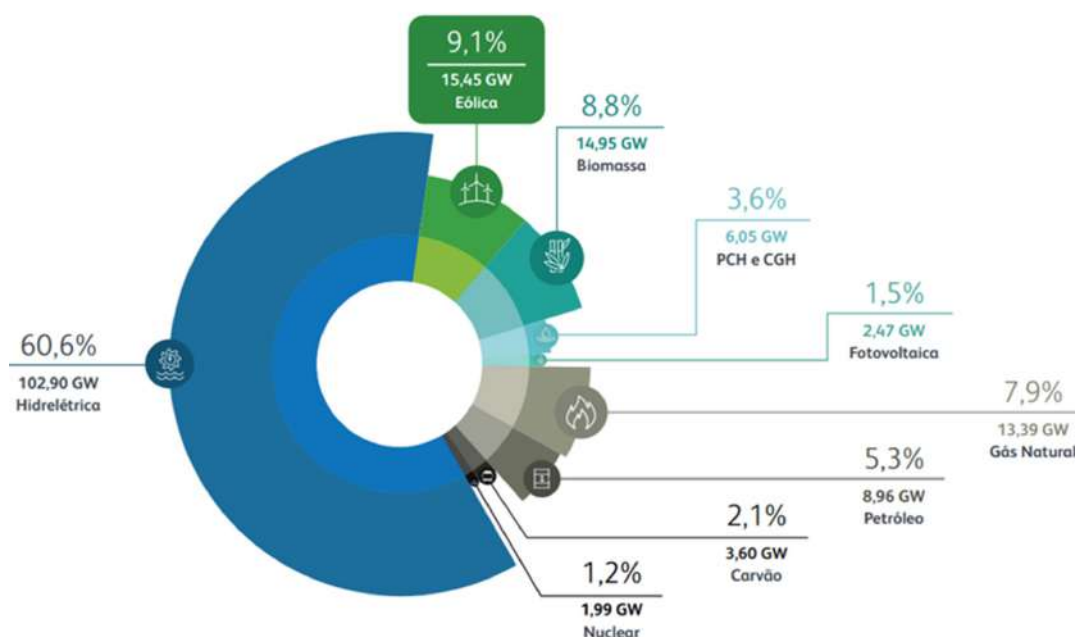
Este capítulo apresenta os aspectos construtivos e de operação do hidrogerador e aerogerador, as principais variáveis do sistema e como estes dados são adquiridos pelo sistema SCADA e pela plataforma IIoT.

O processo de otimização da manutenção em ativos de geração é uma prioridade constante. As iniciativas neste segmento buscam minimizar as ocorrências de manutenção corretiva. As novas tecnologias prometem auxiliar neste ponto, como será introduzido neste estudo.

Segundo a (IEA, 2020, *online*), a projeção de produção de energia por hidrelétricas, considerando um cenário de desenvolvimento sustentável, prevê um aumento de 16% para 2025 e de 33% para 2030. Já para geração eólica terrestre (*on-shore*) é estimado um aumento de 85% e 188%, respectivamente.

Recentemente, o nível de geração eólica ultrapassou a geração por biomassa no Brasil, segundo a (ABEEÓLICA, 2020, *online*), a capacidade instalada em 2019 foi de 15,45GW. A Figura 1 apresenta a composição da matriz energética brasileira para 2019. No site da associação, em outubro de 2020 está sendo registrado a construção de 2,0GW em geração eólica.

Figura 1 – Matriz energética brasileira.



Fonte: (ABEEÓLICA, 2020, *online*).

Conforme o Operador Nacional do Sistema (ONS), o perfil de geração hidráulica (Figura 2) e eólica (Figura 3) são complementares, obviamente que a eólica em menor proporção. Com destaque para o mês de agosto, onde ocorre a mínima na geração hidráulica e máxima na eólica.

Figura 2 – Geração de energia (MWmed) hidrelétricas em 2019.



Fonte: (ONS, 2020, *online*).

Figura 3 – Geração de energia (MWmed) eólica em 2019.



Fonte: (ONS, 2020, *online*).

2.1 HIDROELÉTRICAS

De acordo com a (ANEEL, 2020, *online*), as Centrais Geradoras Hidráulicas (CGHs) são usinas com potência máxima de até 5 MW, já as Pequenas Centrais Hidrelétricas (PCHs) são usinas com reservatório de até treze quilômetros quadrados e potência instalada entre 5 e 30 MW. Para potência superior a 30MW é considerado UHE, usina hidrelétrica.

A Figura 4 apresenta um hidrogerador com turbina Francis em operação. Nesta imagem é possível visualizar o tamanho do mancal dianteiro, que é do tipo de deslizamento com escora e contra escora para suportar o empuxo axial exercido pela turbina.

Figura 4 – Hidrogerador em operação.



Fonte: cortesia WEG, 2020.

2.1.1 Aspectos Construtivos

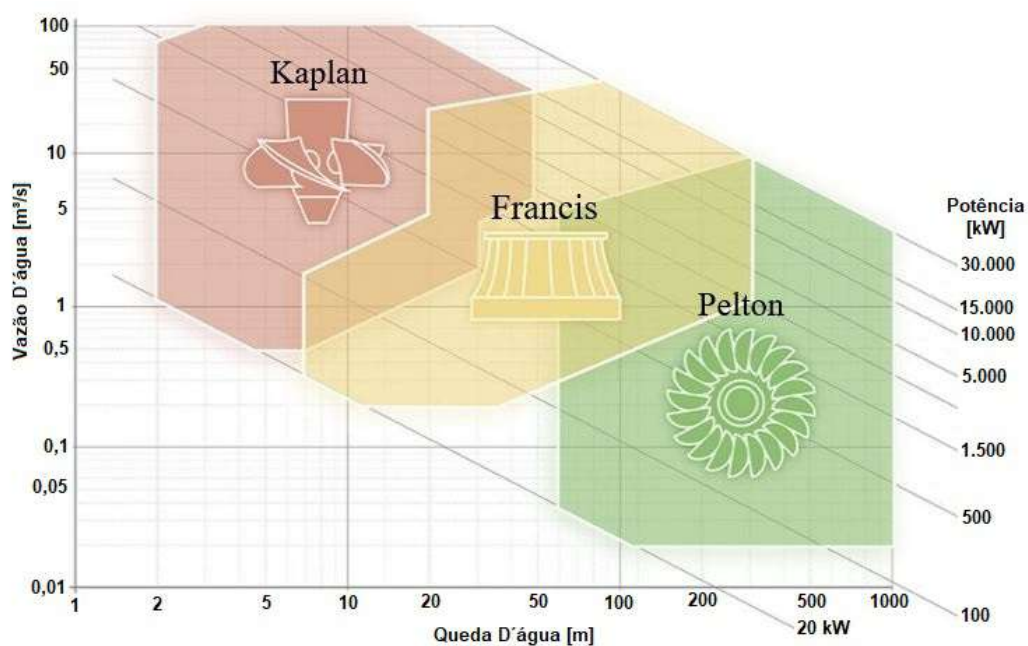
Existem basicamente três tipos de turbinas hidráulicas: Francis, Pelton e Kaplan (Propeller). A escolha da turbina é feita observando a relação entre queda líquida (metros) e vazão (m^3/s). Conforme (ASLAN; ARSLAN; YASAR, 2008), a potência gerada nos terminais do gerador depende diretamente da vazão d'água, da queda líquida d'água e de aspectos construtivos da turbina e gerador, que resultam no rendimento do conjunto.

Na Figura 5 é possível visualizar a faixa de aplicação para cada tipo de turbina, para potência até 30MW. Já na Figura 6, é apresentada a curva de rendimento por potência ativa para turbinas hidráulicas com queda fixa. Os principais parâmetros que influenciam no rendimento do conjunto turbina e gerador são, basicamente, o nível de potência da máquina, o tipo de turbina, o ponto de operação e a tecnologia empregada na construção da turbina.

A Figura 7 apresenta alguns detalhes e aspectos construtivos das turbinas hidráulicas Pelton, Francis e Kaplan.

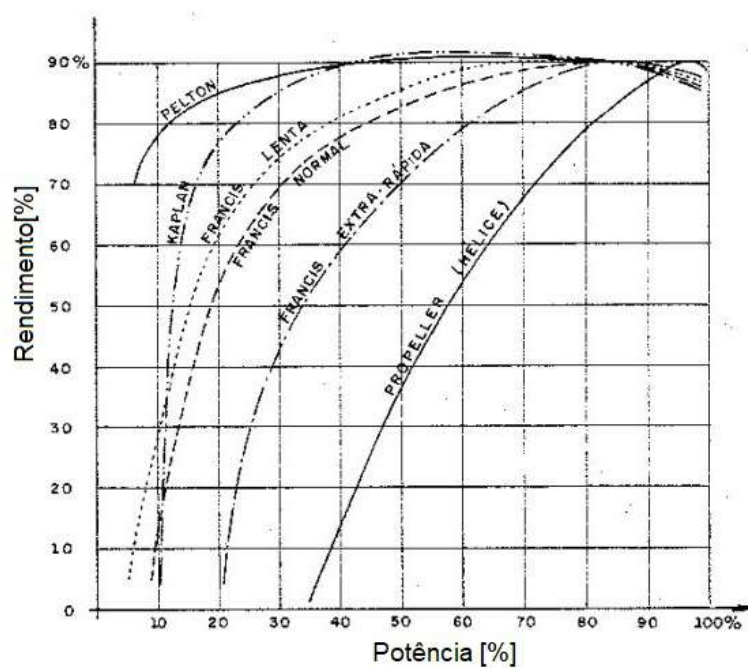
Conforme (WEG, 2016, *online*), grande parte dos geradores síncronos de médio porte possuem excitação *brushless*, as máquinas com rotação elevada são de polos lisos e as máquinas lentas de polos salientes. Em geral, o rotor dos hidrogeradores (Figura 8) são construídos com polos salientes, para operação com tensão de até 1kV em corrente contínua. Já o estator (armadura) dos geradores, são concebidos para operar em média tensão, até 15kV.

Figura 5 – Aplicação de turbinas hidráulicas até 30MW.



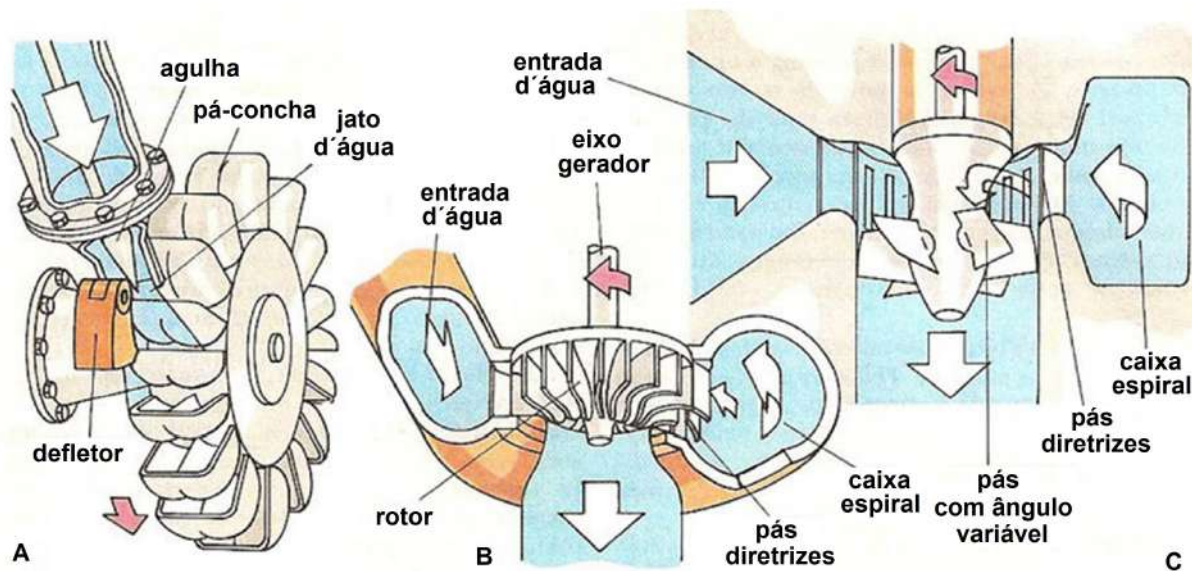
Fonte: Adaptado de (HIDROENERGIA, 2020, *online*).

Figura 6 – Curva de rendimento para turbinas hidráulicas em função da potência.



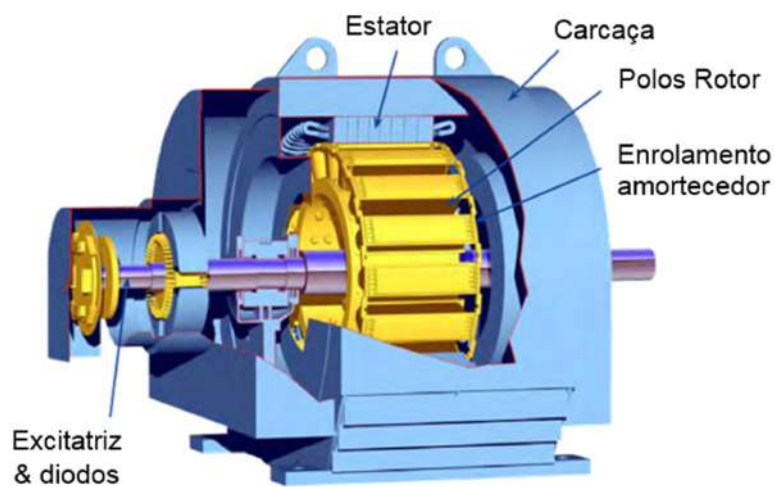
Fonte: (UNESP, 2020, *online*).

Figura 7 – Turbinas Pelton (A), Francis (B) e Kaplan (C).



Fonte: Adaptado de (OSU, 2020, *online*).

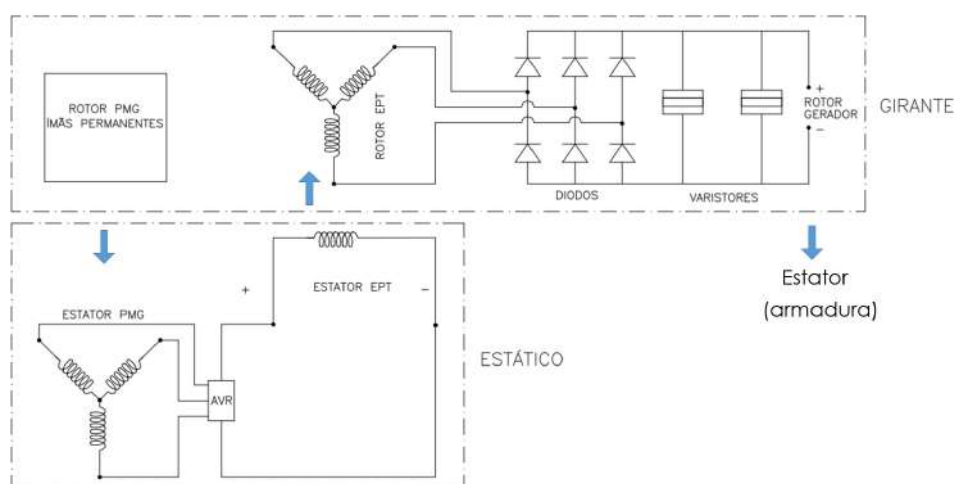
Figura 8 – Gerador de polos salientes com excitação brushless.



Fonte: cortesia WEG, 2020.

A Figura 9 apresenta o diagrama elétrico do gerador. Nela é possível observar como o rotor principal recebe a corrente de excitação proveniente do regulador de tensão (AVR, *automatic voltage regulator*). Este tipo de gerador não precisa de fonte externa para excitação. O fluxo de energia de excitação acontece dos ímãs permanentes para o estator PMG (*permanent magnetic generator*, excitatriz auxiliar); do AVR para excitatriz EPT (excitatriz principal trifásica) por meio do acoplamento magnético; e da retificação com seis diodos de potência (ponte de Graetz) para o rotor principal do gerador.

Figura 9 – Diagrama elétrico do gerador com excitação brushless.



Fonte: (WEG, 2016, *online*).

Conforme observado na equação (1), a velocidade de rotação do rotor depende basicamente do número de polos e da frequência da rede elétrica, conforme exemplos apresentados na Tabela 1. Por este motivo, os geradores convencionais operam com velocidade fixa, para manter a frequência da rede no valor desejado (50Hz ou 60Hz).

$$n_s = \frac{120f}{p} \quad (1)$$

Onde:

n_s = rotação síncrona;

f = frequência elétrica;

p = número de polos físicos.

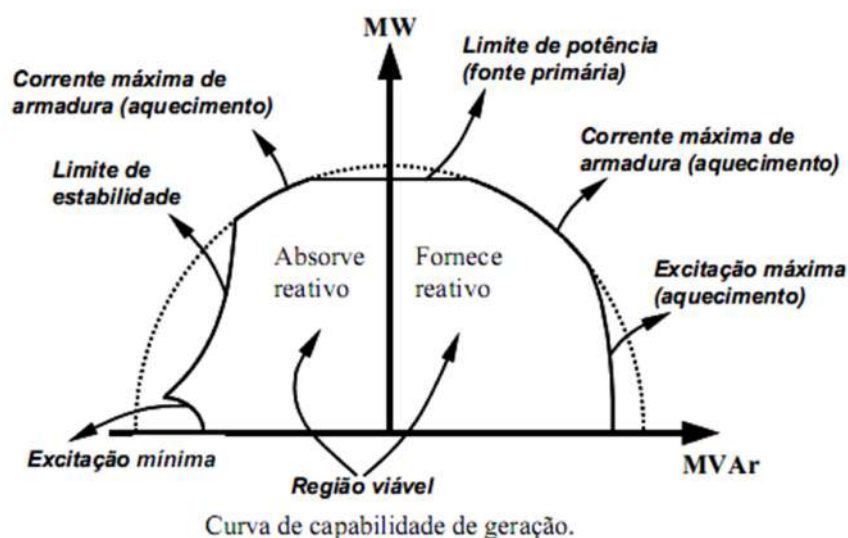
Tabela 1 – Rotação gerador em função da frequência e número de polos.

Número de Polos	Rotação em rpm	
	50Hz	60Hz
2	3000	3600
4	1500	1800
6	1000	1200
8	750	900
10	600	720

Fonte: elaborado pelo autor, 2020.

A curva de capacidade (Figura 10) de um gerador mostra a região operacional viável e segura em termos de potência ativa (W) e reativa (VAr). Se houver a operação além do limite (linha contínua), o gerador corre o risco de perder o sincronismo com a rede (limite de estabilidade), de sobreaquecer o estator (armadura) ou sobreaquecer o rotor.

Figura 10 – Curva de capacidade do gerador e limites de operação.

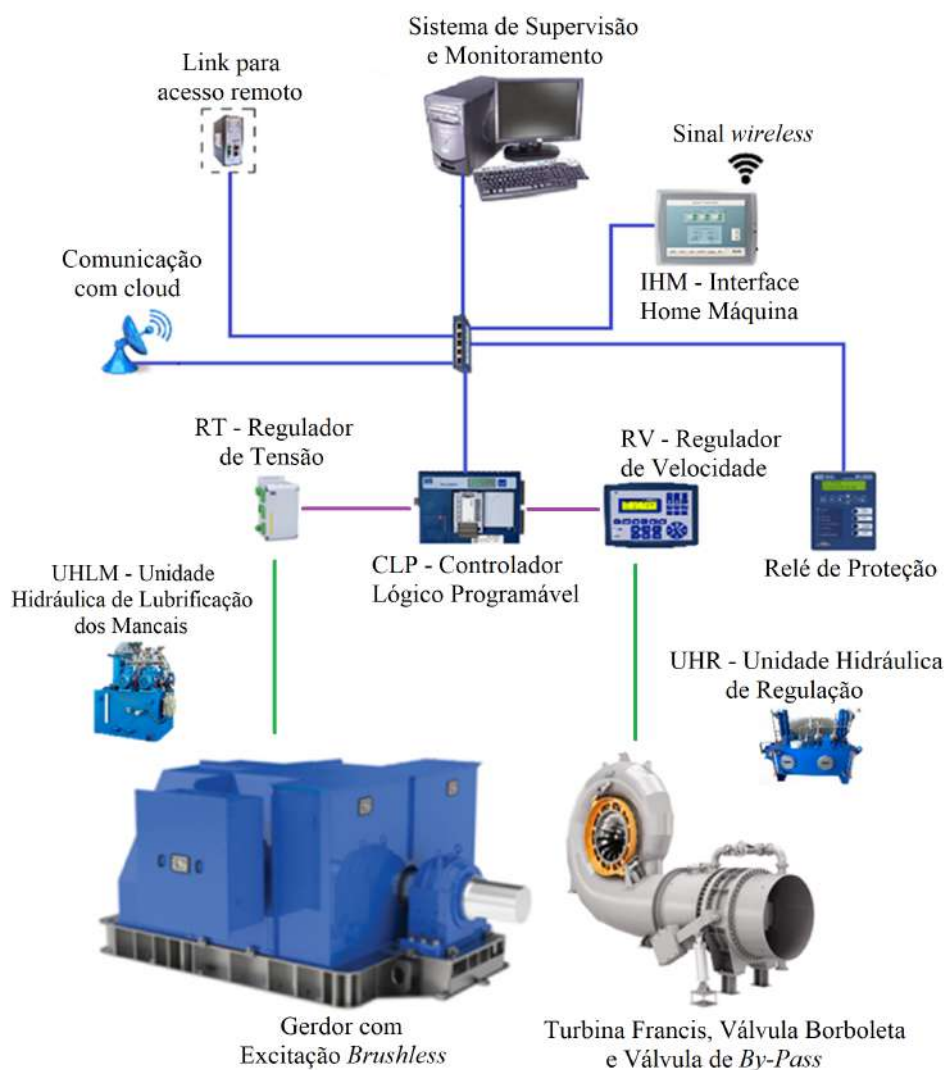


Fonte: (ALMEIDA, 2020, online).

O rendimento do gerador elétrico depende diretamente da potência gerada e do fator de potência durante a operação, além dos aspectos construtivos obviamente. O maior rendimento é observado durante a operação com fator de potência unitário (WEG, 2016, online).

A Figura 11 apresenta o diagrama simplificado do sistema de automação tipicamente utilizado em plantas de hidrogeração. O controle e operação do sistema é realizado por meio do sistema SCADA (Figura 12). Em sistemas modernos, o nível de automação e digitalização dos sensores é elevado, em alguns casos o monitoramento e completa operação da planta acontece de maneira remota.

Figura 11 – Diagrama simplificado do sistema de automação em uma hidroelétrica.



Fonte: cortesia WEG, 2020.

Em geral, os sistemas de automação utilizam protocolos e meios físicos de comunicação variados, a escolha de qual utilizar depende da aplicação e do orçamento disponível. Na Figura 11 as linhas em azul representam a comunicação de alto nível, como por exemplo as redes Ethernet, Profibus, CANopen e outros. As linhas em violeta representam a comunicação com estes protocolos ou ainda por sinais elétricos digitais ou analógicos. Já a interação entre os elementos de controle e os ativos de geração acontece por meio de sinais elétricos (digitais e analógicos) e mecânicos (pneumáticos e hidráulicos), trata-se das linhas em verde.

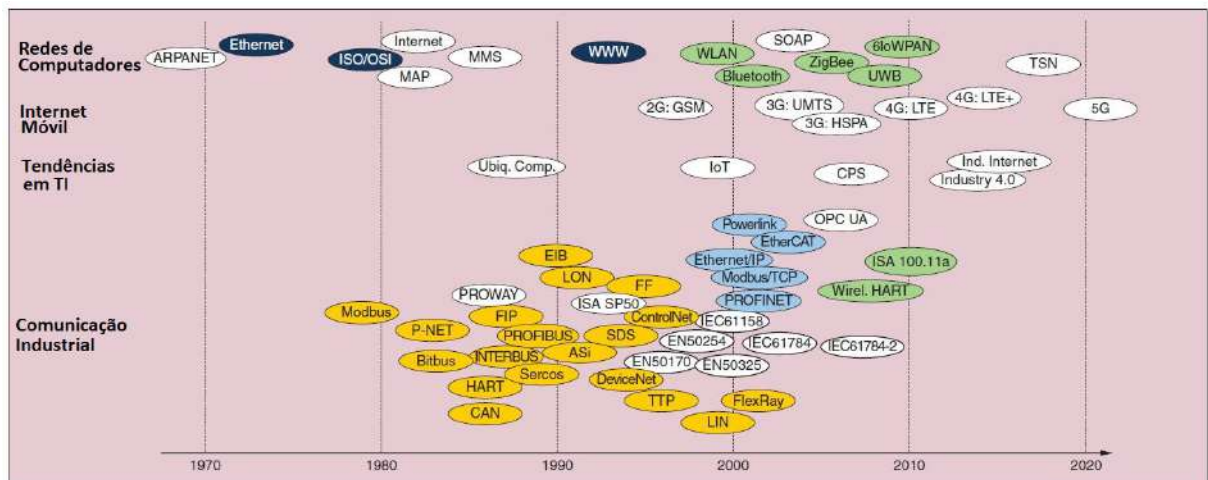
A Figura 13 mostra que o universo de comunicação é vasto, tanto no segmento industrial como no IoT tradicional. Entre a década de 80 e meados dos anos 2000 é notada a criação e evolução de muitos protocolos industriais. Já para o segmento de IoT tradicional, o marco de evolução pode ser considerado o nascimento do *Bluetooth*, nos anos 2000.

Figura 12 – Sistema SCADA para planta de hidrogeração.



Fonte: cortesia WEG, 2020.

Figura 13 – Evolução da comunicação industrial e tecnologias relacionadas.

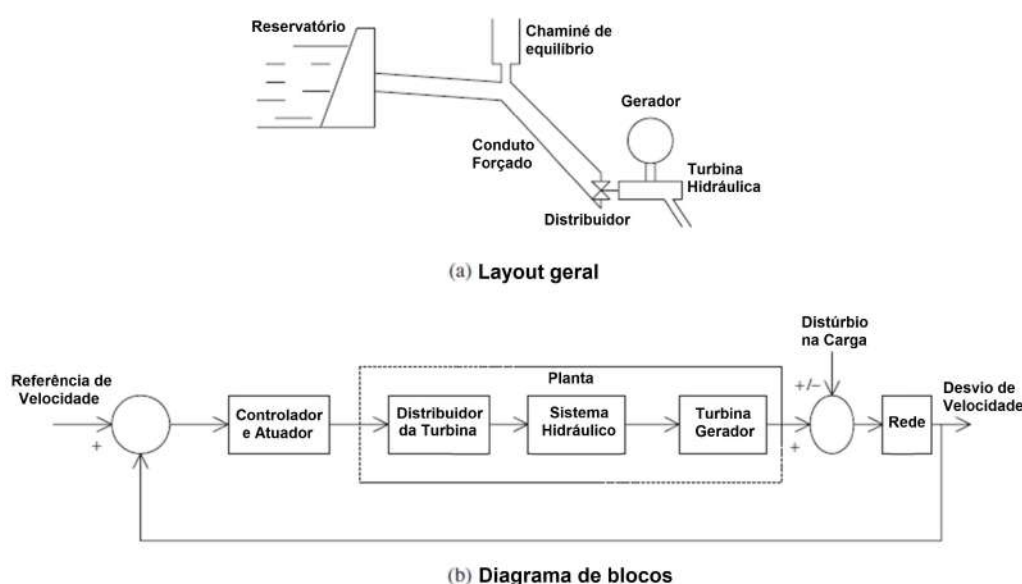


Fonte: (WOLLSCHLAEGE; SAUTER; JASPERNEITE, 2017).

Conforme observado em (KISHOR; SAINI; SINGH, 2007) e (KUNDUR; BALU; LAUBY, 1994), os controles dinâmicos presentes em plantas de geração acontecem no regulador de velocidade (turbina, Figura 14) e no regulador de tensão (gerador, Figura 9).

O controle de velocidade e potência da turbina é realizado por meio da abertura das pás diretrizes (*guide vane*s), também conhecido como distribuidor ou, no caso da turbina Pelton, na abertura da(s) agulha(s). Este elemento é responsável por aumentar ou diminuir o fluxo de água sobre o rotor da máquina. Em turbinas Kaplan, como as pás do rotor são móveis, também é possível controlar o seu ângulo. O controle da turbina possui, basicamente, duas malhas: a malha de controle de velocidade e de carga. A primeira atua quando a máquina está em modo isócrono (isolado) ou com sistema desconectado da rede. Nesta malha, o sistema é responsável por manter a rotação/frequência fixa, no valor de set-point desejado (Tabela 1). Já a malha de controle de carga, esta regula a potência ativa (kW) despachada pelo gerador na rede de transmissão ou distribuição, ou seja, na barra infinita (tensão e frequência fixa).

Figura 14 – Esquemático planta de hidrogeração (a) layout geral e (b) diagrama de blocos.

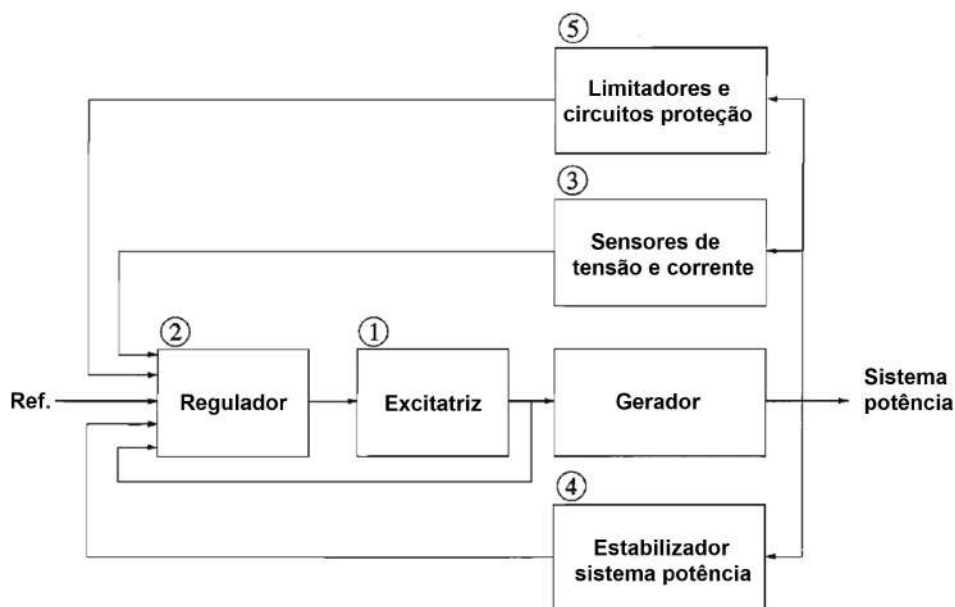


Fonte: Adaptado de (KISHOR; SAINI; SINGH, 2007).

Conforme visto anteriormente, no gerador o controle acontece na corrente de excitação do rotor principal. Há basicamente quatro malhas de controle: corrente de excitação, tensão terminal, fator de potência e potência reativa. A malha de corrente de excitação é usada somente no comissionamento, a malha de tensão terminal é utilizada na operação em modo isócrono (isolado) ou com sistema desconectado da rede. A malha de controle de fator de potência e de potência reativa é utilizada quando o gerador está conectado na rede ou em paralelo com outro gerador.

Normalmente, em CGHs e PCHs o sistema de arrefecimento do gerador é feito pelo ar da casa de força. A circulação de ar é feita por ventiladores instalados no eixo do gerador, tratando-se de um sistema simples e que geralmente não apresenta problemas.

Figura 15 – Diagrama de blocos do sistema de excitação do gerador.

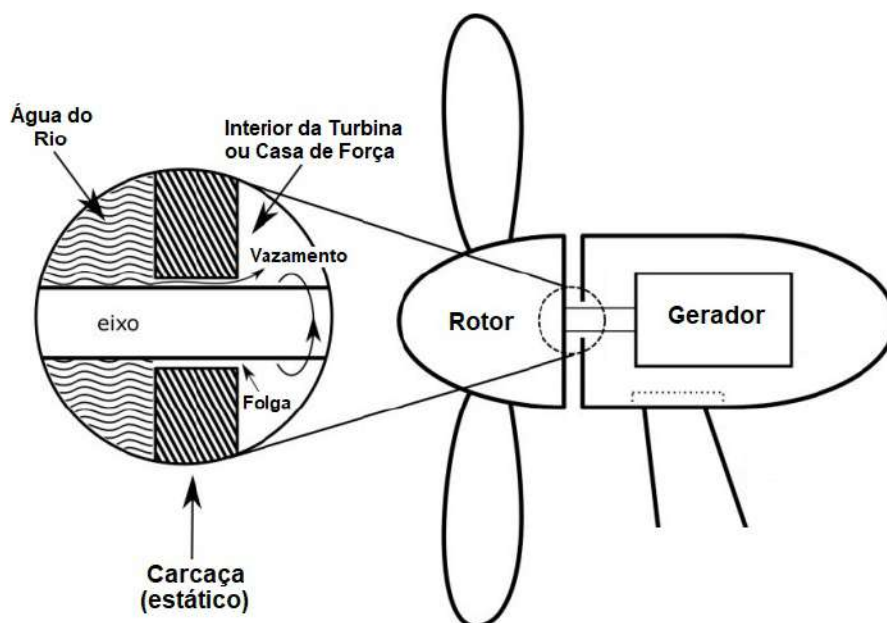


Fonte: Adaptado de (KUNDUR; BALU; LAUBY, 1994).

Em turbinas e bombas hidráulicas, o elemento de vedação do eixo desempenha um papel muito importante, impedindo a passagem do fluido (água) pelas folgas existentes entre as peças estáticas e rotacionais do conjunto. Conforme (QUIXADÁ, 2014), tais vedações mecânicas podem ser fabricadas em diversas geometrias, materiais (ou combinações de materiais) e, dependendo de sua aplicação, podem apresentar características especiais. A Figura 16 apresenta uma turbina do tipo bulbo, no detalhe é mostrado o local de instalação do sistema de vedação do eixo (gaxeta), entre o eixo e a carcaça da turbina. Quando a gaxeta está adequadamente montada e especificada, o fluxo d'água para o compartimento do gerador é mínimo. Uma questão relevante na utilização das gaxetas é que elas precisam de um pequeno volume de fluido no local de interface com eixo, para lubrificação da região de atrito. Isso auxilia no resfriamento do conjunto, removendo parte do calor gerado na fricção.

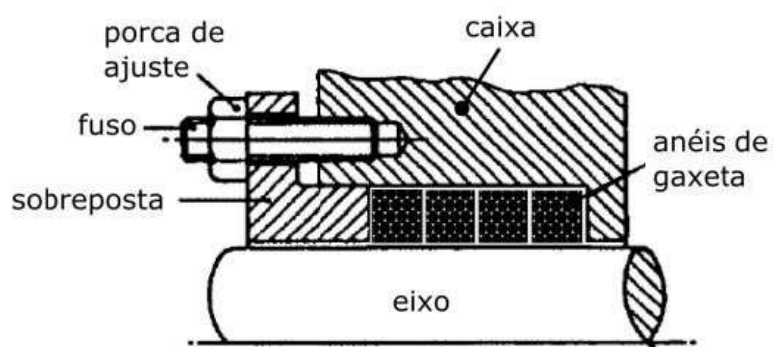
A Figura 17 mostra detalhes da vedação no eixo em uma turbina hidráulica. A referência indica que este sistema de vedação possui uma série de detalhes operacionais e, quando algum desvio acontece, geralmente é observada a elevação de temperatura da gaxeta ou o aumento da vazão do fluido responsável pela lubrificação. O hidrogerador utilizado neste estudo possuía um sensor de temperatura instalado na caixa deste sistema de vedação. Tais consequências são indesejadas e são motivo de constante preocupação pela equipe de O&M.

Figura 16 – Ponto de instalação do sistema de vedação do eixo.



Fonte: Adaptado de (QUIXADÁ, 2014)

Figura 17 – Detalhes do sistema de vedação do eixo com anéis de gaxeta



((a) Configuração típica da vedação



(b) Tipos de anéis de gaxeta.

Fonte: (QUIXADÁ, 2014)

2.1.2 Principais Variáveis de Operação e Aquisição

Em um sistema de hidrogeração, a quantidade de sinais digitais está diretamente relacionada com a idade e com tamanho da planta. Em geral, quanto maior e mais recente a planta, maior será o nível de digitalização do processo de geração. Na Tabela 2 são apresentadas algumas variáveis digitais e analógicas que podem ser monitoradas e usadas como inputs em um algoritmo.

Tabela 2 – Variáveis presentes em um hidrogerador.

Variáveis/Grandeza	Unidade
Disjuntor principal	status digital
Contato de campo (excitação)	status digital
Relé de bloqueio elétrico, mecânico e hidráulico (86E, 86M e 86H)	status digital
Vibração mancal dianteiro e traseiro	RMS mm/s
Corrente das fases do estator	Aca RMS
Tensão das fases do estator	Vca RMS
Corrente de excitação	Adc
Tensão de excitação	Vdc
Frequência	Hz
Fator de potência	-
Potência ativa, reativa e aparente	kW, kVAr e kVA
Energia ativa, reativa e aparente	kWh, kVARh, kVAh
Velocidade do eixo	rpm
Temperatura das fases	°C
Temperaturas do mancal: escora, contra escora, guia e óleo	°C
Temperaturas do mancal traseiro: guia e óleo	°C
Abertura do distribuidor	%
Temperatura da gaxeta de vedação do eixo	°C
Vazão de água do conduto principal	m³/s
Pressão fluido de lubrificação	bar
Pressão fluido de regulação	bar
Pressão sobre válvula borboleta (montante e jusante)	bar

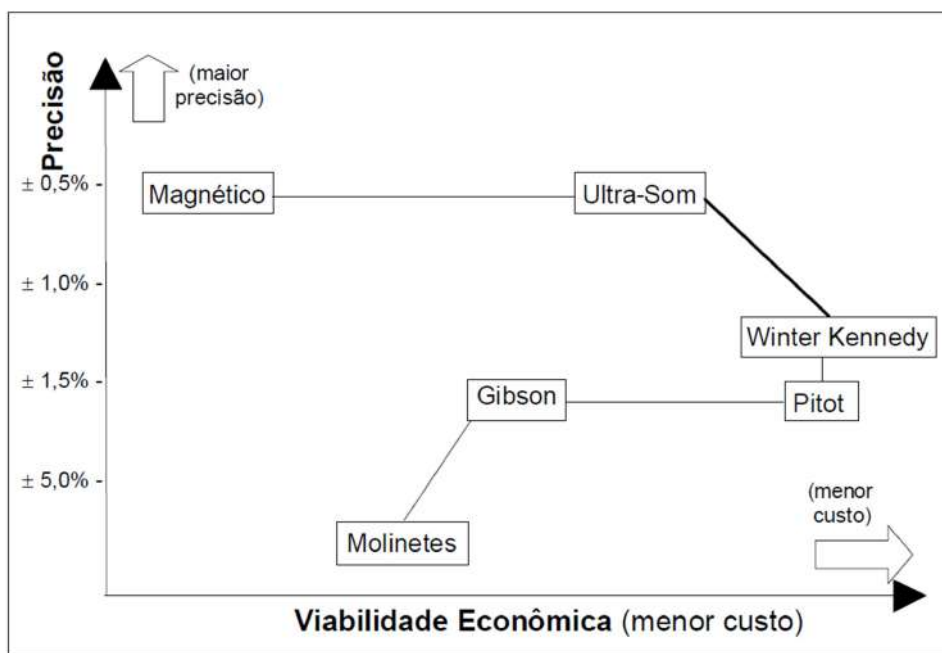
Fonte: elaborado pelo autor, 2020.

Em plantas de pequeno porte, por motivos de economia, algumas projetos são concebidos apenas com os elementos essenciais, dispensando sensores relevantes para o monitoramento da condição, como por exemplo: sensor de vibração dos mancais, sensor de filtro saturado, sensores de temperatura e fluxo de óleo, vazão de água do conduto principal, etc.

Conforme (ANDRADE et al., 2002), para avaliar o rendimento de uma turbina hidráulica é imprescindível a medição da vazão d'água no conduto principal. Valores de rendimento abaixo do esperado podem inviabilizar economicamente o empreendimento. Além disso, o conhecimento do rendimento em tempo real das máquinas hidráulicas pode permitir a otimização dos despachos

de energia elétrica pelos centros de operação do sistema, no entanto tal medição representa um gasto de investimento no sistema de medição. Conforme apresentado na Figura 18, sistemas de medição com precisão próxima de 0,5% geralmente possuem custo financeiro elevado, o que pode inviabilizar sua implementação em plantas de médio e pequeno porte.

Figura 18 – Aplicação dos sensores de vazão.



Fonte: (ANDRADE et al., 2002).

2.2 AEROGERADORES

A Figura 19 apresenta um parque eólico no nordeste do Brasil, com aerogeradores WEG de modelo AGW 110/2.1 (2.1MW), com tecnologia *direct-drive*⁵. Conforme (BARANKIEVICZ et al., 2018), a empresa catarinense produziu cerca de 308 aerogeradores, totalizando 646MW.

2.2.1 Aspectos construtivos

A vista interna da nacele do aerogerador é apresentada na Figura 20. A parte frontal é equipada com cubo que recebe as pás e o sistema de controle de *pitch*. O cubo é diretamente conectado ao rotor de ímãs permanentes do gerador. O controle de *pitch* é responsável pelo controle angular das pás. Na parte estática próxima do cubo há o sistema de frenagem do rotor, destinado para condições de manutenção ou de segurança do equipamento. A frenagem em condição dinâmica é realizada de maneira aerodinâmica, com ajuste no ângulo das pás.

O estator com enrolamento trifásico do gerador possui uma serpentina por todo seu perímetro e nela circula o fluido de arrefecimento. Este enrolamento é conectado no conversor de

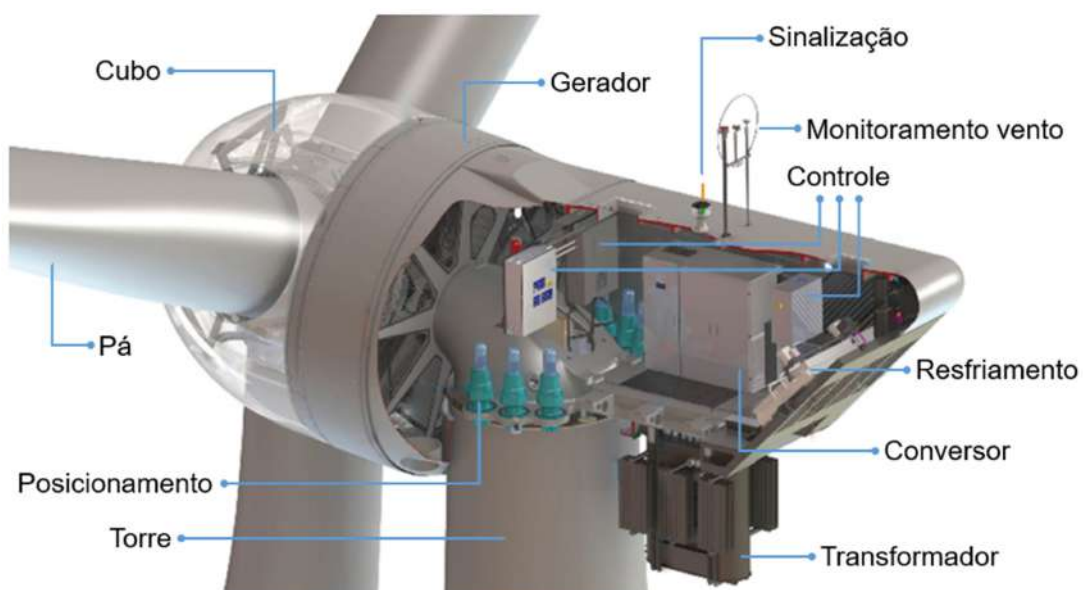
⁵ Possui conversor de potência para conexão direta à rede elétrica. Dispensa o uso de multiplicador de velocidade.

Figura 19 – Parque de geração eólico com AGW 110/2.1.



Fonte: cortesia WEG, 2020.

Figura 20 – Visão interna da nacele do aerogerador.



Fonte: cortesia WEG, 2020.

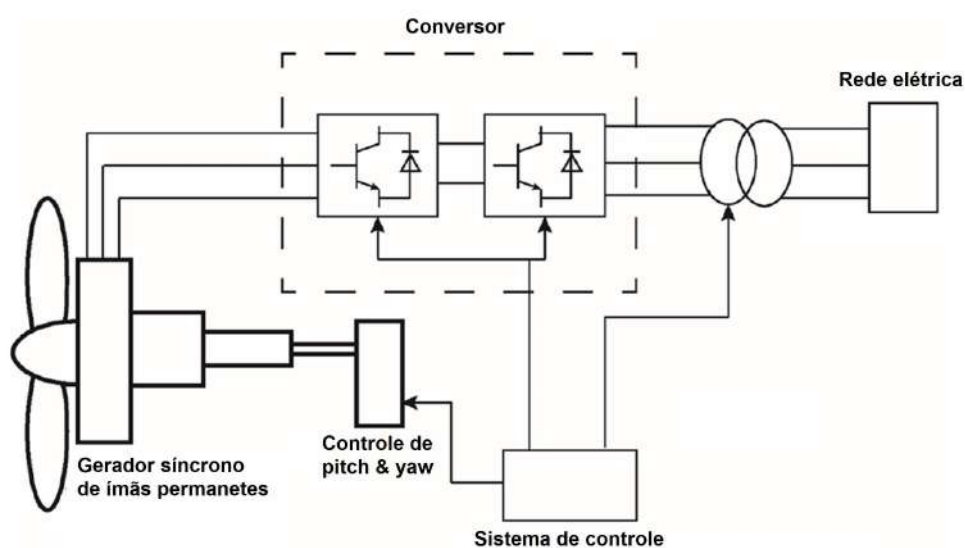
potência que possui um estágio de retificação e outro de inversão. O resfriamento dos elementos semicondutores do conversor é realizado com a circulação do mesmo fluido de arrefecimento, mas nos dissipadores.

O topo da nacele é equipado com sinalizador externo e com a estação meteorológica. Na parte interna da traseira há o sistema de controle geral do aerogerador, os inversores para acionamento dos motores de posicionamento da nacele (yaw^6) e o sistema de arrefecimento, que possui um radiador com ventilação forçada, uma bomba centrífuga e uma válvula proporcional.

Os detalhes do circuito hidráulico do sistema de arrefecimento são apresentados no Anexo B. O algoritmo principal desenvolvido neste estudo busca detectar o desvio operacional na bomba e na válvula proporcional. O resultado obtido (em caráter preliminar) com tal algoritmo é apresentado no Caso 2, na Seção 5.2.

Na parte traseira inferior é instalado o transformador elevador. O lado de baixa tensão é conectado no conversor (inversor) e a parte de alta é conectada na rede de distribuição do parque, conforme mostrado na Figura 21.

Figura 21 – Esquemático simplificado do aerogerador direct-drive.



Fonte: (WANG et al., 2020).

Para o funcionamento do aerogerador *direct-drive* são necessários basicamente quatro malhas de controle: controle de *pitch* (ângulo das pás), controle de posicionamento (*yaw*), controle de potência/torque do conversor retificador e controle de potência no conversor inversor. Estes controles devem interagir entre si de modo natural, com objetivo de atender os requisitos de operação segura.

Na Figura 22 é apresentada a curva típica de potência em relação a velocidade do vento. O controle de *yaw* é independente e busca o melhor alinhamento da nacele com vento. Conforme (BARANKIEVICZ et al., 2018), na região I o vento não é suficiente para operação.

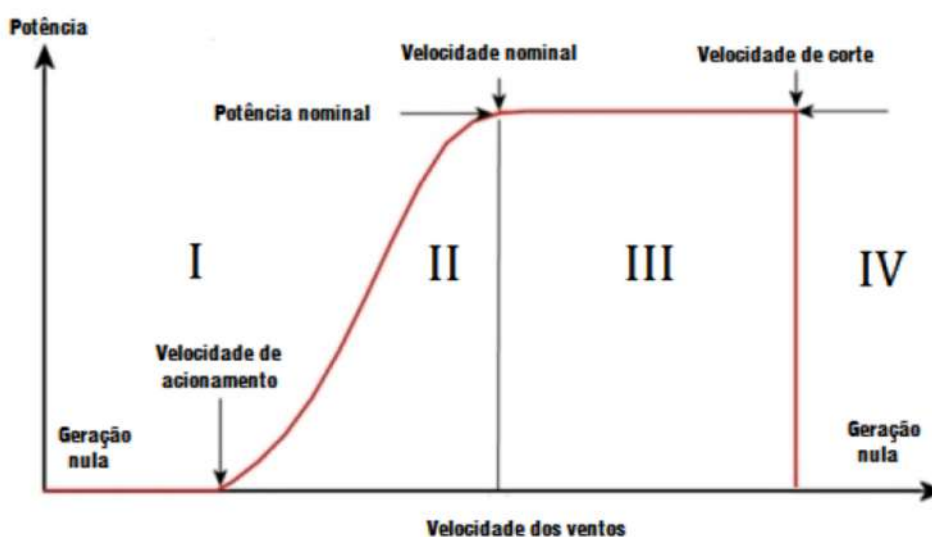
⁶ Sistema responsável por alinhar a nacele ao vento, para o melhor aproveitamento energético.

Na região II, com ventos medianos, o controle de *pitch* mantém o ângulo de ataque das pás para máxima absorção de energia, cerca de zero graus. Neste ponto, o conversor controla o torque do gerador para atender os critérios de máxima velocidade do conjunto e buscando o máximo coeficiente de potência (C_p , relação entre a potência extraída e a potência total do vento).

Na região III o vento é suficiente para extrair a potência nominal e além. Nesta condição o controle de *pitch* atua para diminuir a velocidade do rotor (Figura 23), seguindo o menor coeficiente de potência (C_p) que conserve a velocidade nominal e garanta uma operação segura. A fronteira entre II e III é crítica, pois é onde os controles de *pitch* e torque comutam entre si.

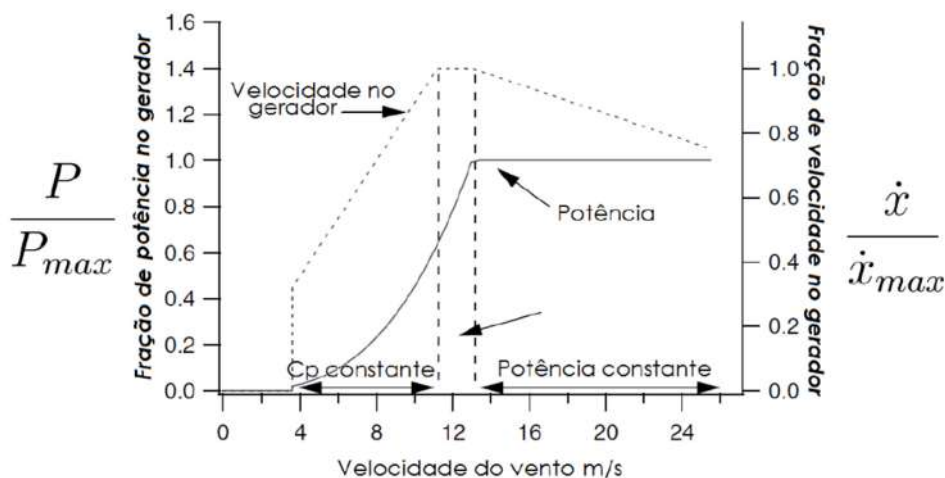
Mesmo com o controle de *pitch* ampliando a faixa de operação, há um limite onde o aerogerador deve ser retirado de funcionamento, a região IV. Uma operação além da velocidade de corte pode oferecer risco ao ativo, por conta dos esforços estruturais envolvidos.

Figura 22 – Regiões da curva de potência do aerogerador.



Fonte: (BARANKIEVICZ et al., 2018).

Figura 23 – Curva de potência, vento e velocidade do aerogerador.



Fonte: (BARANKIEVICZ et al., 2018).

2.2.2 Principais variáveis de operação e aquisição

Em geral, os aerogeradores *direct-drive* são modernos e possuem um elevado nível de sensoriamento e digitalização das variáveis de operação. A Tabela 3 apresenta as variáveis analógicas utilizadas no Caso 2 (Seção 5.2) deste estudo.

Os dados armazenados na plataforma IIoT (*cloud*) são obtidos por meio do sistema SCADA do parque, com amostragem a cada 10 minutos. Estes dados são obtidos por meio do cálculo de média sobre os dados instantâneos (amostragem de 1 segundo) de cada sensor e subsistema. Para atender os termos de sigilo solicitados pela WEG, os detalhes do sistema SCADA e das variáveis do sistema foram suprimidos deste estudo.

Tabela 3 – Variáveis de um aerogerador utilizadas no estudo.

Variáveis/Grandeza	Unidade
Potência ativa	W
Temperatura interna da nacele	°C
Temperatura externa da nacele	°C
Temperatura do estator (múltiplos sensores)	°C
Abertura da válvula proporcional do sistema arrefecimento	%
Velocidade da bomba do fluido do sistema arrefecimento	%

Fonte: elaborado pelo autor, 2020.

2.2.3 Estudo de caso sobre a estimativa de temperatura

Para representar a complexidade envolvida no processo de estimativa de temperatura de um aerogerador, o trabalho de (NIENHAUS; HILBERT, 2012) é utilizado. Nele é abordado a criação e otimização dos parâmetros de um modelo térmico físico (Figura 24) para estimar a

temperatura do estator de um gerador eólico de 2MW, modelo com multiplicador (*gearbox*) e excitação. O resultado alcançado por Nienhaus é apresentado na Figura 25, a margem de desvio é de $\pm 5^\circ\text{C}$, na cor vermelha.

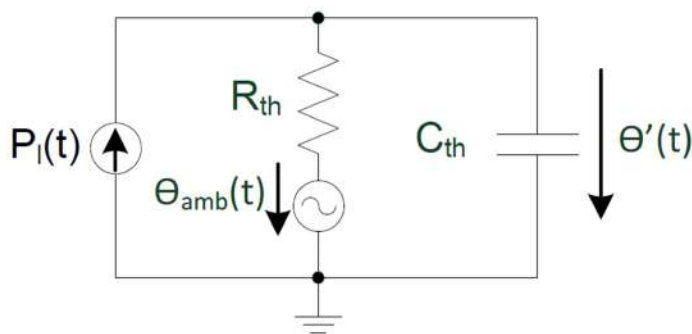
O modelo físico proposto por (NIENHAUS; HILBERT, 2012) atende a condição de potência constante (sem partida e parada). Requer ajustes dos parâmetros por otimização e necessita alguns parâmetros da máquina.

O modelo em questão foi criado para ser implementado no controlador lógico programável do aerogerador, no entanto, antes de colocar o sistema em produção, o autor aponta a necessidade de utilizar métodos de otimização para ajustar os parâmetros do modelo. O autor finaliza informando que o modelo apresentado é a base para trabalhos futuros de pesquisa na área e poderá ser utilizado como parte de um sistema de monitoramento da condição.

(NIENHAUS; HILBERT, 2012) utiliza uma abordagem física para modelar o comportamento da temperatura do estator de um aerogerador. O ajuste dos parâmetros do modelo é feito por processo de otimização, com a variação dos parâmetros do modelo físico (R_{th} e C_{th}) e observação da sua resposta. Sobre o período de treino/teste, a abordagem de Nienhaus não exige muitos dados, uma vez que o processo de aprendizado acontece por meio da otimização dos parâmetros do modelo (R_{th} e C_{th}).

A perda do gerador $P_l(t)$ é diretamente proporcional a potência gerada, já que a parcela principal da perda está relacionado com efeito joule no enrolamento do estator. O elemento R_{th} e C_{th} trata das características construtivas do gerador e a resistência térmica (R_{th}) com os materiais aplicados, formato e composição do componente. Por fim, a capacidade térmica (C_{th}) com volume e massa do componente, o que resulta na inércia térmica do sistema.

Figura 24 – Modelo térmico aerogerador proposto por Nienhaus.



Fonte: (NIENHAUS; HILBERT, 2012).

Onde:

$P_l(t)$ – Perdas no gerador Watts;

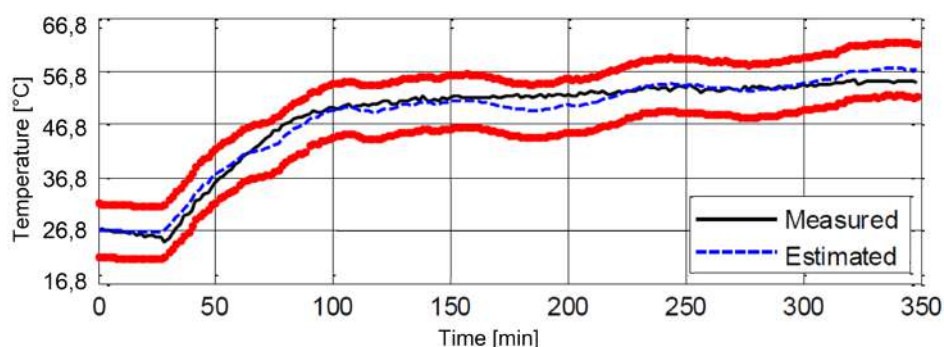
$\Theta_{amb}(t)$ – Temperatura ambiente em $^\circ\text{C}$;

R_{th} – Resistência térmica em $^\circ\text{C}/\text{W}$;

C_{th} – Capacidade térmica em $\text{Ws}/^\circ\text{C}$;

$\Theta'(t)$ – Temperatura estimada do elemento em $^\circ\text{C}$.

Figura 25 – Resultado do modelo térmico de Nienhaus.



Fonte: Adaptado de (NIENHAUS; HILBERT, 2012).

2.3 ASPECTOS RELEVANTES PARA O&M

Conforme (KAA et al., 2020), a realização de manutenção programada de rotina não é necessariamente custosa do ponto de vista financeiro. A manutenção imprevista e reparos não programados, como resultado de baixa confiabilidade, podem consumir tempo e resultar em contas elevadas de manutenção. Em parques eólicos, uma preocupação constante são os reparos que dependem da mobilização de um guindaste, o que representa uma boa quantia financeira e tempo.

A implementação de tecnologias como monitoramento inteligente, aumentará a confiabilidade geral e diminuirá os custos de O&M. Os tópicos apresentados a seguir foram coletados em pesquisas bibliográficas, em entrevista com duas empresas de O&M no segmento de CGHs e PCHs e com o time de O&M da WEG.

Os times consultados são responsáveis pelo O&M de 10 plantas de hidrogeração e 308 turbinas eólicas. A Tabela 4 relaciona alguns aspectos relevantes para respectiva gestão.

A Figura 26 apresenta as ações habilitadas pelas soluções digitais no segmento O&M. A atividade desenvolvida neste estudo pode se encaixar em quatro quadrantes: maximizar o MTBI/MTBF, diagnóstico de falha, diagnóstico e análise de falha remotamente e monitoramento da condição do componente. Estes itens são destacados com retângulo tracejado na cor preta.

Conforme apresentado na introdução deste estudo, uma solução IIoT pode auxiliar a equipe de O&M com relatórios de nível *Data-Informed* e *Data-Driven*. A solução final deve apresentar múltiplas perspectivas, suprimindo as demandas do time de O&M, do proprietário do ativo, parceiros, consultores, OEM e outros. Tais perspectivas devem ser usadas para orientar a criação da solução IIoT, dos *dashboards*, relatórios, notificações, APIs, análise da frota, *insights* sobre a operação e *troubleshooting*.

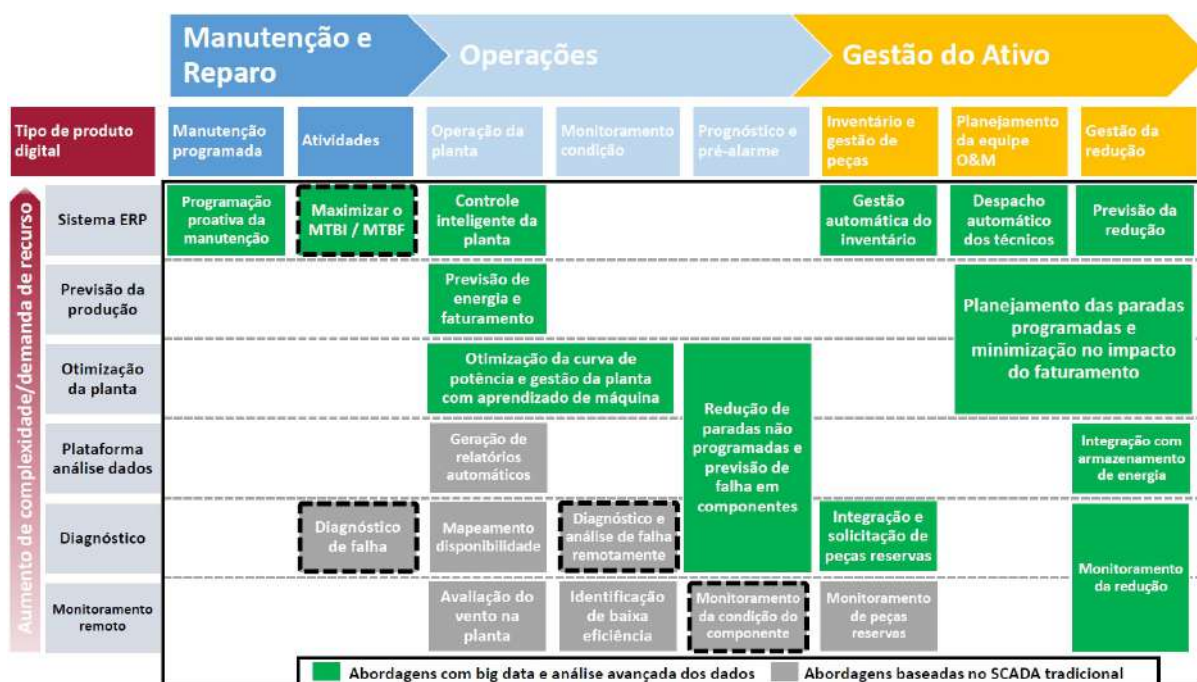
Alguns *insights* do nível *Data-Driven* podem ser criados utilizando a orientação dos times de OEM e O&M. Essa abordagem busca gerar *insights* relevantes sobre a condição operacional do ativo, o que pode contribuir com a diminuição das paradas inesperadas.

Tabela 4 – Aspectos relevantes para O&M.

Aspecto	Aplicação
Conhecer o OEE ou KPI	Aero / Hidro
Maximizar o MTBI/MTBF	Aero / Hidro
Conhecer o aproveitamento eólico/hídrico	Aero / Hidro
Conhecer o nível de degradação das pás do rotor	Aero / Hidro
Descrever corretamente a operação do ativo com dados	Aero / Hidro
Estimar quando acontecerá a próxima falha (probabilidade)	Aero / Hidro
Conhecer o nível de obstrução da grade na tomada d'água	Hidro
Conhecer o comportamento do rio e reservatório em variadas condições meteorológicas e operacionais	Hidro
Registrar eventos relevantes como sincronismos na rede, atuação das proteções, horas de operação da máquina/auxiliares, etc.	Aero / Hidro
Recomendar as próximas ações de manutenção, com base na condição operacional do ativo (horas, carga, vibração, temperatura, etc) e conforme orientação no manual do OEM	Aero / Hidro
Detecção autônoma de desvio operacional com antecedência suficiente para evitar as paradas não programadas, planejar a mobilização da equipe/recursos e providenciar as peças de reposição	Aero / Hidro

Fonte: elaborado pelo autor, 2020.

Figura 26 – Ações habilitadas pelas soluções digitais no segmento O&M.



Fonte: Adaptado de (WOODMACKENZIE, 2015).

Um exemplo de *insight* pode ser o Caso 2 apresentado na Seção 5.2, onde a saída da rede neural pode ser associada a nível de alarme ou pré-alarme para equipe de O&M.

Conforme observado neste estudo, os algoritmos com aprendizado de máquina para detecção de desvio operacional podem operar com inúmeras configurações, que dependerão das variáveis disponíveis e da natureza do problema.

(WELTE et al., 2018) aborda alguns aspectos relevantes sobre a manutenção em hidroelétricas. O autor comenta sobre a complexidade na abordagem de análise baseada pela física ou pelos dados, conforme apresentado na Figura 27. Por exemplo, a detecção de falha no isolamento do rotor (Caso 1 da Figura 27) deve ser tratado com uma abordagem física, ou seja, depende de medição específica no enrolamento do rotor. Já para análise de ruído (Caso 4 da Figura 27), uma abordagem com análise avançada dos dados pode ser mais vantajosa.

(WELTE et al., 2018) comenta as possíveis abordagens para análise, do nível simples ao avançado: visualização dos dados no tempo, modelos simples (e.g. duração de partida e parada de sequências, abertura de válvulas, etc.), modelos estatísticos (e.g. correlações, tendências, etc), modelos avançados (e.g. análise estatística avançada, análise em frequência, aprendizado de máquina, etc.) e por fim, o gêmeo digital⁷.

Figura 27 – Relação entre eventos de manutenção, complexidade e abordagem.

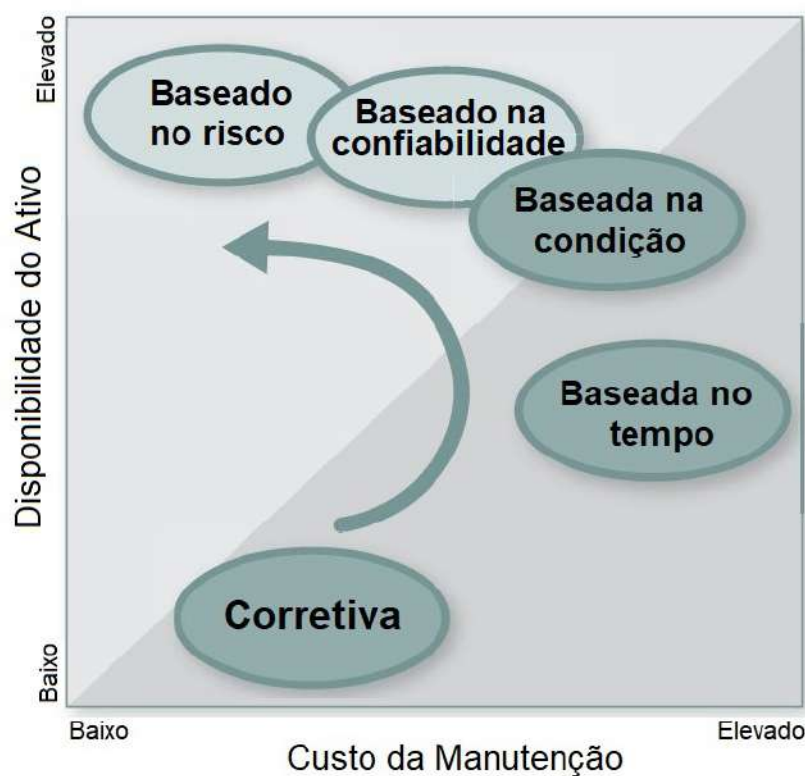


Fonte: (WELTE et al., 2018).

Historicamente, os avanços tecnológicos observados nos centros acadêmicos e nos departamentos de desenvolvimento e inovação de alguns fabricantes e fornecedores buscam oferecer equipamentos cada vez mais sofisticados, com objetivo de disponibilizar ferramentas para orientar a manutenção baseada na condição e elevar o índice de disponibilidade do ativo. A estratégia de manutenção está diretamente relacionada com o custo despendido para manter o ativo em operação e com nível de disponibilidade desejado para o mesmo. Isso pode ser observado na Figura 28. Em geral, busca-se o quadrante superior, onde a disponibilidade é maior.

⁷ Um gêmeo digital é a geração ou coleta de dados digitais que representam um objeto físico.

Figura 28 – Estratégias de manutenção em função da disponibilidade e custo.



Fonte: (KURT et al., 2011, *online*).

2.3.1 Indicadores e estatísticas para O&M

A terceira revolução industrial está relacionada com a utilização da eletrônica e computação para automatização de processos, foi iniciada em meados de século XX e está presente até hoje. As soluções obtidas nesta era foram responsáveis por modernizar e agilizar o processo fabril, mas com limitações para armazenar os dados de operação e geração de indicadores de desempenho em tempo real.

Para se obter os indicadores de operação de um determinado ativo, geralmente é necessário o envolvimento da equipe de O&M, acesso às informações de operação e ordens de serviço. São poucos casos que um sistema ou método estabelecido gere tais indicadores ou relatórios de modo automático. Uma das promessas da quarta revolução é a entrega de tais métricas à equipe de O&M, de modo automático e em tempo real.

A proposta de monitoramento contínuo e de modo autônomo dos indicadores operacionais é de grande importância para as plantas industriais, especialmente aquelas com elevado número de ativos e com sistemas industriais de média ou alta complexidade.

Conforme (KURT et al., 2011, *online*) e (TORELL; AVELAR, 2004), alguns indicadores utilizados pelos times de O&M são: KPI (*key performance indicator*) ou OEE (*overall equipment effectiveness*), LCC (*life-cycle cost*), MTBF (*mean time between failures*), MTTR (*mean time to repair*), disponibilidade, confiabilidade, custo de manutenção/faturamento, *backlog*, custo de

manutenção/custo do ativo novo estimado, custo de manutenção/unidade produzida, hora-homem empregado por tipo de manutenção, entre outros.

Para o cálculo do KPI ou OEE, algumas literaturas apresentam este indicador como o produto de três elementos: disponibilidade, desempenho e qualidade. O artigo “*Key Technical Performance Indicators for Power Plants*”, de (OPREA; BÂRA, 2017), apresenta uma abordagem diferente para plantas de geração.

As métricas mencionadas acima são genéricas e podem ser aplicadas praticamente em qualquer equipamento. Em ativos de geração, pode-se ainda acompanhar alguns indicadores específicos, como por exemplo: ranking de falhas que causaram desligamento repentino da rede, ranking de falhas que causaram indisponibilidade não programada, ranking de falhas que causaram indisponibilidade programada, número de sincronismos realizados/números de tentativa de sincronismo, tempo de partida e sincronismo do gerador na rede, número de desligamentos normais, número de desligamentos por falha elétrica, mecânica ou hidráulica (reles de bloqueio 86E, 86M e 86H⁸ respectivamente), contagem e características dos desligamentos por curto-circuito, entre outros parâmetros e métricas.

É importante mencionar que os sistemas industriais existentes, na maioria dos casos, precisam sofrer alguns ajustes ou aprimoramentos para viabilizar a implantação de uma solução IIoT.

2.4 IIOT E INDÚSTRIA 4.0

A academia alemã Acatech (SCHUH et al., 2017) apresenta a solução IIoT de modo mais amplo, com seis estágios de desenvolvimento, conforme exposto na Figura 29. O primeiro é a computação e automatização dos processos, o segundo é a conectividade entre os elementos, o terceiro a visibilidade dos processos em termos de dados e relatórios, o quarto é a compreensão e entendimento aprofundado dos motivos que levam o processo a operar do modo que está, o quinto é a capacidade de identificar desvios operacionais incipientes ou prever um determinado cenário ou ocorrência. Por último, a capacidade do sistema de se ajustar, com objetivo de atender níveis mais elevados de resultado.

Baseado na Figura 29 cria-se a Figura 30, com os três níveis básicos de uma solução IIoT. A base é a aquisição dos dados, armazenamento e visualização simples. Os projetos de IIoT devem focar inicialmente nos sensores e dados digitais disponíveis (em verde), para minimizar o investimento inicial e acelerar a implantação do sistema. Durante a vida da solução, o modo de aquisição pode ser aprimorado e novos sensores podem ser adicionados (em violeta). A aquisição pode ser realizada pelo sistema de automação existente e/ou com sistemas dedicados, isso depende da aplicação e da planta.

O segundo nível pode ser chamado de solução *Data-Informed*, conforme visto em algumas literaturas. Neste nível os dados são explorados de maneira específica, com visualização

⁸ Códigos da tabela ANSI para sistema de proteção.

Figura 29 – Estágios no desenvolvimento das soluções na Indústria 4.0.

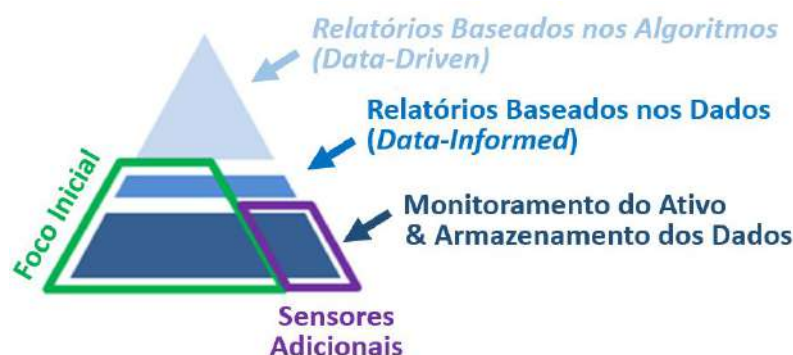


Fonte: Adaptado de (SCHUH et al., 2017).

customizada, descrevendo de maneira apropriada a operação do ativo e auxiliando no processo de *troubleshooting*. Nesta etapa, a solução está isenta de elementos avançados, mas viabiliza a identificação dos primeiros *insights* via análise humana, por meio dos relatórios baseados nos dados.

No terceiro nível, a solução *Data-Driven*, o conhecimento especialista humano e as experiências anteriores, oriundas da atividade de *troubleshooting* são utilizadas para criação de um sistema avançado, com algoritmos de aprendizado de máquina. Tal sistema pode ser utilizado para identificar de modo autônomo algumas ocorrências incipientes conhecidas e, em alguns casos, ocorrências inéditas também. A saída destes algoritmos podem ser utilizada para compor um relatório específico ao time de O&M, com foco na manutenção preditiva.

Figura 30 – Níveis básicos para uma solução IIoT.



Fonte: Adaptado de (SCHUH et al., 2017).

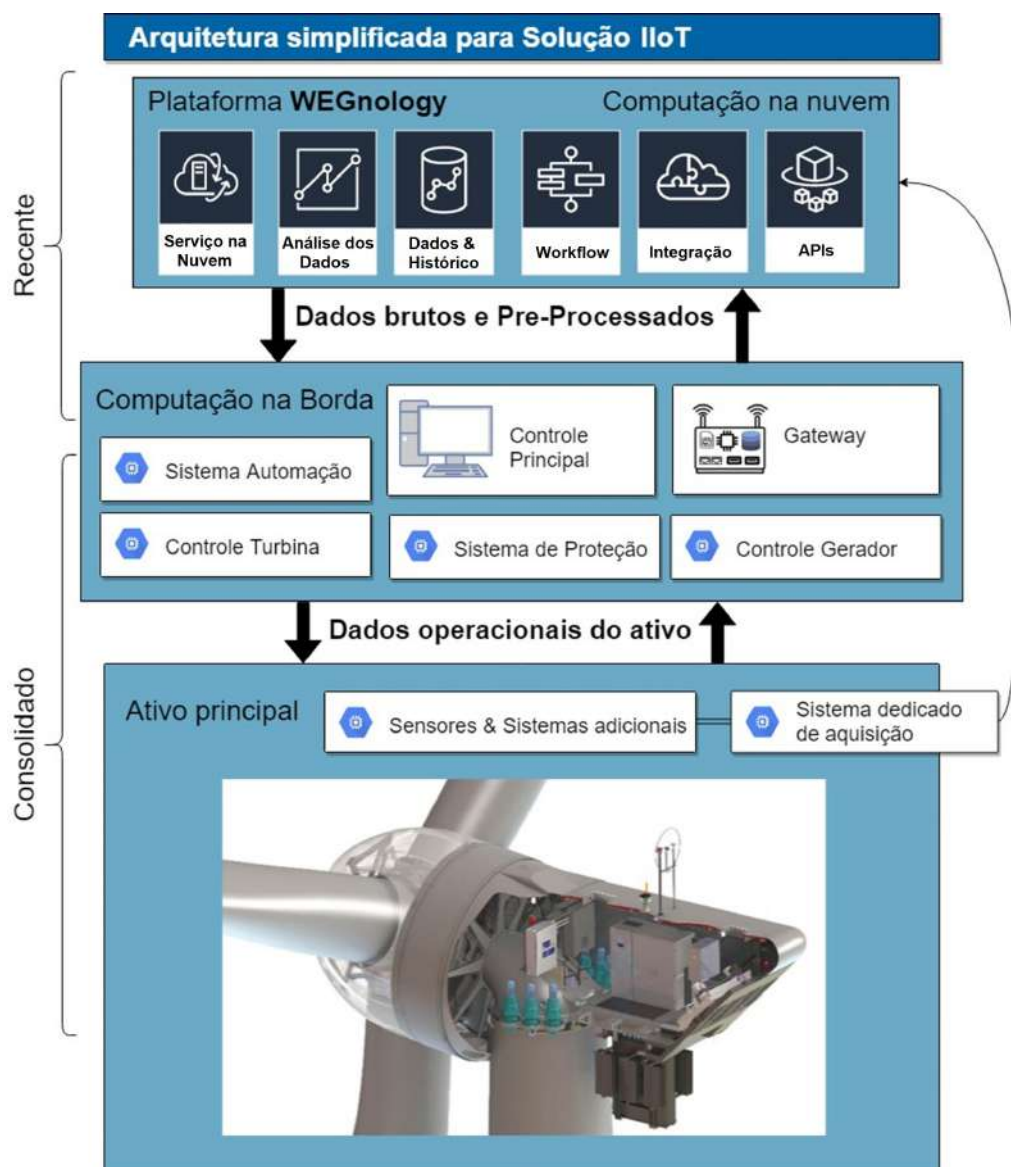
A Figura 31 apresenta a arquitetura simplificada de uma solução IIoT em um aerogerador. Em aplicações tradicionais, os dispositivos de borda são responsáveis pela operação do sistema, são eles: controle do gerador, controle da turbina, sistema de automação, sistema de proteção e o

sistema de controle geral. Para um sistema industrial se conectar a uma plataforma IIoT (*cloud*) é necessário acrescentar um elemento novo, de interface, o *gateway*.

Os sensores, os subsistemas e o *gateway* desempenham um papel crucial dentro da solução IIoT, eles são responsáveis por adquirir os dados do ativo, criar o *payload*, armazenar os dados durante o período *off-line* e enviá-los para plataforma quando conectados. Em sistemas mais sofisticados, o *gateway* também desempenha a função de computação na borda (*edge computing*).

A aquisição de dados pode ser realizada, basicamente, em três níveis, conforme mostra a Tabela 5. Para os dois casos abordados neste estudo, dados foram adquiridos com valores médios e amostragem a cada 10 minutos, ou seja, baixa frequência.

Figura 31 – Arquitetura simplificada de uma solução IIoT.



Fonte: elaborado pelo autor, 2020.

Tabela 5 – Níveis de aquisição de dados para ativos industriais.

Amostragem	Dado	Custo estimado	Dispositivo de aquisição	Análise
Baixa Frequência [<10Hz]	Valor RMS	\$	Gateway simples	Análise em regime. (horas e dias)
Média Frequência [10Hz <1kHz]	Valor RMS	\$\$\$	Gateway especial ou hardware dedicado	Análise de transitório. (fração de seg.)
Alta Frequência [>1kHz]	Forma de onda	\$\$\$\$\$	Hardware dedicado	Análise de transitório e de espectro.

Fonte: elaborado pelo autor, 2020.

Em ativos críticos e de valor agregado elevado, sistemas dedicados de monitoramento podem ser necessários para aquisição de dados específicos, com taxa de aquisição em média ou alta frequência. Neste caso, sensores e sistemas de aquisição especiais são instalados de modo complementar, conforme apresentado na Figura 31. São inúmeros os equipamentos que podem ser adicionados para monitorar uma máquina elétrica girante, como por exemplo: registrador digital das grandezas elétricas, vibração da cabeça de bobina, vibração do mancal, temperatura do estator/rotor, deslocamento do eixo, descargas parciais, fluxo, isolamento do estator, isolamento do rotor, tensão e corrente de eixo, medição do entreferro, etc.

Os detalhes sobre a aquisição dos dados como formato, amostragem, sensores, métodos de pré-processamento e análise avançada são cruciais e deverão ser selecionados conforme o nível desejado de solução IIoT e *troubleshooting*.

2.5 EXPLORAÇÃO COMERCIAL E PATENTES

Na criação de uma solução IIoT, inúmeras atividades de desenvolvimento de hardware e software são necessárias. Na perspectiva do hardware, há a criação e instalação de novos sensores, sistemas de aquisição, *gateways*, etc. Em alguns casos, para minimizar o investimento, o hardware existente recebe upgrade para enviar os *payloads* à plataforma IIoT escolhida.

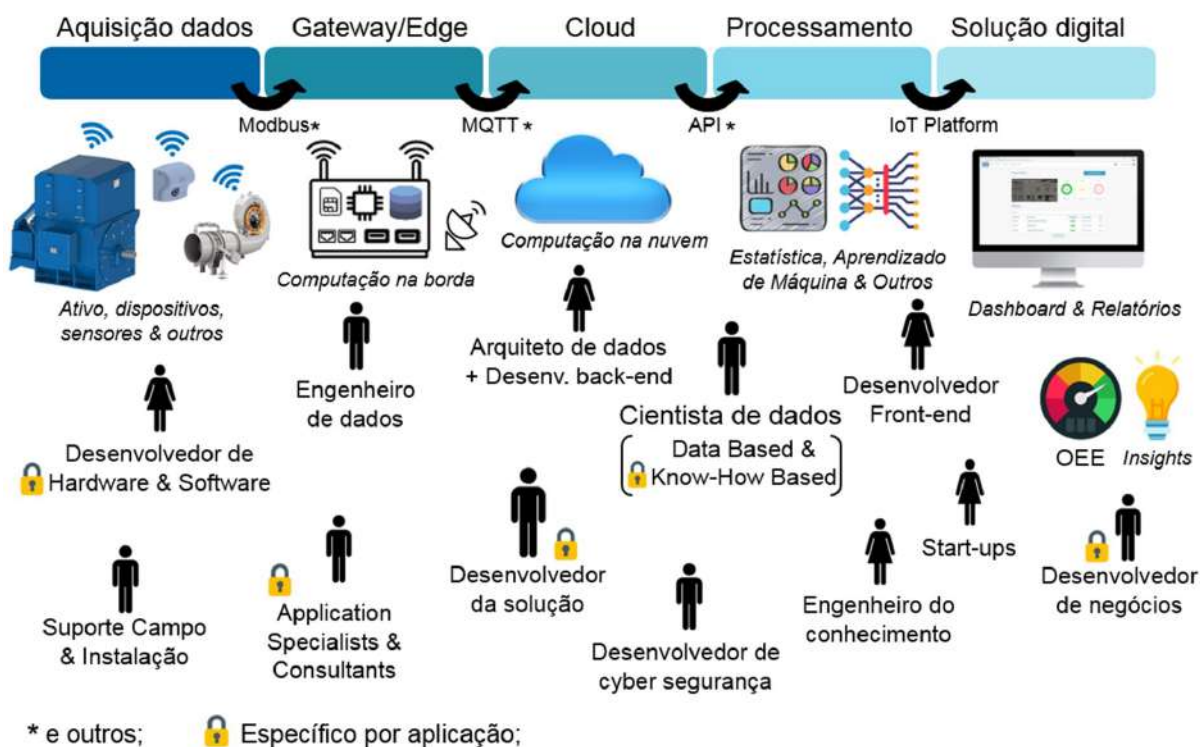
Neste tipo de solução, os desafios de software são variados. Esta área requer múltiplas habilidades profissionais como: arquiteto de dados, desenvolvedor de *back-end* e *front-end*, analista de *cyber* segurança, analista do conhecimento, analista de dados, cientista de dados, etc.

A Figura 32 mostra alguns personagens necessários para o desenvolvimento de uma solução IIoT, alguns são itinerantes entre os segmentos de aplicação e outros são específicos. Por exemplo, o desenvolvedor de *hardware* para aquisição de dados de aerogeradores dificilmente possui o conhecimento necessário para realizar tal tarefa em uma turbina hidráulica. No caso do desenvolvedor de *back-end*, este poderá atuar nas duas aplicações, uma vez que a especialidade deste profissional independe da aplicação final.

A aquisição dos dados deve observar os requisitos do algoritmo de análise. O elemento responsável pela aquisição dos dados é o *gateway*, sendo importante que o mesmo esteja apto

para atender os requisitos de *cyber* segurança, sincronização dos dados, rápido processamento de dados e a habilidade de realizar atualização remotamente (FOTA, *firmware over-the-air*).

Figura 32 – Recursos necessários em uma solução IIoT.



Fonte: elaborado pelo autor, 2020.

Segundo (KALABIN et al., 2018), as variedades de soluções em manutenção preditiva estão em rápido crescimento e os elementos principais de um sistema de monitoramento com aprendizado de máquina são:

- *Servidor*, local ou remoto/*cloud*, para armazenar os dados brutos e poder computacional para processar os algoritmos;
- *Software* com aprendizado de máquina (algoritmo) e interface gráfica para interação com os dados sob análise;
- *Dataset* ou *database*, trata-se do histórico de dados, disponível para o processamento do algoritmo;
- Integrador de dados, necessários para que o *dataset* disponível seja adequado e utilizado pelo algoritmo;

A pesquisa realizada por (KALABIN et al., 2018) avaliou 10 empresas com solução digital disponível no mercado: Maintpartner INtelligence®, NEC SIAT, Avantis® PRiSM, Uptake™, Siemens (Plant Monitor e MindSphere), GE (SmartSignal e Predix), ABB Ability™

(IBM Watson e MS Azure), C3 IIoT Platform™, Seeq® e SAP. A comparação geral de recursos é apresentada na Tabela 6.

Tabela 6 – Comparação geral dos recursos nas soluções de mercado.

	1. INtelligence	2. SIAT	3. PRISM	4. Uptake	5. Siemens	6. GE	7. ABB	8. C3 IoT	9. Seeq	10. SAP
A family of products available	N	N	Y	N	Y	Y	Y	N	N	Y
Several algorithms employed	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
Deep learning ANN	N	N	N	N	Y	N	Y ³⁰	N	N	Y ³¹
Lifecycle estimation (Asset health)	N	N	N	Y	N	Y	Y	N	N	Y
Tablet/phone application	N ³²	N	Y	N	Y ³³	Y ³⁴	N	N	N	Y
Custom applications (Open API)	N	N	Y	N	Y	Y	N ³⁵	Y	Y	Y
Additional security solution	N	Y	N	Y	N	N	Y ³⁶	Y	N	Y
Optimization option	Y	N	Y	N	Y	Y	Y ³⁷	N	N	N ³⁸

Table 1 Comparable feature overview for the solutions analyzed

³⁰ at the moment, the direct application for PP process analysis is undisclosed

³¹ application of ANN is likely to be tangentially related to PP processes

³² web-based application is planned for release in the future

³³ third party, access to the predictive analysis is unclear

³⁴ access only to process data, not the predictive analysis [116]

³⁵ mentioned for other product, unrelated to PPs

³⁶ exists as an additional network intrusion monitoring option for Symphony Plus

³⁷ more of a smart grid-oriented solution rather than PP performance optimizing one

³⁸ existing optimization solutions seem to be unrelated to PPs directly [117]

Fonte: (KALABIN et al., 2018).

No estudo de (KALABIN et al., 2018) é apresentada a duração estimada para as etapas de implantação de uma solução *Data-Driven*. De modo geral, gasta-se aproximadamente 1/3 do tempo na análise exploratória e preparação dos dados e outro 1/3 na criação, treino, teste e ajuste fino do modelo. O tempo restante é utilizado na etapa inicial de definição do problema e na etapa final (*deploy*). A Tabela 7 apresenta um resumo dos tempos.

No estudo de (KALABIN et al., 2018) são citadas algumas patentes no segmento de solução digital e sistema com aprendizado, voltado para a manutenção preditiva. As 13 patentes mencionadas são apresentadas na Tabela 8 e foram utilizadas para uma pesquisa de patentes. Para isso foi utilizado o software Orbit Intelligence (Questel), uma ferramenta de busca e análise de patentes. A busca realizada por similaridade encontrou 1623 registros e foi feita em setembro de 2020.

Como o período de sigilo de um novo depósito é de 18 meses, é sabido que a análise realizada revela adequadamente o cenário até a data de fevereiro de 2019. As principais conclusões notadas na análise foram:

Tabela 7 – Etapas e duração estimada para implantação da solução *Data-Driven*.

Atividade	Duração estimada
1. Definição do problema e avaliação inicial do sistema/dados	2...4 semanas
2. Conexão e preparação dos dados	2...4 semanas
3. Interpretação, limpeza e preparação dos dados	4...8 semanas
4. Criação, treino e teste do modelo (contato ativo com cliente)	4...8 semanas
5. Validação, <i>tuning</i> e monitoramento do modelo no cliente	4...8 semanas
6. Ativação da solução <i>Data-Driven</i> e serviço SaaS (<i>software as a service</i>)	2...4 semanas

Fonte: adaptado de (KALABIN et al., 2018).

- Predomínio de depósito nos Estados Unidos, seguido pela China, Europa, Japão, Canada, WO⁹, Alemanha, etc.
- A empresa GE demonstra pioneirismo no segmento, com sólido domínio tecnológico, 47 depósitos concedidos, 19 pendentes e 22 inativos. Cita poucas empresas e é comumente citada pelas demais;
- A empresa IBM também se destaca no segmento com 32 depósitos concedidos, 16 pendentes e 10 inativos;
- As empresas Hitachi, ABB, Siemens se destacam como fabricantes de máquinas elétricas, além da GE obviamente.
- No cenário avaliado, 78.3% pertence ao grupo G, relacionado com física, 11.8% ao grupo H, da eletricidade, 8.8% ao grupo Y, das novas tecnologias em criação e 1,1% ao grupo A, das necessidades humanas;
- Primeiro depósito observado em 1991;
- Entre 1991 e 2006 a taxa média de depósitos foi de 22 depósitos/ano.
- Entre 2007 e 2015 de 73 depósitos/ano;
- Nota-se um aumento significativo nos depósitos a partir de 2016, 144 depósitos/ano;
- Até o momento, o auge de depósito aconteceu em 2018, com 168 aplicações.

O tema propriedade intelectual é naturalmente amplo, os temas relacionados com a proteção no segmento de IIoT e sistemas com aprendizado de máquina é recente. O objetivo deste tópico busca apresentar uma introdução ao tema, sem se aprofundar. Os principais códigos IPC encontrados na pesquisa são apresentados na Tabela 9.

⁹ Termo simplificado de WIPO, *World Intellectual Property Organization*, Organização Mundial da Propriedade Intelectual, www.wipo.int

Tabela 8 – Patentes na área de solução digital e análise avançada.

Patente	Empresa	Ano	Estado Legal	Título
US9379951	Avantis	2014	Ativo	Method and apparatus for detection of anomalies in integrated parameter systems
US20180060703	Uptake	2016	Ativo	Detection of Anomalies in Multivariate Data
US20160371584	Uptake	2015	Ativo	Local Analytics at an Asset
US7509235	GE	2005	Ativo	Method and system for forecasting reliability of assets
US9665843	ABB	2013	Ativo	Industrial asset health profile
US20140365271	ABB	2013	Ativo	Industrial asset health model update
US20140156225	ABB	2013	Ativo	Assessment of power system equipment for equipment maintenance and/or risk mitigation
US20170006135	C3ai Inc	2016	Pendente	Systems, methods, and devices for an enterprise internet-of-things application development platform
US20160238640	C3ai Inc	2015	Ativo	Systems and methods for regression-based determination of expected energy consumption and efficient energy consumption
US20160359895	C3ai Inc	215	Ativo	Systems and methods for providing cybersecurity analysis based on operational technologies and information technologies
US20140201115	SAP	2013	Abandonada	Determining software object relations using neural networks
US20160062876	SAP	2014	Abandonada	Automated software change monitoring and regression analysis
EP3029614	SAP	2014	Ativo	Parallel development and deployment for machine learning models

Fonte: adaptado de (KALABIN et al., 2018).

Na pesquisa realizada até o momento, não ficou claro como as empresas mencionadas exploram comercialmente as soluções envolvendo algoritmos de inteligência, como pretendem monetizar e tornar o sistema autossustentável, tão pouco como é o modelo de negócio adotado.

Na Figura 33 a Gartner¹⁰ mostra os principais *players* em plataformas IIoT. Como destaque, pode-se apontar as empresas Hitachi, PTC e Microsoft que são os únicos *players* a figurar no quadrante de líder.

Os *players* de mercado estão investindo no segmento IIoT e sistemas com inteligência. A breve pesquisa realizada corrobora com esta afirmação, o segmento está em crescimento.

¹⁰ Gartner Group, famosa empresa de consultoria americana fundada em 1979.

Tabela 9 – Principais códigos IPC da pesquisa realizada.

Código IPC	Descrição	Registros na pesquisa
G06Q	<i>Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for.</i>	828
G06F	<i>Electric digital data processing (computer systems based on specific computational models G06N).</i>	670
G05B	<i>Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements.</i>	266
H04L	<i>Transmission of digital information.</i>	251
G06N	<i>Computer systems based on specific computational models.</i>	226
G16H	<i>Healthcare informatics, i.e. Information and communication technology [ict] specially adapted for the handling or processing of medical or healthcare data.</i>	164
Y02P	<i>Climate change mitigation technologies in the production or processing of goods.</i>	113

Fonte: (WIPO, 2020, online).

Figura 33 – Quadrantes mágicos Gartner para Plataformas IIoT.



Fonte: (GARTNER, 2020, online).

2.6 CONSIDERAÇÕES

Este Capítulo mostra a importância do conhecimento especialista, tanto do OEM como do O&M, e o quão vasto e aprofundado o conteúdo sobre um determinado ativo pode ser.

Foram apresentados detalhes do sistema de vedação (gaxeta) do eixo do hidrogenador e do sistema de arrefecimento do aerogerador (no Anexo B) para melhor contextualização do Caso 1 (Seção 5.1) e Caso 2 (Seção 5.2), respectivamente.

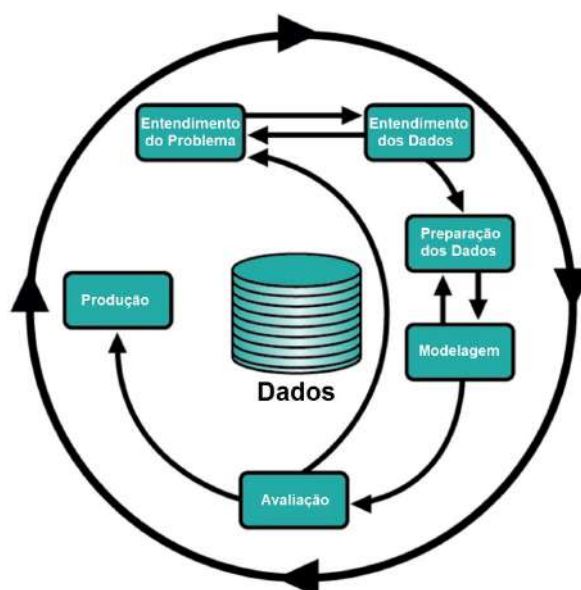
O Capítulo 3 introduz o tema ciência de dados, apresenta uma breve análise exploratória sobre os dados reais de operação do hidrogenador e mostra como a rede neural TLFN focada é configurada para aplicação neste estudo.

3 CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA

A metodologia CRISP-DM (WIRTH; HIPPI, 2000) é comumente citada nos trabalhos de ciência de dados. De acordo com (CHAPMAN et al., 1999), esta metodologia é composta por 6 fases, organizadas de maneira cíclica e com fluxo direcional. O ciclo inicia no entendimento do negócio ou problema (*Business Understanding*), para isso é necessário realizar o entendimento dos dados (*Data Understanding*), buscando identificar se os dados disponíveis são capazes de descrever o problema.

Na sequência, os dados são preparados (*Data Preparation*) para etapa de modelagem (*Modeling*). Como o sistema pode contar com mais de um algoritmo, a preparação dos dados deve observar os requisitos de cada um. Além disso, os dados deverão ser segregados em pacotes para o treino, teste e avaliação. Após treinar os algoritmos de aprendizado, é realizada a etapa de avaliação (*Evaluation*). Se for comprovado o bom funcionamento, o mesmo poderá ser implementado na solução, ou seja, colocado em produção (*Deployment*).

Figura 34 – CRISP-DM.



Fonte: (WIRTH; HIPPI, 2000).

Para uma abordagem com ativo industrial, é crucial contar com envolvimento de um especialista do ativo, principalmente nas etapas iniciais do processo CRISP-DM, onde os ganhos e valores da solução proposta devem estar bem definidos.

O conceito CRISP-DM apresentado na Figura 34 é utilizado para criar o *workflow* proposto neste estudo, que é apresentado na Figura 69, no Capítulo 4.

3.1 SÉRIE TEMPORAL

Para (HAYKIN, 2007), o tempo constitui um ingrediente essencial no processo de aprendizagem, que pode ser contínuo ou discreto. Independente da sua forma, o tempo é uma entidade ordenada que é básica para muitas tarefas cognitivas encontradas na prática. É por meio da incorporação do tempo na operação de uma rede neural que ela é capacitada a seguir as variações estatísticas em processos não-estacionários, como em sinais provenientes de um determinado ativo.

Ainda conforme (HAYKIN, 2007), para incorporar o elemento temporal em uma rede neural há duas possibilidades: a representação implícita, onde a estrutura temporal do sinal é inserida na estrutura espacial da rede; e a representação explícita, onde a estrutura temporal do sinal é inserida na camada de entrada da rede.

Conforme (FAWAZ et al., 2019), nas últimas duas décadas, a classificação de séries temporais tem sido considerado um dos maiores desafios na área de ciência de dados. Com aumento da disponibilidade de dados temporais, centenas de modelos foram propostos desde 2015. Devido à ordenação natural dos registros pelo tempo, dados de séries temporais estão presentes em muitas tarefas que requerem algum tipo de processo cognitivo humano. Segundo (BROWNLEE, 2020, *online*), muitos modelos numéricos clássicos abordam a extração de características das séries temporais para determinar um padrão e gerar uma predição, tanto para dados univariados como para dados multivariados.

3.1.1 Univariado e multivariado

A exploração de dados univariados é muito comum no mercado de ações. Os fatores que influenciam o comportamento de uma ação são inúmeros e muitas vezes são desconhecidos. Em alguns casos, o comportamento da própria ação (série univariada) é utilizado para extrair suas características e criar um modelo.

Segundo (BROWNLEE, 2020, *online*), algumas técnicas que podem ser empregadas em série temporal univariada são: autorregressão (AR), média móvel (MA), média móvel autorregressiva (ARMA), média móvel integrada autorregressiva (ARIMA), média móvel integrada autorregressiva sazonal (SARIMA), média móvel integrada autorregressiva sazonal com regressores exógenos (SARIMAX), suavização exponencial simples (SES) e suavização exponencial de *Holt Winter* (HWES). Obviamente que há outras técnicas para modelagem.

Dados multivariados são encontrados em muitas aplicações, alguns exemplos são: processos produtivos, automóveis, meteorologia, ativos industriais como motores elétricos, bombas, máquinas, etc.

Conforme (BROWNLEE, 2020, *online*), algumas técnicas que podem ser empregadas em série temporal multivariada são: auto regressão vetorial (VAR), *extreme gradient boosting* (XGBoost), rede neural recorrente *Long Short-Term Memory* (LSTM), entre outras. Segundo (HAYKIN, 2007), uma rede neural recorrente pode ter seus neurônios de entrada arranjados para receber

uma série temporal multivariada. Este tipo de modelo é chamado de TLFN focada (*focused time lagged feedforward network*), que será explorado na Subseção 3.6.5 e aplicado no Caso 2 (Seção 5.2). Semelhante ao caso univariado, as técnicas para dados multivariados apresentadas aqui representam uma fração das opções de modelagem disponíveis.

3.2 CORRELAÇÃO E CAUSALIDADE

Correlação não implica em causalidade. O fato de dois sinais possuírem elevado grau de correlação não significa que irão possuir alguma relação de causa e efeito (causalidade). Uma correlação é considerada forte quando seu coeficiente for acima de 0.8 ou inferior a 0.8 negativo, neste caso o resultado mostra que o comportamento estatístico entre duas variáveis possuem forte relação. Conforme observado em algumas referências, determinar o nível de causalidade entre variáveis não é trivial.

Para definir as variáveis utilizadas como entrada e saída no algoritmo proposto, é observado o modelo físico do sistema de arrefecimento (apresentado na Seção 2.2.1 e Anexo B), que naturalmente traz variáveis com grau relevante de causa e efeito. O modo como essas variáveis são arranjadas para uso no algoritmo, ou seja, no vetor de entrada, é apresentado na Subseção 3.6.6.

Para ciência de dados, os fatores de correlação são muito importantes e dominam os modelos de aprendizado de máquina. Na perspectiva da física, a causalidade é o fator chave que descreve o comportamento operacional de um ativo. Um bom projeto de aprendizado de máquina para ativos industriais deve combinar adequadamente a correlação e causalidade, do contrário, modelos fracos podem ser criados.

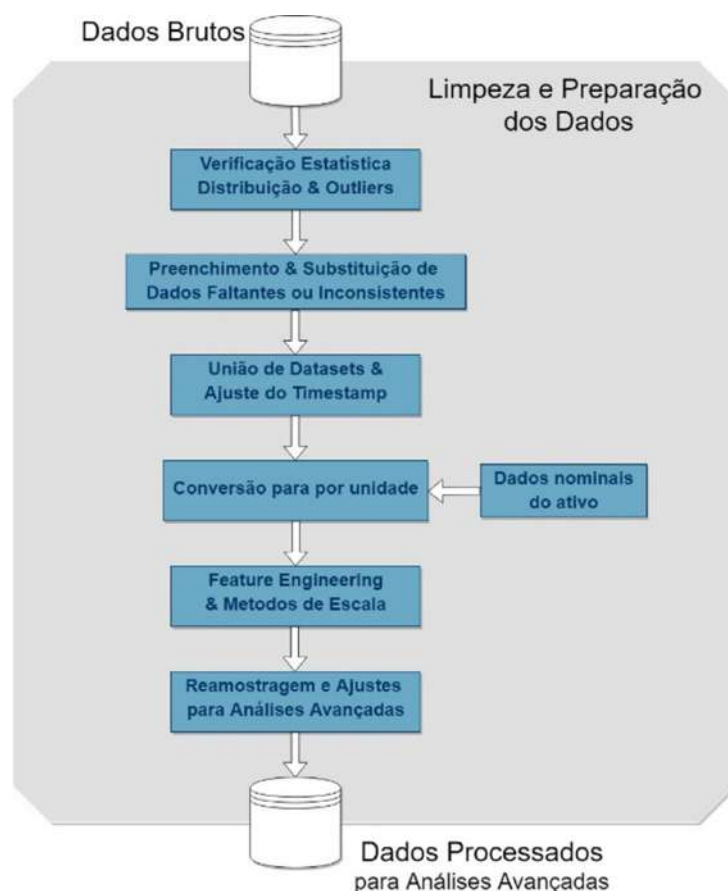
3.3 PREPARAÇÃO DOS DADOS

Antes de qualquer análise, o *dataset* deve passar por uma limpeza e preparação. A Figura 35 apresenta o *workflow* utilizado em tal atividade. Trata-se de uma etapa importante que antecede a criação dos algoritmos, especialmente em sistemas com elevado número de dados. Em alguns casos, esta etapa pode ser relativamente extensa.

Conforme apresentado na Figura 35, as principais atividades desta etapa são: verificação estatística como distribuição e *outliers*, o preenchimento ou remoção de dados faltantes ou inconsistentes, a união de *datasets* proveniente de diferentes fontes, o ajuste e ordenação do índice em modo *timestamp*, a remoção dos índices repetidos, a conversão dos valores para pu (por unidade, *per unit*), identificação do status operacional do ativo, a criação de variáveis virtuais (*feature engineering*) e, se necessário, o ajuste dos dados para um formato e escala adequado para os algoritmos de aprendizado. Por fim, observando o tipo de análise avançada e o período de interesse, o *dataset* é reamostrado e o tipo do dado é ajustado para tal.

Em alguns casos, sinais lógicos da planta podem ser utilizados para seleção dos períodos de maior interesse, como por exemplo, o sinal do disjuntor principal do gerador, que indica

Figura 35 – Exemplo de *workflow* para limpeza e preparação dos dados.



Fonte: elaborado pelo autor, 2020.

quando o gerador está sincronizado na rede e possui operação com carga.

Conforme (BROWNLIE, 2020, *online*) e demais referências, os modelos numéricos possuem uma série de particularidades e requisitos. Em geral, os modelos com rede neural devem receber como entrada valores entre 0 e 1.

Para conversão dos dados para pu, são utilizados os valores nominais da aplicação (Tabela 15, em Anexo B). Neste estudo, tal conversão é considerado a técnica de adequação de escala do *dataset*. Para as variáveis de temperatura e vibração foram utilizados os valores de 100°C e 2.5mm/s como valores de base, respectivamente. Esta abordagem é empregada para permitir o uso de um mesmo algoritmo em diferentes ativos de uma frota, o que não seria possível com os métodos *minmaxscaler*¹¹ e *standardscaler*¹², segundo os testes realizados.

Como um algoritmo de detecção de desvio operacional pode ser compartilhado entre máquinas, é importante que o processo de preparação dos dados obedeça ao mesmo padrão, independente da máquina sob análise. Após estabelecer o processo de limpeza/preparação dos dados, o mesmo pode ser colocado em um *pipeline* bem definido. Esta prática é importante para

¹¹ Método de pré-processamento do Sklearn para colocar os dados na mesma escala considerando o valor máximo e mínimo do *dataset* original, podendo resultar em um *dataset* de saída com valores entre 0 e 1, por exemplo.

¹² Método de pré-processamento do Sklearn para colocar os dados na mesma escala, removendo a média e escala a variância, podendo resultar em um *dataset* de saída com valores de média zero e desvio padrão unitário.

padronizar tal atividade e garantir que o algoritmo de aprendizado receba um pacote de dados adequado, independente da origem ou conteúdo numérico presente no *dataset*.

O *feature engineering* utiliza o conhecimento especialista do ativo e as habilidades do analista e cientista de dados para criar dados virtuais úteis a partir dos dados reais. Tanto os dados reais quanto os virtuais podem ser utilizados como entrada no algoritmo de aprendizado. Em muitos casos, sem uma boa preparação de dados e/ou um bom *feature engineering*, o resultado do algoritmo/modelo pode ser insignificante.

3.4 ORIGEM E REMOÇÃO DE *OUTLIERS*

Para estatística, *outlier* é uma observação que apresenta um grande afastamento das demais amostras da série ou que é inconsistente segundo um determinado critério. Tipicamente, a existência de *outliers* causa prejuízos para a interpretação do contexto apresentado pelos dados.

Após troca de experiências com o time de O&M WEG, nota-se que os principais motivos para geração de dados inconsistentes, *outliers* e dados faltantes em uma solução IIoT são:

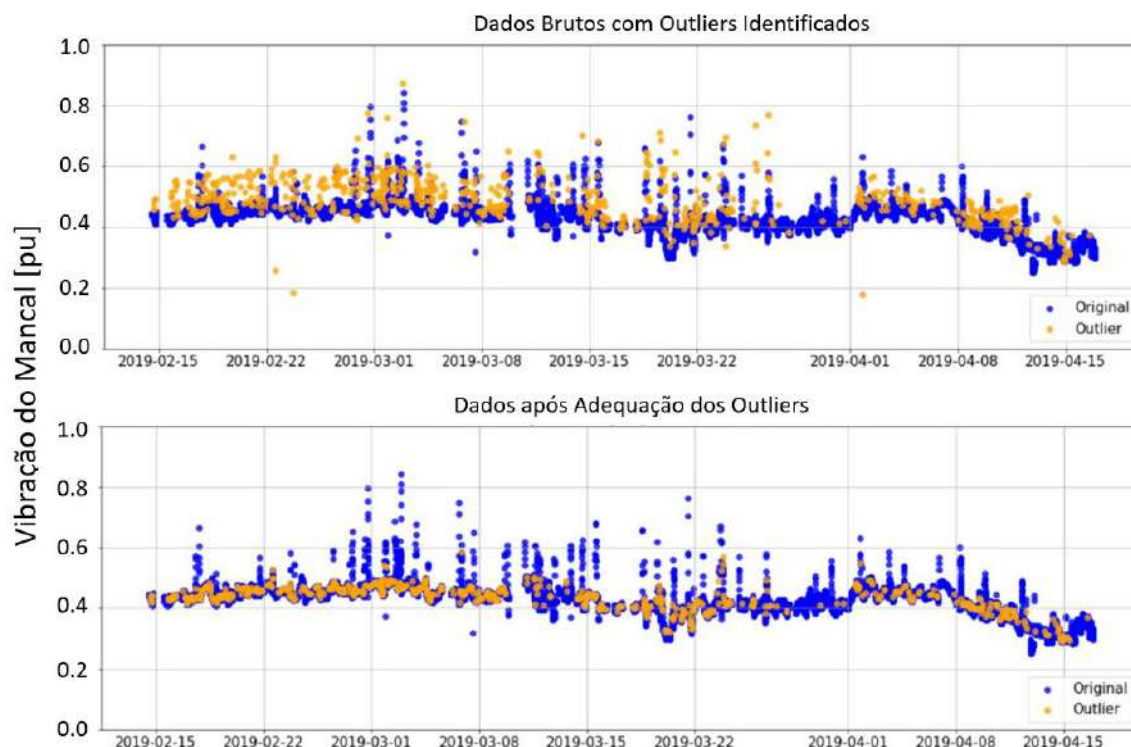
- Falhas de comunicação com *cloud* por tempo demasiado;
- *Gateway* sem armazenamento de dados para o período *off-line*;
- Inconsistência no armazenamento local dos dados no período *off-line*;
- Inconsistência no processo de envio dos dados ao *cloud* após um período *off-line*;
- Falha no sensor, cabeamento e sistema de aquisição;
- Detalhes de instalação, configuração e ajuste dos sensores;
- Falhas de comunicação temporárias entre o PLC e o SCADA;
- Intervenção humana no processo.

Para remoção dos *outliers*, foram utilizados dois métodos distintos. O primeiro e mais simples é o corte de registros para valores acima de 2pu e abaixo de -0.5pu. O corte está relacionado com a remoção completa do registro naquele determinado tempo (*timestamp*).

O segundo método utiliza uma janela deslizante com cálculo e avaliação do *z-score* sobre a série temporal univariada. Nesta abordagem, os parâmetros de ajuste do método são: o número de amostras para janela deslizante, número de *loops* sobre a série temporal, o tipo de operação sobre o *outlier* e o nível de corte (*z-score*). No exemplo abaixo, é adotado uma janela de 10 amostras, 10 *loops*, substituição do *outlier* com função de interpolação linear e corte com *z-score* de 2.3.

Os testes realizados no Caso 1 (Seção 5.1) utilizam o segundo método, já o Caso 2 (Seção 5.2) utiliza o primeiro método, com corte em 2pu e -0.5pu.

Figura 36 – Exemplo de identificação e adequação dos *outliers* com segundo método.



Fonte: elaborado pelo autor, 2020.

3.5 ANÁLISE EXPLORATÓRIA

A análise exploratória em contexto estatístico busca verificar o comportamento e distribuição dos dados de um modo geral, com objetivo de identificar as características da variável e a presença de *outliers*. Já a análise exploratória no contexto operacional, esta busca compreender os aspectos de O&M e *troubleshooting* presentes nos dados de um determinado ativo.

3.5.1 Contexto estatístico dos dados

A avaliação do contexto estatístico é importante para identificar a qualidade dos dados presentes no *dataset*. Neste processo é verificado a quantidade de dados inconsistentes (*outliers*) e de dados faltantes. Esta etapa não deve ser utilizada para remoção dos *outliers* ou preenchimento das lacunas de dados, mas para compreender a condição do *dataset* e indicar a melhor estratégia de limpeza e preparação dos dados. A análise do contexto estatístico possui dados em pu.

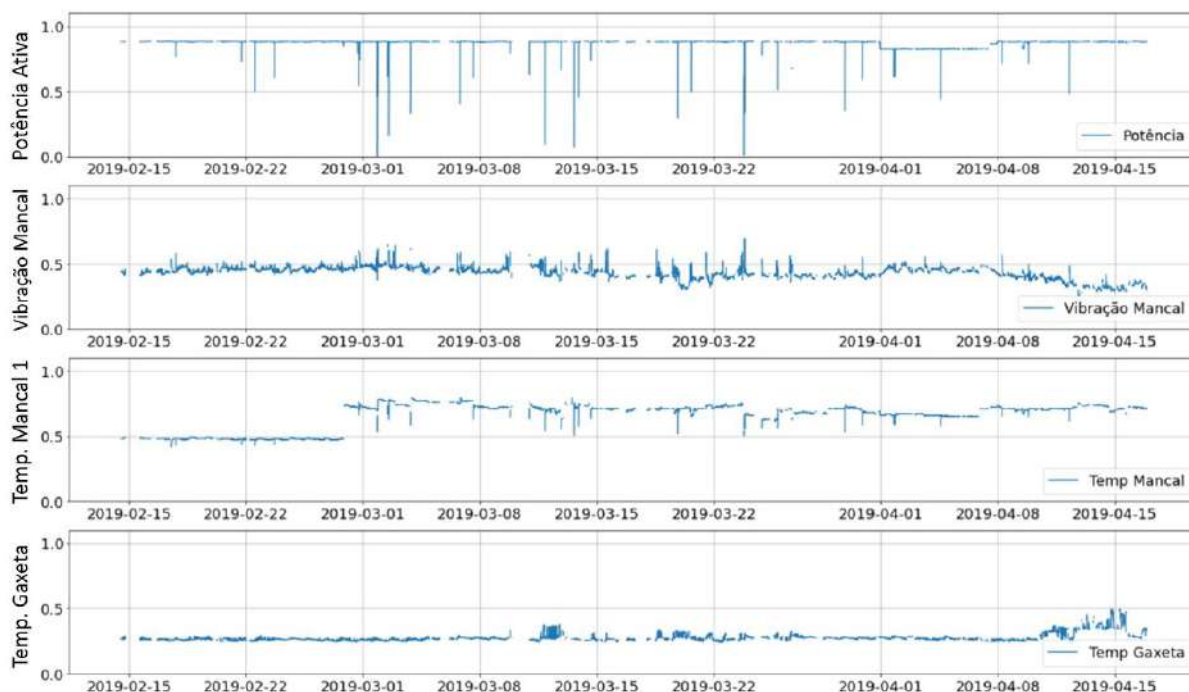
Para análise de dados tabulares genéricos, pode-se utilizar a biblioteca *pandas.profling*¹³, que gera um relatório de exploração dos dados no contexto estatístico. Até a data de publicação deste estudo não foi encontrada uma ferramenta semelhante para série temporal multivariada.

A primeira etapa é a visualização de cada variável no tempo, conforme exemplo apresentado na Figura 37. Em seguida, são verificados alguns indicadores estatísticos, como por

¹³ Método do Pandas para uma análise exploratória rápida. Com poucas linhas de código é possível visualizar e identificar a distribuição de cada variável.

exemplo: valores máximos, mínimos, média, desvio padrão, curtose e assimetria. Outros testes de hipóteses podem auxiliar neste processo.

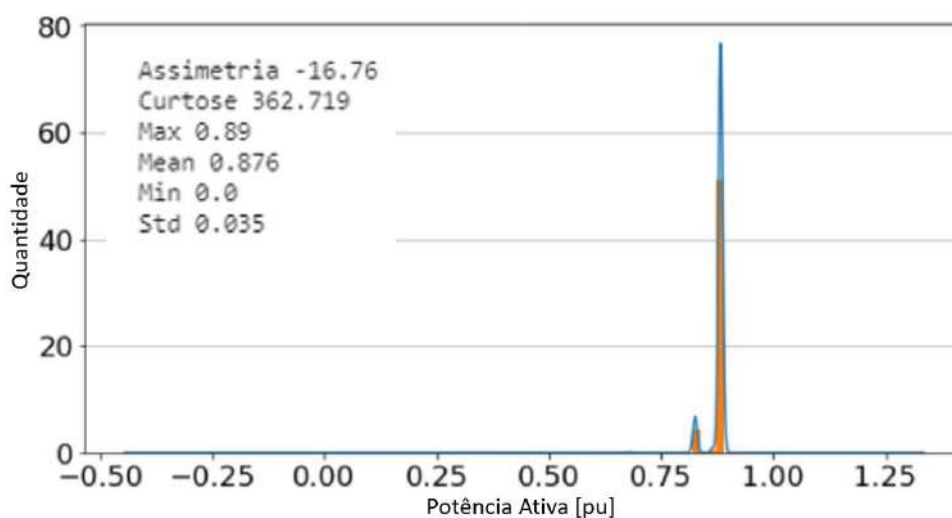
Figura 37 – Dados brutos de fevereiro a abril de 2019 - caso 1 hidrogerador.



Fonte: elaborado pelo autor, 2020.

A distribuição dos dados e a informação estatística para potência ativa do hidrogerador em estudo é apresentada na Figura 38. Este parâmetro apresenta um comportamento praticamente fixo em 0.88pu.

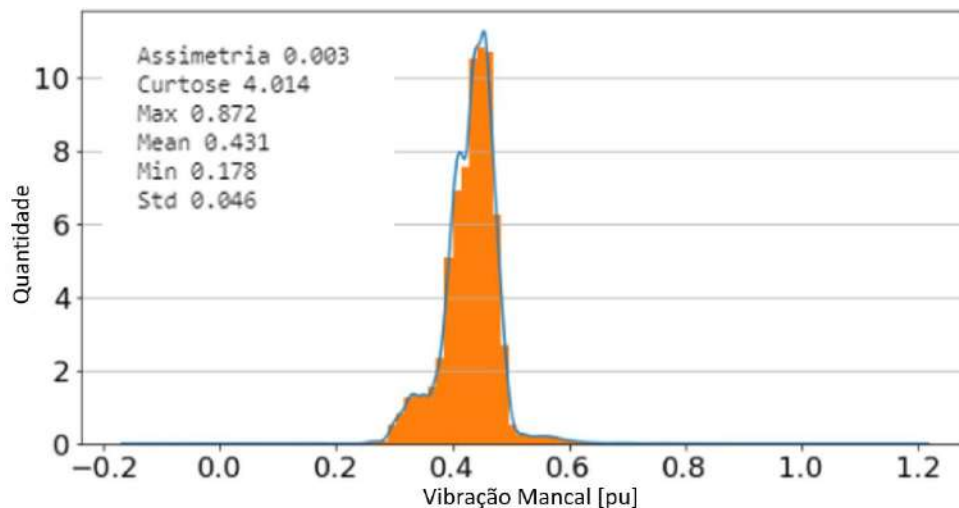
Figura 38 – Informação estatística e histograma da potência ativa.



Fonte: elaborado pelo autor, 2020.

A Figura 39 apresenta a distribuição dos dados de vibração do mancal, com distribuição próxima da normal. O valor médio deste parâmetro está em 0.43pu.

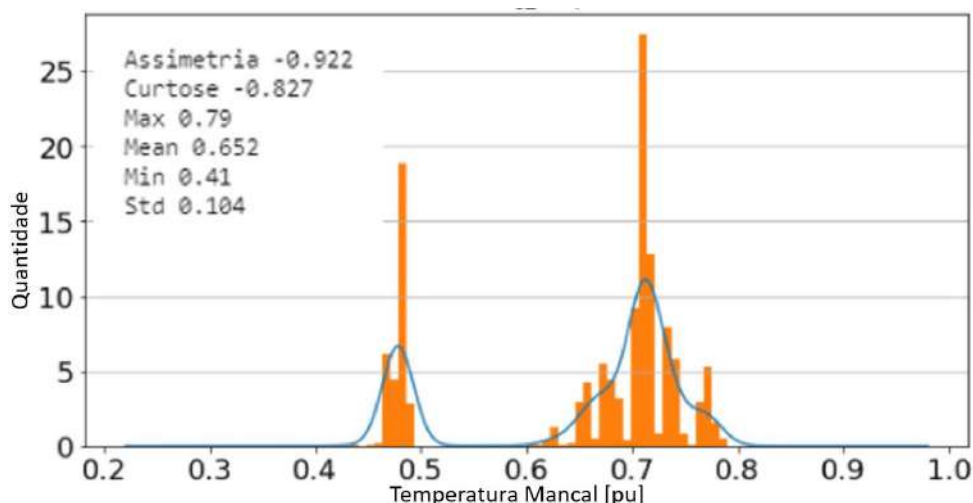
Figura 39 – Informação estatística e histograma da vibração do mancal.



Fonte: elaborado pelo autor, 2020.

Já na Figura 40 é apresentada a distribuição dos dados de temperatura do mancal. Nota-se que tanto na visualização em série temporal como na distribuição é possível observar dois patamares, um deles em 0.48pu e outro em 0.7pu. Tudo indica que o sensor estava instalado de maneira inadequada e operou deste modo até o final de fevereiro. Após a correção, o sinal apresentou média de aproximadamente 0.71pu.

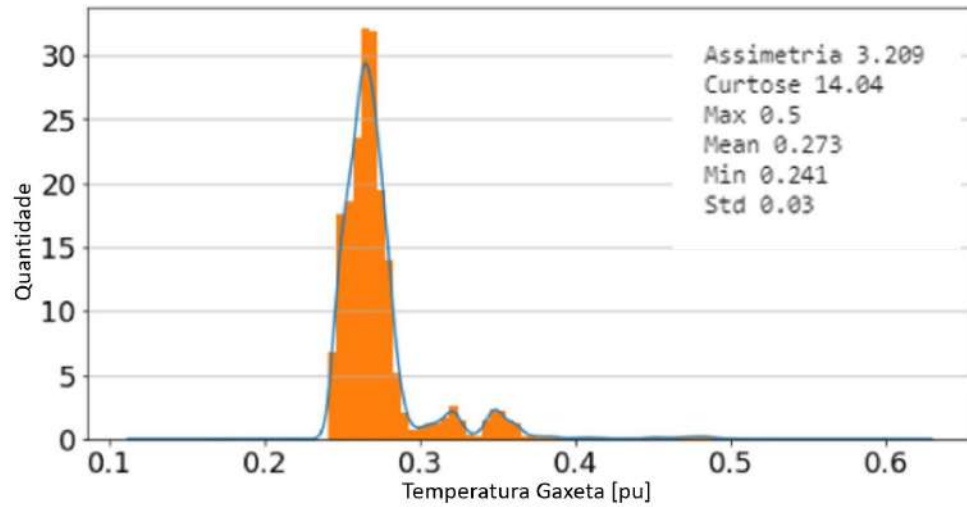
Figura 40 – Informação estatística e histograma da temperatura mancal.



Fonte: elaborado pelo autor, 2020.

A distribuição dos dados de temperatura da gaxeta é apresentada na Figura 41. Observa-se que o comportamento normal deste parâmetro fica em torno de 0.27pu (27°C). Analisando a série temporal e a distribuição, é possível notar que em alguns casos a temperatura da gaxeta sai do padrão, com operação acima de 0.3pu.

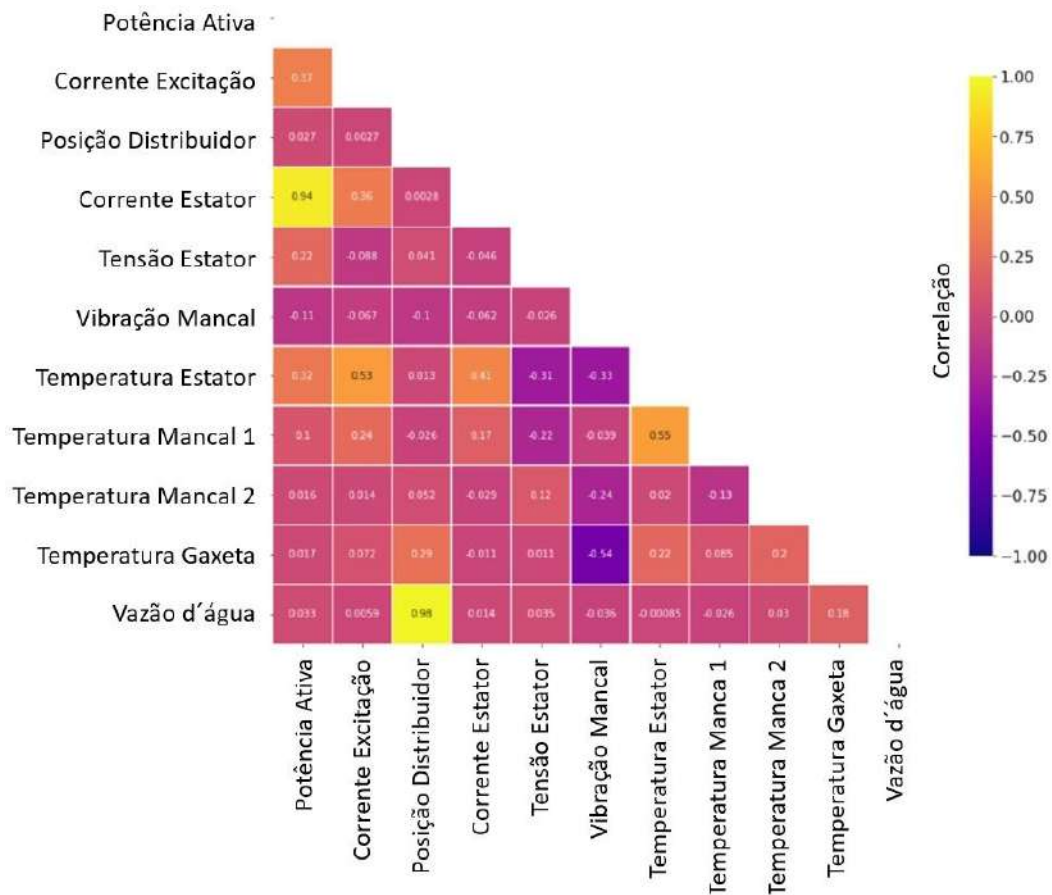
Figura 41 – Informação estatística e histograma da temperatura da gaxeta.



Fonte: elaborado pelo autor, 2020.

A Figura 42 apresenta o gráfico de correlação entre um grupo de variáveis do hidrogerador, para um período de operação de 60 dias.

Figura 42 – Mapa de correlação entre um grupo de variáveis do hidrogerador.

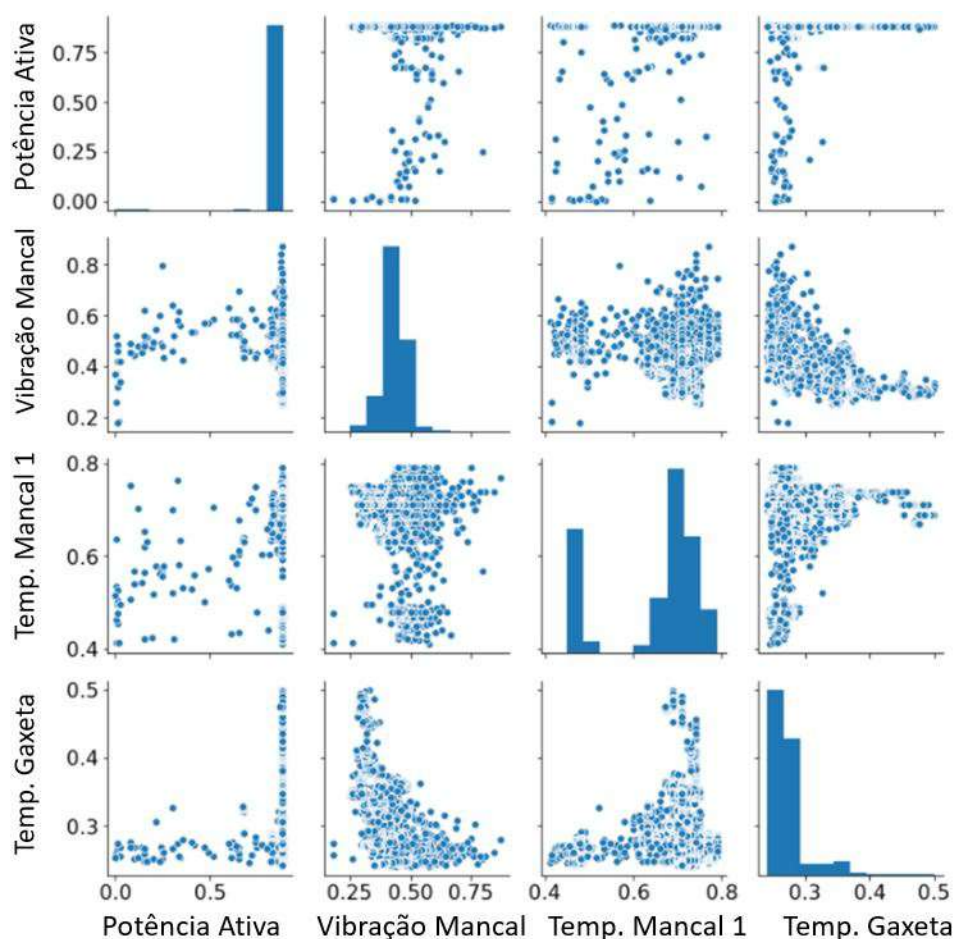


Fonte: elaborado pelo autor, 2020.

O período observado possui operação com potência gerada fixa, conforme apresentado anteriormente. Como esperado, a corrente de armadura do gerador possui elevado grau de correlação com a potência gerada, com a abertura do distribuidor e com a vazão de água na turbina.

Para avaliar a dispersão entre algumas variáveis de interesse é utilizado o método *pairplot* da biblioteca *Seaborn*. Esta abordagem é útil para visualizar (Figura 43) o comportamento entre as variáveis e compreender melhor como se dá a correlação entre elas. Observa-se que a diagonal deste gráfico é o próprio histograma da variável.

Figura 43 – Pairplot das grandezas.

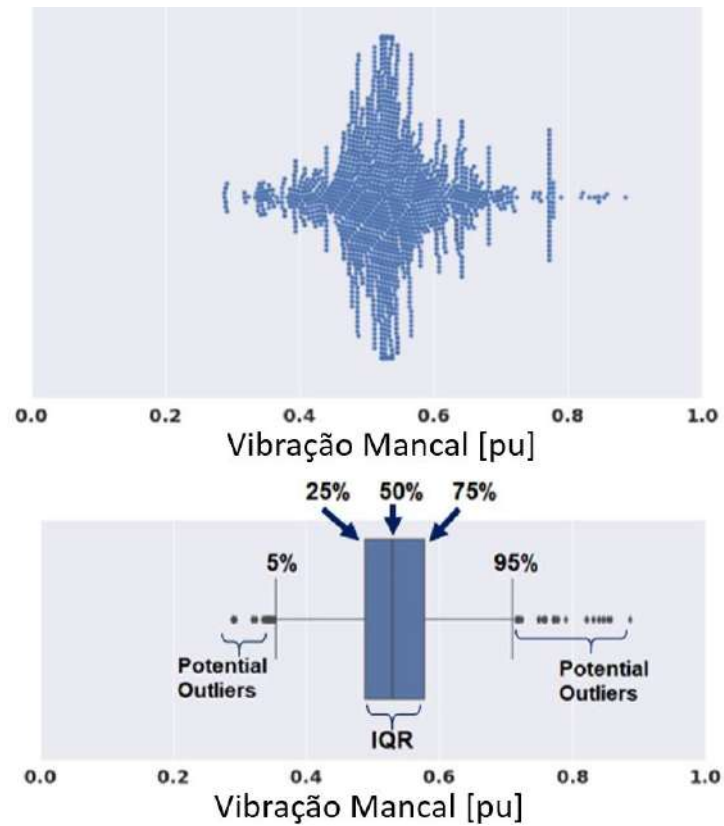


Fonte: elaborado pelo autor, 2020.

Outra abordagem interessante para visualização de séries temporais é o uso do *boxplot*, semelhante ao que é praticado no mercado de ações. A Figura 44 apresenta todos os dados de um sensor de vibração e seu respectivo boxplot para um período de 60 dias.

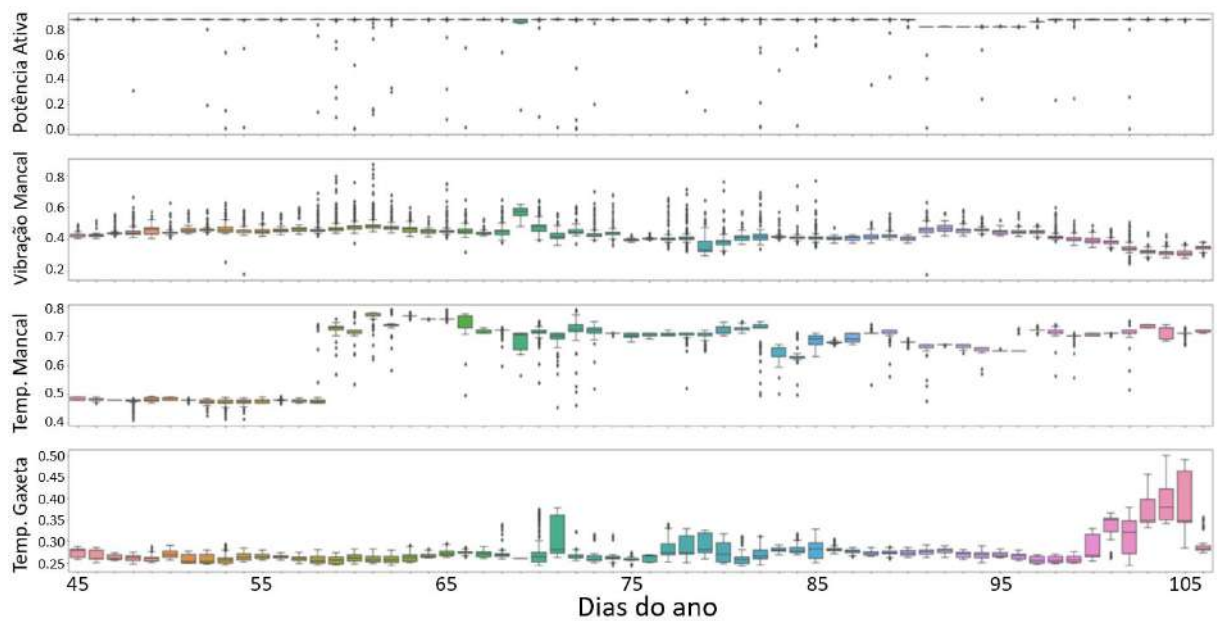
A Figura 45 apresenta o comportamento de quatro variáveis, sendo que cada *boxplot* representa um dia de operação. Esta abordagem pode facilitar o processo de visualização e interpretação dos dados. O primeiro sinal é a potência ativa e, como esperado, o interquartil é pequeno.

Figura 44 – Dados do sensor de vibração e o respectivo *boxplot*.



Fonte: elaborado pelo autor, 2020.

Figura 45 – Visualização de série temporal multivariada com *boxplot*(valores em pu).



Fonte: elaborado pelo autor, 2020.

O segundo sinal é a vibração do mancal, com valor estável em torno de 0.42pu até o 99° dia. A partir desta data observa-se uma pequena diminuição.

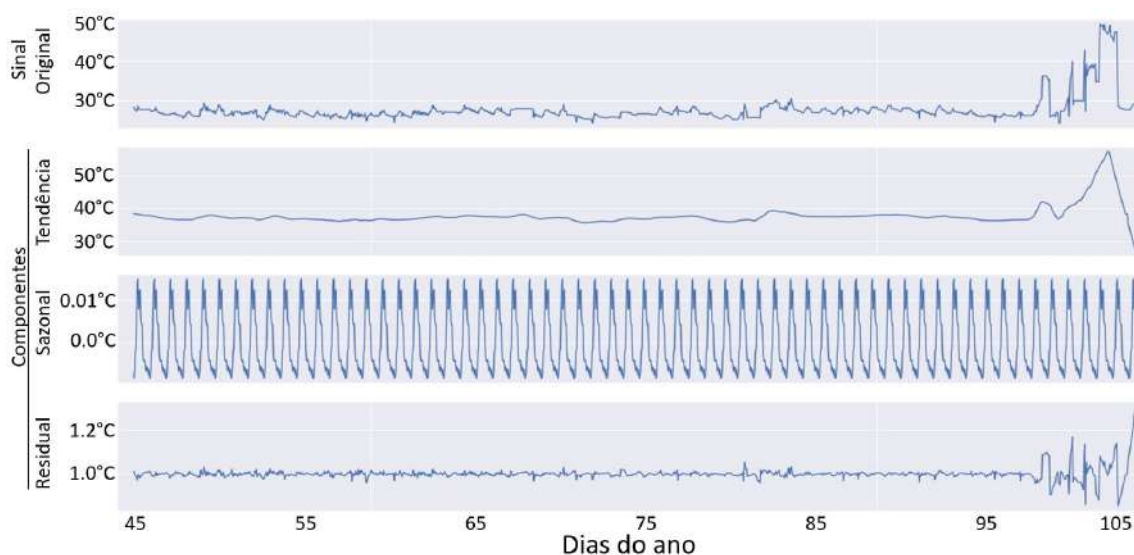
O terceiro sinal apresenta a temperatura do mancal. Nota-se que a partir do 58° dia o comportamento se altera. O quarto sinal apresenta a temperatura da gaxeta, onde se percebe que a partir do 100° dia o comportamento deste sinal é nitidamente alterado.

De modo geral, pode-se afirmar que há quatro padrões de operação neste período de 60 dias. Até o 58° dia, observa-se a operação inadequada da temperatura do mancal; entre o 58° e o 91° dia, operação normal; entre 91° e 97° dia, operação com uma sensível diminuição de potência; e a partir do 100° dia, operação com temperatura da gaxeta em valor acima do padrão. Tal informação é relevante para o Caso 1, apresentado na Seção 5.1.

Outra ferramenta importante que pode ser utilizada na avaliação de uma série temporal é a decomposição do sinal. A Figura 46 apresenta tal operação no sinal de temperatura da gaxeta. O primeiro sinal são os dados observados, o segundo sinal a linha de tendência, o terceiro sinal a parcela sazonal e o último sinal a componente residual.

Para análise do comportamento de um ativo, em geral, a tendência é o elemento de maior relevância. No caso apresentado, a sazonalidade está relacionada com a temperatura ambiente, que apresenta quadros de elevação de temperatura durante o dia e diminuição da temperatura durante a noite. A componente residual é aquela que não se enquadra na linha de tendência e sazonalidade.

Figura 46 – Decomposição da série temporal sensor de vibração.



Fonte: elaborado pelo autor, 2020.

3.5.2 Contexto operacional dos dados

Na construção de um relatório *Data-Informed*, parte dos dados e gráficos são obtidos na análise exploratória de contexto estatístico e parte da análise exploratória de contexto operacional.

Para compreender o contexto operacional do ativo, uma abordagem interessante é a combinação de visualização dos dados de séries temporais de operação com o limite operacional e valores nominais do ativo em linhas de tendências, gráficos como histogramas, gráficos de dispersão, *heatmaps* e outros. Esta etapa ganha maior significado se realizada com auxílio do conhecimento especialista.

Na Figura 47 é apresentado um gráfico tradicional no tempo para tensão e corrente, durante a operação do hidrogerador. Conforme esperado, a tensão de operação de um gerador fica um pouco acima do valor nominal (linha em verde). Isso acontece, pois o sistema de geração deve suprir as perdas do sistema.

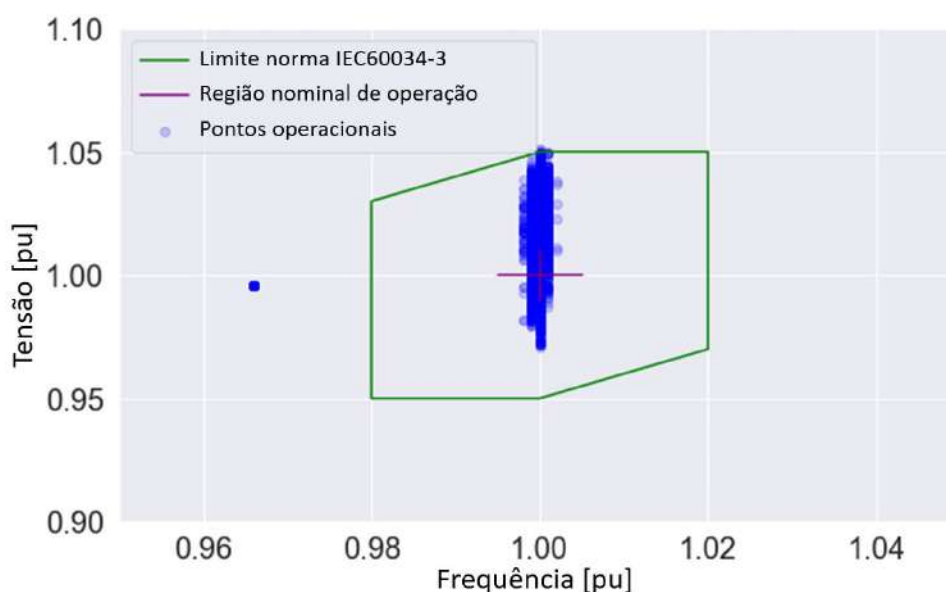
Figura 47 – Tensão e corrente terminal do hidrogerador.



Fonte: elaborado pelo autor, 2020.

Um ponto relevante para operação do gerador é o comportamento Volts/Hertz, por este motivo é criado o gráfico da Figura 48 com limite operacional conforme norma (IEC60034-3, 2020). Nota-se claramente que o gerador em questão operou dentro do limite.

Figura 48 – Gráfico Volts/Hertz com limite operacional IEC60034-3.

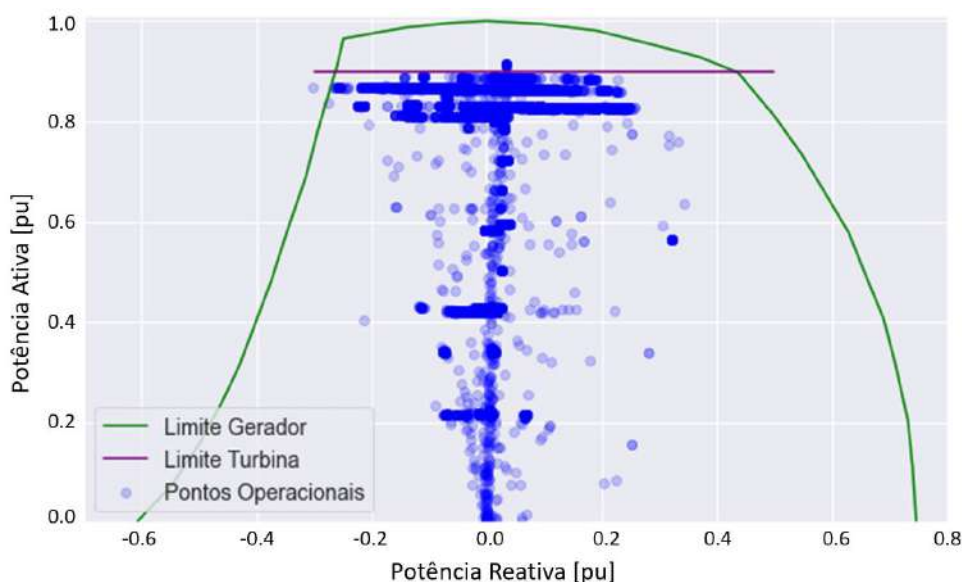


Fonte: elaborado pelo autor, 2020.

Pode-se concluir que o sistema onde o gerador está conectado possui boa estabilidade em frequência, uma vez que o sinal fica muito próximo de 1.0pu. Em geral, o comportamento da tensão terminal de um gerador está relacionado, basicamente, com sistema de regulação da excitação e com sistema elétrico onde o gerador está conectado. Com este gráfico, pode-se concluir que, para o período analisado, a tensão do gerador está dentro dos limites estabelecidos pela norma.

A Figura 49 apresenta um gráfico de dispersão com potência ativa no eixo y e potência reativa no eixo x. A linha verde apresenta o limite construtivo do gerador, informação fornecida pelo fabricante da máquina elétrica (curva de capacidade do gerador, Figura 10). A linha em violeta representa o limite da máquina primária, que neste caso é uma turbina hidráulica.

Figura 49 – Curva de capacidade do gerador com dados reais.



Fonte: elaborado pelo autor, 2020.

Para avaliar a operação do equipamento nas faixas de potência é utilizado um histograma tradicional, na Figura 50. Nota-se que, para o caso em questão, o gerador permaneceu parado por aproximadamente 600 horas e operou com potência entre 0.8pu e 0.88pu por um período um pouco superior a 1200 horas. O limite da máquina primário é apresentado com a linha verde.

A visualização com histograma, para grandezas como potência, temperatura e vibração, podem auxiliar na avaliação do ativo, planejamento da manutenção e estimativa de desgaste de algumas peças, principalmente da máquina primária ou carga. Este tipo de visualização também é importante para avaliar o perfil de operação de uma frota de geradores e/ou ativos.

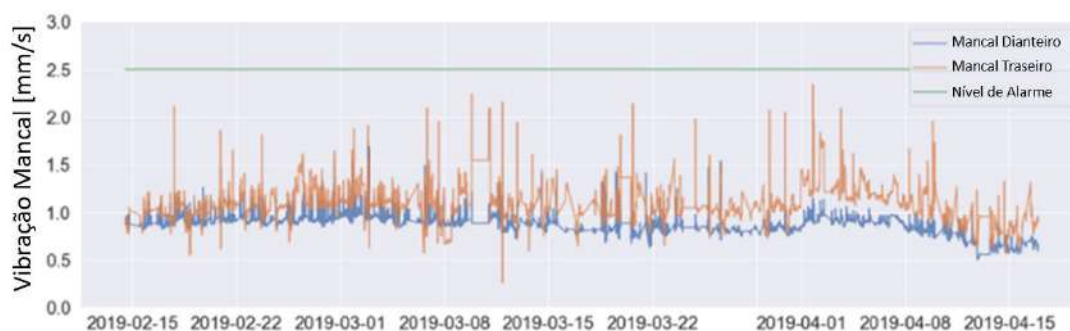
Figura 50 – Histograma de potência do gerador.



Fonte: elaborado pelo autor, 2020.

Um gráfico tradicional de vibração do mancal dianteiro e traseiro no tempo é mostrado na Figura 51. Nota-se que o valor de vibração está abaixo do limite estabelecido como limite de alarme. A norma de referência para vibração em partes não girantes é a ISO 10816-5.

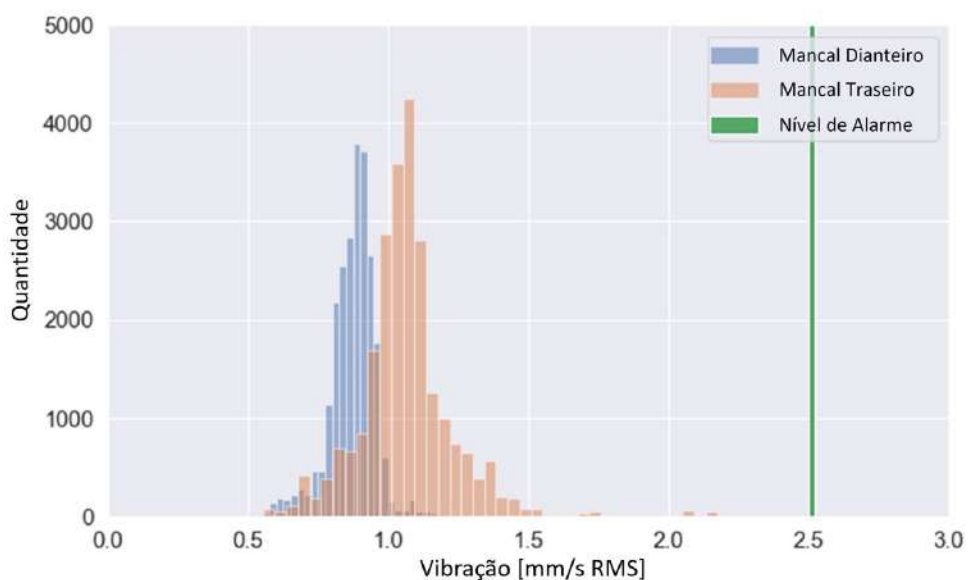
Figura 51 – Vibração em mm/s RMS dos mancais no tempo.



Fonte: elaborado pelo autor, 2020.

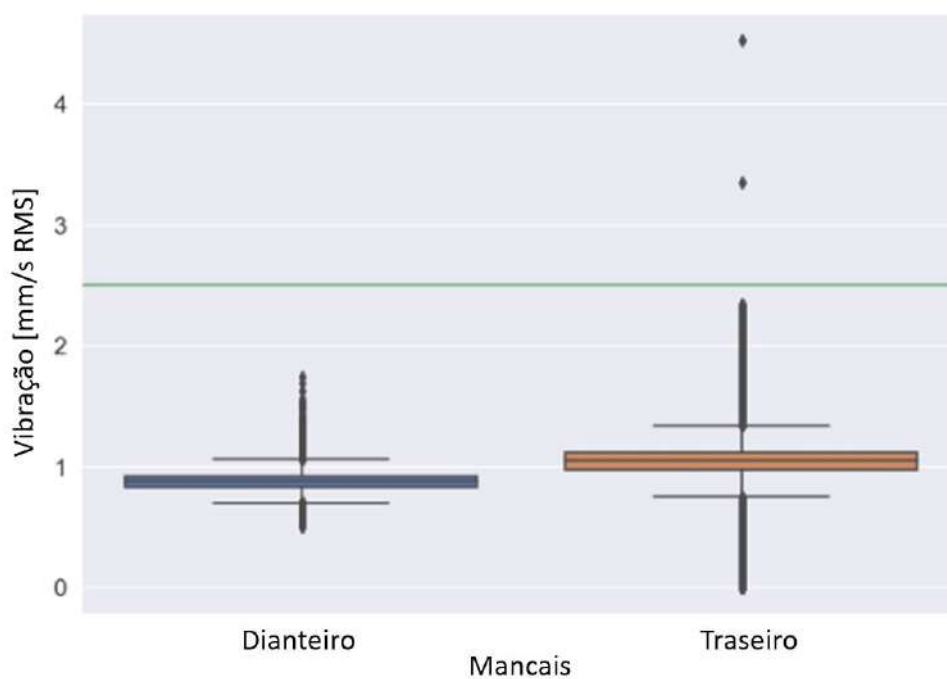
A Figura 52 apresenta o histograma e a Figura 53 o *boxplot* de vibração para os mancais. Os dados utilizados para compor estes gráficos são os mesmos da Figura 51. Com estes dois gráficos é possível identificar o nível predominante de vibração no referido período.

Figura 52 – Histograma de vibração para os mancais.



Fonte: elaborado pelo autor, 2020.

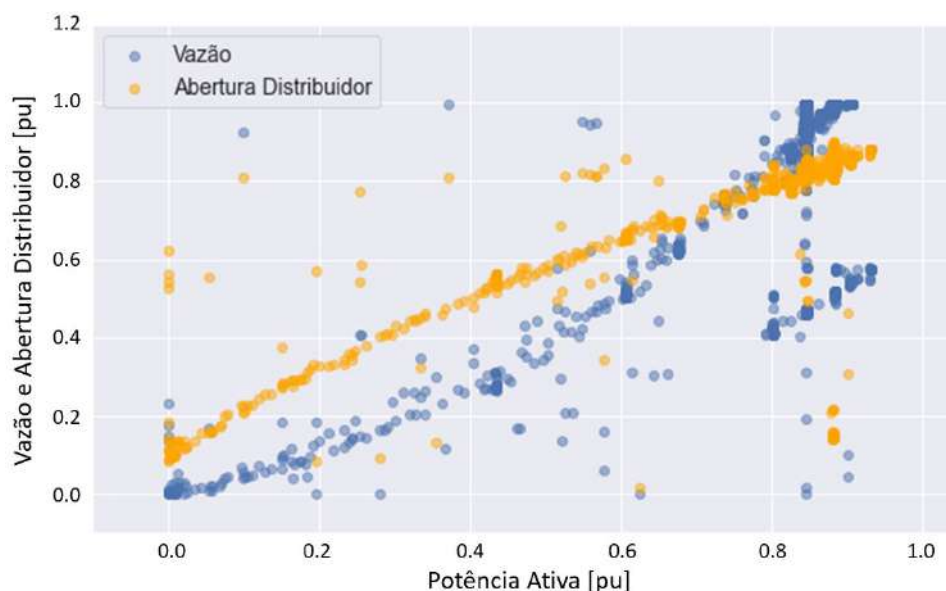
Figura 53 – Boxplot da vibração para os mancais.



Fonte: elaborado pelo autor, 2020.

Para avaliar o comportamento da vazão e abertura do distribuidor em função da potência gerada é criado o gráfico da Figura 54. Neste caso, é selecionado o período de gerador sincronizado na rede, utilizando o status do disjuntor principal. Nota-se que ambos sinais apresentam um comportamento próximo do linear, com algumas ocorrências fora da tendência.

Figura 54 – Gráfico de dispersão entre potência ativa, vazão e abertura distribuidor.



Fonte: elaborado pelo autor, 2020.

Para os gráficos da Figura 55 e Figura 56 é comentado sobre a operação estratégica de hidrelétricas, conforme o artigo de (SILVA et al., 2016). Uma alternativa para tentar alcançar um valor de geração mais próximo ao da garantia física mensal, é utilizar um deplecionamento diário, de pelo menos um metro (depende da planta) junto a um modelo de previsão chuva-vazão. Na prática os operadores devem baixar os reservatórios antes das chuvas, de forma a diminuir o vertimento. É estimado que tal abordagem gere um ganho de aproximadamente 5%.

Conforme glossário (ONS, 2020, *online*), o deplecionamento significa o rebaixamento do nível de água de um reservatório ou diminuição do volume de água armazenado. A Figura 55 apresenta o comportamento da queda líquida para turbina 1 e 2 em função da potência gerada, para o referido período de 60 dias e operação não simultânea das turbinas. A Figura 56 apresenta o *boxplot* da queda líquida na condição de turbinas ligada e desligada. Tal recurso pode auxiliar no processo de avaliação do comportamento do reservatório e na estratégia de operação e deplecionamento.

Figura 55 – Gráfico de dispersão potência ativa e queda líquida.

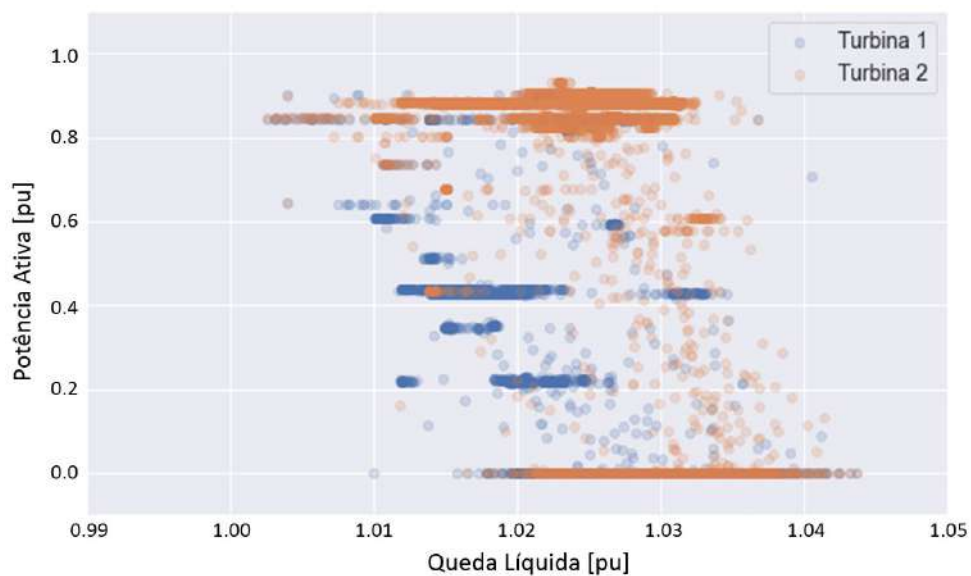
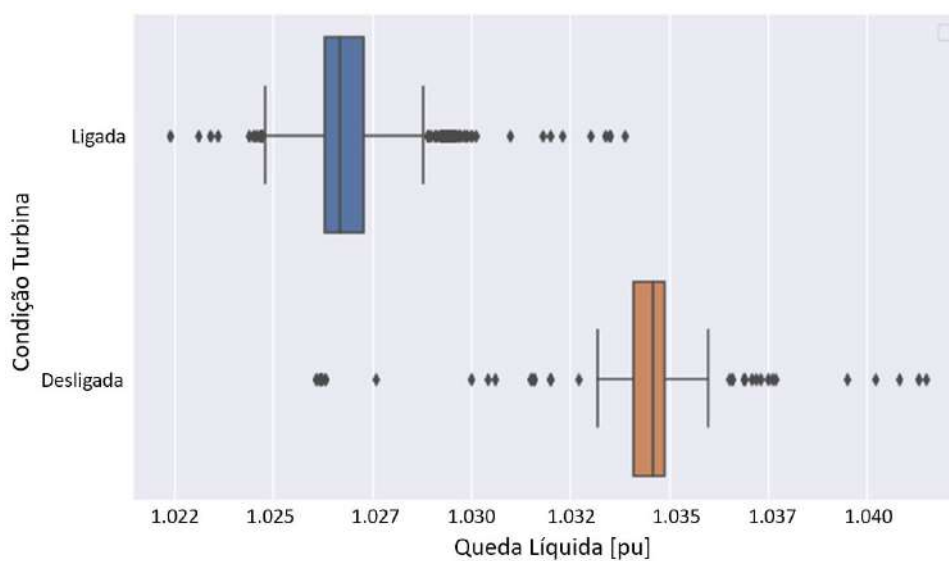


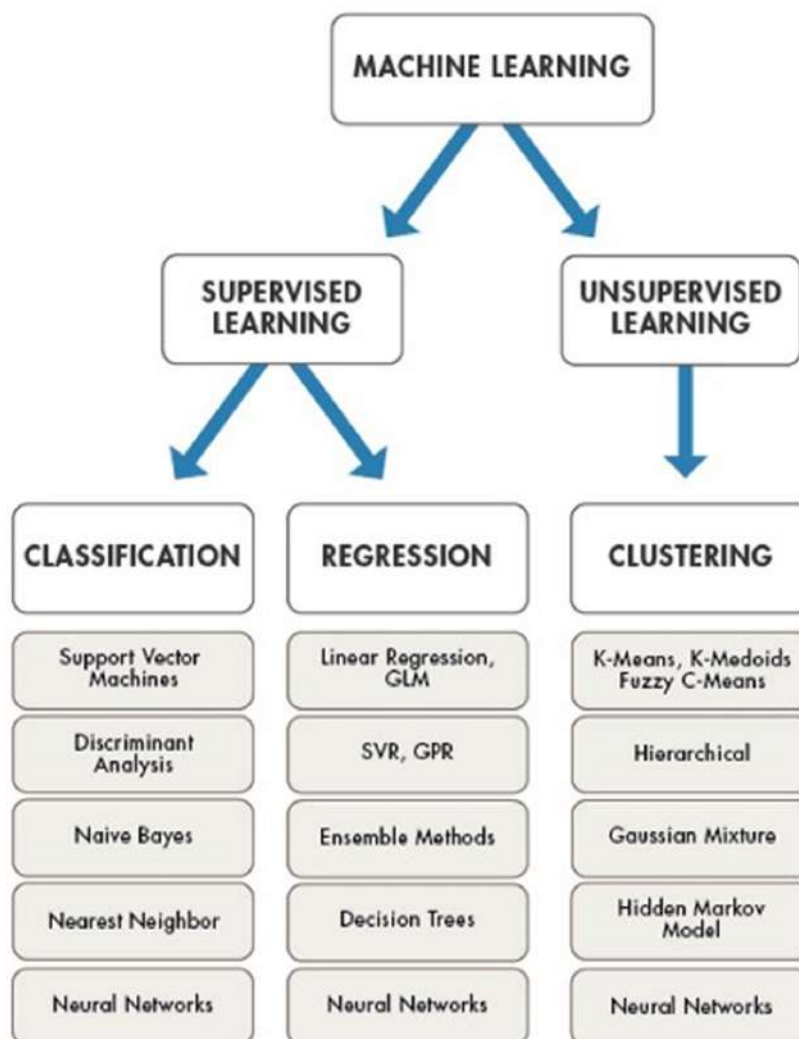
Figura 56 – *Boxplot* queda líquida para condição de turbinas ligada e desligada.



3.6 APRENDIZADO DE MÁQUINA

Há inúmeros modelos de aprendizado de máquina disponíveis (Figura 57), cada um com suas particularidades, prós, contras e tipo de aprendizado. Estes modelos podem ser aplicados para solucionar problemas de regressão ou classificação, com o aprendizado supervisionado ou não supervisionado.

Figura 57 – Técnicas de aprendizado de máquina.



Fonte: (PANT, 2019, *online*).

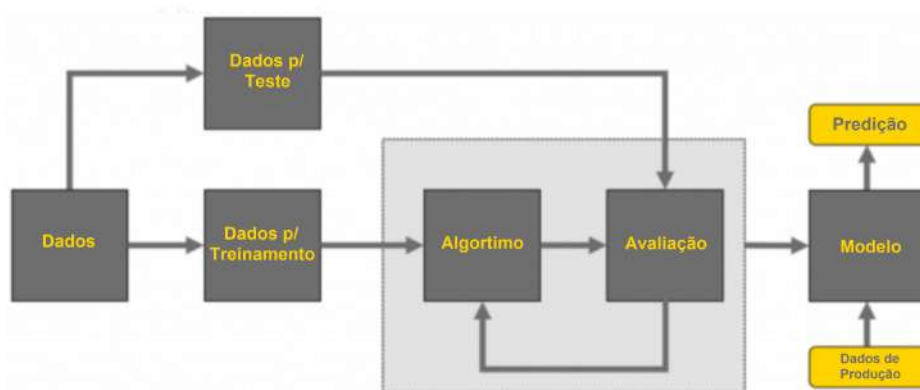
Conforme demonstrado por (ALRAYESS et al., 2018), no artigo “*Using Machine Learning Techniques and Deep Learning in Forecasting The Hydroelectric Power Generation in Almus Dam, Turkey*”, de 2018, escolher um modelo de aprendizado de máquina para resolver uma determinada problemática é algo pouco trivial. Neste artigo os autores exploraram três técnicas distintas de aprendizado e utilizaram 20 anos de dados mensais.

O processo de aprendizado de um modelo depende de vários pontos, os principais são: a seleção e configuração do método, a quantidade e qualidade dos dados, como os dados são limpos, como o vetor de entrada do modelo é criado entre outros.

Considerando uma perspectiva simplista, o processo de aprendizado é o ajuste dos hiperparâmetros do modelo, que ocorre de modo iterativo com dados de treino. No caso de uma rede neural, estes hiperparâmetros conectam os neurônios de entrada aos neurônios ocultos e/ou neurônios de saída. A Subseção 3.6.5 mostra em detalhes o funcionamento da rede neural.

A Figura 58 apresenta um *workflow* de referência para o processo de treinamento e teste dos algoritmos/modelos com aprendizado de máquina. Este fluxo também é considerado durante a elaboração do *workflow* proposto neste estudo (Figura 69).

Figura 58 – *Workflow* para construção dos algoritmos com aprendizado de máquina.



Fonte: (PANT, 2019, *online*).

O processo de treinamento, teste e validação dos algoritmos é um trabalho amplo e iterativo, demanda habilidades de programação, infraestrutura e recurso computacional. Antes de realizar a liberação de um modelo é necessário avaliar se o desempenho e acurácia estão dentro do mínimo aceitável. A efetividade do algoritmo deve ser revisada e, se necessário, o modelo deve ser treinado novamente ou o processo de criação deve ser revisto.

Uma preocupação constante durante a criação de um algoritmo é a capacidade de generalização do modelo, observando as questões de sobretreino (*overfitting*), treino insuficiente (*underfitting*), variância e viés (*bias*). Tal abordagem não é trivial.

3.6.1 Modelo físico versus modelo baseado em dados

Segundo (ZHOU; HU; TOMLIN, 2016), o modelo baseado em dados é mais acessível devido à sua menor complexidade, e poderia servir como um substituto para o modelo físico, que geralmente é complexo. Adicionalmente, um modelo semi paramétrico baseado em dados pode ser facilmente utilizado em um aplicação com uma exigência modesta de dados registrados, já o modelo físico requer dados construtivos detalhados do ativo, que na prática, muitas vezes, estão sujeitos a imprecisão e dificuldades para obtenção. O estudo de (ZHOU; HU; TOMLIN, 2016) aborda a estimativa da temperatura de ambientes para efeitos de controle da climatização.

Já para (JI et al., 2012), um modelo é usado para previsão e aviso de cheias. Os modelos físicos de chuva-escoamento para o processo convencional de previsão de cheias requer muitas

informações e dados, e incluem incertezas que podem acumular erros durante a modelagem. Por outro lado, o modelo baseado em dados combinando rede neural e técnicas Fuzzy pode diminuir a quantidade de dados físicos necessários para construção do modelo.

Há tempos que a comunidade acadêmica vem experimentando modelos baseados em dados para complementar ou substituir modelos físicos. Em geral, estudos como (JI et al., 2012) e (ZHOU; HU; TOMLIN, 2016) mostram que tal abordagem não é trivial e ambos métodos possuem vantagens e desvantagens. A Tabela 10 apresenta um resumo deste comparativo.

Tabela 10 – Comparativo entre o modelo físico e modelo baseado em dados.

Modelo físico	Modelo baseado em dados
Comportamento explicável por equações bem conhecidas.	“Caixa preta”, pouco explicável por conta da dinâmica de aprendizado.
Requer parâmetros construtivos do ativo.	Capacidade de abstrair os parâmetros construtivos.
Em geral, necessita de poucos dados de operação do ativo, se comparado com modelo baseado em dados.	Em geral, necessita de grande quantidade de dados.
Requer poder computacional para o processo de otimização dos modelos.	Requer elevado poder computacional para o processo de treinamento dos modelos.
Requer conhecimento especialista aprofundado.	Requer o envolvimento do conhecimento especialista.
Atualmente é artesanal e continuará sendo.	Atualmente é bastante artesanal, mas poderá ser automatizado.
-	Requer infraestrutura em <i>cloud</i> .

Fonte: adaptado de (JI et al., 2012) e (ZHOU; HU; TOMLIN, 2016)

Em alguns casos, a abordagem híbrida pode ser uma opção interessante, um modelo físico combinado com modelo baseado em dados pode descrever adequadamente um determinado fenômeno aproveitando as vantagens que cada método oferece.

Conforme (GREENLAND et al., 2016), a saída dos modelos de aprendizado não deverão ser utilizados como um veredicto final. Devem ser tratados como uma ferramenta que auxilia na tomada de decisão, apontando as principais hipóteses para uma investigação humana. A epígrafe deste estudo também reforça esse conceito: “All models are wrong, but some are useful.”, em tradução livre, “Todos os modelos estão errados, mas alguns são úteis.”, por George E.P. Box.

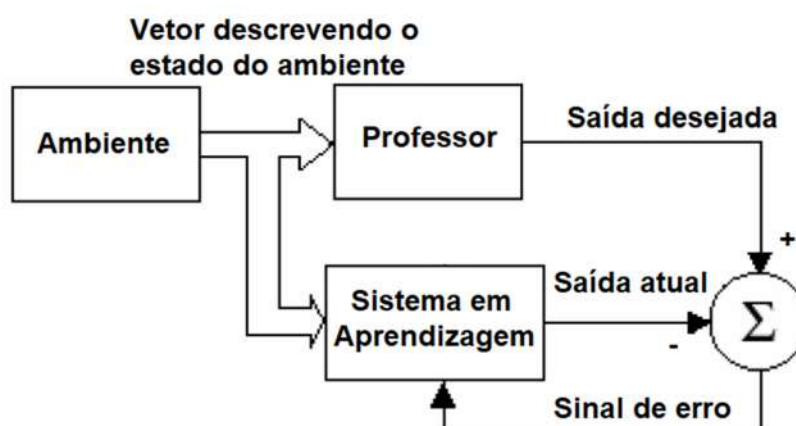
3.6.2 Aprendizado supervisionado e não-supervisionado

O aprendizado supervisionado ou aprendizado com professor é representado pelo diagrama da Figura 59. Conforme referência (PUCRIO, 2020, *online*), trata-se de um paradigma de aprendizagem, no qual o professor possui conhecimento sobre o ambiente em que a rede irá operar. Este conhecimento está representado sob a forma de um conjunto de amostras de entrada-saída (Figura 68).

O aprendizado ocorre por meio do ajuste dos pesos sinápticos (hiperparâmetros) do modelo. Desta forma, o conhecimento do ambiente disponível ao professor é transferido para a rede por meio do treinamento. Quando esta condição é alcançada, o professor é dispensado e a rede pode lidar com o ambiente de maneira autônoma (HAYKIN, 2007).

Do ponto de vista de aplicação, o aprendizado supervisionado tem como desvantagem a necessidade de conhecer a saída, o que nem sempre é obtido facilmente. A vantagem está na funcionalidade do sistema aprender um comportamento físico ou uma função cognitiva humana e permitir a reprodução destas habilidades em ambiente computacional.

Figura 59 – Representação do aprendizado supervisionado.



Fonte: adaptado de (PUCRIO, 2020, *online*).

No aprendizado não supervisionado a rede ou modelo não utiliza nenhum valor para ser comparado na saída. O sistema se auto organiza para identificar as entradas mais semelhantes e vai agrupá-las, criando classes. Uma vez treinada, quando um sinal é enviado para o modelo, ele identifica a qual classe esse novo sinal pertence e aciona a respectiva saída. Se esse sinal não pertencer a nenhuma classe anteriormente definida por ela, uma nova classe é criada ((MAFRA, 2019) e (KARTALOPOULOS, 1997)).

Do ponto de vista de aplicação, o aprendizado não supervisionado apresenta a vantagem de dispensar o conhecimento prévio da saída e o poder de proporcionar resultados sem tal informação prévia.

3.6.3 Classificadores

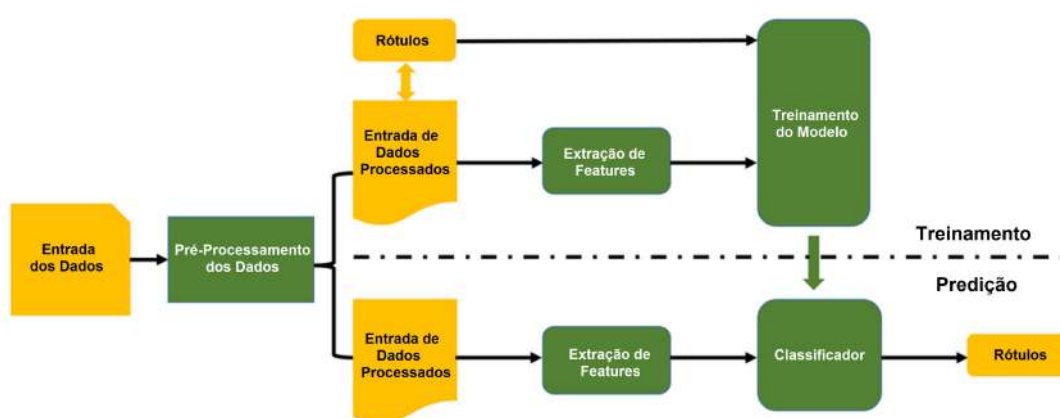
Conforme (ZAKI; MEIRA, 2014), a tarefa de classificação é prever o rótulo ou classe de um determinado ponto não rotulado. Formalmente, um classificador é um modelo ou função M que prediz a classe ou rótulo y para um determinado input x , resultando em $y = M(x)$, onde $y \in \{c_1, c_2, \dots, c_k\}$ e cada c_i é uma classe ou rótulo, como por exemplo: operação normal ou anormal.

Para construir o modelo é necessário um *dataset* de treino, com os pontos de entrada x e seus respectivos rótulos y . Após treinar o modelo M , pode-se prever a classe ou rótulo y de um ponto x inédito para o modelo. Muitos tipos de modelos de classificação vêm sendo propostos, como por exemplo árvores de decisão, classificador probabilístico, *support vector machines*, entre outros.

Conforme (FAWAZ et al., 2019), a diversidade de tipos de *datasets* no arquivo UCR/UEA, o maior repositório de dados de séries temporais, mostra as diferentes aplicações do problema envolvendo a classificação de séries temporais. Considerando a crescente necessidade de classificar séries temporais com precisão, pesquisadores vem propondo centenas de métodos para solucionar esta tarefa. Uma das mais populares e tradicionais abordagens para classificação de série temporal univariada é o uso do classificador nearest neighbor acoplado com uma função de distância. Particularmente, a função de distância *Dynamic Time Warping* (DTW) combinada com classificador *nearest neighbor* vem se mostrando como uma sólida referência para série univariada.

A Figura 60 apresenta o processo de treino e teste de um modelo de classificação. Nota-se que para o processo de treino é necessário conhecer o rótulo das amostras.

Figura 60 – Processo de treino e teste de um modelo de classificação.



Fonte: (JING, 2020, online).

Classificação e clusterização são dois métodos de aprendizado que podem ser utilizados para detecção de padrões. Apesar de ambos métodos possuírem uma similaridade, a principal diferença entre eles está no fato de que o classificador utiliza classes ou rótulos pré-definidos, ou seja, aprendizado supervisionado.

A clusterização identifica a similaridade entre os objetos ou amostras sem qualquer classe ou rótulo, agrupando os objetos ou amostras conforme suas características em comum, e diferenciando dos demais grupos (*clusters*).

3.6.4 Regressores

A tarefa de regressão é prever um determinado valor numérico com base em um determinado cenário numérico. Formalmente, um regressor é um modelo ou função N que prediz o valor de y para um determinado valor de entrada x , resultando em $y = N(x)$.

Um exemplo aplicável pode ser a estimativa de temperatura de operação de um gerador para uma determinada condição de potência gerada. Para construir um modelo é necessário um *dataset* com os pontos de x e seu respectivo resultado em y , que é chamado de *dataset* de treino.

Com modelo N treinado, pode-se prever o resultado y para um cenário x inédito. Conforme apresentado na Subseção 3.1.1, há inúmeras técnicas de regressão, tanto para séries temporais univariadas como multivariadas. Neste estudo, é utilizado a rede neural TLFN focada em modo regressor, com 21 pontos (ou neurônios) na entrada e 1 na saída.

3.6.5 Redes neurais artificiais e TLFN

Os créditos relacionados à invenção das redes neurais são variados, mas muitos atribuem a teoria de Warren McCulloch e Walter Pitts, que representaram matematicamente o funcionamento de um neurônio com a característica do “tudo-ou-nada”, com artigo “*A Logical Calculus of the Ideas Immanent in Nervous Activity*” de 1943.

Em 1949, o psicólogo Donald Hebb publicou “*The Organization of Behavior: A Neuropsychological Theory*”, que resultou na regra de Hebb, que descreve como a cadeia de neurônio é excitada, conforme suas próprias palavras: ‘*neurons that fire together wire together*’.

O trabalho da dupla McCulloch e Pitts influenciou Frank Rosenblatt, que publicou o artigo “*The perceptron: Perceiving and Recognizing Automation*” em 1957, onde descreve a construção dos perceptrons e como a rede neural artificial pode aprender dos dados. Rosenblatt é creditado por criar o aprendizado supervisionado, onde o neurônio altera seus próprios pesos (sinápticos) com base na sua precisão.

Em 1975, o interesse em redes neurais e aprendizado de máquina é renovado por Paul Werbos, que utilizou sua dissertação para introduzir o treinamento prático de redes multicamadas. Por meio do algoritmo de retropropagação (algoritmo *backpropagation*), que utiliza o erro para modificar os pesos sinápticos em cada nó.

A exploração de modo significativo das redes neurais aconteceu após a década de 90, com o advento da computação distribuída e avançada. A partir de 2010, as primeiras abordagens com aprendizado profundo (*deep learning*) são notadas.

Conforme (MAFRA, 2019), o estudo em redes neurais tem sido motivado desde o começo pelo conceito de que o cérebro humano processa informações de uma forma diferente do computador convencional. O cérebro é um sistema de processamento de informações altamente complexo, não-linear e paralelo. Ele tem a capacidade de organizar seus constituintes estruturais (neurônios) para realizar um processamento avançado de modo muito ágil.

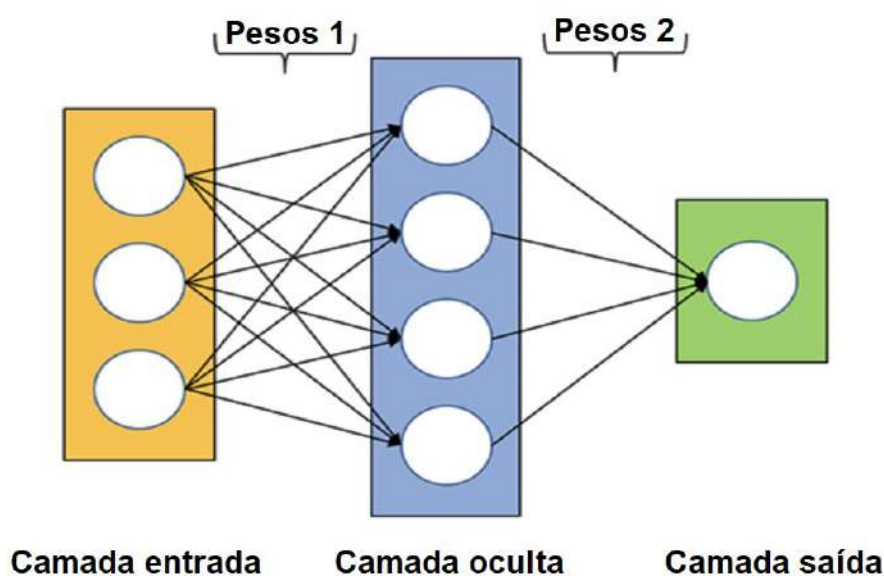
Para (HAYKIN, 2007) é evidente que o poder computacional da rede está relacionado com sua estrutura altamente paralela e distribuída. Outro ponto relevante é a habilidade da rede aprender e generalizar. A generalização está relacionada com a capacidade de a rede neural gerar saídas apropriadas para entradas ou cenários não presentes no *dataset* de treino.

Conforme apresentado por (HENAO et al., 2014), modelos baseados em dados são muito promissores, pois dispensam qualquer conhecimento sobre o equipamento e seu modelo físico.

Conforme (GUPTA et al., 2014), as redes neurais artificiais são muito flexíveis, podem ser utilizadas para tarefas de regressão ou classificação, podem adotar inúmeros pontos/neurônios na camada de entrada, nas camadas ocultas e na camada de saída. A Figura 61 apresenta uma arquitetura simples, com três elementos na camada de entrada, quatro neurônios na camada oculta e um ponto/neurônio na camada de saída.

Nota-se que nesta combinação existem 12 pesos sinápticos (pesos 1) entre a camada de entrada e a camada oculta e 4 pesos sinápticos (pesos 2) entre a camada oculta e o ponto/neurônio da camada de saída. Os elementos *bias* não são apresentados nesta imagem.

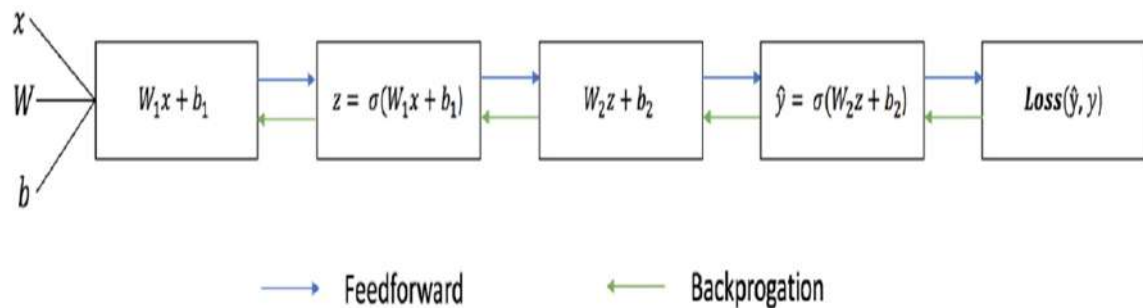
Figura 61 – Arquitetura simplificada de uma rede neural.



Fonte: (GUPTA et al., 2014).

Para o exemplo apresentado na Figura 61, do ponto de vista matemático, há basicamente dez etapas envolvidas no processo de treinamento, cinco no processo de avanço (*Feedforward*) e outras cinco no processo de retro-propagação, isso mostrado na Figura 62.

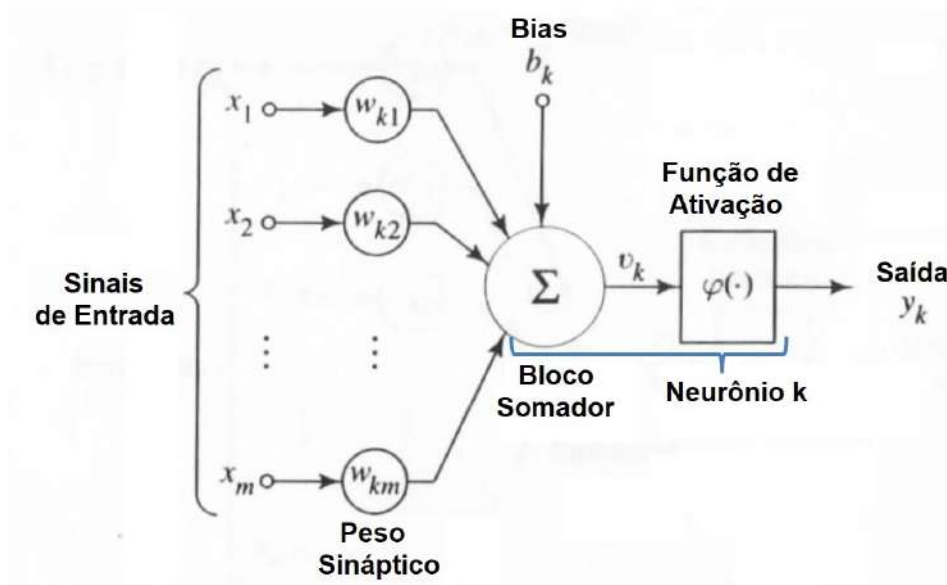
Figura 62 – Processo *Feedforward* e *Backpropagation* de uma rede neural.



Fonte: (GUPTA et al., 2014).

A Figura 63 apresenta a representação de um neurônio artificial. À medida que mais neurônios são combinados em camadas e colunas, obtém-se uma rede neural artificial multicamada, conforme apresentado na Figura 61.

Figura 63 – Representação de um neurônio artificial.



Fonte: (MAFRA, 2019).

Conforme exposto por (HAYKIN, 2007), há quatro elementos básicos no modelo apresentado na Figura 63:

1. Um conjunto de sinapses, trata-se dos elos de conexão entre a camada de entrada (sinais de entrada) e o(s) neurônio(s). O sinal de entrada x_j da sinapse j é conecta ao neurônio k por meio do peso sináptico w_{kj} . Os pesos sinápticos de um neurônio artificial podem ser iniciados com valor randômico e, após o primeiro ciclo (época) de treinamento, podem assumir valores positivos e negativos, decorrente do processo de retro-propagação.

2. *Bias* b_k , algumas redes utilizam este recurso. O *bias* tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação. Durante o processo de treinamento, este parâmetro sofre ajuste de modo semelhante aos pesos sinápticos.

3. O bloco somador agrega o sinal de cada entrada $\{x_1, x_2, \dots, x_m\}$ considerando o respectivo peso sináptico w_{kj} e incluindo o *bias* b_k .

4. Por fim, é aplicado a função de ativação (ϕ) para restringir a amplitude da saída do neurônio y_k . Tipicamente, o intervalo normalizado da amplitude de saída de um neurônio é escrito como o intervalo unitário fechado $[0, 1]$ ou alternativamente $[-1, 1]$.

O resultado matemático do modelo apresentado na Figura 63 é mostrado na equação (2). Este processo/cálculo para obter o elemento de saída da rede ($\hat{y}_k$) é chamado de propagação de avanço ou *Feedforward*.

$$\hat{y}_k = \phi \left(\sum_{j=0}^n w_{kj} x_j + b_k \right) \quad (2)$$

Onde:

x_j – sinal de entrada da sinapse j ;

w_{kj} – peso da sinapse j para o neurônio k ;

b_k – sinal de *bias* para o neurônio k ;

ϕ – função de ativação (ex. linear, sigmoid, Relu, etc);

$\hat{y}_k$ – sinal de saída do neurônio k .

A função de ativação tem um papel muito relevante no funcionamento da rede neural. Algumas opções de função de ativação encontradas na literatura são apresentadas na Figura 64.

Nas referências estudadas sobre redes neurais, há poucas recomendações sobre a escolha dos parâmetros e a topologia da rede, como: número de neurônios, número de camadas, função de ativação, *bias*, etc. Grande parte destas definições são feitas de modo empírico, seguindo a experiência do projetista. Até o momento, não se conhece uma fórmula ou metodologia bem definida para especificação de tais parâmetros.

Tratando-se da função de ativação, deve-se optar por uma função derivável (diferenciável) e observar o objetivo da aplicação. Em alguns casos, realizar um experimento com diferentes tipos de função de ativação pode ser uma alternativa.

Como uma rede neural pode possuir vários neurônios, a partir deste ponto as variáveis são apresentadas como matrizes, logo cada posição corresponde a um conjunto de pesos sinápticos ou neurônios. Esta representação é feita com os símbolos em caixa alta. Conforme apresentado por (GUPTA et al., 2014), após a primeira iteração da rede neural, obtém-se o valor estimado na saída ($\hat{Y}$), conforme equação (2) e com este resultado é calculado o erro em relação ao valor real esperado Y , por meio da equação (3).

$$ERRO = \hat{Y} - Y \quad (3)$$

Figura 64 – Funções de ativação.

Função ativação	Equação	Gráfico
Função degrau	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	
Sigmoid	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	
Linear	$\phi(z) = z$	
Logística	$\phi(z) = \frac{1}{1 + e^{-z}}$	
Tangente hiperbólica	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	

Fonte: Adaptado de (RASCHKA, 2020, *online*).

Na sequência é realizado o processo de retro-propagação para ajuste dos pesos sinápticos W . Para isso é calculado a variação necessária do peso (ΔW), com a equação (4), onde o elemento (δ) representa a derivada. Para definir o novo valor do peso sináptico W' , é utilizado a equação (5).

$$\Delta W = X^T 2ERRO \delta \phi(\hat{Y}) \quad (4)$$

$$W' = W - \alpha \Delta W \quad (5)$$

Onde:

ΔW – Matriz com delta de correção para o peso sináptico;

X^T – Matriz transposta da camada de entrada;

δ – Derivada;

W – Matriz com peso sináptico na iteração atual;

W' – Matriz com peso sináptico para próxima iteração;

α - Passo de aprendizado (*step size* ou *learning rate*).

O peso sináptico W' carrega consigo o aprendizado e isso acontece porque a função *backpropagation* busca minimizar o erro (*loss function*) de modo iterativo, como é ilustrado na Figura 65. Nota-se que a velocidade de minimização do erro e o ponto de "acomodação" do erro depende do passo de aprendizado (α).

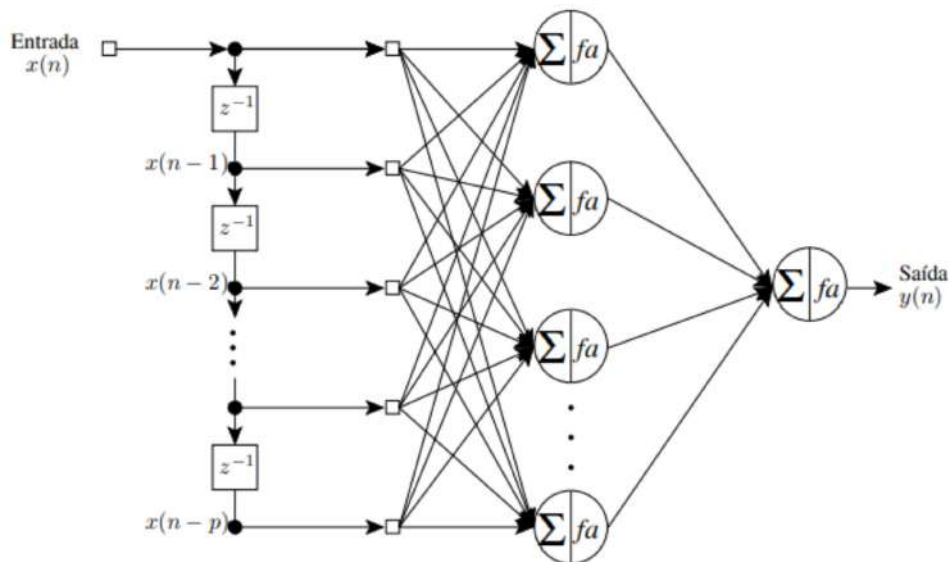
Figura 65 – Gradiente descendente e minimização do erro.



Fonte: (CUI, 2018)

Conforme (HAYKIN, 2007), o reconhecimento temporal de padrões requer o processamento de padrões que evoluem no tempo, com a resposta em um instante particular de tempo dependendo não apenas do valor presente da entrada, mas também de seus valores passados. Por este motivo, a camada de entrada contém o valor presente $x(n)$ e p valores passados $x(n-1), x(n-2), \dots, x(n-p)$, armazenados em uma memória de linha de atraso de ordem p . Tal estrutura pode ser implementada em uma rede neural, referida como rede neural alimentada adiante focada atrasada no tempo (*focused time lagged feedforward network*, TLFN focada) apresentada na Figura 66.

Figura 66 – Rede neural TLFN focada.



Fonte: (NIED, 2007).

3.6.6 Criação do vetor de entrada do modelo

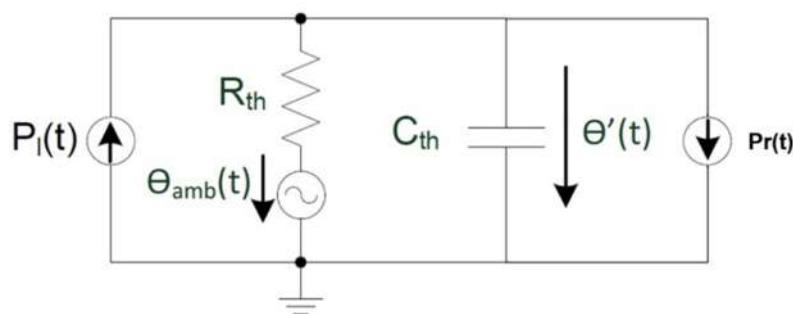
A Figura 66 apresenta uma TLFN focada com a variável de entrada $x(n)$, com atraso de ordem p e saída $y(n)$. O termo n pode ser observado como um instante de tempo t . Conforme

mentionado anteriormente, para escolha das variáveis de entrada no modelo observou-se o modelo físico de Nienhaus.

A partir do modelo proposto por (NIENHAUS; HILBERT, 2012) e com base no diagrama do sistema de arrefecimento (Figura 77, em Anexo B), é criado o modelo físico simplificado do aerogerador WEG (Figura 67). A principal mudança está na inclusão do elemento $P_r(t)$ [Watts], que representa a capacidade de arrefecimento do sistema e está diretamente relacionado com a velocidade da bomba do fluido de arrefecimento, com a abertura da válvula proporcional, com a velocidade do ventilador do radiador e com a temperatura ambiente.

Para criação do vetor de entrada do Caso 2 (Seção 5.2), foram observados os elementos $P_l(t)$ e $P_r(t)$. Para $P_l(t)$ é considerado a potência gerada pelo aerogerador (*var1*), já para o elemento $P_r(t)$ é utilizado a abertura da válvula proporcional (*var2*) e a velocidade da bomba (*var3*), ambos do sistema de arrefecimento. Sendo assim, o vetor de entrada recebeu três variáveis distintas e um atraso de ordem 6. Como a amostragem utilizada é de 10 minutos, o período observado pelo modelo é de 60 minutos.

Figura 67 – Modelo físico térmico simplificado do sistema de arrefecimento.

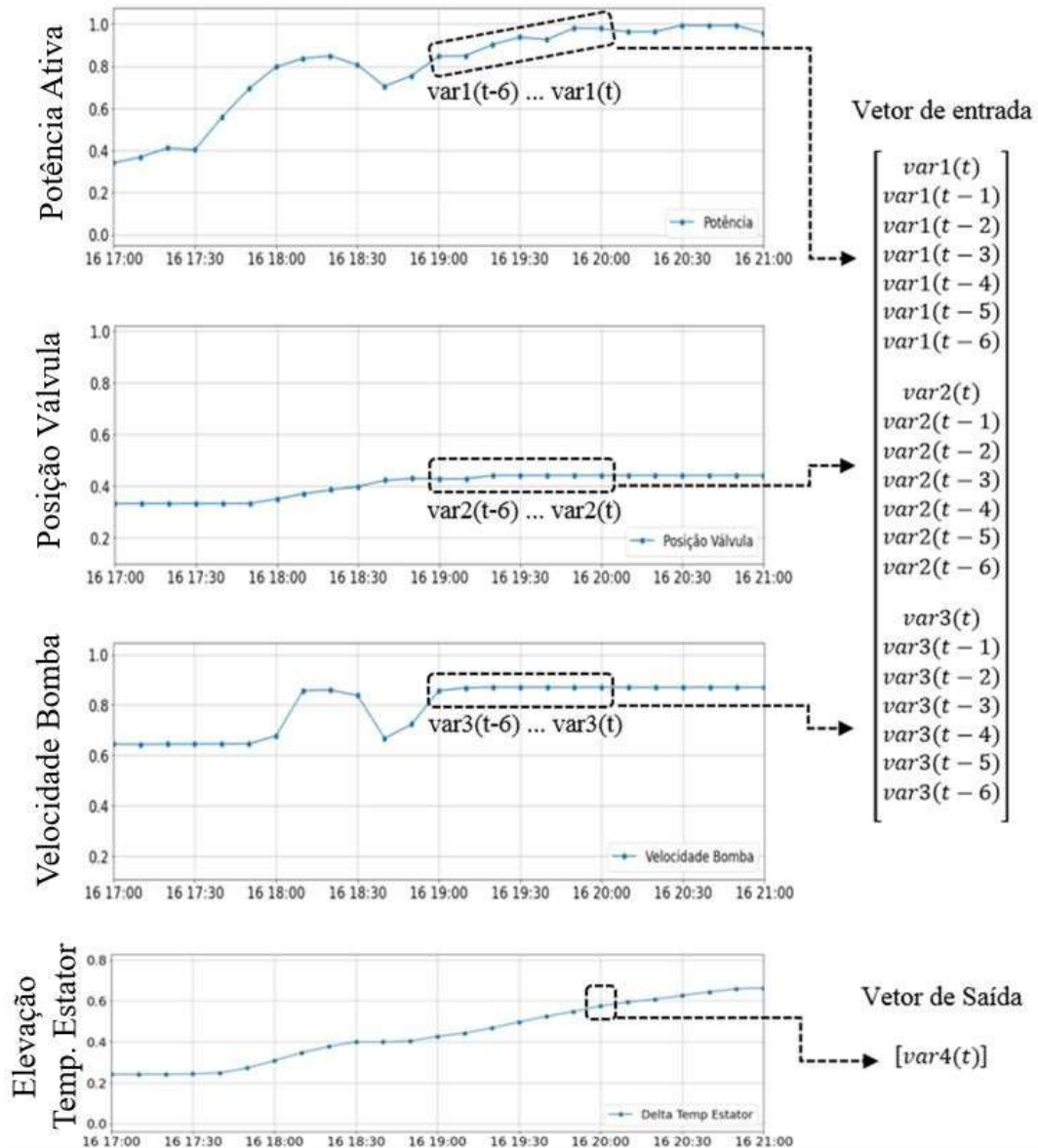


Fonte: adaptado de (NIENHAUS; HILBERT, 2012)

Para o atributo meta ou variável de saída $\theta'(t)$, é considerado a elevação de temperatura no estator. A variável de saída do modelo é a elevação de temperatura do estator do gerador (*var4*). Deste modo, conforme observado na Figura 68, obtêm-se 21 elementos na camada de entrada e 1 elemento na camada de saída. Os detalhes da topologia, parâmetros utilizados e resultados obtidos com este caso e modelo são apresentados na Seção 5.2.

É importante notar que, durante o processo de criação do vetor de entrada, o código responsável por isso obedece a sequência de amostras na série temporal. Logo, o vetor de entrada será criado apenas se todos os registros entre $x(t)$ e $x(t - p)$ estiverem presentes no *dataset*. Por este motivo é importante que o *dataset* possua poucos dados faltantes, do contrário o vetor de entrada possuirá um elevado número de lacunas.

Figura 68 – Composição do vetor de entrada e saída do modelo.



Fonte: elaborado pelo autor, 2020.

3.6.7 Períodos de treinamento e teste do modelo

Antes de determinar o período de treino e teste, é necessário identificar um caso real com operação normal e com desvio no sistema de arrefecimento. Para isso, foi realizada uma pesquisa na lista de ocorrências e manutenções em um dos parques eólicos gerido pelo time de O&M WEG.

Após identificar algumas ocorrências para o estudo, foi realizada uma análise exploratória dos dados com foco no contexto operacional, o que resultou em três casos candidatos. Na sequência, a equipe de O&M foi consultada para realizar uma análise sobre os casos e indicar qual seria o mais relevante do ponto de vista de O&M. Ao final deste processo, foi escolhido um candidato, que foi explorado por este estudo e é apresentado na Seção 5.2.

De posse da informação da engenharia sobre como o sistema de arrefecimento deve operar (detalhes no Anexo B) e após a análise da equipe de O&M, foi possível apontar quatro cenários distintos de operação no caso escolhido. A Tabela 16 (no Anexo B) apresenta os períodos indicados pela equipe de O&M. O período “Treino” escolhido é aquele onde o sistema apresenta operação normal, já os três períodos subsequentes foram considerados para "Teste" do algoritmo/modelo, nestes o sistema apresenta algum tipo de desvio.

3.6.8 Métricas para avaliação

Conforme (ZAKI; MEIRA, 2014) e a documentação do *Scikit Learn*, para avaliação de um modelo de classificação, pode-se empregar a matriz de confusão (Tabela 11). A partir desta matriz pode-se calcular métricas relevantes como: *precision*, *recall* e *F1_{score}*, por meio das equações (6), (7) e (8), respectivamente.

Tabela 11 – Matriz de confusão.

Predicted Class	True Class	
	Positive (c_1)	Negative (c_2)
Positive (c_1)	True Positive (<i>TP</i>)	False Positive (<i>FP</i>)
Negative (c_2)	False Negative (<i>FN</i>)	True Negative (<i>TN</i>)

Fonte: [(ZAKI; MEIRA, 2014).

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1_{score} = 2 \left(\frac{Precision \cdot Recall}{Precision + Recall} \right) \quad (8)$$

É importante citar o risco envolvido no que tange as questões de falso positivo (FP) e falso negativo (FN). No ambiente industrial (e na vida), a experiência com falso indicativo pode ser altamente comprometedora, o que pode abalar a confiança sobre um determinado sistema ou dispositivo.

As métricas acima não são aplicadas para casos de regressão, deste modo, para avaliação dos modelos de aprendizado do tipo regressão pode-se utilizar métricas como *R2_{score}* e *RMSE* (*root mean square error* ou raiz quadrada do erro médio), que são apresentadas na equação (9) e equação (10), respectivamente. Adicionalmente, o erro máximo pode ser utilizado como métrica

também. Notar que quando a predição está adequada o valor $R2_{score}$ se aproxima do unitário e o $RMSE$ se aproxima de zero.

$$R2_{score} = 1 - \frac{\sum_{i=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2} \quad (10)$$

Conforme observado durante o desenvolvimento deste estudo, avaliar o desempenho de um modelo de regressão é uma tarefa pouco trivial, não significando que uma resposta com $R2_{score}$ próximo do unitário ou $RMSE$ próximo de zero são bons modelos.

Durante a elaboração deste estudo, em discussão com colegas que atuam na área e após a avaliação de alguns modelos de regressão, notou-se que um modelo sobre treinado (com *overfitting*) apresenta bom resultado em termos de métrica, mas não necessariamente apresenta um bom desempenho ou capacidade de generalização.

Para avaliação dos resultados, além das métricas foi necessária uma análise humana, para efetivamente identificar qual algoritmo foi capaz de representar a generalização esperada. Dois pontos relevantes para avaliação de regressores com série temporal e métricas são: o período e contexto observado. Quanto maior o período e mais variado for o contexto, menos confiáveis as métricas serão.

3.7 CONSIDERAÇÕES

Este Capítulo mostra a complexidade por trás das atividades do analista de dados e do cientista de dados. Fica claro a importância da interação entre os times de dados e do conhecimento especialista, sem essa cooperação a solução em construção pode ser fraca do ponto de vista da aplicação final.

O Capítulo 4 apresenta o *workflow* proposto neste estudo para construção de um algoritmo com aprendizado de máquina para aplicação industrial. Adicionalmente, algumas recomendações com base nos aprendizados obtidos neste estudo são listas.

O *dataset* para treino e teste são eventos históricos adequadamente rotulados pela equipe de O&M ou OEM. Se isso for feito, o *dataset* se torna uma valiosa "memória", útil para construção dos modelos, ao invés de um simples banco de dados sem contexto operacional ou significado.

Para obtenção de resultado satisfatório é importante adotar uma abordagem analítica bem estruturada. Alguns aspectos importantes que devem ser observados são: criação do banco de dados rotulados, priorizar a qualidade dos dados do que a quantidade, priorizar causalidade do que a correlação e cuidado com a sincronização nos dados proveniente de diferentes fontes.

Do ponto de vista de ciência de dados, existem dados estruturados e não estruturados. Para ativos industriais e soluções IIoT, a maioria dos dados são coletados de maneira sistemática, o que naturalmente geram dados estruturados.

4.1 CONSTRUÇÃO DO ALGORITMO

Os códigos foram feitos em Python 3, por meio da interface JupyterLab e Visual Studio. As principais bibliotecas utilizadas foram: *pandas*, *numpy*, *datetime*, *matplotlib*, *plotly*, *seaborn*, *sklearn* (*scikit-learn*) e *Tensorflow.keras*. O TensorFlow é uma ferramenta de aprendizado de máquina desenvolvida pela Google Brain, e o Keras, permite o seu uso de modo facilitado, modular e extensível; trata-se de uma biblioteca que permitir a experimentação rápida de redes neurais. Todas as ferramentas utilizadas são *open source*.

Para o modelo de classificação é utilizado o método *Random forest* do *Sklearn* e para o modelo de regressão, que explora o uso de uma rede neural, é utilizado o *Keras* e *TensorFlow*. Os códigos criados processaram os dados apenas em ambiente local, sem a necessidade de nuvem ou servidor.

Antes de aplicar o *Keras* e *TensorFlow*, foi utilizado um exemplo de rede neural utilizando apenas as bibliotecas *pandas* e *numpy*, para melhor entendimento dos detalhes e dinâmicas internas de uma rede neural. Após este processo, ficaram claras as inúmeras vantagens de uma biblioteca dedicada para construção, treino, teste e avaliação de uma rede neural.

Conforme mencionado anteriormente, antes de iniciar a construção do algoritmo é importante realizar a análise exploratória com foco principal no contexto operacional do ativo. Para criação do código, foram seguidas as etapas apresentadas na Tabela 12.

Espera-se que o algoritmo de regressão treinado possa ser utilizado para indicar o grau de semelhança com um determinado cenário conhecido (treino) e o algoritmo de classificação para indicar com qual cenário específico os dados inéditos mais se assemelham.

Antes do processo de treinamento e teste do algoritmo, algumas práticas foram seguidas, como por exemplo: armazenar separadamente o índice com *timestamp* do vetor de entrada/saída para posterior utilização, já que os modelos utilizam apenas as variáveis *var1*, *var2*, *var3* e *var4*, sem o índice. Outra medida adotada para atender um dos requisitos do modelo é a conversão do vetor de entrada e saída para *numpy array float32*.

Tabela 12 – Principais etapas adotadas na criação do algoritmo.

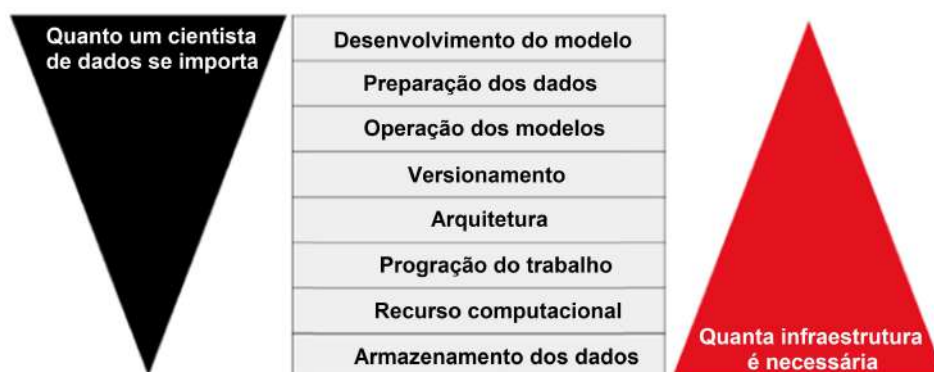
Etapa	Referência
Leitura dos dados e rótulos da operação	Seção 2.3
Preparação dos dados	Seção 3.3
Conversão para pu	Tabela 15
Remoção de <i>outliers</i>	Seção 3.4
Criação do vetor de entrada	Subseção 3.6.6
Criação do modelo	Subseção 3.6.3 e 3.6.5
Treinamento do modelo	Subseção 3.6.7
Teste do modelo	Subseção 3.6.8
Avaliação por métricas e visual	Subseção 3.6.8
Escolha do modelo e preparação para produção	Capítulo 4

Fonte: elaborado pelo autor, 2020.

Nos experimentos realizados durante este estudo ficou claro que o modelo é sensível à estrutura recebida no vetor de entrada e saída. Para um processo de aprendizado adequado, a problemática que se deseja resolver deve estar adequadamente representada no vetor de entrada e saída, considerando os critérios e dinâmicas de treinamento do algoritmo/modelo, do contrário não haverá aprendizado.

Além do Python, outras ferramentas disponíveis na internet podem ser utilizadas para acelerar o processo de análise exploratória e criação dos modelos. Alguns exemplos são: Power BI da Microsoft, o Watson da IBM, o Weka da Universidade de Waikato (Nova Zelândia), o KNIME da Universidade de Konstanz (Alemanha), entre outras. Para auxiliar na organização e padronização das pastas dos projetos em ciência de dados, o *Cookiecutter Data Science* pode ser uma opção. Na Figura 70, o time de dados do *Netflix* apresenta a importância e necessidades para ciência de dados.

Figura 70 – Importância e necessidades para o time de dados.



Fonte: (BERG et al., 2020, *online*).

Outra ferramenta que pode auxiliar no processo de criação de soluções com ciência de dados é o *Metaflow*, criado pelo *Netflix* em dezembro de 2019. Trata-se de uma ferramenta de infraestrutura para criação do *pipeline* de análise e aprendizado de máquina, que permite gerenciar os experimentos e facilitar o processo de escala e *deploy*.

4.2 RECOMENDAÇÕES PARA IMPLEMENTAÇÃO DO ALGORITMO EM PRODUÇÃO

Disponibilizar um algoritmo com aprendizado de máquina em produção não é trivial, requer equipe especializada e uma infraestrutura em nuvem ou servidor local apta para tal operação. Como o foco principal deste estudo é a abordagem conceitual de um algoritmo de aprendizado de máquina, algumas recomendações de *software* e de aquisição de dados devem ser observadas antes de colocar esta abordagem em produção.

Na terminologia IIoT, o processo de preparação e ativação de um sistema em produção é chamado de *deploy*. O nível de esforço necessário nesta etapa depende diretamente da experiência da equipe e do nível de tecnologia da plataforma IIoT utilizada. Por isso, utilizar uma plataforma moderna pode significar simplificação e agilidade no processo de *deploy*.

Na criação de um sistema especialista nota-se a importância do envolvimento direto dos especialistas humanos, O&M e OEM. Este item é muito relevante, pois a abordagem de aprendizado de máquina defendida por este estudo considera o tipo supervisionado, logo a matéria prima de entrada no algoritmo são dados rotulados, que dependem do O&M e OEM.

Para avaliação do regressor, pode-se utilizar as métricas R^2_{score} e $RMSE$, com tamanho de janela e contexto adequados. Já para classificação, pode-se utilizar a *precisão*, *recall* e $f1_{score}$.

Durante a análise exploratória em contexto operacional, que obteve a contribuição do time de O&M, observou-se que alguns sensores ou subsistema podem sofrer alteração operacional de modo deliberado, pela própria equipe de O&M e por outros motivos diversos. Esta informação é relevante e deve ser levada em conta durante a criação do algoritmo. Do contrário, uma ação deliberada da equipe de O&M pode ser interpretada pelo algoritmo como desvio operacional em potencial.

Adicionalmente, outros pontos devem ser considerados durante o trabalho de criação do algoritmos, como por exemplo: ganho nos sensores, configuração diferente dos equipamentos na mesma frota, condição de desgaste dos componentes e, em particular, os detalhes construtivos de cada ativo.

Para colocar o algoritmo criado neste estudo em produção em uma frota de geração eólica, uma série de aprimoramentos no código e na aquisição de dados deverão ser feitas, conforme apresentado na Tabela 13. Além destes pontos, testes avançados com diferentes cenários são necessários, isso é essencial para tornar o sistema robusto e confiável.

As ações listadas tem por objetivo criar um algoritmo único, apto para operação na frota de aerogeradores. Esta abordagem é altamente desejável para o time de dados, do ponto de vista de implementação e manutenção da solução.

Tabela 13 – Aprimoramentos de *software* e aquisição de dados necessários para tornar o estudo uma solução.

Aprimoramento	Local
Combinar o modelo de regressão (rede neural TLFN focada) e o modelo de classificação em um (<i>Random Forest</i>) único algoritmo com indicador de desvio operacional.	Código
Adequação do código para minimizar o erro na saída do modelo decorrente dos instantes com dados faltantes;	Código
Treinamento, teste e ajuste fino do algoritmo com <i>dataset</i> adequado;	Código
Adequação do código para o vetor de entrada do modelo abstrair a medição pontual de cada sensor de temperatura;	Código
Inclusão das variáveis velocidade da bomba e temperatura de água fria do sistema de arrefecimento no sistema de aquisição de cada aerogerador;	Aquisição de dados
Adequação do código para "rodar" na plataforma IIoT ou no servidor local do parque.	Sistema
Criação do <i>dataset</i> com rótulos para o treino e teste do modelo, com múltiplos aerogeradores e com período mínimo de doze meses de geração;	Sistema

Fonte: elaborado pelo autor, 2020.

4.3 CONSIDERAÇÕES

Conforme (CAMBUIM, 2020, *online*) e notado durante o desenvolvimento deste estudo, é nítido que extrair o conhecimento especialista, transferi-lo para um algoritmo e aplicá-lo em ambiente industrial de modo relevante não é trivial.

Durante o desenvolvimento do algoritmo ficou claro a importância de utilizar uma ferramenta de infraestrutura dedicada para a avaliação de desempenho dos modelos em criação. Tal ferramenta deve atender as necessidades do time de dados, facilitar a criação, o *deploy* e proporcionar a escalabilidade da solução de modo sustentável.

O Capítulo 5 apresenta uma amostra do resultado obtido em dois casos, um com algoritmo de classificação (*Random Forest*) e outro de regressão (rede neural TLFN focada). Mesmo tratando-se de uma abordagem conceitual, tal estudo é muito importante para compreender os detalhes envolvidos em uma aplicação industrial com aprendizado de máquina.

5 RESULTADOS E DISCUSSÕES

Devido a algumas limitações técnicas e de tempo não foi possível avançar na criação e avaliação dos modelos de modo aprofundado, por este motivo os resultados apresentados aqui estão em nível conceitual.

No Caso 1, uma série temporal multivariada, a partir dos dados reais de um hidrogerador, é utilizada como entrada em um modelo de classificação com método *Random Forest*, neste caso é realizado apenas uma avaliação qualitativa, sem a utilização das métricas. Já para o Caso 2, uma série temporal multivariada, a partir dos dados reais de um aerogerador, é utilizada como entrada em um modelo de regressão com rede neural. A avaliação do modelo é feita de modo qualitativo e quantitativo por meio das métricas apresentadas na Subseção 3.6.8.

5.1 CASO 1 – HIDROGERADOR, TEMPERATURA DA GAXETA

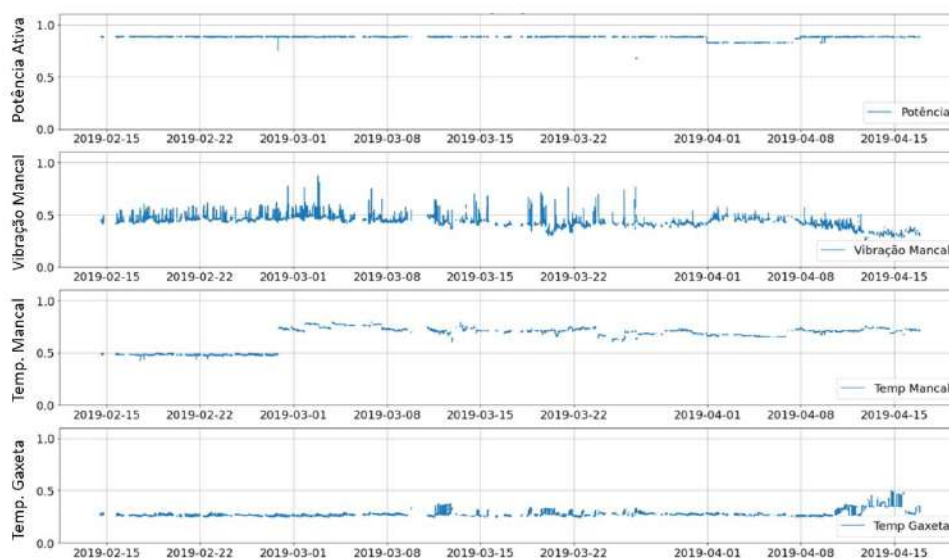
A Figura 37 apresenta os dados brutos coletados pelo sistema de aquisição no ativo. Observa-se a presença de alguns *outliers* e algumas lacunas com dados faltantes.

No período de 60 dias, a unidade geradora operou a maior parte do tempo com potência ativa de 0.88pu e, por uma semana, em 0.86pu. A vibração ficou a maior parte do tempo abaixo de 0.5pu, ou seja, 1.25mm/s *RMS*.

No início do monitoramento, o sensor de temperatura do mancal apresenta valor estável de 50°C, algo diferente do esperado. Tudo indica que isso ocorreu devido a uma irregularidade na instalação do elemento sensor. Ao final de fevereiro, o valor passou a apresentar comportamento coerente, com valores entre 60°C e 80°C.

Os dados utilizados como entrada no modelo são apresentados na Figura 71. Devido ao processo de limpeza e criação do vetor de entrada, algumas lacunas (dados faltantes) são geradas.

Figura 71 – Dados preparados de fevereiro a abril de 2019 - hidro.

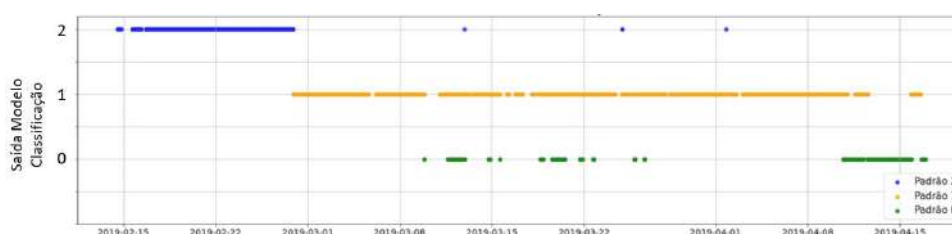


Fonte: elaborado pelo autor, 2020.

O sensor de temperatura da gaxeta apresenta comportamento normal, durante grande parte do tempo, com valor na ordem de 27°C . Há dois momentos que este sensor apresenta uma medição anormal, na primeira quinzena de março e no final da série, Figura 37. Para o treinamento do modelo é realizada a limpeza e a organização dos dados em uma janela de trinta minutos, para criação do vetor de entrada para o modelo de classificação. Este vetor é construído de modo semelhante ao apresentado na Subseção 3.6.6.

A saída do modelo de classificação com três padrões é apresentada na Figura 72. O padrão zero, em verde, pode ser relacionado com desvio operacional apresentado pelo sensor de temperatura da gaxeta. O padrão um, em laranja, pode ser relacionado com a operação normal, que trata-se da maior parte do tempo. O padrão dois, em azul, pode ser relacionado com desvio operacional no sensor de temperatura do mancal.

Figura 72 – Saída do modelo de classificação com 3 padrões.



Fonte: elaborado pelo autor, 2020.

A Figura 73 apresenta a saída do modelo com quatro padrões, o padrão zero e dois são semelhantes ao da saída anterior, a diferença está no padrão um. Neste caso, parte das ocorrências se mantiveram no padrão um e parte foram para o padrão três, na cor preta. Parte das ocorrências como padrão um podem ser relacionadas com a operação na potência gerada de 0.88pu, já o padrão três, com a potência de 0.86pu.

Figura 73 – Saída do modelo de classificação com 4 padrões.



Fonte: elaborado pelo autor, 2020.

O método *Random Forest* pode ser uma opção para classificar a operação de um determinado ativo. Para um modelo identificar adequadamente os padrões de operação, os dados devem estar escalados e organizados adequadamente no vetor de entrada. A medida que o número de classes aumenta, mais dados são necessários e maior será o trabalho de avaliação do algoritmo.

5.2 CASO 2 – EÓLICO, SISTEMA DE ARREFECIMENTO

Este caso aborda a identificação de desvio operacional no sistema de arrefecimento do aerogerador, de modo mais específico, na velocidade da bomba do fluido ou na abertura da válvula proporcional.

A Figura 78 (no Anexo B) apresenta os dados de operação de 20 meses do aerogerador em estudo. No primeiro *subplot* na parte superior é apresentada a potência ativa gerada em azul, e sua média em 24 horas, na cor laranja; no segundo, a abertura da válvula proporcional; no terceiro, a velocidade da bomba do sistema de arrefecimento; e no último, a temperatura interna da nacele na cor laranja e a temperatura média do estator do gerador na cor azul. Nota-se que a variedade de dados e o período utilizados para o processo de treino do algoritmo estão aquém do ideal. Essa afirmação é feita, pois o período de maior geração no Brasil acontece, em geral, no segundo semestre do ano. Por este motivo, o modelo possuirá experiência operacional limitada.

Observando a Figura 79 (no Anexo B), nota-se que a partir da metade de abril de 2020, a válvula proporcional apresentou desvio operacional, com abertura entre 0,5pu e 1pu. Após verificar os registros de manutenção, foi constatado que uma manutenção ocorreu próximo desta data. Tudo indica que tal intervenção possa ter relação com a alteração de comportamento.

Além do desvio operacional na válvula proporcional, também é notado um desvio na velocidade da bomba. Avaliando os dados de operação da bomba é possível notar que a velocidade da bomba sai do comportamento normal na metade de março, com elevação gradual da velocidade até a metade de abril. Percebe-se ainda que a bomba opera com velocidade próxima de 1pu por mais de quatro semanas consecutivas. Como o reparo no sistema foi realizado na metade de maio, pode-se afirmar que o sistema operou fora do padrão operacional normal por aproximadamente seis semanas.

Para o treinamento do modelo, o vetor de entrada é criado com uma janela específica, conforme apresentado na Subseção 3.6.6. O vetor de entrada possui 21 pontos/neurônios na camada de entrada. Para o cálculo das métricas de avaliação, foram estabelecidos três períodos de interesse, conforme apresentado na Tabela 10 (no Anexo B).

A Figura 80 (no Anexo B) apresenta os dados utilizados como entrada no modelo. Nota-se que há algumas lacunas de dados e isso acontece por dois motivos: pelo processo de remoção de *outliers* e por conta dos períodos com dados ausentes, causados pelo processo de aquisição e criação do vetor de entrada. As variáveis de entrada do modelo foram: potência gerada, abertura da válvula proporcional e velocidade da bomba de arrefecimento. O atributo meta é a elevação de temperatura média do estator, ou seja, a diferença entre a temperatura média do estator e a temperatura interna da nacele.

Para avaliação da rede neural foram criadas duas redes, a primeira rede possui apenas 1 neurônio e a segunda 30 neurônios. A rede 1 possui 1 neurônio na camada oculta e 1 neurônio na camada de saída. Esta composição resulta em um modelo com 24 parâmetros/pesos sinápticos treináveis. A segunda rede é mais ampla, possui 20 neurônios na primeira camada oculta,

10 neurônios na segunda camada oculta e 1 neurônio na camada de saída. Esta composição resulta em um modelo com 661 parâmetros/pesos sinápticos treináveis. A Tabela 14 apresenta a configuração da rede 1 e 2, de modo resumido.

Tabela 14 – Configuração da Rede 1 e Rede 2.

Parâmetros	Rede 1	Rede 2
Número de neurônios na camada de entrada	21	21
Número de neurônios na camada oculta (1ª)	1	20
Camada de neurônios na camada oculta (2ª)	-	10
Número de neurônios na camada de saída	1	1
Total de parâmetros/pesos sinápticos treináveis	24	661

Fonte: elaborado pelo autor, 2020.

Durante a etapa de exploração da rede neural foram experimentadas diferentes tipos de função de ativação como: linear, sigmoide, softmax e relu. A função linear apresentou o melhor resultado para o caso em análise. Outro parâmetro investigado é o lote de treinamento. Foram realizados experimentos com lotes de 4, 32 e 256. O resultado final pouco se altera, no entanto, o tempo de treino da rede é significativamente afetado. Para um lote de 4 amostras é observado um período de 105 segundos; para 32 amostras, 12 segundos; e para 256 amostras, 4 segundos.

No caso da rede 1 e 2 o vetor de entrada é o mesmo, ou seja, 21 neurônios. Para as duas redes treinadas é adotada a função de ativação linear, tamanho do lote de treino de 32 (*batch size*), bias ativado, otimizador ADAM (sigla derivada de *adaptive moment estimation*), taxa de aprendizado padrão do Tensorflow, ciclos de treinamento limitado em 50 épocas e função de perda *MSE* (erro médio quadrado).

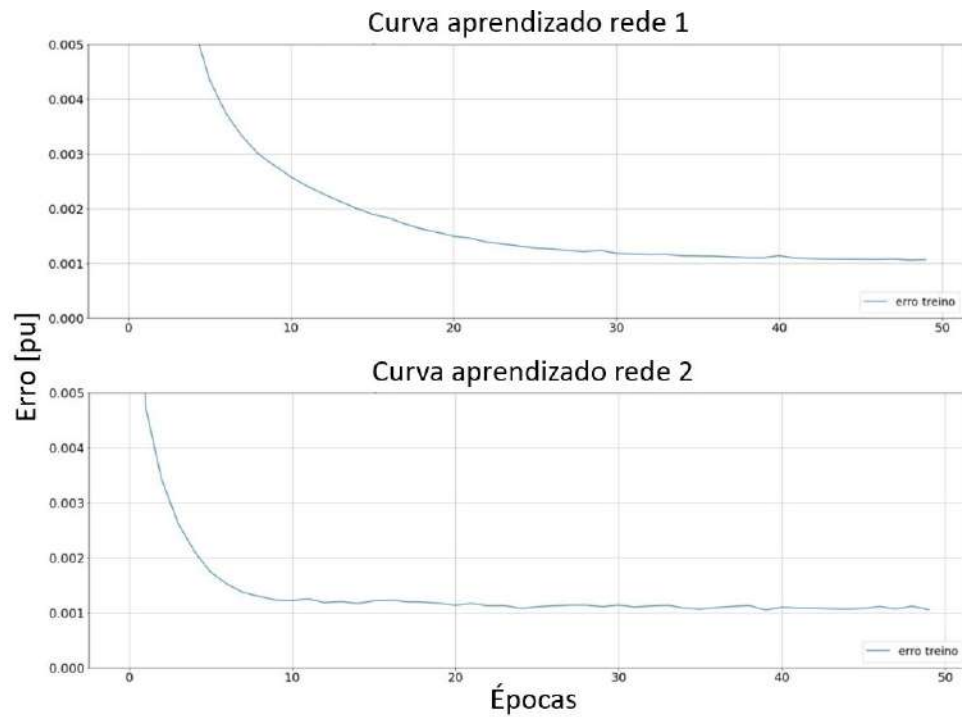
Conforme observado na Figura 74, a rede 2 apresenta uma curva de aprendizado mais acelerada que a rede 1. O processo completo de preparação dos dados, treino do modelo, cálculo da saída e métricas é de aproximadamente 16 segundos, para ambas as redes. O erro *MSE* residual também ficou semelhante entre as redes, algo em torno de 0.1°C.

A partir da quinquagésima época de treino, a diminuição do erro é quase nula para ambas as redes, ou seja, não é notado melhora no aprendizado. Este comportamento pode estar relacionado com a quantidade limitada de dados para treino.

O atributo meta do modelo é a elevação média de temperatura do estator. A Figura 75 (em azul) apresenta o erro entre o valor previsto pelo modelo e o valor real. Para suavizar a saída do modelo e evidenciar a linha de tendência, é aplicado uma média móvel de 6 horas, o que resulta no indicador de desvio operacional proposto neste estudo. Segundo (NIENHAUS; HILBERT, 2012), uma margem de erro de 5°C pode ser considerado satisfatório. Neste estudo é definido uma margem de erro de 3°C (linha amarela).

Para o período de Treino (8977 pontos), espera-se um erro próximo de zero, já que os dados possuem somente operação normal e trata-se de um cenário conhecido para o modelo.

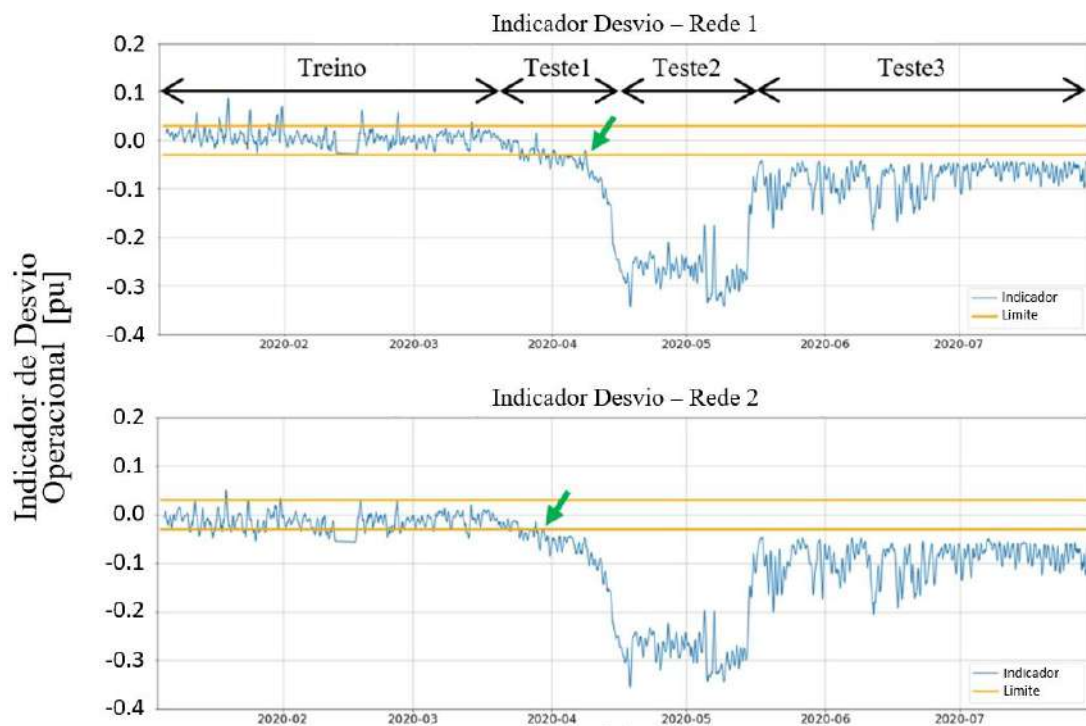
Figura 74 – Curva de aprendizado do modelo com a rede 1 e 2.



Fonte: elaborado pelo autor, 2020.

Neste período, nota-se um $RMSE$ de 1.8°C e $2,4^{\circ}\text{C}$, erro máximo de 8.7°C e 5.6°C e $R2_{score}$ de 0.966 e 0.963, respectivamente para rede 1 e 2.

Figura 75 – Saída modelo para rede 1 e rede 2, indicador desvio operacional.



Fonte: elaborado pelo autor, 2020.

No período Teste 1 (3440 pontos), a bomba apresenta elevação gradual da velocidade, ou seja, trata-se de um desvio operacional ocorrendo de modo evolutivo, algo que se deseja detectar com o algoritmo proposto. Neste cenário, a saída do modelo é sensibilizada pelo desvio operacional mencionado, notando-se um *RMSE* de 4.8°C e 6.3°C, erro máximo de 13.3°C e 15.7°C e R2 score de 0.867 e 0.789, respectivamente para rede 1 e 2. No entanto, a diferença mais significativa está na observação de rompimento do indicador de desvio operacional além do limite estabelecido de 3°C (linha amarela). Nota-se que a rede 2 rompe o limite na última semana de março, enquanto que na rede 1, isso acontece apenas no início da segunda semana, uma diferença de aproximadamente 10 dias.

No período Teste 2 (4055 pontos), além da alteração na velocidade da bomba, a válvula proporcional também apresenta um comportamento com desvio. Conforme esperado, neste instante a saída do modelo é significativamente afetada, resultando um *RMSE* de 27.3°C e 27.9°C, erro máximo de 34.3°C e 35.6°C e R2 score de -2.131 e -2.27, respectivamente para rede 1 e 2.

Ao final do período de Teste 2, o ativo recebeu reparo e a velocidade da bomba retorna ao padrão de operação normal, porém a válvula proporcional continua com desvio.

Durante o período Teste 3 (10422 pontos), a saída da rede também é alterada pelo desvio operacional do sistema e isso pode ser notado no valor *RMSE* de 8.3°C e 9.7°C, no erro máximo de 18.5°C e 20.6°C e R2 score de 0.725 e 0.627, respectivamente para rede 1 e 2. As métricas observadas são apresentadas na Tabela 17 (no Anexo B).

De modo geral, pode-se afirmar que a rede neural atendeu ao propósito em vista, observa-se que a saída do modelo é sensibilizada de modo diferente para cada um dos três cenários de teste, mas em especial para o período de Teste 1, onde o desvio na velocidade da bomba aconteceu de modo evolutivo e gradual, com duração de três semanas. A abordagem com rede neural criada em nível conceitual conseguiu identificar o desvio operacional com antecedência. Tal identificação é desejável, pois neste caso o evento de manutenção poderia ser antecipado em seis semanas.

5.3 CONSIDERAÇÕES

Ficou nítido que a aplicação com aerogerador possui uma dinâmica e variação de operação bem maior que aquela notada em um hidrogerador, isso acontece por conta da natureza da energia e dos controles aplicados no ativo.

O modelo de maior interesse explorado neste estudo é a rede neural TLFN focada. O resultado, em nível prova de conceito, mostra que esta abordagem pode ser uma opção de algoritmo para identificação do desvio operacional no sistema de arrefecimento do aerogerador, mais especificadamente na velocidade da bomba e na abertura da válvula proporcional. Ficou claro que para colocar este modelo em produção é necessário realizar uma série de aprimoramentos, validações e investimentos. Tais melhorias são apresentadas na Seção 4.2.

6 CONCLUSÃO

Este estudo busca identificar e compreender os principais passos necessários para construção de um algoritmo com aprendizado de máquina, a partir dos dados (série temporal multivariada) do sistema SCADA. Desta forma, concluiu-se que o *workflow* proposto (Figura 69) e a rede neural TLFN focada utilizada, mesmo em caráter conceitual, podem ser uma opção de metodologia e algoritmo para detecção de desvio operacional em um subsistema industrial, principalmente para ocorrências de evolução gradual, contexto complexo e dados multivariados.

Esta abordagem é interessante, pois alguns eventos indesejados na indústria não são capturados pelos dispositivos tradicionais de monitoramento. O caso em estudo focou na identificação de desvio operacional no sistema de arrefecimento do aerogerador, mais especificadamente na velocidade da bomba e na abertura da válvula proporcional. Ficou claro que para colocar este sistema em produção é necessário realizar uma série de aprimoramentos, validações e alguns investimentos (Seção 4.2).

O algoritmo de maior interesse explorado neste estudo é a rede neural TLFN focada. Experimentou-se duas configurações, com um neurônio (24 pesos sinápticos) e trinta neurônios (661), conforme apresentado no Caso 2 (Seção 5.2). Ambas redes identificaram o desvio, mas a rede com maior número de neurônios apresentou um comportamento levemente superior.

Para extrair bons resultados de tal tecnologia, ficou claro que parte do desafio está no processo de obtenção dos dados, na criação do vetor de entrada para o algoritmo (Subseção 3.6.6) e na seleção e configuração do algoritmo de aprendizado de máquina. Este processo devem observar a problemática sob investigação e contar com a experiência do time de dados.

É evidente que a criação de uma solução IIoT com valor agregado (inteligência) não é trivial, requer múltiplas competências para o desenvolvimento e uma plataforma IIoT robusta. Há dois personagens cruciais para construção desta camada de inteligência, o especialista do ativo (OEM e O&M) e o time de dados (analista e cientista de dados). O primeiro é responsável por identificar um desvio operacional passível de aplicação via solução IIoT e o segundo é responsável por analisar os dados e criar um algoritmo/modelo apto para detectar tal desvio. Durante esta atividade, a cooperação entre estas duas competências é fundamental, do contrário a solução pode ser fraca do ponto de vista da aplicação final, diminuindo o valor agregado. O *workflow* proposto neste estudo sugere uma abordagem considerando tal contexto colaborativo. O resultado deste *workflow* está em um primeiro momento nos relatórios *Data-Informed* e seguidamente nos relatórios *Data-Driven*, elementos relevantes em uma solução IIoT.

Conforme observado na pesquisa de patentes e no relatório Gartner sobre atuação em IIoT e sistemas inteligentes, os *players* de mercado estão investindo no segmento. As empresas Hitachi, ABB, Siemens se destacam como fabricantes de máquinas elétricas, além da GE obviamente. Sobre o pioneirismo em patentes sobre plataformas IoT, destacam-se a GE e a IBM.

Conforme apresentado na Tabela 10, tanto o modelo físico quanto o modelo baseado em dados possuem vantagens e desvantagens. Recentemente, os modelos baseado em dados e

com aprendizado de máquina vêm recebendo uma atenção especial, por vários motivos, mas principalmente devido à sua capacidade de abstrair alguns parâmetros físicos e pela habilidade de replicar o poder cognitivo humano em ambiente computacional. Durante o desenvolvimento do algoritmo ficou claro a importância de utilizar uma ferramenta de infraestrutura dedicada para a avaliação de desempenho dos modelos em criação. Tal ferramenta deve atender as necessidades do time de dados, facilitar a criação, o *deploy* e proporcionar a escalabilidade da solução de modo sustentável.

Ao que tudo indica, os testes realizados e resultados alcançados mostraram que a abordagem explorada neste estudo pode ser aplicável em outros sistemas presentes na indústria, como ferramenta de detecção do desvio operacional de modo autônomo e computacional, a partir da série temporal multivariada proveniente do sistema SCADA. Os artigos, dissertações, teses, fóruns, patentes e outros estudos utilizados mostram que tal tecnologias irá impactar positivamente a gestão dos ativos industriais.

Por fim, nota-se que a implementação de solução com algoritmos de aprendizado de máquina em ambiente industrial não é trivial. Tal abordagem demanda equipe multidisciplinar, acesso ao conhecimento especialista, quantidade significativa de dados (com qualidade), infraestrutura para treinamento e processamento dos algoritmos e uma metodologia (*workflow* proposto) para tornar o banco de dados (*big data*) em informação útil, tanto para equipe de O&M como para exploração via algoritmos.

REFERÊNCIAS

- ABEEÓLICA. **Boletim anual de geração eólica**. 2020, *online*. Disponível em: <<http://abeeolica.org.br/>>. Citado na página 21.
- ALMEIDA, Antonio Tadeu Lyrio. **Apostila Máquinas Síncronas**. 2020, *online*. Disponível em: <<https://pt.scribd.com/document/239437700/Maquinas-Sincronas-Apostila-EFEI>>. Citado na página 27.
- ALRAYESS, Hesham et al. Using machine learning techniques and deep learning in forecasting the hydroelectric power generation in almus dam, turkey. **SAFETY**, 2018. Citado na página 72.
- ANDRADE, Luiz Augusto de et al. Estudo comparativo dos métodos de medição de vazão—uma aplicação em comissionamento de turbinas hidráulicas. **Centro de Pesquisas em Hidráulica e Recursos Hídricos—CPH, Grupo de Turbomáquinas, Universidade Federal de Minas Gerais—UFMG**, 7p, 2002. Citado 2 vezes nas páginas 33 e 34.
- ANEEL. Resolução normativa nº 875, de 10 de março de 2020. 2020, *online*. Disponível em: <<http://www2.aneel.gov.br/cedoc/ren2020875.pdf>>. Citado na página 22.
- ASLAN, Yilmaz; ARSLAN, Oguz; YASAR, Celal. A sensitivity analysis for the design of small-scale hydropower plant: Kayabogazi case study. **Renewable Energy**, Elsevier, v. 33, n. 4, p. 791–801, 2008. Citado na página 23.
- BARANKIEVICZ, Victor Skaetta et al. Controle de sobrefrequência por droop simulado em fast para aerogeradores agw 110/2.1. Florianópolis, SC., 2018. Citado 4 vezes nas páginas 34, 36, 37 e 38.
- BERG, David et al. **Open sourcing metaflow a human centric framework for datascience**. 2020, *online*. Disponível em: <<https://netflixtechblog.com/open-sourcing-metaflow-a-human-centric-framework-for-data-science-fa72e04a5d9>>. Citado na página 89.
- BROWNLEE. **Time series analysis**. 2020, *online*. Disponível em: <https://machinelearningmastery.com/?s=time+series&post_type=post&submit=Search>. Citado 2 vezes nas páginas 55 e 57.
- CAMBUIM, Lucas. **Sistema Especialista**. 2020, *online*. Disponível em: <<https://www.cin.ufpe.br/~lfsc/cursos/introducaoainteligenciaartificial/IA-Aula8-SistemasEspecialistas.pdf.pdf>>. Citado na página 91.
- CHAPMAN, Pete et al. The crisp-dm user guide. In: **4th CRISP-DM SIG Workshop in Brussels in March**. [S.l.: s.n.], 1999. v. 1999. Citado na página 54.
- CUI, Nan. Applying gradient descent in convolutional neural networks. **Journal of Physics: Conference Series**, v. 1004, p. 012027, 04 2018. Citado na página 82.
- EUA, Departamento de Energia dos. **Energy Report**. 2020, *online*. Disponível em: <<https://www.energy.gov/>>. Citado na página 16.
- FAWAZ, Hassan Ismail et al. Deep learning for time series classification: a review. **Data Mining and Knowledge Discovery**, Springer, v. 33, n. 4, p. 917–963, 2019. Citado 3 vezes nas páginas 19, 55 e 76.

GARTNER. **Magic Quadrant for Industrial IoT Platforms**. 2020, *online*. Disponível em: <<https://braincube.com/resource/download-the-gartner-2020-magic-quadrant-for-industrial-iot-platforms/>>. Citado na página 52.

GREENLAND, Sander et al. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. **European journal of epidemiology**, Springer, v. 31, n. 4, p. 337–350, 2016. Citado na página 74.

GUPTA, Ankit et al. Performance analysis of neural network and fuzzy logic based mppt techniques for solar pv systems. In: IEEE. **2014 6th IEEE Power India International Conference (PIICON)**. [S.l.], 2014. p. 1–6. Citado 4 vezes nas páginas 19, 78, 79 e 80.

HAYKIN, Simon. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007. Citado 6 vezes nas páginas 19, 55, 75, 78, 79 e 82.

HENAO, Humberto et al. Trends in fault diagnosis for electrical machines: A review of diagnostic techniques. **IEEE industrial electronics magazine**, IEEE, v. 8, n. 2, p. 31–42, 2014. Citado na página 78.

HIDROENERGIA. **Tipos de Turbinas Hidráulicas**. 2020, *online*. Disponível em: <<http://www.hidroenergia.com.br/tipos-de-turbinas-hidraulicas/>>. Citado na página 24.

IEA. **Data and statistics**. 2020, *online*. Disponível em: <<https://www.iea.org/data-and-statistics>>. Citado na página 21.

IEC60034-3. Rotating electrical machines – part 3. IEC, 2020. Citado na página 66.

Ji, J et al. Comparison of a data-driven model and a physical model for flood forecasting. **WIT Transactions on Ecology and the Environment**, WIT Press, v. 159, p. 133–142, 2012. Citado 3 vezes nas páginas 19, 73 e 74.

JING, Huayi. **Registrant Classification using Machine Learning**. 2020, *online*. Disponível em: <<https://blog.nzrs.net.nz/registrant-classification/>>. Citado na página 76.

KAA, Geerten van de et al. Wind turbine technology battles: Gearbox versus direct drive-opening up the black box of technology characteristics. **Technological Forecasting and Social Change**, Elsevier, v. 153, p. 119933, 2020. Citado na página 40.

KALABIN, Stanislav et al. Machine learning solutions for maintenance of power plants. 2018. Citado 4 vezes nas páginas 48, 49, 50 e 51.

KARTALOPOULOS, Stamatios V. **Understanding neural networks and fuzzy logic: basic concepts and applications**. [S.l.]: Wiley-IEEE Press, 1997. Citado na página 75.

KISHOR, Nand; SAINI, RP; SINGH, SP. A review on hydropower plant models and control. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 11, n. 5, p. 776–796, 2007. Citado na página 30.

KUNDUR, Prabha; BALU, Neal J; LAUBY, Mark G. **Power system stability and control**. [S.l.]: McGraw-hill New York, 1994. v. 7. Citado 2 vezes nas páginas 30 e 31.

KURT, O et al. **Reaching peak O&M performance in power generation**. 2011, *online*. Disponível em: <<https://pt.scribd.com/document/413528650/Reaching-Peak-OM-Performance-in-Power-Generation>>. Citado na página 43.

MAFRA, A. C. Redes neurais artificiais aplicadas ao sensoriamento não invasivo de temperatura de um motor bldc. UDESC, 2019. Citado 4 vezes nas páginas 19, 75, 77 e 79.

NIED, Ademir. Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa. Universidade Federal de Minas Gerais, 2007. Citado 2 vezes nas páginas 19 e 82.

NIENHAUS, Karl; HILBERT, Marc. Thermal analysis of a wind turbine generator by applying a model on real measurement data. In: IEEE. **2012 IEEE International Workshop on Applied Measurements for Power Systems (AMPS) Proceedings**. [S.l.], 2012. p. 1–5. Citado 6 vezes nas páginas 19, 38, 39, 40, 83 e 95.

ONS. **Resultados da operação - Histórico da Operação**. 2020, *online*. Disponível em: <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx>. Citado 2 vezes nas páginas 22 e 70.

OPREA, Simona Vasilica; BÂRA, Adela. Key technical performance indicators for power plants. **Recent Improvements of Power Plants Management and Technology**, BoD–Books on Demand, p. 9, 2017. Citado na página 44.

OSU. **Types of turbine**. 2020, *online*. Disponível em: <<http://rivers.bee.oregonstate.edu/book/export/html/35>>. Citado na página 25.

PANT, Ayush. **Workflow of a Machine Learning project**. 2019, *online*. Disponível em: <<https://towardsdatascience.com/workflow-of-a-machine-learning-project-ecldba419b94>>. Citado 2 vezes nas páginas 72 e 73.

PUCRIO. **Redes Neurais Artificiais**. 2020, *online*. Disponível em: <https://www.maxwell.vrac.puc-rio.br/5566/5566_6.PDF>. Citado 2 vezes nas páginas 74 e 75.

QUIXADÁ, Ricardo Ferreira. Estudo de selo mecânico para turbinas hidrocínéticas. 2014. Citado 2 vezes nas páginas 31 e 32.

RASCHKA, Sebastian. **How was classification, as a learning machine, developed?** 2020, *online*. Disponível em: <<https://www.quora.com/How-was-classification-as-a-learning-machine-developed>>. Citado na página 81.

SCHUH, Günther et al. Industrie 4.0 maturity index. **Managing the digital transformation of companies**. Munich: Herbert Utz, 2017. Citado 3 vezes nas páginas 16, 44 e 45.

SILVA, Marina Luiza Bortoli da et al. Análise da metodologia do mme/aneel para o cálculo dos montantes de energia assegurada e garantia física de energia de pchs: Estudo de caso de uma pch. Florianópolis, SC., 2016. Citado na página 70.

TORELL, Wendy; AVELAR, Victor. Mean time between failure: Explanation and standards. **white paper**, v. 78, 2004. Citado na página 43.

UNESP. **Turbinas Hidráulicas**. 2020, *online*. Disponível em: <<http://www.dem.feis.unesp.br/intranet/capitulo8.pdf>>. Citado na página 24.

WANG, Xian et al. Condition monitoring of wind turbines based on analysis of temperature-related parameters in supervisory control and data acquisition data. **Measurement and Control**, SAGE Publications Sage UK: London, England, v. 53, n. 1-2, p. 164–180, 2020. Citado na página 36.

WEG. **Manual Hidrogeradores**. 2016, *online*. Disponível em: <<https://static.weg.net/medias/downloadcenter/hcc/h6a/WEG-hidrogeradores-sem-escovas-11969326-manual-portugues.pdf>>. Citado 3 vezes nas páginas 23, 26 e 27.

WELTE, Thomas Michael et al. Monitorx—experience from a norwegian-swedish research project on industry 4.0 and digitalization applied to fault detection and maintenance of hydropower plants. **Hydro 2018-Progress through partnership, Gdansk, Poland, 15-17 October 2018**, 2018. Citado na página 42.

WIPO. **IPC Publication**. 2020, *online*. Disponível em: <<http://ipc.inpi.gov.br/classifications/ipc>>. Citado na página 52.

WIRTH, Rüdiger; HIPPE, Jochen. Crisp-dm: Towards a standard process model for data mining. In: SPRINGER-VERLAG LONDON, UK. **Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining**. [S.l.], 2000. p. 29–39. Citado 2 vezes nas páginas 19 e 54.

WOLLSCHLAEGER, Martin; SAUTER, Thilo; JASPERNEITE, Juergen. The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0. **IEEE industrial electronics magazine**, IEEE, v. 11, n. 1, p. 17–27, 2017. Citado na página 29.

WOODMACKENZIE. Wood Mackenzie Report, 2015. Citado na página 41.

ZAKI, Mohammed J; MEIRA, Wagner. **Data mining and analysis: fundamental concepts and algorithms**. [S.l.]: Cambridge University Press, 2014. Citado 2 vezes nas páginas 75 e 85.

ZHOU, Datong; HU, Qie; TOMLIN, Claire J. Model comparison of a data-driven and a physical model for simulating hvac systems. **arXiv preprint arXiv:1603.05951**, 2016. Citado 3 vezes nas páginas 19, 73 e 74.

GLOSSÁRIO

Desvio operacional: Cenário operacional do ativo industrial fora do padrão conhecido como normal.

Dataset ou Dataframe: Grupo de dados brutos ou processados. Gateway Dispositivo da borda para coleta dos dados.

Edge computing: Processamento realizado na borda, útil para compactar o *payload* ou gerar ação na borda sem a necessidade de conexão com a nuvem.

Payload: *Dataset* oriundo do *gateway*.

Dados do ativo: Informação (dados) proveniente dos sensores.

Feature engineering: Criação de variáveis virtuais (*features*) a partir do *dataset* e/ou adequação das mesmas para atender as exigências dos modelos.

Insight: Percepção sobre um determinado contexto.

Solução Data-Informed: Reporta o histórico de operação do ativo, com visualização simples dos dados, e geração dos *insights* iniciais.

Solução Data-Driven: Reporta os eventos importantes de operação do ativo, com visualização avançada dos dados e análise avançada, para geração de *insights* relevantes para equipe O&M e o proprietário.

Greenfield: Ativo ou planta nova.

Brownfield: Ativo ou planta existente.

Full Stack: Infraestrutura completa para soluções digitais: *back-end*, *front-end*, armazenamento dos dados, processamento na nuvem, etc;

Pipeline: Processamento de dados de modo estruturado.

Conhecimento especialista: Especialista sobre o ativo e sua aplicação, é o *know-how* do OEM e O&M para construção, *troubleshooting* e operação do ativo.

Troubleshooting: Processo de investigação de uma falha, com intuito de compreender, de modo detalhado, os elementos de causa e efeito na ocorrência.

Deploy: Colocar um sistema em produção.

ANEXOS

ANEXO A – CARTA DE AUTORIZAÇÃO E AVISO LEGAL

Para atender algumas exigências de sigilo e anonimato impostas pela empresa WEG, algumas informações confidenciais e sensíveis foram deliberadamente modificadas¹⁴ e/ou colocadas no Anexo B, sendo que o Anexo B será suprimido da versão pública.

Figura 76 – Carta de autorização WEG para utilização dos dados.



Fonte: cortesia WEG, 2020.

Aviso Legal - As opiniões e raciocínios apresentados neste trabalho constituem o julgamento do autor na data de publicação. Trata-se de conteúdo exclusivo para fins de informação e estão sujeitos a alterações sem aviso prévio. Este material não possui relação direta ou indireta com os produtos e serviços da WEG. O desempenho e métodos apresentados não são indicativos de resultados futuros.

¹⁴ Como medida adicional, alguns valores foram deliberadamente alterados para reforçar o aspecto de anonimato e sensibilidade da informação.

ANEXO B – INFORMAÇÕES CONFIDENCIAIS

Para acesso ao Anexo B, entrar em contato (nicoladelli@weg.net).

Conteúdo do Anexo B:

Figura 77 – Sistema de arrefecimento do aerogerador.

Figura 78 – Dados brutos de janeiro de 2019 a agosto de 2020 - aerogerador.

Tabela 15 – Valores nominais para conversão do *dataset* para pu.

Tabela 16 – Períodos de treino e teste do modelo - Caso 2.

Figura 79 – Dados brutos de janeiro de 2020 a agosto de 2020 - aerogerador.

Figura 80 – Dados preparados para utilização no modelo - Caso 2.

Tabela 17 – Períodos de treino e teste do modelo, status do equipamento e métricas para avaliação da rede 1 e 2.

Pesquisa de patentes:

Figura 81 – *Players* por status legal.

Figura 82 – Publicações por ano.

Figura 83 – Tecnologias e aplicações.

Figura 84 – Dependências entre os *players* por citações.