



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
COLEGIADO DO CURSO DE ENGENHARIA MECÂNICA

Pablo Louzada Schiavo

**Diagnóstico automático de falhas em
máquinas rotativas utilizando algoritmo de
inteligência artificial**

Vitória, ES

2024

Pablo Louzada Schiavo

Diagnóstico automático de falhas em máquinas rotativas utilizando algoritmo de inteligência artificial

Monografia apresentada ao Colegiado do Curso de Engenharia Mecânica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Bacharel em Engenharia Mecânica.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Colegiado do Curso de Engenharia Mecânica

Orientador: Prof. Dr. Luis Gustavo Giacon Villani

Coorientador: Prof. Dr. Arnaldo Leal Junior

Vitória, ES

2024

Pablo Louzada Schiavo

Diagnóstico automático de falhas em máquinas rotativas utilizando algoritmo de inteligência artificial/ Pablo Louzada Schiavo. – Vitória, ES, 2024-

53 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Luis Gustavo Giacon Villani

Monografia (PG) – Universidade Federal do Espírito Santo – UFES
Centro Tecnológico
Colegiado do Curso de Engenharia Mecânica, 2024.

1. Banco de dados. 2. Classificação. I. Pablo Louzada Schiavo. II. Universidade Federal do Espírito Santo. IV. Diagnóstico automático de falhas em máquinas rotativas utilizando algoritmo de inteligência artificial

Pablo Louzada Schiavo

Diagnóstico automático de falhas em máquinas rotativas utilizando algoritmo de inteligência artificial

Monografia apresentada ao Colegiado do Curso de Engenharia Mecânica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Bacharel em Engenharia Mecânica.

Trabalho aprovado. Vitória, ES, 11 de julho de 2024:

Prof. Dr. Luis Gustavo Giacon Villani
Orientador

Prof. Dr. Lucas Silveira Campos
Convidado 1

Mestre Leandro Cassa Macêdo
Convidado 2

Vitória, ES
2024

Agradecimentos

Gostaria de expressar minha profunda gratidão a todos aqueles que, de uma forma ou de outra, contribuíram para a realização deste trabalho de conclusão de curso. Primeiramente, agradeço a Deus, cuja orientação e força me sustentaram em cada etapa desta jornada. Sem a Sua presença e graça, nada disso seria possível.

Aos meus familiares, Anisio, Madalena e Pollianny que sempre estiveram ao meu lado, oferecendo amor incondicional, compreensão e apoio emocional. A vocês, minha eterna gratidão por acreditarem em mim e por me incentivarem a seguir em frente, mesmo nos momentos mais desafiadores.

Aos meus amigos, que foram pilares de suporte e encorajamento, compartilhando risos, angústias e conquistas. Obrigado por estarem presentes e por suas palavras de incentivo que tanto me motivaram.

Um agradecimento especial ao meu professor orientador Luis Gustavo Giacon Villani, cujo conhecimento, paciência e dedicação foram fundamentais para a realização deste trabalho. Sua orientação criteriosa e disponibilidade para esclarecer dúvidas foram imprescindíveis para o meu crescimento acadêmico e profissional.

Aos demais professores do Departamento de Engenharia Mecânica, que, ao longo do curso, compartilharam seu conhecimento e experiência, contribuindo significativamente para minha formação. Suas aulas e conselhos foram valiosos e inspiradores.

A todos, meu sincero agradecimento. Este trabalho é fruto do esforço conjunto de cada um de vocês.

Resumo

A manutenção preditiva está se consolidando cada vez mais como uma metodologia essencial, permitindo otimizar o número de paradas com diagnósticos precisos de falhas. Em ambientes industriais críticos, onde há impactos significativos em componentes, humanos ou ambientais, o monitoramento contínuo é crucial. Uma das técnicas amplamente adotadas na indústria para monitorar e diagnosticar falhas é a análise de assinaturas de vibração. Especificamente para máquinas rotativas, objeto deste estudo, o mapeamento do comportamento esperado diante de diversos tipos de falhas representa uma base sólida para aplicar modelos de aprendizado de máquina para classificação de falhas. Nesse contexto, o presente trabalho utiliza os dados coletados em uma bancada de simulação de falhas disponível no Laboratório de Dinâmica de Sistemas e Estruturas Mecânicas da Universidade Federal do Espírito Santo, com o objetivo de desenvolver e aprimorar um modelo preditivo capaz de classificar defeitos, enfrentando o desafio da utilização de dados de teste obtidos após remontagens da máquina, realizadas em diferentes velocidades de rotação. Espera-se, ao final deste estudo, extrair conclusões que contribuam para a construção de modelos mais robustos e generalizáveis, fundamentais para melhoria contínua em processos industriais e na segurança operacional.

Palavras-chaves: Manutenção preditiva. Vibração. Máquinas rotativas. Classificação de falha. Random Forest Classifier.

Lista de ilustrações

Figura 1 – Formas de desbalanceamento	15
Figura 2 – Comparação entre condição saudável e desbalanceada a 26,7 Hz	16
Figura 3 – Partes do rolamento de esferas	17
Figura 4 – Conjunto de treino rotulado para aprendizado supervisionado	18
Figura 5 – Conjunto de treino não rotulado para aprendizado não supervisionado	19
Figura 6 – Aprendizado semi-supervisionado	19
Figura 7 – Aprendizado por reforço	20
Figura 8 – Comparação entre algoritmos de classificação	21
Figura 9 – Árvore de decisão	22
Figura 10 – classificação de floresta aleatória	23
Figura 11 – <i>Machinery Fault Simulator</i> da SpectraQuest	27
Figura 12 – Montagem da bancada de aquisição utilizada	29
Figura 13 – Desbalanceamento com massa de 12 gramas	30
Figura 14 – Configurações consideradas nos experimentos	31
Figura 15 – Pontos de medição de vibração na carcaça da caixa de mancal	32
Figura 16 – Diagrama de fluxo de processos para monitoramento de condição	33
Figura 17 – Exemplo dos dados por condição para velocidade de rotação de 37 Hz	34
Figura 18 – Sinais no tempo e histogramas por eixo, para a condição saudável do equipamento a 37 Hz de rotação	36
Figura 19 – Matriz de confusão utilizando todas as características	42
Figura 20 – Correlação entre as características selecionadas	43
Figura 21 – Matriz de confusão utilizando características selecionadas	44

Lista de tabelas

Tabela 1 – Especificações da bancada MFS	28
Tabela 2 – Especificação do Inversor de frequência utilizado	30
Tabela 3 – Especificações do Acelerômetro Piezoelétrico utilizado	32
Tabela 4 – Quantidade de teste por condição	35
Tabela 5 – Características selecionadas para cada acelerômetro	43
Tabela 6 – Seleção de hiper-parâmetros pelo método de <i>GridSearchCV</i>	44
Tabela 7 – Métricas por seleção de características	45
Tabela 8 – Métricas por localização dos acelerômetros	46
Tabela 9 – Métricas por eixo de medição	47
Tabela 10 – Métricas para os casos testados	47

Lista de abreviaturas e siglas

ACC	Acelerômetro
BIFI	Frequência para dano na pista interna de mancal de rolamento
BPFO	Frequência para dano na pista externa de mancal de rolamento
BSF	Frequência para dano nas esferas
CF	Fator adimensional de crista (<i>Crest Factor</i>)
FFC	Centroide espectral
FFT	Fast Fourier Transform
FKUR	Curtose no espectro de frequência
FSKEW	Assimetria no espectro de frequência
FVAR	Variância do espectro de frequência
IF	Fator adimensional de impulso (<i>Impulse Factor</i>)
KF	Fator adimensional de curtose (<i>Kurtosis Factor</i>)
KV	Curtose (<i>Kurtosis</i>)
MF	Fator adimensional de marge (<i>Margin Factor</i>)
MFS	Machinery Fault Simulators
PPV	Valor pico-a-pico (<i>Peak-to-Peak Value</i>)
RMS	Valor efetivo normalizado (<i>Root Mean Square</i>)
SF	Fator adimensional de formato (<i>Shape Factor</i>)
SRA	Raiz quadrada da amplitude (<i>Square Root of Amplitude</i>)
SV	Assimetria (<i>Skewness</i>)
UFES	Universidade Federal do Espírito Santo

Sumário

1	INTRODUÇÃO	11
1.1	Objetivos	12
1.1.1	Objetivo geral	12
1.1.2	Objetivos específicos	12
1.2	Estrutura do trabalho	12
2	REFERENCIAL TEÓRICO	14
2.1	Assinaturas de vibração	14
2.1.1	Desbalanceamento	15
2.1.2	Falhas em mancais	16
2.2	Aprendizado de máquina	18
2.2.1	Classificadores supervisionados	20
2.2.1.1	Classificação de floresta aleatória	21
2.2.2	Métricas de avaliação e otimização	23
2.2.2.1	<i>Overfitting</i>	24
2.2.2.2	Otimização	25
2.3	Predições de falha em máquinas rotativas	26
3	CONSTRUÇÃO DA BASE DE DADOS	27
3.1	Aquisição de dados experimentais	27
3.1.1	Aquisição do sinal de vibração	31
3.1.2	Resumo dos dados	34
3.2	Extração das características	35
3.2.1	Características no domínio do tempo	35
3.2.2	Características no domínio da frequência	38
3.2.3	Base de dados	40
4	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	41
4.1	Comparação entre escolha das característica dos dados	41
4.2	Comparação entre a localização dos sensores	45
4.3	Comparação entre eixos de medição	46
4.4	Resumo dos resultados	47
5	CONSIDERAÇÕES FINAIS	49
5.1	Trabalhos futuros	50

1 Introdução

Máquinas e equipamentos em diversas indústrias podem apresentar falhas precoces se não houver manutenção adequada. Identificar o momento ideal para realizar a manutenção não é uma tarefa fácil, pois é necessário alinhar um intervalo adequado na linha de produção com a real necessidade de manutenção. Devido a isso, técnicas e métodos de monitoramento baseado em condição têm sido amplamente desenvolvidos. Esta estratégia de manutenção preditiva tem se mostrado mais eficiente em uma vasta gama de casos (RANDALL, 2011).

A evolução da manutenção até o momento atual com técnicas avançadas de manutenção preditiva pode ser dividida em algumas fases. Inicialmente, as máquinas operavam até que um de seus componentes falhasse de maneira catastrófica. Apesar de alcançar um tempo máximo de funcionamento entre paradas de produção, os custos associados eram elevados, uma vez que poderiam causar danos severos a bens materiais, humanos ou ambientais. Na fase seguinte, iniciaram-se paradas para manutenção em intervalos regulares para evitar falhas. Contudo, ainda havia um custo considerável devido ao aumento do uso de peças sobressalentes e à quantidade de paradas. Nesse contexto, a manutenção preditiva visa explorar ao máximo a vida útil do componente, realizando intervenções apenas quando necessário para evitar quebras e de maneira eficiente ao pré-diagnosticar possíveis problemas (RANDALL, 2011; KARDEC; NASCIF, 2009).

De acordo com [Associacao Brasileira de Normas Tecnicas \(1994\)](#), a manutenção baseada na condição (ou preditiva) é caracterizada pela aplicação sistemática de técnicas de análise, utilizando meios de supervisão centralizados ou de amostragem. Existem diversas formas de acompanhar a condição dos equipamentos em operação, como análise de vibração, análise de óleo e detritos de desgaste, ultrassom, termografia, entre outros. Para máquinas rotativas, a técnica mais eficaz é a análise por vibração, pois a condição da máquina está diretamente ligada com a forma do sinal ([SCHEFFER; GIRDHAR, 2004](#)).

No caso específico da análise por vibração, foco deste trabalho, essas técnicas envolvem medições periódicas da resposta dinâmica do sistema por meio de transdutores de vibração. A etapa seguinte é a extração de parâmetros sensíveis aos danos a partir dos dados coletados e, então, a avaliação da condição. Nesses casos, a associação de danos a parâmetros constitui uma base adequada para a aplicação de técnicas de aprendizado de máquina, que auxiliam na classificação dos danos. Segundo [Manyika \(2011\)](#), um relatório do McKinsey Global Institute afirma que o aprendizado de máquina será o impulsionador da próxima grande onda de inovação em diversas áreas.

A aplicação dessas técnicas já ocorre na indústria. Setores como o de geração de energia e o (petro)químico, onde as máquinas operam a velocidade quase constante e com carga estável,

foram pioneiros na adoção. À medida que as técnicas evoluíram e os custos diminuíram, a manutenção baseada na condição foi estendida para outros setores industriais, onde as máquinas operam em condições mais adversas (RANDALL, 2011).

Os modelos de aprendizado de máquina no contexto de classificação por sinal de vibração visam organizar uma estrutura de decisão a partir de características extraídas do sinal. Nesse contexto, alguns algoritmos se mostram mais eficientes que outros, como o de floresta aleatória. Em particular, esse modelo destaca-se por sua boa capacidade de generalização, facilidade de implementação, simplicidade no ajuste para otimização, além de apresentar alta acurácia em diversos contextos (SCIKIT LEARN, 2024d; LIMA; ZAMBRANO-LÓPEZ; NETTO, 2016).

Seguindo nessa linha, este trabalho propõe a construção de um classificador para problemas de diferentes níveis de complexidade e com ocorrência frequente em máquinas rotativas em operação em diversos ramos industriais. Para isso, será explorada também a construção de uma base de dados e os demais processos anteriores ao modelo preditivo final.

1.1 Objetivos

1.1.1 Objetivo geral

Objetiva-se desenvolver um banco de dados a partir da medição de vibração coletada no *Machinery Fault Simulator*, fabricado pela Spectra Quest, para as condições de operação saudável, desbalanceamento e danos nos componentes do mancal de rolamento. Em seguida, pretende-se criar um modelo de aprendizado de máquina para o diagnóstico automático de falhas.

1.1.2 Objetivos específicos

Para a realização do objetivo geral será buscada a realização das seguintes etapas e objetivos complementares:

- abordar a metodologia para aquisição de dados;
- definir, extrair e analisar características relevantes dos sinais de vibração;
- discutir as etapas de um projeto de aprendizado de máquina aplicado ao diagnóstico de falhas e apresentar trabalhos de referência;
- testar o desempenho do classificador utilizando diferentes combinações de sensores;

1.2 Estrutura do trabalho

O trabalho está dividido conforme segue:

- Referencial teórico: Apresenta-se as assinaturas esperadas para os defeitos estudados neste trabalho. Além disso, são apresentados os tipos de algoritmos de aprendizado de máquina com enfase no método de classificação de floresta aleatória e as métricas de desempenho para classificadores. Por fim são mencionadas as abordagens exploradas em trabalhos anteriores de aquisição de dados e de aprendizado de máquina no diagnóstico de falhas;
- Construção da base de dados: são discutidas com maior detalhe a bancada de aquisição utilizada e o processo de construção do banco de dados. Também, são descritas as técnicas de processamento de sinais e de extração de características que foram utilizadas como parâmetros de entrada para o algoritmo de aprendizado de máquina;
- Apresentação e discussão dos resultados: são expostos os resultados para diferentes combinações de sensores e do conjunto de características; e
- Considerações Finais: são destacados as contribuições do trabalho, as suas limitações e são sugeridos aprofundamentos no estudo do diagnóstico de defeitos a partir de bancadas de vibração.

2 Referencial teórico

Na transição da manutenção preventiva para a manutenção preditiva, técnicas de monitoração da condição de equipamentos sem intervenção na produção foram desenvolvidas. Nesse contexto, a medição de vibração ganhou destaque, pois equipamentos mecânicos apresentavam uma assinatura específica conforme a condição do sistema. Assim, a vibração tornou-se o parâmetro de condição mais tradicionalmente medido, especialmente em máquinas rotativas ([KARDEC; NASCIF, 2009; ADAMS, 2009](#)).

A vibração de uma máquina ou equipamento é um movimento físico em resposta a uma ou mais fontes de excitação e à influência do caminho até o ponto de medição. A vibração é mensurável por meio de parâmetros cinemáticos como velocidade, deslocamento e aceleração, sendo que os transdutores convertem esse movimento em sinal elétrico. Equipamentos para medir esses parâmetros já são largamente utilizados e têm sido cada vez mais aprimorados ([SCHEFFER; GIRDHAR, 2004; KARDEC; NASCIF, 2009](#)).

Para máquinas rotativas, as frequências de ressonância não aparecem no espectro de resposta a menos que sejam diretamente excitadas por uma frequência específica. Essas frequências não são fixas, variando de acordo com a velocidade de rotação da máquina. Além disso, na literatura especializada, é comum encontrar correlações entre problemas de vibração e assinaturas de vibração esperadas para esse tipo de equipamento. Isso se deve ao fato de que máquinas operando em velocidade constante geralmente são excitadas periodicamente por sinais senoidais, que geram respostas também senoidais na mesma frequência. No entanto, devido às condições não ideais do equipamento em operação, essas respostas podem ser ligeiramente distorcidas, incluindo múltiplos harmônicos da frequência de excitação ([ADAMS, 2009; RANDALL, 2011](#)).

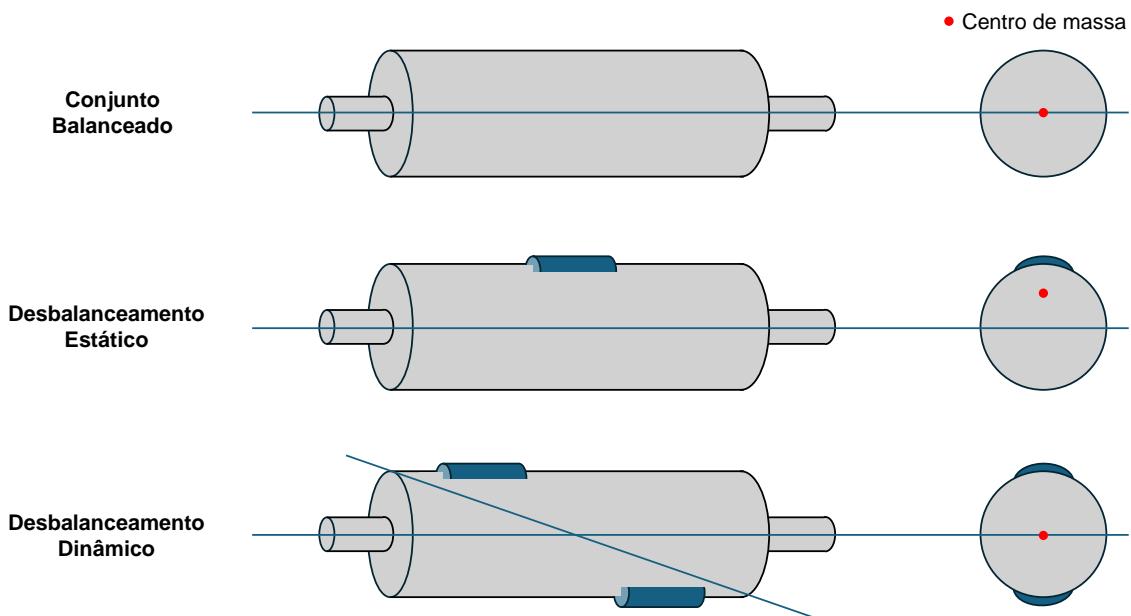
2.1 Assinaturas de vibração

Para uma gestão eficiente da manutenção, mapear as principais causas de falhas em componentes críticos é essencial para evitar quebras e paradas indesejadas. Nesse sentido, a [Abecom \(2022\)](#) listou as 10 principais causas de falha em sistemas rotativos: desbalanceamento, desalinhamento, empenamento de eixo, falha em rolamentos, folga mecânica, roçamento, falhas em engrenamento, ressonância, lubrificação inadequada e falhas elétricas. A seguir, serão apresentadas as assinaturas de vibração esperadas para falhas de desbalanceamento e falhas em rolamentos, que são abordadas para previsão neste trabalho.

2.1.1 Desbalanceamento

O desbalanceamento em máquinas rotativas pode ocorrer de maneira estática ou dinâmica, sendo, de maneira geral, causado por desproporcionais na distribuição de massa nos componentes rotativos. Para o caso estático, o centro de massa do rotor é diferente do centro de rotação. Já para o caso dinâmico, a linha de centro de sua massa não coincide com a linha geométrica de rotação, independente do desbalanceamento estático (ABECOM, 2022). A figura 1 apresenta o comparativo entre as formas.

Figura 1 – Formas de desbalanceamento



Fonte: Autoria própria

Para essa condição de falha, uma força de desbalanceamento surge no equipamento quando operando em velocidade constante. Esta força, para o caso estático, é descrita pela equação 2.1, onde m_o representa a massa desbalanceada, e é o deslocamento radial do centro de massa e ω_r^2 é a velocidade angular de rotação (INMAN, 2013). Para o caso dinâmico, utiliza-se a equação 2.2, onde m_d representa a massa de desbalanceamento dinâmico, r_d é o deslocamento radial entre a massa e o centro de massa, l é a distância entre as massas e ω_r^2 é a velocidade angular de rotação (ADAMS, 2009).

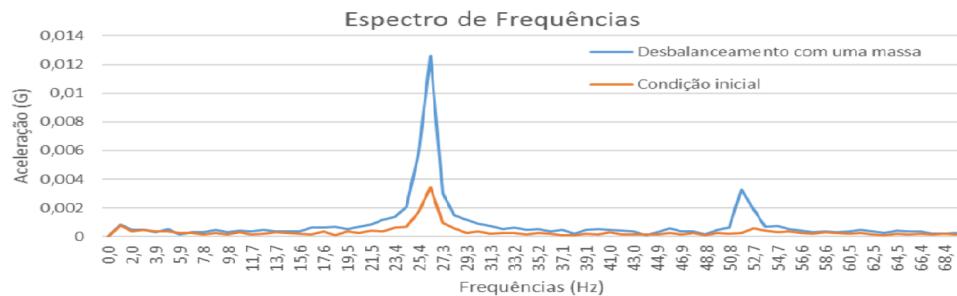
$$F_o = m_o e \omega_r^2 \quad (2.1)$$

$$F_d = m_d r_d l \omega_r^2 \quad (2.2)$$

Pode-se observar que, tanto na equação 2.1 quanto na equação 2.2, a força devido ao desbalanceamento é proporcional à frequência de rotação. Isso contribui significativamente para

o padrão característico da assinatura de vibração de desbalanceamento no equipamento. Em outras palavras, o desbalanceamento gera um pico no espectro de frequência na velocidade angular do eixo, predominantemente na direção vertical (RANDALL, 2011; ADAMS, 2009). A figura 2 ilustra o comportamento do desbalanceamento no espectro de frequência.

Figura 2 – Comparação entre condição saudável e desbalanceada a 26,7 Hz



Fonte: Bassani et al. (2018)

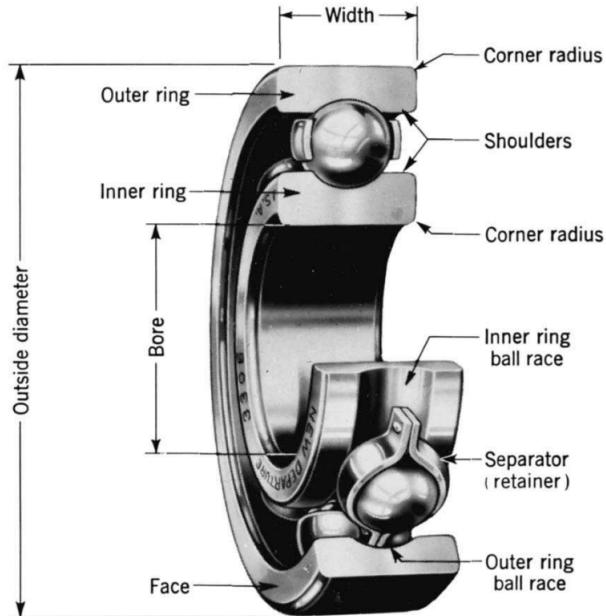
2.1.2 Falhas em mancais

Falha em mancais é um termo genérico, uma vez que este elemento mecânico, dependendo do grau de complexidade da construção, pode apresentar defeitos em diversos componentes. Os mancais, são elementos de máquinas destinados a suportar e guiar eixos rotativos ou oscilantes, reduzindo o atrito e o desgaste entre as partes móveis. Eles podem ser construídos com partes rolantes ou por deslizamento (INMAN, 2013).

Mancais de rolamento, utilizados neste trabalho, utilizam elementos rolantes como esferas ou rolos para manter a separação entre as partes móveis do rolamento, minimizando o atrito e facilitando o movimento suave. A figura 3 ilustra um exemplo de rolamento de esferas, um tipo comum de mancal de rolamento que utiliza esferas para minimizar o atrito entre o anel interno e o anel externo.

As falhas neste tipo de componente podem ocorrer por diversas causas, como carga excessiva, vedações ineficientes, lubrificação e montagem incorreta. Contudo, cada um desses fatores pode provocar um tipo específico de dano ao equipamento. Esse componente danificado, por sua vez, apresentará uma assinatura de vibração específica (ABECOM, 2022).

Figura 3 – Partes do rolamento de esferas



Fonte: [Budynas e Nisbeth \(2016\)](#)

A assinatura dos defeitos em rolamento, diferentemente do desbalanceamento, não são pícos na frequência de rotação ou ainda, em seus harmônicos. Entretanto, a resposta é função da velocidade de rotação do equipamento. Para os três casos de danos abordados neste trabalho, as equações 2.3, 2.4 e 2.5 são apresentadas na literatura para descrever o comportamento no espectro em frequência desses defeitos. Vale ainda ressaltar que o espectro de vibração dos defeitos em mancais de rolamento possui assinatura relacionada a severidade do dano ([SCHEFFER; GIRDHAR, 2004](#)).

$$\text{BPFI} = \frac{N_b}{2} \left(1 + \frac{B_d}{P_d} \cos \theta \right) \times \text{rpm} \quad (2.3)$$

$$\text{BPFO} = \frac{N_b}{2} \left(1 - \frac{B_d}{P_d} \cos \theta \right) \times \text{rpm} \quad (2.4)$$

$$\text{BSF} = \frac{P_d}{2B_d} \left[1 - \left(\frac{B_d}{P_d} \right)^2 \cos^2 \theta \right] \times \text{rpm} \quad (2.5)$$

Nas equações N_b refere-se ao número de esferas ou rolos no rolamento, B_d é o diâmetro das esferas ou rolos e P_d indica o diâmetro do passo do rolamento. θ representa o ângulo de contato em graus. BPFI refere-se à frequência para dano na pista interna, BPFO refere-se à frequência para dano na pista externa e BSF refere-se à frequência para dano nas esferas.

2.2 Aprendizado de máquina

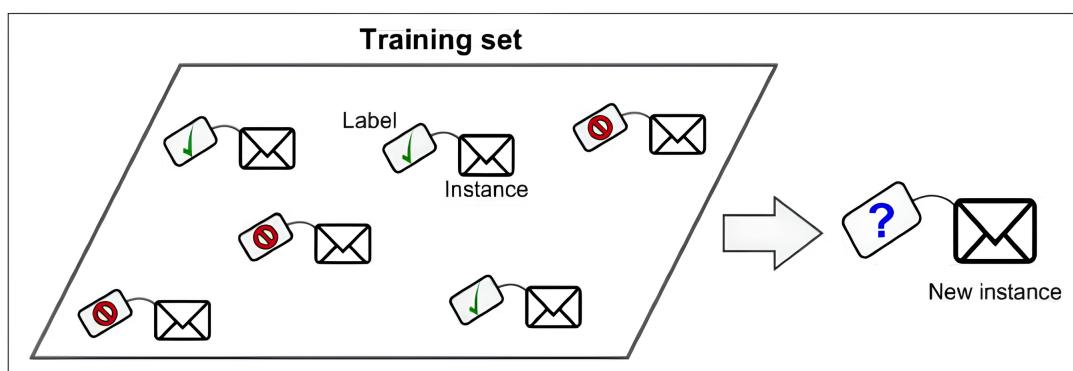
Aprendizado de máquina é a ciência de programar computadores para aprender a partir de conjuntos de dados (GÉRON, 2019). Esta é frequentemente uma alternativa muito atraente em comparação à construção manual desses programas, e, na última década, o uso do aprendizado de máquina se espalhou rapidamente pela ciência da computação e além. O aprendizado de máquina é usado em diversas atividades do cotidiano, sendo a um exemplo o filtro automático de spam na caixa de e-mail (DOMINGOS, 2012).

Ao analisar grandes volumes de dados operacionais em tempo real, algoritmos de aprendizado de máquina, fazendo uso de uma base estatística, constroem modelos capazes de identificar padrões e anomalias que antecedem falhas potenciais ou o diagnóstico da condição. Com o aprendizado contínuo dos algoritmos, a precisão das previsões e recomendações se torna cada vez mais robusta, trazendo significativos benefícios econômicos e operacionais para as empresas.

Os sistemas de aprendizado de máquina podem ser classificados de acordo com a quantidade e o tipo de supervisão que recebem durante o treinamento. Existem quatro categorias principais: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado por reforço (GÉRON, 2019). Eles podem ser definidos como:

- Aprendizado supervisionado: é um tipo de aprendizado de máquina onde o algoritmo é treinado com um conjunto de dados rotulados, ou seja, cada instância de treinamento é acompanhado da resposta correta (rótulo). A figura 4 ilustra este modelo;

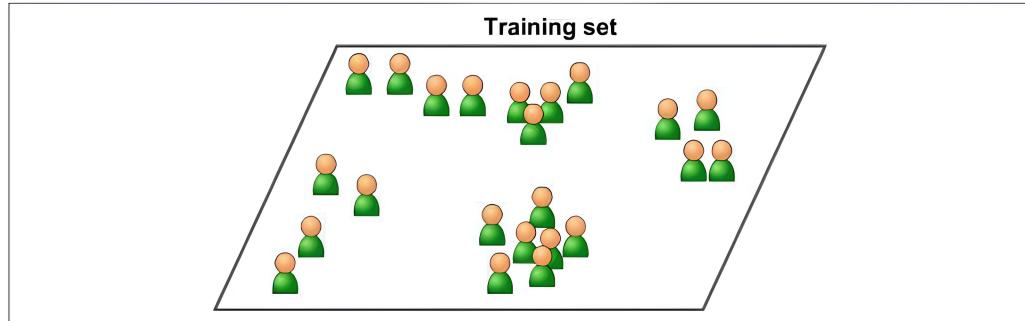
Figura 4 – Conjunto de treino rotulado para aprendizado supervisionado



Fonte: Géron (2019)

- Aprendizado não supervisionado: é um tipo de aprendizado de máquina onde o algoritmo é treinado com dados que não possuem rótulos ou respostas pré-definidas. dessa forma, busca-se identificar padrões, estruturas ou agrupamentos ocultos nos dados (*cluster*). A figura 5 ilustra este modelo;

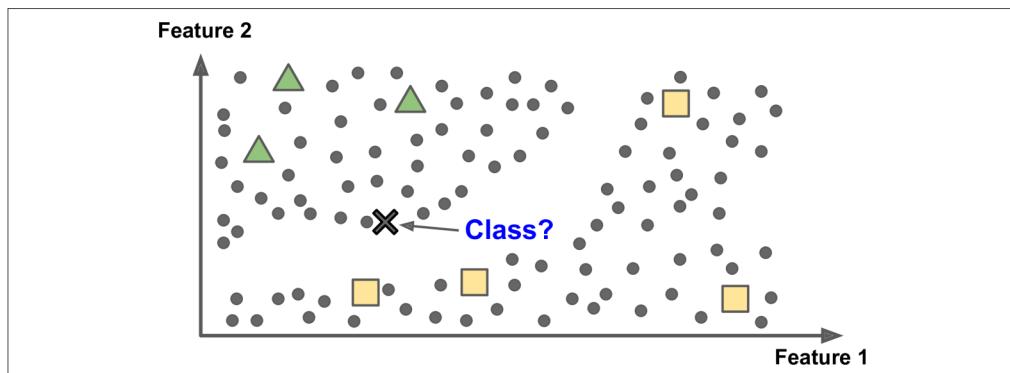
Figura 5 – Conjunto de treino não rotulado para aprendizado não supervisionado



Fonte: Géron (2019)

- Aprendizado semi-supervisionado: é um tipo de aprendizado de máquina que utiliza uma combinação de dados rotulados e não rotulados para treinamento. Esse método aproveita a pequena quantidade de dados rotulados para guiar o aprendizado, enquanto explora a grande quantidade de dados não rotulados para melhorar a precisão e a generalização do modelo. A figura 6 ilustra este modelo;

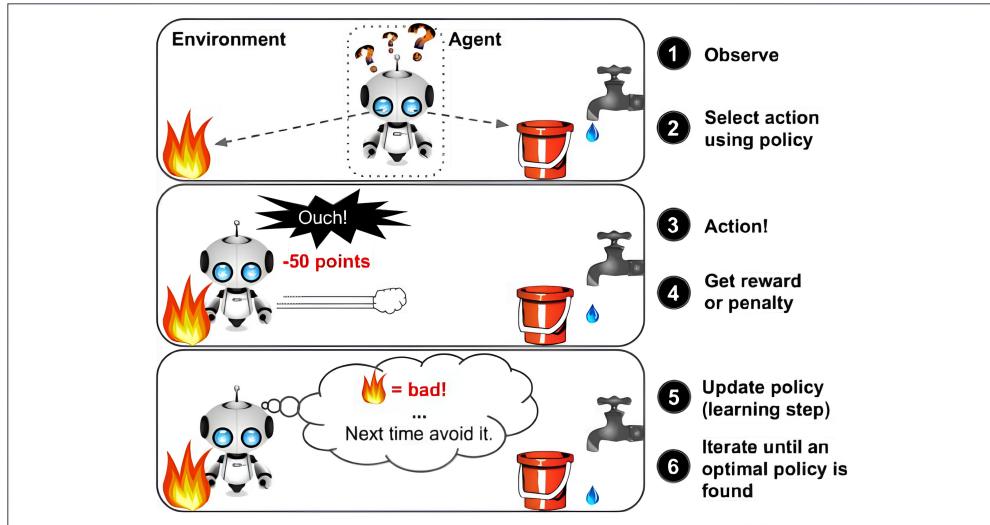
Figura 6 – Aprendizado semi-supervisionado



Fonte: Géron (2019)

- Aprendizado por reforço: é uma abordagem de aprendizado de máquina em que um agente aprende a tomar ações em um ambiente para maximizar uma recompensa cumulativa ou minimizar as penalidades. O agente interage com o ambiente de maneira iterativa de modo a aprender a realizar ações que levem a recompensas mais altas ao longo do tempo, aprendendo por tentativa e erro. A figura 7 ilustra este modelo;

Figura 7 – Aprendizado por reforço



Fonte: Géron (2019)

Além disso, é crucial ter um entendimento claro do problema ao qual as técnicas de aprendizado de máquina serão aplicadas, para escolher os algoritmos mais adequados. Isso se deve à distinção entre problemas de classificação, que visam identificar a categoria à qual uma instância pertence, e problemas de regressão, cujo objetivo é prever um valor contínuo para um determinado atributo da instância.

Como a base de dados construída neste trabalho pôde ser rotulada devido ao conhecimento das classes de problemas inseridas na bancada, o modelo a ser utilizado é um classificador supervisionado.

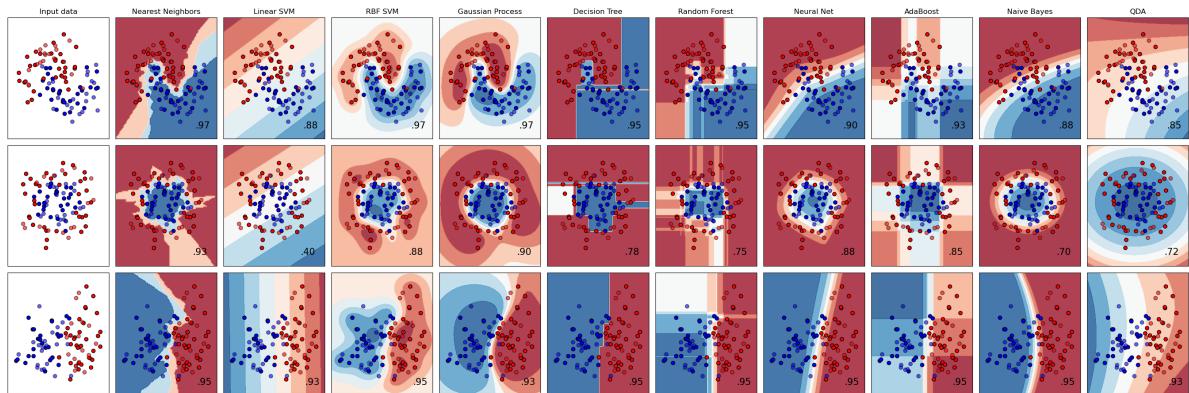
2.2.1 Classificadores supervisionados

A maioria dos problemas abordados com ferramentas de aprendizado de máquina são problemas de classificação supervisionada, conforme exposto por Domingos (2012). Esses métodos são aplicáveis a uma ampla variedade de situações, desde a determinação de doenças em indivíduos com base em exames sanguíneos até o diagnóstico de falhas em máquinas rotativas. A capacidade de classificar com precisão dados rotulados torna esses algoritmos indispensáveis em inúmeros contextos, proporcionando soluções eficientes e com maior confiabilidade aos sistemas mecânicos.

Embora o aprendizado supervisionado possa oferecer vantagens às empresas, como conclusões profundas de dados e automação aprimorada, existem alguns desafios ao criar modelos sustentáveis de aprendizado supervisionado (IBM, 2024). Os principais desafios são: a criação e manutenção dos modelos podem exigir certos níveis de especialização; os modelos podem ser muito demorados; qualidade e disponibilidade de dados para treino; dentre outros desafios.

Existem diversos algoritmos baseados em abordagens matemáticas e estatísticas específicas para criar modelos capazes de classificar dados a partir de um conjunto de treino rotulado. Esses algoritmos variam em termos de complexidade, desempenho e velocidade de decisão. A figura 8 apresenta uma ilustração dos principais tipos de classificadores em um conjunto de dados sintéticos.

Figura 8 – Comparação entre algoritmos de classificação



Fonte: [Scikit Learn \(2024a\)](#)

A escolha do classificador adequado é crucial para o sucesso do modelo preditivo. Considerando fatores como a natureza dos dados, a necessidade de interpretabilidade do modelo e os recursos computacionais disponíveis, um modelo pode ser mais apropriado que outro. Portanto, generalizar todos os problemas em um único classificador pode levar a decisões equivocadas.

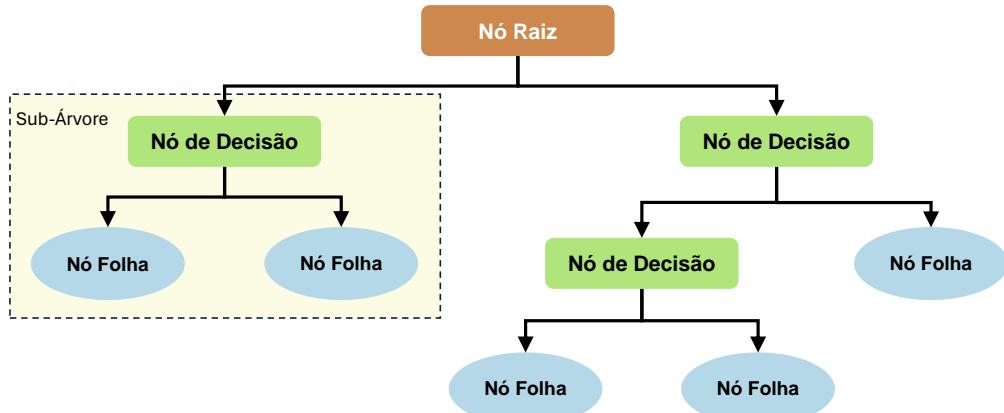
Nesse sentido, este trabalho utilizará o algoritmo de classificação por floresta aleatória (*Random Forest Classifier*), conhecido por sua capacidade de generalização e habilidade de absorver novos dados. Portanto, esse modelo pode oferecer uma boa capacidade preditiva para os dados após a remontagem, além de ser um algoritmo com alto índice de acertos em diversos casos.

2.2.1.1 Classificação de floresta aleatória

As árvores de decisão são um dos métodos poderosos frequentemente empregados em diversos campos, tais como aprendizado de máquina, processamento de imagens e identificação de padrões ([CHARBUTY; ABDULAZEEZ, 2021](#)).

Esse algoritmo é composto de nós de decisão e nós folha. Cada nó de decisão corresponde a um teste básico sobre um único atributo dos dados de entrada, no qual uma característica numérica é comparada a um valor limite em cada teste. Essa árvore possui um número de ramos, onde cada um dos quais lida com um resultado do teste. Cada nó folha representa uma classe que é o resultado da decisão para um ramo ([STEIN et al., 2005; CHARBUTY; ABDULAZEEZ, 2021](#)). A figura 9 ilustra um algoritmo genérico de árvore de decisão.

Figura 9 – Árvore de decisão



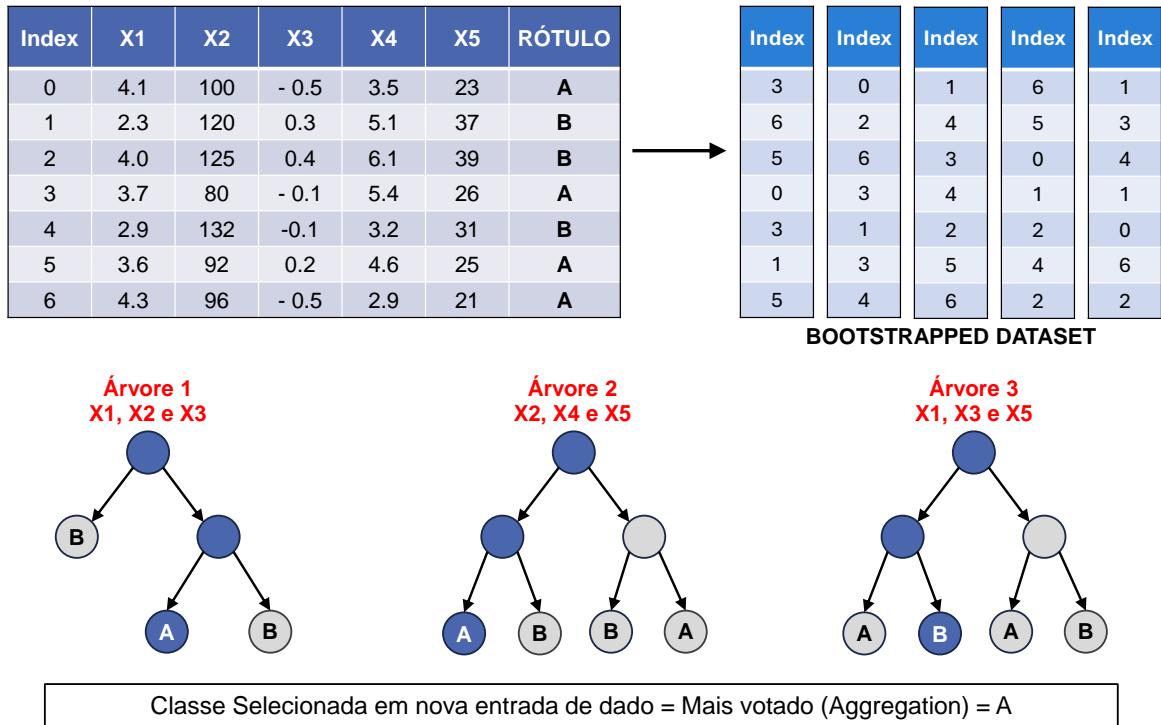
Fonte: Adaptado de [Charbuty e Abdulazeez \(2021\)](#)

A partir dessa metodologia básica de classificação, novos algoritmos mais robustos foram desenvolvidos para melhorar a capacidade e superar as limitações do modelo inicial. Entre esses modelos derivados, o modelo por floresta aleatória (*random forest*) destaca-se como o mais popular.

Uma floresta aleatória é um modelo que ajusta vários classificadores de árvore de decisão em subamostras do conjunto de dados e utiliza a média para melhorar a precisão preditiva e controlar o *overfitting* ([SCIKIT LEARN, 2024d](#)).

Para isso, subconjuntos de dados são criados de maneira aleatória a partir da base de dados inicial, mantendo a mesma dimensão, em um processo conhecido como *Bootstrapped Datasets*. Em seguida, várias árvores de decisão são geradas utilizando um percentual das características disponíveis nos dados de entrada e treinadas com esses subconjuntos. Finalmente, ao testar o classificador, cada árvore indica uma classe, sendo que o resultado final da classificação é determinado pela classe de maior incidência entre todas as árvores. Esse processo de combinar os resultados das diversas árvores de decisão é chamado de *Aggregation*. A figura 10 apresenta de maneira simplificada os processos de *Bootstrap* e *Aggregation*.

Figura 10 – classificação de floresta aleatória



Fonte: Autoria própria

2.2.2 Métricas de avaliação e otimização

Para avaliar o desempenho do modelo e compará-lo com outros classificadores e suas modificações, é essencial utilizar parâmetros numéricos que possam representar essas informações com precisão. Esses parâmetros se baseiam em fundamentos matemáticos sólidos, fornecendo métricas quantificáveis para a análise comparativa. Através dessas métricas, podemos identificar a eficácia de um modelo em tarefas de classificação, além de verificar possibilidades de melhorias.

A maneira mais intuitiva de visualizar o resultado é por meio de uma matriz de confusão. Nessa representação, os resultados preditos são organizados em uma matriz onde as linhas representam os rótulos reais e as colunas os rótulos previstos pelo modelo. Assim, um resultado perfeito corresponde a uma matriz com 100% dos valores na diagonal principal, indicando apenas verdadeiros positivos e verdadeiros negativos (GÉRON, 2019).

Além disso, métricas são utilizadas para avaliar quantitativamente o desempenho global do modelo. Os principais parâmetros incluem acurácia, precisão, recall e a métrica F1, todos derivados da matriz de confusão. Cada um desses parâmetros oferece uma perspectiva diferente sobre o desempenho do modelo:

Acurácia: Corresponde à proporção de previsões corretas em relação ao total de casos avaliados. A equação 2.6 apresenta essa métrica, onde TP corresponde a classificação verdadeira

positiva, TN à verdadeiro negativo, FP à falso positivo e FN à falso negativo.

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.6)$$

Precisão: Corresponde à proporção de previsões verdadeiro positivo corretamente classificados e o número total de exemplos identificados como positivos pelo classificador. A equação 2.7 representa essa métrica.

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2.7)$$

Recall (ou Sensibilidade): Corresponde à proporção de verdadeiros positivos corretamente classificados e o número total de positivos reais. A equação 2.8 representa essa métrica.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.8)$$

F1-Score: Corresponde a média harmônica entre precisão e do recall, proporcionando uma única métrica que prestigia ambos os aspectos. A equação 2.9 representa essa métrica.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (2.9)$$

Além dessas métricas básicas, outras técnicas avançadas como a área sob a curva ROC (AUC-ROC) também podem ser utilizadas para fornecer uma visão mais detalhada do desempenho do modelo. Contudo, a mais utilizada para a problemática tratada neste trabalho é a acurácia, conforme pode-se observar nos trabalhos citados na seção 2.3.

Outro método utilizado para avaliar o modelo é a validação cruzada. Nessa abordagem, o conjunto de dados é dividido em K partições. Em cada iteração de treino, uma partição é usada como dados de teste, enquanto as demais são usadas como dados de treino. Esse processo é repetido K vezes, permitindo que cada partição seja utilizada uma vez como dados de teste. Isso proporciona uma avaliação mais robusta do desempenho do modelo, reduzindo a variabilidade associada a uma única divisão de treino/teste (NILSSON, 1998).

2.2.2.1 *Overfitting*

Overfitting, por vezes traduzido como sobre-ajuste, é um problema significativo no aprendizado de máquina. Ocorre quando o modelo se ajusta excessivamente aos dados de treino, mas tem um desempenho insatisfatório em novos dados de teste. Nesses casos, o modelo não consegue generalizar adequadamente para situações reais, limitando sua eficácia como ferramenta preditiva.

Esse tipo de falha nos modelos geralmente ocorre devido a um conjunto de dados de

treino insuficiente, sendo intensificado quando há um viés para os dados de treino que não são hipóteses consistentes para o universo completo do problema ([NILSSON, 1998](#)).

2.2.2.2 Otimização

O desempenho dos modelos de aprendizado de máquina está geralmente relacionado à qualidade dos dados de treino e à combinação de hiperparâmetros utilizados. Nesse sentido, existem métodos que permitem otimizar o sistema e obter um ganho adicional na performance.

Uma maior quantidade de características em um banco de dados não necessariamente resulta em melhores generalizações ao criar árvores de decisão. Duas características altamente correlacionadas podem atribuir um peso excessivo a uma delas, inclinando a decisão de maneira equivocada. Por outro lado, uma característica com baixa correlação com a classificação pode direcionar as árvores de decisão para ramos incorretos. Portanto, é crucial selecionar cuidadosamente as características relevantes e aplicar técnicas de pré-processamento para evitar esses problemas.

Uma maneira comum de realizar esse pré-processamento é analisar a matriz de correlação na construção do modelo, eliminando características que apresentam correlações indesejáveis. A matriz de correlação oferece uma visão global da relação entre todas as características da base de dados, com o coeficiente de correlação variando de 1 a -1, conforme descrito a seguir ([GÉRON, 2019](#)):

- Valores próximos a 1 indicam uma forte correlação positiva, o que significa que o aumento do valor de um parâmetro está associado ao aumento do parâmetro correlacionado.
- Valores próximos a -1 indicam uma forte correlação negativa, o que significa que o aumento do valor de um parâmetro está associado à redução do parâmetro correlacionado.
- Valores próximos a 0 indicam que não há correlação linear, ou seja, o aumento do valor de um parâmetro não está associado à variação positiva ou negativa do parâmetro correlacionado.

Outra abordagem, não excludente, é utilizar algoritmos de seleção de características e/ou redução de dimensionalidade. Um desses algoritmos é o *Tree-based feature selection* do Sklearn, no qual estimadores baseados em árvores de decisão são usados para calcular as importâncias das características. Essas técnicas ajudam a aprimorar a precisão e a eficiência do modelo, evitando problemas de multicolinearidade e irrelevância ([SCIKIT LEARN, 2024b](#)).

Para a seleção da melhor combinação de hiperparâmetros, existem diversas técnicas que estimam boas combinações. Uma delas é o *GridSearchCV* do Sklearn, que realiza uma série de testes com validação cruzada a partir de um intervalo de valores para cada parâmetro. Por fim, retorna a combinação com o melhor desempenho ([SCIKIT LEARN, 2024c](#)).

2.3 Predições de falha em máquinas rotativas

Devido à capacidade de diagnosticar defeitos em máquinas rotativas a partir da assinatura típica de vibração, muitos estudos exploram a evolução desse sistema ao combinar classes de assinaturas com algoritmos de aprendizado de máquina. Um aspecto crucial para o desenvolvimento de modelos preditivos eficazes é a construção de um banco de dados a partir de ensaios de vibração realizados em bancadas de teste. Esses ensaios proporcionam uma base rica e diversificada para treinar e validar os modelos de aprendizado de máquina ([SPECTRAQUEST, 2024](#)).

A Universidade Federal do Rio de Janeiro disponibiliza um banco de dados de sinais de vibração com 1951 séries temporais, compreendendo diversas condições do equipamento, coletados em uma bancada da SpectraQuest ([MAFALDA, 2021](#)), similar à utilizada neste trabalho. Utilizando esses dados, diversos trabalhos de predição por aprendizado de máquina foram realizados, como o de [Ramos \(2020\)](#). Nesse projeto de graduação, foi utilizado o algoritmo de máquina de vetores de suporte (*Support Vector Machine*) para predizer falhas em algumas combinações de sensoriamento. No melhor dos casos, obteve-se uma acurácia superior a 98%.

[Paudyal \(2019\)](#), por sua vez, também adquiriu dados de vibração em uma bancada semelhante à utilizada neste trabalho, com ênfase em desalinhamento, desbalanceamento e eixo danificado, além da condição saudável. Após o processamento dos dados, foram aplicadas as técnicas de diagrama de plano de fase, lógica fuzzy e o método de aprendizado de máquina dos k-vizinhos mais próximos (KNN) como classificador das condições experimentais. Em todos os casos, obteve-se grande sucesso na classificação dos defeitos analisados.

Ao abordar o mesmo algoritmo de predição utilizado neste trabalho, [Lima, Zambrano-López e Netto \(2016\)](#) aplicaram-no para a classificação de nove tipos de defeitos em motores elétricos, a partir de dados de vibração medidos em uma bancada experimental. Como resultado, foi alcançada uma acurácia superior a 97%.

Outros trabalhos exploram as melhores características a se extrair do sinal de vibração para a classificação dos defeitos. Nessa linha, [Silva e Souto \(2024\)](#) exploram a capacidade dos parâmetros de fator de crista e curtose, para caracterizar falhas aplicada em defeitos de rolamentos.

Por fim, destaca-se também o uso comercial dessa abordagem preditiva. Empresas como a Pruftechnik e a Tractian oferecem equipamentos ou softwares para o medição de sinal, processamento de dados e diagnóstico de condição de sistemas mecânicos em operação ([TRACTIAN, 2021; PRUFTECHNIK, 2024](#)).

3 Construção da base de dados

Este trabalho usará uma base de dados obtidas na bancada *Machinery Fault Simulator* da SpectraQuest entre março e junho de 2024 no Laboratório de Dinâmica de Sistemas e Estruturas Mecânicas localizado no Centro Tecnológico da Universidade Federal do Espírito Santo. Ela consiste em uma coleção de sinais de vibração gerados cuja construção, composição e descrição desta base de dados serão discutidas nesse capítulo.

3.1 Aquisição de dados experimentais

As bancadas de testes desempenham um papel crucial no cenário dinâmico das atividades de pesquisa e desenvolvimento (P&D). Elas constituem uma infraestrutura fundamental que permite a avaliação, validação e refinamento de uma ampla gama de produtos e tecnologias em diferentes estágios do processo de inovação.

Ao fornecer um ambiente controlado e replicável, as bancadas de testes possibilitam a análise detalhada do desempenho, confiabilidade e eficácia de novos conceitos, protótipos e sistemas. Nesse sentido, o simulador de falhas (MFS), mostrado na figura 11, auxilia no estudo das assinaturas de falhas comuns em máquinas. O resumo das principais especificações está descrito na tabela 1.

Figura 11 – *Machinery Fault Simulator* da SpectraQuest



Fonte: [SpectraQuest \(2024\)](#)

Tabela 1 – Especificações da bancada MFS

Especificação	Valor
Motor	1/2 HP - Modelo DVM 56T34F5301JP
RPM Base	3450 rpm
Diâmetro do rotor	152,4 mm
Diâmetro do eixo	5/8 pol.
Comprimento do eixo	520 mm
Distância entre mancais	360 mm
<i>Mancais de Rolamento</i>	
Modelo	ER10K fabricado pela MB Manufacturing
Diâmetro externo	1,85 pol.
Velocidade máxima	10000 rpm

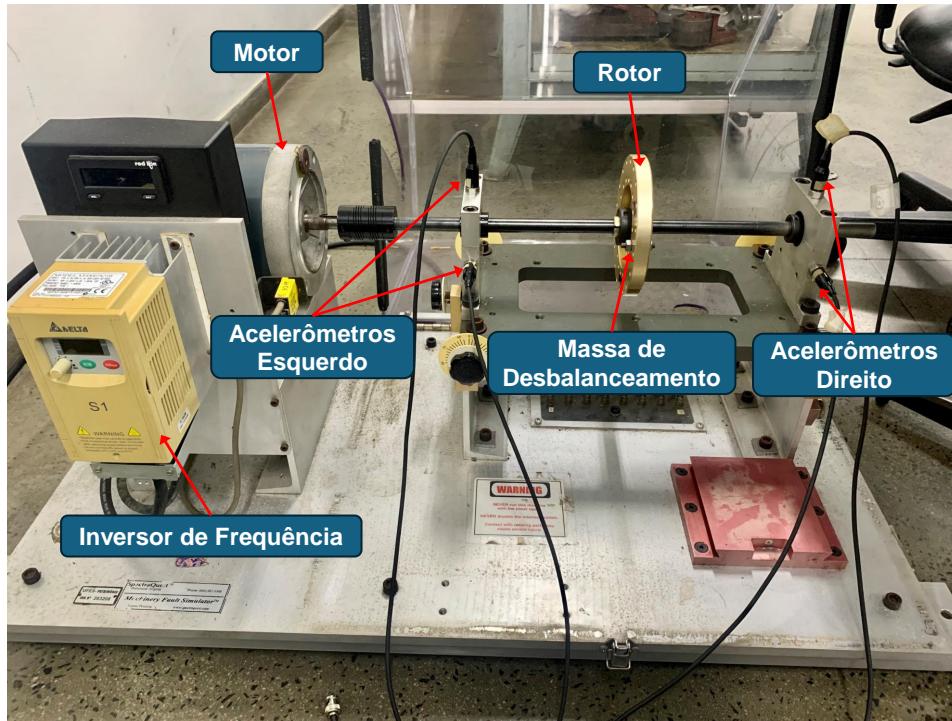
Fonte: [SpectraQuest \(2024\)](#), [Motion \(2024\)](#), [Rexnord \(2024\)](#)

O Simulador MFS é altamente modular, oferecendo a flexibilidade necessária para variar entre diferentes condições de saúde do rolamento, alinhamento, desbalanceamento e rigidez dos acoplamentos. Além disso, ele permite a integração de uma variedade de acessórios adicionais, ampliando consideravelmente a gama de simulações possíveis.

Para a construção da base de dados, padronizou-se o simulador com uma estrutura básica onde o motor está ligado por meio de um acoplamento à um eixo com rotor central, sustentado por um par de mancais de rolamento. A figura 12 apresenta a montagem utilizada. Para minimizar a influência dos componentes secundários e concentrar a análise na região de interesse, a bancada foi montada diretamente no chão. Adicionalmente, os acelerômetros foram conectados diretamente à placa de aquisição, eliminando a necessidade de um circuito intermediário.

O ajuste entre as configurações de dano ou condição saudável é feito ao desmontar a bancada de modo a desconectar o eixo principal, a carcaça dos rolamentos e os mancais de rolamento, que são então preparados de acordo com a condição desejada. Os rolamentos danificados foram sempre inseridos no lado direito do eixo. Cada condição foi replicada várias vezes por meio de remontagens e de forma não sequencial, ou seja, com pelo menos três condições diferentes entre cada repetição da mesma condição.

Figura 12 – Montagem da bancada de aquisição utilizada

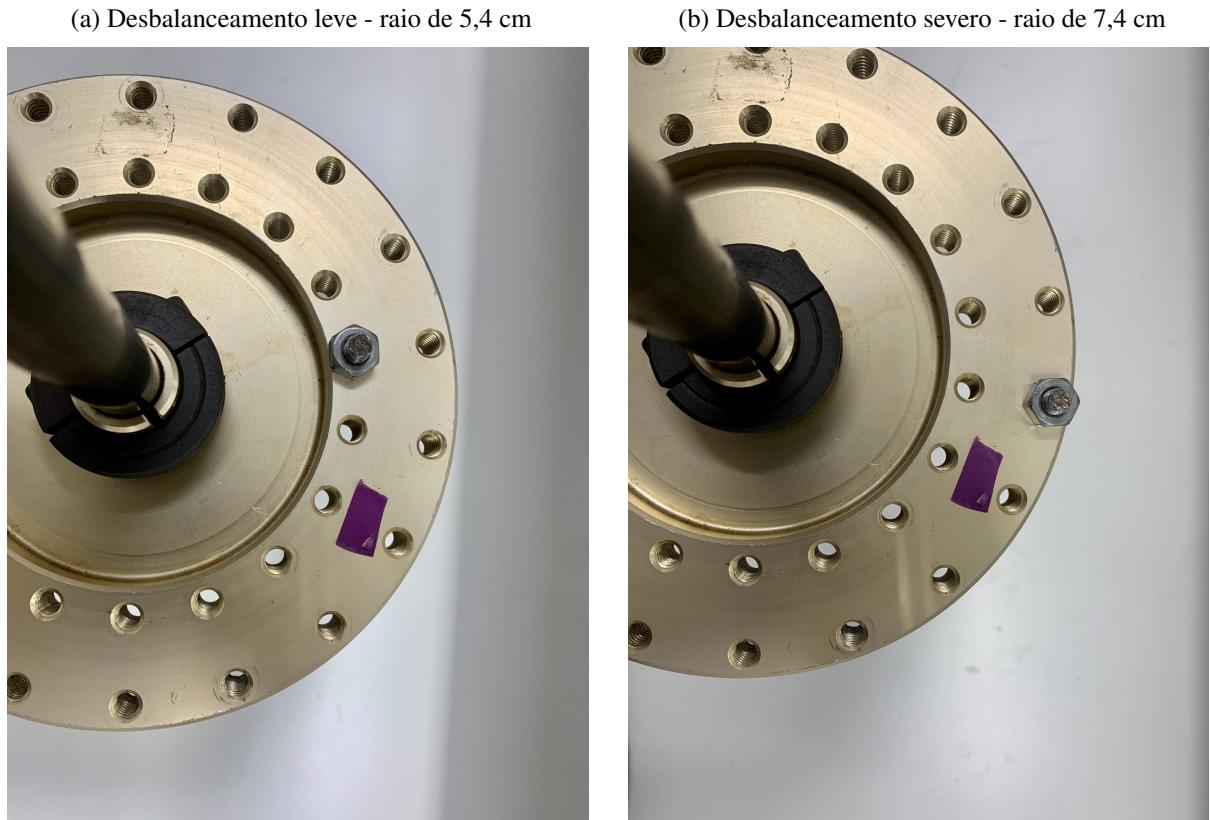


Fonte: Autoria própria

Essa abordagem permite a construção de uma base de dados que aproxima os testes de um ambiente industrial real. Na prática, as máquinas frequentemente passam por reparos, montagem e desmontagem, o que pode gerar variações sutis ou significativas nos sinais coletados. Ao replicar essas condições no banco de dados, é possível capturar uma variedade de cenários que refletem melhor os desafios enfrentados em operações industriais cotidianas. Isso aumenta a robustez e a precisão dos modelos preditivos, garantindo que eles sejam capazes de identificar e diagnosticar falhas com maior confiabilidade. Além disso, uma base de dados que incorpora essas variações permite que os algoritmos de aprendizado de máquina aprendam a reconhecer padrões sutis e inconsistências nos dados de vibração. Dessa forma, o modelo pode ser aplicado de maneira mais eficaz em ambientes industriais reais, contribuindo para a manutenção preditiva e a redução do tempo de inatividade das máquinas.

Para simular a condição de desbalanceamento estático, foi utilizado um parafuso associado a uma porca em cada lado do rotor, com uma massa total de 12 gramas. O desbalanceamento leve foi obtido posicionando o parafuso na pista interna (raio de 5,4 cm), como mostrado na figura 13a. Já o desbalanceamento severo foi obtido posicionando o parafuso na pista externa (raio de 7,4 cm), conforme ilustrado na figura 13b.

Figura 13 – Desbalanceamento com massa de 12 gramas



Fonte: Autoria própria

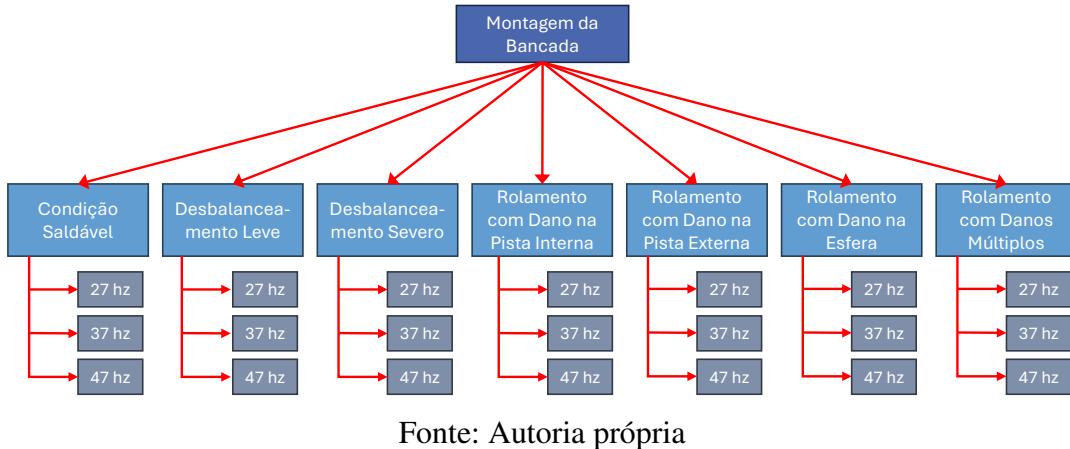
O fluxograma da figura 14 apresenta as diferentes condições ensaiadas, bem como as diferentes velocidades de rotação do motor elétrico, ajustadas por meio do inversor de frequência citado na tabela 2. As velocidades de rotação do motor utilizadas neste trabalho foram: 27 Hz, 37 Hz e 47 Hz. Essas velocidades foram escolhidas de modo a operar o equipamento em condições que não são próximas à frequência natural do equipamento nem próximas entre si.

Tabela 2 – Especificação do Inversor de frequência utilizado

Inversor de Frequência Delta Modelo VFD007S21B	
Corrente de entrada	9,7 A monofásico / 5,1 A trifásico
Tensão de entrada	200 a 240 V
Corrente de saída	4,2 A
Tensão de saída	Trifásica proporcional à entrada
Potência Máxima de Saída	0,75 kW (1 HP)
Frequência de saída	0,1 a 400 Hz

Fonte: [ManualsLib \(2024\)](#)

Figura 14 – Configurações consideradas nos experimentos



3.1.1 Aquisição do sinal de vibração

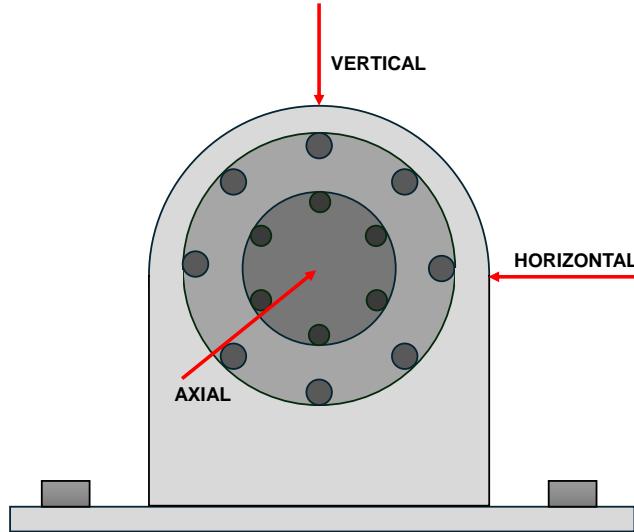
Conforme apresentado por Kardec e Nascif (2009), para definir qual a melhor maneira para medir a vibração em uma máquina ou numa estrutura, alguns questionamentos devem ser feitos: qual é o tipo da máquina; como é sua construção; qual o propósito da medição; o que objetiva-se observar; e qual a faixa de frequência.

A maioria das medições de vibração atualmente é feita através da medição da aceleração. Para calcular a velocidade ou o deslocamento, basta realizar uma ou duas integrações, respectivamente, ou ainda utilizar um algoritmo de Transformada de Fourier para o sinal em frequência. Existem diversos tipos de acelerômetros, mas todos eles compartilham um princípio comum: um sistema de massa-mola, classificado como transdutor sísmico.

O tipo mais comum, utilizado também neste trabalho para a MFS, é o piezoelétrico. Este consiste em um ou mais cristais piezoelétricos, tensionados por uma massa e encapsulados em uma carcaça. Quando sujeitos a uma deformação mecânica, esses cristais geram uma corrente elétrica proporcional. Esse processo de transdução converte a energia mecânica da vibração em um sinal elétrico, captado posteriormente por uma placa de aquisição.

A localização dos transdutores está diretamente relacionada a capacidade de medir um sinal que permita identificar falha nos componentes. Isso deve-se ao comportamento de alguns componentes do conjunto que atuam como amortecedores de vibração. Para as tipos de falhas abordadas neste trabalho e a característica física do sistema em questão, Kardec e Nascif (2009) recomenda que as medições sejam feitas na carcaça dos mancais. Dessa forma, há três direções possíveis de medição, conforme mostrado na figura 15, sendo elas: vertical, horizontal e axial.

Figura 15 – Pontos de medição de vibração na carcaça da caixa de mancal



Fonte: Adaptada de [Kardec e Nascif \(2009\)](#)

Quanto a montagem dos acelerômetros, [Scheffer e Girdhar \(2004\)](#) introduz que o método de montagem por parafuso é o melhor método disponível para aplicações de montagem permanente em acelerômetros piezoelétricos, oferecendo a resposta em frequência mais ampla e a fixação mais segura e confiável. Este método é realizado parafusando o sensor em um pino ou um bloco usinado, permitindo que o transdutor meça a vibração da maneira mais ideal e deve ser usado sempre que possível. Os outros métodos, como montagem adesiva ou montagem magnética, reduzem a faixa de frequência superior do sensor.

Foram utilizados quatro acelerômetros fixados por parafusos na carcaça do mancal, sendo eles posicionados da seguinte forma: um na direção horizontal do mancal esquerdo; um na direção vertical do mancal direito; um na direção horizontal do mancal direito; e um na direção vertical do mancal direito. As especificações do acelerômetro piezoelétrico utilizado estão descritas na tabela 3.

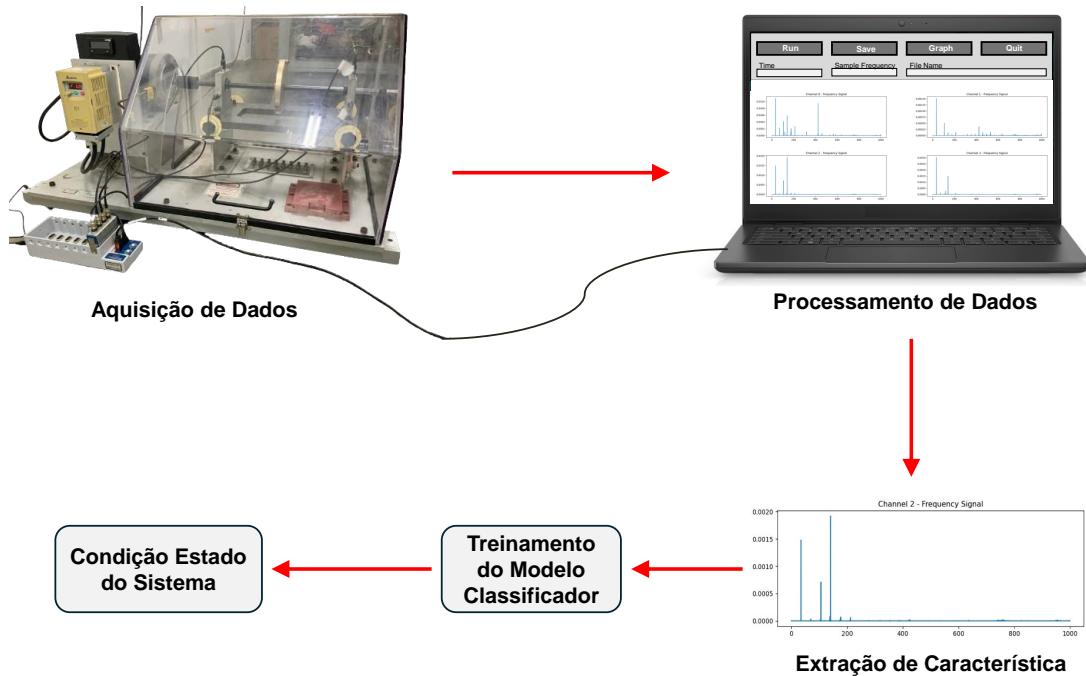
Tabela 3 – Especificações do Acelerômetro Piezoelétrico utilizado

Acelerômetro Modelo 608A11	
Fabricante	IMI
Sensibilidade	100 mV/g \pm 15% 10,2 mV/(m/s ²) \pm 15%
Faixa de frequência	0,5 a 10000 Hz \pm 3 dB
Faixa de medição	\pm 50 g \pm 490 m/s ²

Fonte: [Controls \(2024\)](#)

Os sensores mencionados foram integrados à placa de aquisição, onde ocorre a conversão dos sinais analógicos em sinais digitais. Nos experimentos, utilizou-se a placa NI 9233 da National Instruments, dotada de 4 canais de entrada e capaz de operar com uma taxa de aquisição de até 50,0 kHz por canal, associada ao chassis NI cDAQ 9178, responsável pela alimentação de energia e pela conexão com o computador através da interface USB. A figura 16 representa o processo de aquisição de dados e as etapas que se sucedem.

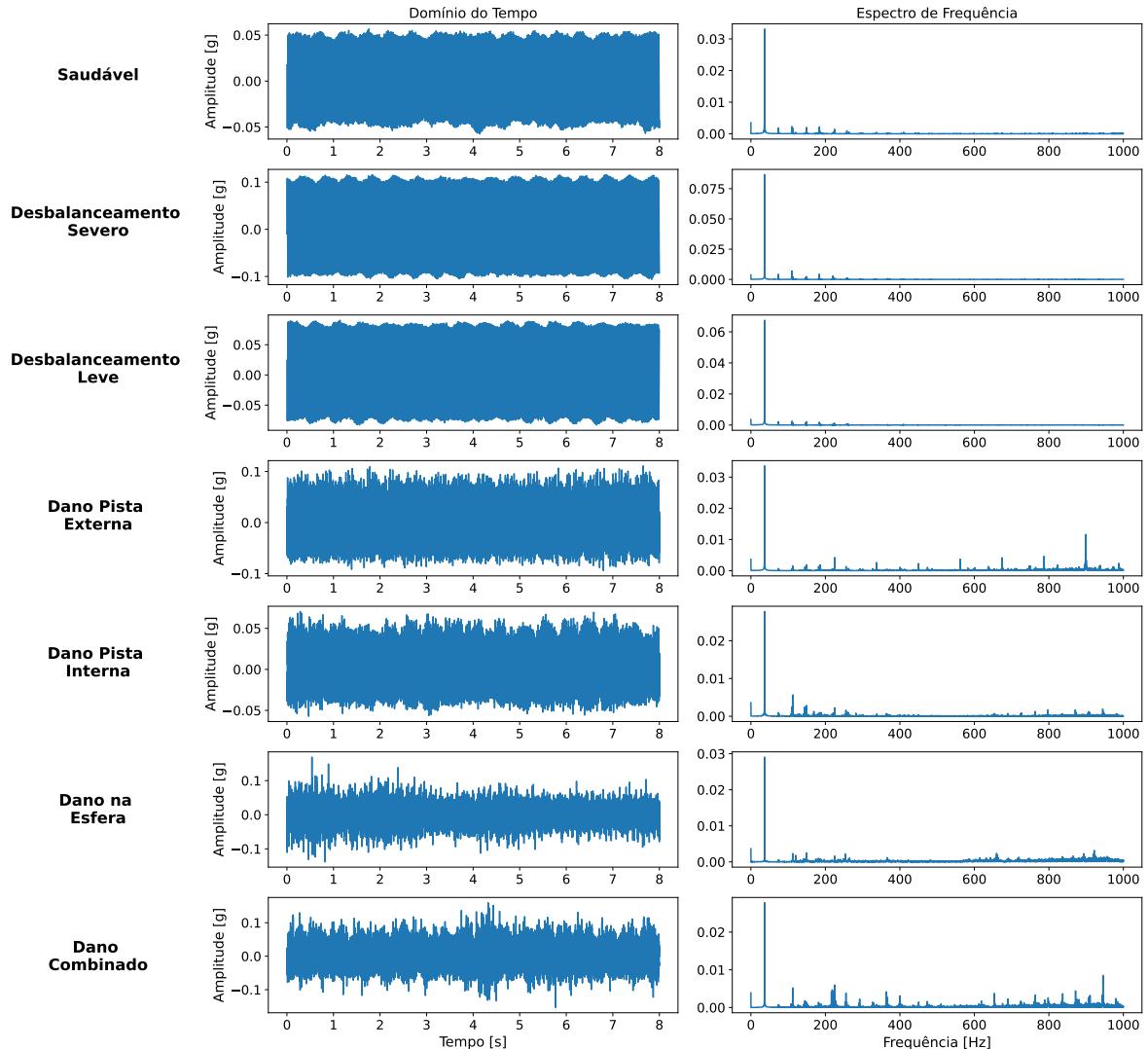
Figura 16 – Diagrama de fluxo de processos para monitoramento de condição



Fonte: Autoria própria.

No computador, foram realizados o processamento, a visualização e o armazenamento dos sinais obtidos por meio da integração do drive da placa com uma interface desenvolvida em *python*. Definiu-se via interface o tempo de aquisição como 8 segundos com uma taxa de amostragem de 2000 Hz para cada conjunto de dados. A figura 17 apresenta um exemplo dos dados no domínio do tempo e seu correspondente espectro de vibração para o sensor posicionado na horizontal do mancal de rolamento direito a velocidade de 37 Hz.

Figura 17 – Exemplo dos dados por condição para velocidade de rotação de 37 Hz



Fonte: Autoria própria.

A aquisição dos dados teve início após a definição da frequência de rotação através do inversor de frequência, seguida pelo estabelecimento da bancada em regime estacionário. Uma vez alcançada esta condição, adquiriu-se 10 conjuntos de dados. Posteriormente, mantendo a bancada operacional, procedeu-se à alteração da frequência de rotação e à aquisição de outros 10 conjuntos de dados após a estabilização do equipamento. Este procedimento foi repetido para as frequências de 27 Hz, 37 Hz e 47 Hz. Cada um dos acelerômetros supracitados corresponde a uma coluna no arquivo gerado.

3.1.2 Resumo dos dados

A tabela 4 apresenta o resumo apresentando a quantidade de testes realizados por condição.

Tabela 4 – Quantidade de teste por condição

Condição	Quantidade de testes na condição	Quantidade total de arquivos de dados
Saudável	4	120
Desbalanceamento leve	4	120
Desbalanceamento severo	4	120
Rolamento com dano na pista externa	6	180
Rolamento com dano na pista interna	6	180
Rolamento com dano na esfera	6	180
Rolamento com dano combinado	6	180
Total	36	1080

Fonte: Autoria própria

3.2 Extração das características

A qualidade dos dados é essencial para a construção de um bom modelo de aprendizado de máquina. Dados bem estruturados e com formatação consistente facilitam o pré-processamento e o treinamento do modelo, resultando em melhores previsões e generalizações.

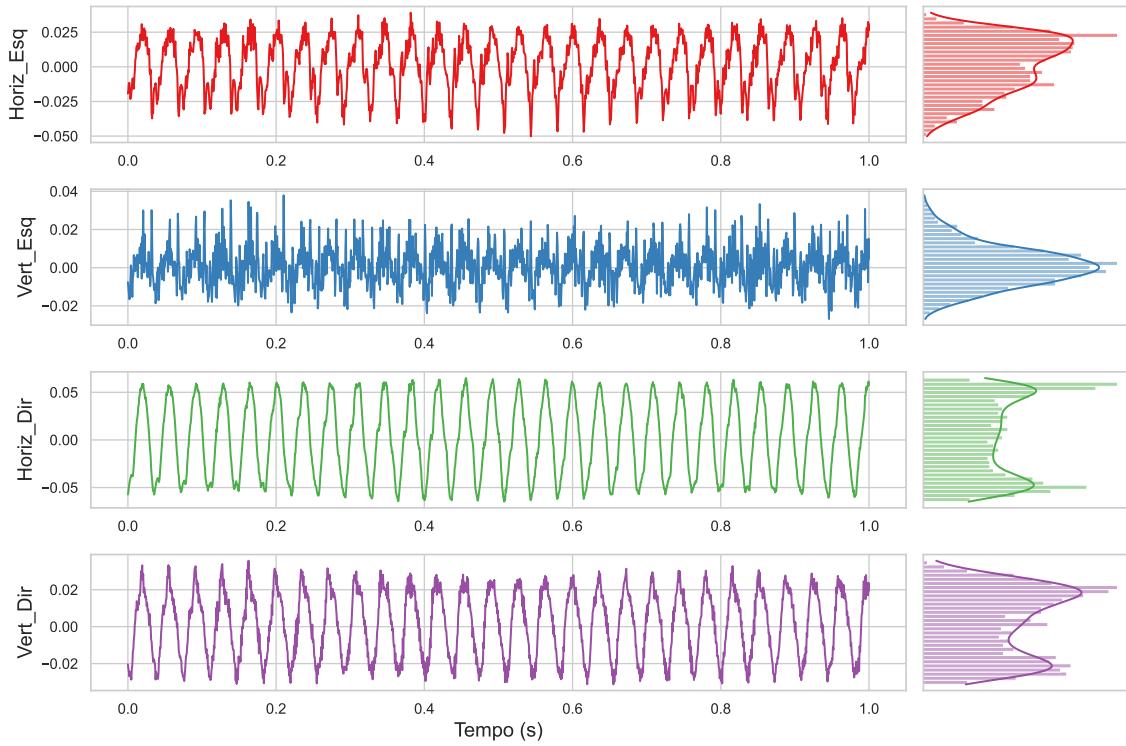
O sinal de vibração, como é recebido após a aquisição, não é adequado para ser utilizado diretamente na criação de modelos de aprendizado de máquina, pois é pouco estruturado, extenso e tem pouco valor quando considerado uma amostra isoladamente. Assim, busca-se extrair características dos registros de dados que, de forma agregada, indiquem as propriedades do sistema.

A escolha das características a serem extraídas do sinal é fundamental, pois algumas transmitem mais informação do que outras. Na literatura, há recomendações de características que podem ser extraídas do sinal de vibração para a predição de falhas, sendo estas no domínio do tempo ou no domínio da frequência.

3.2.1 Características no domínio do tempo

Ao tratar-se do domínio do tempo, uma série de medidas podem ser consideradas para caracterizar o sinal, em geral fazendo-se uso de teorias estatísticas. As medidas estatísticas são usadas para criar descritores que representam os sinais de maneira eficaz para classificação e identificação. A figura 18 apresenta um recorte de uma das condições e seus respectivos histogramas, indicando individualidade da assinatura do sinal medido por cada acelerômetro.

Figura 18 – Sinais no tempo e histogramas por eixo, para a condição saudável do equipamento a 37 Hz de rotação



Fonte: Autoria própria

No domínio do tempo, diversas medidas são utilizadas para caracterizar sinais. Abaixo estão algumas das principais medidas acompanhadas de suas definições ([RAUBER; BOLDT; VAREJAO, 2015](#); [XIA et al., 2012](#)):

- Valor Efetivo (RMS - *Root Mean Square*) Normalizado: Medida da amplitude eficaz do sinal, normalizada para facilitar comparações entre diferentes sinais. A equação 3.1 apresenta o cálculo da medida, onde x_i é cada amostra do sinal medido, N é o número total de amostras e f é a frequência de rotação naquele sinal.

$$RMS = \frac{1}{f} \left(\frac{1}{N} \sum_i^N x_i^2 \right)^{\frac{1}{2}} \quad (3.1)$$

- Raiz Quadrada da Amplitude (SRA - *Square Root of Amplitude*): Raiz quadrada da média das amplitudes dos valores absolutos dos pontos do sinal. A equação 3.2 apresenta o cálculo da medida, onde x_i é cada amostra do sinal medido e N é o número total de amostras.

$$SRA = \left(\frac{1}{N} \sum_i^N \sqrt{|x_i|} \right)^2 \quad (3.2)$$

- Assimetria (SV - *Skewness*): Indica a assimetria da distribuição dos dados do sinal em relação à média. Valores positivos indicam que a distribuição tem mais valores diferentes à direita da média do que à sua esquerda, enquanto valores negativos indicam uma cauda mais longa à esquerda. A equação 3.3 apresenta o cálculo da medida, onde x_i é cada amostra do sinal medido, N é o número total de amostras e σ é o desvio padrão.

$$SV = \left(\frac{1}{N} \sum_i \frac{x_i - \bar{x}}{\sigma} \right)^3 \quad (3.3)$$

- Curtose (KV - *Kurtosis*): Medida da forma da distribuição dos dados do sinal. Indica ocorrência de maior ou menor valores discrepantes em comparação com uma distribuição normal. A equação 3.4 apresenta o cálculo da medida, onde x_i é cada amostra do sinal medido, N é o número total de amostras e σ é o desvio padrão.

$$KV = \left(\frac{1}{N} \sum_i \frac{x_i - \bar{x}}{\sigma} \right)^4 \quad (3.4)$$

- Amplitude Máxima do Sinal pelo Valor Pico-a-Pico (PPV - *Peak-to-Peak Value*): Diferença entre o valor máximo e o valor mínimo do sinal durante um período de tempo específico. A equação 3.5 apresenta o cálculo da medida, onde $\max_i(x)$ é a amplitude máxima do sinal medido e $\min_i(x)$ é a amplitude mínima do sinal.

$$PPV = \max_i(x) - \min_i(x) \quad (3.5)$$

- Fator Adimensional de Crista (CF - *Crest Factor*): Razão entre o valor de pico do sinal e seu valor RMS. Indica a relação entre os picos máximos e a amplitude média do sinal. A equação 3.6 apresenta o cálculo da medida, onde $\max_i(|x|)$ é a amplitude máxima absoluta do sinal medido.

$$CF = \frac{\max_i(|x|)}{RMS} \quad (3.6)$$

- Fator Adimensional de Impulso (IF - *Impulse Factor*): Medida da magnitude dos impulsos no sinal, geralmente expressa como a razão entre o valor de pico do sinal e a média dos valores absolutos das amostras. A equação 3.7 apresenta o cálculo da medida, onde $\max_i(|x|)$ é a amplitude máxima absoluta do sinal medido, N é o número total de amostras e x_i é cada amostra do sinal medido.

$$IF = \frac{\max_i(|x|)}{\frac{1}{N} \sum_i^N |x_i|} \quad (3.7)$$

- Fator Adimensional de Margem (MF - *Margin Factor*): Medida da margem de segurança

entre o sinal e um limiar específico, frequentemente utilizado em análises de tolerância e robustez. A equação 3.8 apresenta o cálculo da medida, onde $\max_i(|x|)$ é a amplitude máxima absoluta do sinal medido.

$$MF = \frac{\max_i(|x|)}{SRA} \quad (3.8)$$

- Fator Adimensional de Formato (SF - *Shape Factor*): Indica a forma geral do sinal, avaliando a proporção entre a amplitude média e a média dos valores absolutos das amostras. A equação 3.8 apresenta o cálculo da medida, onde N é o número total de amostras e x_i é cada amostra do sinal medido.

$$SF = \frac{RMS}{\frac{1}{N} \sum_i^N |x_i|} \quad (3.9)$$

- Fator Adimensional de Curtose (KF - *Kurtosis Factor*): Medida da forma da distribuição dos dados do sinal, sendo calculada como uma relação entre a curtose e valor efetivo do sinal.

$$KF = \frac{KV}{RMS^4} \quad (3.10)$$

3.2.2 Características no domínio da frequência

A principal forma de conversão de um sinal temporal, tal como medida por um acelerômetro, em informação no domínio da frequência é através da transformada de Fourier ou uma série de Fourier. A teoria desenvolvida indica que qualquer função que se repete no tempo pode ser expressa por uma série infinita conforme mostrado na equação 3.11 (INMAN, 2013):

$$F(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega_T t + b_n \sin n\omega_T t) \quad (3.11)$$

$$\omega_T = \frac{2\pi}{T} \quad (3.12)$$

$$a_0 = \frac{2\pi}{T} \int_0^T F(t) dt \quad (3.13)$$

$$a_n = \frac{2}{T} \int_0^T F(t) \cos n\omega_T t dt \quad n = 1, 2, \dots \quad (3.14)$$

$$b_n = \frac{2}{T} \int_0^T F(t) \sin n\omega_T t dt \quad n = 1, 2, \dots \quad (3.15)$$

Os coeficientes a_0 , a_n e b_n representam informações do sinal temporal no domínio da frequência (INMAN, 2013). Esse processo é realizado de maneira simples pela implementação da função *rfft* da biblioteca do *Numpy* em linguagem *Python*. Esta função calcula a transformada

de Fourier discreta (DFT) de n pontos unidimensional de uma matriz de valor real por meio de um algoritmo eficiente chamado *Fast Fourier Transform* (FFT).

Com os dados no domínio da frequência, é possível extrair diversas características. Para isso, utilizou-se o pacote em Python *tsfresh*, que calcula automaticamente um grande número de características a partir de séries temporais, ou seja, da condição base em que o sinal foi adquirido (TSFRESH DEVELOPMENT TEAM, 2024).

As características de frequências, conforme a literatura e sintetizado no capítulo 2, são diretamente ligadas com a classificação do defeito. Dessa forma, quatro características foram selecionadas: centroide espectral, variância, assimetria e curtose do espectro absoluto da transformada de fourier (RAUBER; BOLDT; VAREJAO, 2015; TSFRESH DEVELOPMENT TEAM, 2024).

- Centroide espectral: Corresponde à posição do centroide da frequência, representando o ponto de equilíbrio do espectro de frequências. O centroide espectral é então calculado como a média ponderada das amplitudes, onde a ponderação é feita pelas próprias amplitudes normalizadas, conforme apresentado pela equação 3.16, onde A_i são as amplitudes na i -ésima frequência após a FFT, f_i é a frequência correspondente à i -ésima amplitude, e os somatórios estão sobre todas as frequências consideradas na análise

$$FFC = \frac{\sum_i A_i \cdot f_i}{\sum_i A_i} \quad (3.16)$$

- Variância: Corresponde à dispersão das componentes de frequência em torno da média. Uma alta variância pode indicar a presença de componentes freqüenciais dominantes que podem ser anômalos. A equação 3.17 apresenta essa medida, onde A_i são as magnitudes das frequências obtidas pela FFT, \bar{A} é a média das magnitudes das frequências e n é o número total de amostras.

$$FVAR = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (A_i - \bar{A})^2 \quad (3.17)$$

- Assimetria: Corresponde a medida que indica se as frequências estão inclinadas para a direita ou esquerda. Nesse sentido, pode indicar se há componentes de frequência com magnitudes muito altas em relação à média. A equação 3.18 apresenta essa medida, onde A_i são as magnitudes das frequências obtidas pela FFT, \bar{A} é a média das magnitudes das frequências e n é o número total de amostras (ANDERSON, 2024).

$$FSKEW = \text{Assimetria} = \frac{\frac{1}{n} \sum_{i=1}^n (A_i - \bar{A})^3}{\left(\frac{1}{n} \sum_{i=1}^n (A_i - \bar{A})^2 \right)^{3/2}} \quad (3.18)$$

- Curtose: Corresponde à quanto pontiagudo ou achatamento do espectro de frequências. Ela indica se há mais valores discrepantes (*outliers*) na distribuição do que o esperado pela

distribuição normal. A Curtose pode ser calculado pela equação 3.19, onde A_i são as magnitudes das frequências obtidas pela FFT, \bar{A} é a média das magnitudes das frequências e n é o número total de amostras (ANDERSON, 2024).

$$FKUR = \frac{\frac{1}{n} \sum_{i=1}^n (A_i - \bar{A})^4}{\left(\frac{1}{n} \sum_{i=1}^n (A_i - \bar{A})^2\right)^2} - 3 \quad (3.19)$$

3.2.3 Base de dados

Conforme discutido e resumido na tabela 4, foram adquiridos dez conjuntos de dados para cada frequência em cada condição específica de montagem, com uma taxa de amostragem de 2000 Hz por um período de 8 segundos. Isso resulta em cada arquivo contendo 16000 amostras.

Para aumentar a quantidade total de dados disponíveis para o modelo, os dados foram posteriormente divididos em conjuntos menores de amostras. Foram testadas várias possibilidades de divisão inteira do sinal, com um máximo de partes de 0,5 segundos. O menor tamanho de amostra que representou o sinal sem perder suas características originais, de acordo com as medidas utilizadas, foi o intervalo de 2 segundos, contendo 4000 amostras na mesma taxa de amostragem.

Portanto, o banco de dados resultante das características extraídas é composto por 56 características, distribuídas em 14 características por acelerômetro. Desses, 40 são derivadas de medidas estatísticas no domínio do tempo, enquanto as 16 restantes são baseadas em medidas estatísticas no domínio da frequência. Adicionalmente, o banco de dados foi dividido previamente em conjuntos de treino e teste. Esse procedimento permite que o modelo seja treinado em uma condição específica e testado na mesma condição após a remontagem da bancada. Importante ressaltar que, entre a coleta de dados de uma condição inicial e a coleta de dados da mesma condição após a remontagem da bancada, ocorreram pelo menos outras três montagens distintas.

4 Apresentação e discussão dos resultados

Este capítulo apresentará os resultados e suas respectivas discussões, organizadas de forma comparativa entre três aspectos principais: a escolha do melhor conjunto de características, a localização dos sensores (mancal esquerdo ou direito) e os eixos de medição (vertical ou horizontal).

Antes de prosseguirmos, é importante destacar que a métrica ideal de avaliação de eficiência é de 100% de acerto. No entanto, alcançar essa acurácia é raro. O valor considerado ótimo é variável e depende dos riscos humanos, financeiros e ambientais associados a uma predição errônea.

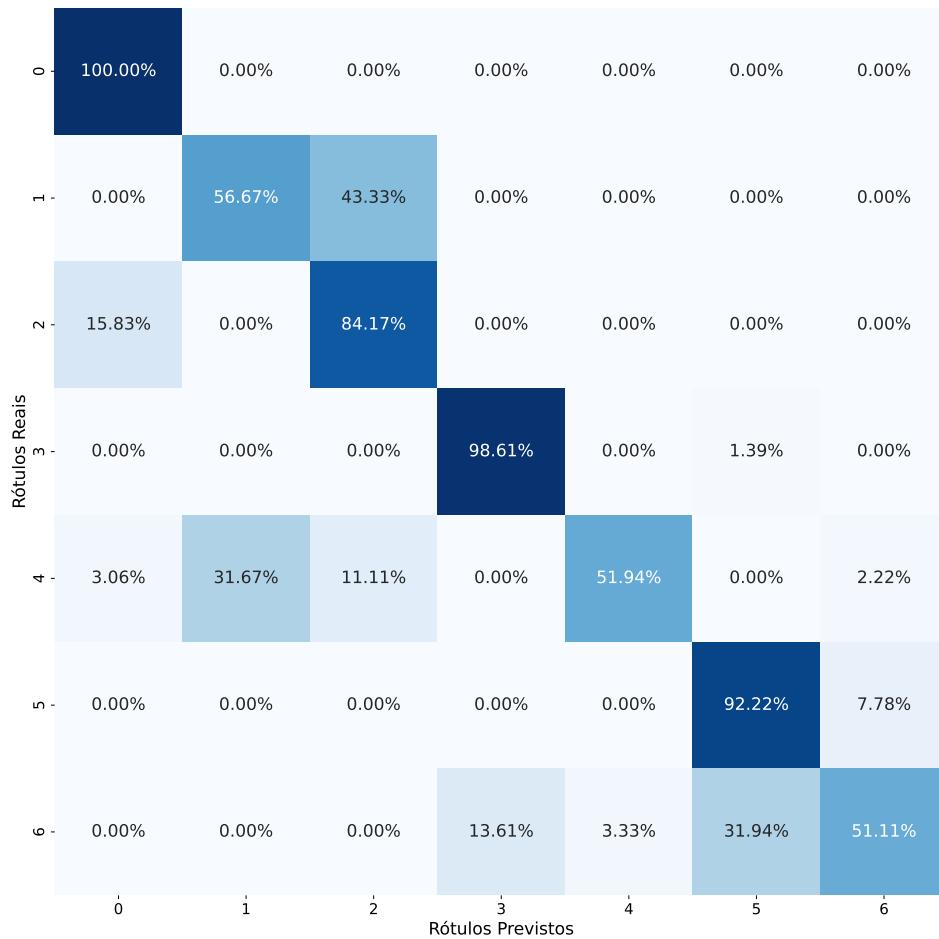
Além disso, é importante ressaltar que, apesar dos dados terem sido medidos em três diferentes velocidades de rotação, a informação da velocidade não faz parte das características utilizadas nos conjuntos de treino ou teste, sendo os valores das características para os três casos adicionados de maneira indistinta na base geral. Outro desafio para a classificação são as remontagens, uma vez que os dados de teste foram coletados de uma condição de montagem que o algoritmo nunca encontrou anteriormente.

4.1 Comparação entre escolha das característica dos dados

A capacidade preditiva do classificador depende fortemente das características utilizadas. Algumas características carregam mais informações relevantes para a classificação, enquanto outras são insensíveis às mudanças no estado de saúde do equipamento. Além disso, mesmo que uma característica possua grande capacidade de classificação, se houver alta correlação com outras classes, o modelo pode se tornar tendencioso em relação a esses parâmetros intercorrelacionados. Isso reduz a sensibilidade a outras características igualmente importantes e, consequentemente, piora o desempenho geral do modelo.

Reunindo as 56 características calculadas na seção 3.2, o modelo demonstrou desempenho, a partir das métricas apresentadas no capítulo 2, conforme mostrado na tabela 7. Também, a figura 19 apresenta a matriz de confusão de classificação para este caso. Nesta matriz, o rótulo 0 representa a condição saudável, 1 desbalanceamento severo, 2 desbalanceamento leve, 3 dano na pista externa do rolamento, 4 dano na pista interna do rolamento, 5 dano nas esferas e 6 danos combinados no rolamento.

Figura 19 – Matriz de confusão utilizando todas as características



Fonte: Autoria própria

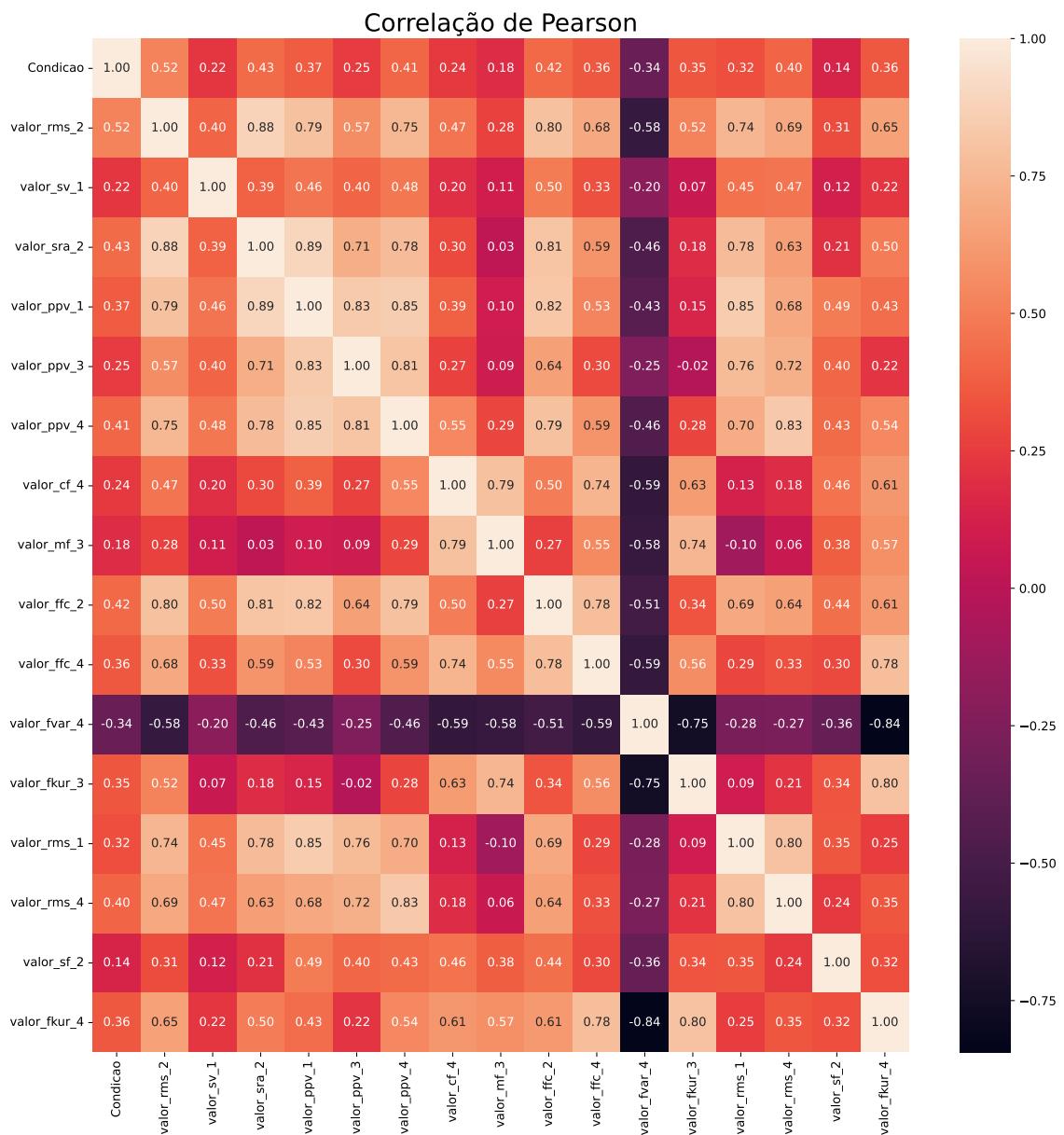
Visando aprimorar o modelo, selecionou-se a melhor combinação de características extraídas dos dados através de uma combinação entre análise de correlação e o método *Tree-based feature selection*. Primeiramente, um pré-conjunto de características foi selecionado com base em sua importância para o algoritmo de floresta aleatória, utilizando a propriedade *feature_importances_* intrínseca ao modelo preditivo ([SCIKIT LEARN, 2024d](#)). Em seguida, as características pré-selecionadas foram analisadas manualmente com base em sua matriz de correlação. Para isso, foram excluídas as características com correlações menores que 0,14 com a condição e também aquelas com correlação acima de 0,9 com outra característica. As características selecionadas para o modelo que utiliza todos os acelerômetros estão listadas na tabela 5, e a matriz de correlação final está apresentada na figura 20.

Tabela 5 – Características selecionadas para cada acelerômetro

Acelerômetro				Característica Selecionada														
Horizontal Esquerdo				RMS, SV e PPV														
Vertical Esquerdo				RMS, SRA, SF e FFC														
Horizontal Direito				PPV, MF e FKUR														
Vertical Direito				RMS, PPV, CF, FFC e FKUR														

Fonte: Autoria própria

Figura 20 – Correlação entre as características selecionadas



Fonte: Autoria própria

Em seguida, utilizou-se o método de otimização de hiper-parâmetros descrito no capítulo

2 para definir a melhor combinação de valores para o treinamento do modelo, a partir de um conjunto amplo de possibilidades. O resultado encontrado para a condição com todos os acelerômetros está listado na tabela 6

Tabela 6 – Seleção de hiper-parâmetros pelo método de *GridSearchCV*

Parâmetro	Valor
bootstrap	False
max_depth	15
max_features	'sqrt'
min_samples_leaf	1
n_estimators	3000
min_samples_split	5
Demais Parâmetros	Valor padrão

Fonte: Autoria própria

Por fim, os resultados obtidos são apresentados na matriz de confusão mostrada na figura 21. Nesta matriz, o rótulo 0 representa a condição saudável, 1 desbalanceamento severo, 2 desbalanceamento leve, 3 dano na pista externa do rolamento, 4 dano na pista interna do rolamento, 5 dano nas esferas e 6 danos combinados no rolamento.

Figura 21 – Matriz de confusão utilizando características selecionadas



Fonte: Autoria própria

Tabela 7 – Métricas por seleção de características

Métrica	Todas as características	Características selecionadas
Acurácia	74,83%	87,17%
Precisão	79,10%	88,18%
Recall	74,83%	87,17%
F1-Score	74,24%	87,03%

Fonte: Autoria própria

Desse modo, observa-se uma melhora significativa do modelo otimizado, com um desempenho geral 10% superior para o caso de características selecionadas. Esse resultado é coerente com o esperado, conforme discutido no capítulo 2.

As principais diferenças na capacidade preditiva encontram-se nas condições de desbalanceamento, onde o modelo com todas as características acertou 56,67% para desbalanceamento severo e 84,17% para desbalanceamento leve. Além disso, no caso com todas as 56 características, houve erro na classificação, indicando uma situação com falha como sendo saudável. Essa indicação é um grande problema, pois pode resultar na decisão de continuar o funcionamento do equipamento tendo como fim uma falha catastrófica inesperada.

A principal deficiência do modelo otimizado foi a classificação incorreta do dano na pista externa do mancal de rolamento, identificando-o como desbalanceamento. Embora outras classes apresentem um percentual maior de erro, esses erros permanecem dentro do mesmo componente defeituoso. Em contraste, classificar erroneamente a falha, como no caso do dano na pista interna, pode resultar em um tempo de parada mais longo para a equipe de manutenção assistida pelo modelo, reduzindo o potencial de ganhos do processo preditivo. No caso com todas as características, essa classificação errônea apresentou porcentagens ainda maiores.

Das 14 características disponíveis, apenas 9 foram selecionadas para uso pelo modelo, conforme mostrado na tabela 5. Essa redução não apenas otimiza o desempenho do modelo, mas também contribui para a diminuição do tamanho do banco de dados. Isso é crucial especialmente em contextos industriais, onde a gestão de bases de dados volumosas pode representar um desafio significativo para a implementação de modelos robustos.

Por fim, observa-se um equilíbrio na quantidade de características entre os acelerômetros, conforme mostrado na tabela 5. Isso destaca a importância de todos os quatro sensores para uma predição precisa. Além disso, é notável a presença repetida dos parâmetros RMS, PPV, FFC e FKUR, os quais também são empregados em produtos comerciais (TRACTIAN, 2021).

4.2 Comparação entre a localização dos sensores

A localização dos transdutores de vibração o mais próximo possível do dano é crucial para uma medição precisa da assinatura de vibração, pois minimiza a influência do caminho

entre a fonte de vibração e o dano na característica do sinal. Além disso, uma localização estratégica permite reduzir a quantidade de sensores necessários, o que é economicamente vantajoso, considerando que transdutores de alta qualidade e outros equipamentos necessários para a aquisição possuem um custo elevado.

Utilizando o mesmo processo de seleção de características e otimização de hiperparâmetros descrito na seção anterior, comparou-se o desempenho dos acelerômetros localizados no mancal de rolamento esquerdo com os localizados no mancal de rolamento direito. Os resultados obtidos estão listados na tabela 8.

Tabela 8 – Métricas por localização dos acelerômetros

Métrica	Acelerômetros à Esquerda	Acelerômetros à Direita
Acurácia	64,72%	65,06%
Precisão	65,69%	74,27%
Recall	64,72%	65,06%
F1-Score	64,02%	66,18%

Fonte: Autoria própria

A análise revela que utilizar apenas os sensores do lado direito resultou em desempenho ligeiramente melhor do que os sensores do lado esquerdo, especialmente em termos de precisão. Isso se deve principalmente à maior precisão na identificação de danos nos rolamentos, dado que os danos estavam localizados exclusivamente no lado direito da máquina. No entanto, a taxa de acerto na detecção de desbalanceamento não ultrapassou os 65% em nenhum dos casos, o que teve um impacto significativo nas métricas gerais. Além disso, em ambos os cenários, a precisão na classificação de condições saudáveis não foi superior a 86%, comparado aos resultados anteriores.

4.3 Comparação entre eixos de medição

A dinâmica do comportamento do sistema varia conforme o tipo de dano, pois determinados danos podem gerar forças na direção vertical em vez da direção axial ou horizontal, por exemplo. Outra situação é que diferenças na rigidez da carcaça onde os transdutores são montados podem influenciar a medição, já que um transdutor pode estar posicionado em uma região mais rígida do que outro em uma direção diferente. Uma análise nesse sentido permite avaliar se a utilização de um acelerômetro em determinada direção é realmente necessária, considerando que equipamentos de alta qualidade para aquisição de dados possuem um custo elevado.

Utilizando a otimização conforme descrito na seção anterior, comparou-se o desempenho dos acelerômetros localizados no eixo horizontal com os localizados no eixo vertical. Os resultados obtidos estão listados na tabela 9.

Tabela 9 – Métricas por eixo de medição

Métrica	Acelerômetros na Horizontal	Acelerômetros na Vertical
Acurácia	51,89%	75,83%
Precisão	59,61%	76,12%
Recall	51,89%	75,83%
F1-Score	51,51%	74,93%

Fonte: Autoria própria

É evidente um desempenho significativamente superior dos acelerômetros no eixo vertical em comparação aos do eixo horizontal. Isso se deve à capacidade superior de detecção de defeitos como o desbalanceamento na direção vertical, como destacado por [Randall \(2011\)](#) e também observado nos resultados de [Ramos \(2020\)](#). No entanto, esse resultado ainda fica aquém do desempenho alcançado pelo uso combinado desses sensores.

4.4 Resumo dos resultados

A tabela 10 apresenta um resumo dos resultados encontrados nas comparações supracitadas neste capítulo.

Tabela 10 – Métricas para os casos testados

Métrica	4 Acc Otimizado	Todas as Características	Acc à Esquerda	Acc à Direita	Acc na Horizontal	Acc na Vertical
Acurácia	87,17%	74,83%	64,72%	65,06%	51,89%	75,83%
Precisão	88,18%	79,10%	65,69%	74,27%	59,61%	76,12%
Recall	87,17%	74,83%	64,72%	65,06%	51,89%	75,83%
F1-Score	87,03%	74,24%	64,02%	66,18%	51,51%	74,93%

Fonte: Autoria própria

Observa-se que a maior quantidade de acelerômetros otimizado apresenta melhor resultado do que para o caso com limitação de sensores. Contudo, essa situação apresenta maior custo agregado na aquisição, sendo de responsabilidade da gestão de manutenção avaliar a necessidade.

Além disso, observa-se uma alta assertividade na distinção entre uma classificação binária de equipamento saudável ou danificado para os casos em que usou-se os 4 acelerômetros e para o caso de acelerômetros na vertical. Este resultado é de grande valor, tornando o modelo uma ferramenta útil para a gestão da manutenção. Ele auxilia na realização de paradas do equipamento no momento adequado, garantindo intervenções mais precisas e eficientes.

A análise dos casos mencionados anteriormente, ilustrada na figura 21, revela que, com exceção dos erros na detecção de danos na pista interna, os demais erros na classificação de danos acertaram a categoria global do defeito. Ou seja, a classificação entre condição saudável,

desbalanceamento e dano em rolamento (sem especificação do local) foi quase totalmente precisa. Isso sugere que o modelo pode efetivamente reduzir o tempo total de manutenção ao direcionar corretamente para o componente que requer reparo, aumentando assim a disponibilidade operacional.

5 Considerações finais

Para reduzir os custos de maneira eficiente garantindo a alta performance, é essencial implementar as melhores práticas de manutenção. Isso inclui a aplicação de técnicas e as evoluções no processo avaliativo, focando na identificação das causas principais das falhas. A qualidade dos serviços é crucial, refletindo-se na diminuição de retrabalho, assim como na qualidade dos materiais e peças sobressalentes utilizados.

Para chegar ao processo final de predição, é essencial utilizar um banco de dados. Nesse sentido, este trabalho conseguiu construir uma base de dados por meio de experimentos padronizados na bancada MFS, a qual suportou a implementação de um algoritmo preditivo.

Conforme discutido no capítulo 2, há uma maturidade considerável no entendimento das assinaturas de vibrações e nos métodos para obtê-las. No entanto, o espectro de vibração apresenta características particulares para cada sistema mecânico. Apesar disso, os algoritmos de aprendizado de máquina, após serem devidamente treinados, podem se tornar ferramentas valiosas na classificação dessas características e sobreposição dessa adjecção. Essa observação é corroborada pelo aumento de artigos, livros e ferramentas comerciais que tratam do assunto.

O diagnóstico correto da falha é crucial, pois representa a fase inicial de uma série de processos subsequentes, seja para intervenção imediata ou para garantir a continuidade das operações (KARDEC; NASCIF, 2009). Nesse contexto, o modelo desenvolvido neste estudo demonstrou um bom desempenho nos testes realizados após a remontagem, alcançando uma acurácia de 87%. Esse resultado pode ser considerado significativo ao compará-lo com sistemas comerciais, como o oferecido por (TRACTIAN, 2021), onde é prometida uma precisão de 89%.

Neste trabalho, foram considerados dados de teste provenientes de uma remontagem diferente daquela utilizada no treinamento do modelo. Essa remontagem pode ser equiparada, na prática, a uma situação de manutenção no equipamento. Com base no desempenho do modelo, isso sugere que não seria necessário realizar um novo processo de treinamento para garantir sua eficácia.

Um dos principais problemas inerentes a monitoração da condição por vibração é diagnosticar se o motivo da alteração das características do espectro está sendo causado pela fonte real ou pelo caminho entre a fonte e o transdutor (RANDALL, 2011). Nesse sentido, o resultado encontrado no capítulo 4, indica que um maior conjunto de sensores, bem como uma localização adequada permite minimizar os efeitos do caminho na análise.

Por fim, destaca-se a importância de softwares livres para a realização de todas as etapas deste trabalho. A implementação de muitos dos métodos e análises foi possível graças ao uso de bibliotecas e ferramentas de código aberto, como o scikit-learn, que proporcionou algoritmos

eficientes de aprendizado de máquina para o processamento e a classificação dos dados. Além disso, os pacotes de integração da National Instruments foram essenciais para a aquisição e manipulação dos dados experimentais, permitindo uma interface robusta e flexível entre os equipamentos de medição e o ambiente de desenvolvimento. A combinação dessas ferramentas livres e acessíveis é crucial para o desenvolvimento e a execução bem-sucedida de trabalhos nessa área. Quanto a este trabalho, os códigos utilizados no desenvolvimento encontram-se disponíveis no repositório [GitHub](#).

5.1 Trabalhos futuros

Este trabalho foi desenvolvido com uma quantidade limitada de dados, utilizando-se apenas 10 conjuntos de dados por condição. Para contornar essa limitação, fez-se necessário a adoção de técnicas para aumentar a quantidade de dados disponíveis. Além disso, a aquisição de dados foi realizada em apenas três níveis de velocidades. Portanto, para trabalhos futuros, recomenda-se a construção de uma base de dados mais robusta, com um maior número de conjuntos de dados por condição e um intervalo mais amplo de velocidades, a fim de melhorar a representatividade e a eficácia dos modelos desenvolvidos.

Outra abordagem para o monitoramento de condição é a comparação entre diferentes métodos de monitoramento para máquinas rotativas. Nesse sentido, recomenda-se, para trabalhos futuros, a análise da ação individual e conjunta do monitoramento de vibração e da corrente elétrica do motor.

Recomenda-se também investigar o problema utilizando dados coletados em datas específicas para o treinamento do modelo e, posteriormente, dados coletados após um intervalo de tempo, como de um ou dois meses, para o teste. Isso porque é esperado que ocorram variações na assinatura do espectro de frequência ao longo do tempo, dificultando a ação do modelo preditivo ([RANDALL, 2011](#)).

Outra direção a seguir é explorar a classificação em tempo real utilizando o modelo preditivo treinado em uma base de dados, além de desenvolver uma interface de usuário que apresente essa ferramenta preditiva de maneira prática e acessível.

Referências

- ABECOM. *Como identificar falhas em sistemas rotativos e prevenir paradas não programadas*. 2022. Acessado em 14 de junho de 2024. Disponível em: <<https://www.abecom.com.br/falhas-em-sistemas-rotativos/>>. Citado 3 vezes nas páginas 14, 15 e 16.
- ADAMS, M. L. *Rotating Machinery Vibration: From Analysis to Troubleshooting, Second Edition*. 2. ed. [S.l.: s.n.], 2009. ISBN 1439807175; 9781439807170. Citado 3 vezes nas páginas 14, 15 e 16.
- ANDERSON, B. *Medidas de Forma*. statorial, 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://statorial.org/pt/medidas-de-forma/>>. Citado 2 vezes nas páginas 39 e 40.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TECNICAS. *NBR 5462: Confiabilidade e manutenabilidade*. [S.I.]: ABNT, 1994. Citado na página 11.
- BASSANI, M. J. et al. Identificação de defeitos em equipamentos rotativos através da análise de vibração. *ORNADA DE INICIACAO CIENTIFICA DA FAACZ*, 2018. Acessado em 14 de junho de 2024. Disponível em: <http://www.faacz.com.br/portal/conteudo/iniciacao_cientifica/programa_de_iniciacao_cientifica/2018/anais/identificacao_de_defeitos_em_equipamentos_rotativos_atraves_da_analise_de_vibracao.pdf>. Citado na página 16.
- BUDYNAS, R.; NISBETH, J. *Elementos de Máquinas de Shigley - 10ª Edição*. McGraw Hill Brasil, 2016. ISBN 9788580555554. Disponível em: <<https://books.google.com.br/books?id=ekN4CwAAQBAJ>>. Citado na página 17.
- CHARBUTY, B.; ABDULAZEEZ, A. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, v. 2, n. 01, p. 20 – 28, Mar. 2021. Disponível em: <<https://www.jastt.org/index.php/jasttpath/article/view/65>>. Citado 2 vezes nas páginas 21 e 22.
- CONTROLS, S. *IMI 608A11 Accelerometer*. 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://www.spartancontrols.com/products/asset-management-reliability/rotating-equipment-condition-monitoring/imi-608a11-accelerometer/>>. Citado na página 32.
- DOMINGOS, P. A few useful things to know about machine learning. *Communications of the ACM*, ACM New York, NY, USA, v. 55, n. 10, p. 78–87, 2012. Citado 2 vezes nas páginas 18 e 20.
- GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. [S.I.]: O'Reilly Media, 2019. ISBN 9781492032618. Citado 5 vezes nas páginas 18, 19, 20, 23 e 25.
- IBM. *O que é aprendizado supervisionado?* 2024. Acessado em 17 de junho de 2024. Disponível em: <<https://www.ibm.com/br-pt/topics/supervised-learning>>. Citado na página 20.
- INMAN, D. J. *Engineering Vibration (4th Edition)*. [S.I.]: Prentice Hall, 2013. Citado 3 vezes nas páginas 15, 16 e 38.

- KARDEC, A.; NASCIF, J. *Manutenção - Função Estratégica*. 3. ed. [S.I.]: QUALITYMARK, 2009. ISBN 9788573033236. Citado 5 vezes nas páginas 11, 14, 31, 32 e 49.
- LIMA, A. de; ZAMBRANO-LÓPEZ, R.; NETTO, S. L. Diagnóstico de falhas em máquinas rotativas utilizando random forest. *XXXIV SIMPOSIOS BRASILEIRO DE TELECOMUNICACOES*, 2016. Disponível em: <<https://www02.smt.ufrj.br/~sergioln/papers/BC28.pdf>>. Citado 2 vezes nas páginas 12 e 26.
- MAFALDA. *Machinery Fault Database*. 2021. Acesso em: 20 de junho de 2024. Disponível em: <<https://www02.smt.ufrj.br/~offshore/mfs/index.html#TOC1>>. Citado na página 26.
- MANUALSLIB. *Delta Electronics VFD007S21B Product Manual*. 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://www.manualslib.com/products/Delta-Electronics-Vfd007s21b-13583726.html>>. Citado na página 30.
- MANYIKA, J. Big data: The next frontier for innovation, competition, and productivity. In: . [s.n.], 2011. Disponível em: <<https://api.semanticscholar.org/CorpusID:166449414>>. Citado na página 11.
- MOTION. *Marathon Electric 56T34F5301J - General Purpose Motor*. 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://www.motion.com/products/sku/03668693>>. Citado na página 28.
- NILSSON, N. J. *Introduction to Machine Learning: An Early Draft of a Proposed Textbook*. Stanford University, 1998. Disponível em: <<https://ai.stanford.edu/~nilsson/MLBOOK.pdf>>. Citado 2 vezes nas páginas 24 e 25.
- PAUDYAL, S. *Classification of Rotating Machinery Fault Using Vibration Signal*. Dissertação (Dissertação de Mestrado) — The University of North Dakota, 2019. Disponível em: <<https://commons.und.edu/theses/2579/>>. Citado na página 26.
- PRUFTECHNIK. *Vibration Measurement Pruftechnik Services in Detail*. 2024. Acessado em 20 de junho de 2024. Disponível em: <<https://www.pruftechnik.com/en-US/Products-and-Services/Services/Vibration-Analysis-and-Balancing-Services/>>. Citado na página 26.
- RAMOS, R. T. *Diagnóstico automático de falhas em máquinas rotativas utilizando máquina de vetores-suporte*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, Vitoria, ES, maio 2020. Citado 2 vezes nas páginas 26 e 47.
- RANDALL, R. B. *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. [S.I.]: John Wiley Sons, 2011. Citado 7 vezes nas páginas 11, 12, 14, 16, 47, 49 e 50.
- RAUBER, T. W.; BOLDT, F. de A.; VAREJAO, F. M. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, v. 62, n. 1, p. 637–646, 2015. Citado 2 vezes nas páginas 36 e 39.
- REXNORD. *ER10K 5/8 Unmounted Ball Bearing*. 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://www.rexnord.com/products/bearings/ball-bearings/unmounted-replacement-bearings/er-k/er10k>>. Citado na página 28.
- SCHEFFER, C.; GIRDHAR, P. *Practical machinery vibration analysis and predictive maintenance*. 1. ed. [S.I.]: Elsevier; Newnes, 2004. (Practical professional books from Elsevier). ISBN 0750662751; 9780750662758. Citado 4 vezes nas páginas 11, 14, 17 e 32.

SCIKIT LEARN. *Classifier comparison*. 2024. Acessado em 12 de junho de 2024. Disponível em: <https://scikit-learn/stable/auto_examples/classification/plot_classifier_comparison.html>. Citado na página 21.

SCIKIT LEARN. *Feature Selection*. 2024. Acessado em 19 de junho de 2024. Disponível em: <https://scikit-learn.org/stable/modules/feature_selection.html>. Citado na página 25.

SCIKIT LEARN. *GridSearchCV*. 2024. Acessado em 19 de junho de 2024. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>. Citado na página 25.

SCIKIT LEARN. *Random Forest Classifier*. 2024. Acessado em 12 de junho de 2024. Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier>>. Citado 3 vezes nas páginas 12, 22 e 42.

SILVA, A. A.; SOUTO, C. d. R. *Métodos de Detecção e Análise de Sinais de Vibração Utilizando Fator de Crista e Curtose para Defeitos em Rolamentos*. Campina Grande, PB, Brasil: [s.n.], 2024. Disponível em: <<https://www.ipen.br/biblioteca/cd/conem/2000/NC8694.pdf>>. Citado na página 26.

SPECTRAQUEST. *Simulador de Falhas de Máquinas*. 2024. <<https://spectraquest.com/machinery-fault-simulator/details/mfs/>>. Acessado em 10 de junho de 2024. Citado 3 vezes nas páginas 26, 27 e 28.

STEIN, G. et al. Decision tree classifier for network intrusion detection with ga-based feature selection. *Proceedings of the 43rd annual Southeast regional conference - Volume 2*, 2005. Disponível em: <<https://api.semanticscholar.org/CorpusID:6681228>>. Citado na página 21.

TRACTIAN. *Vibration Sensor*. 2021. Acessado em 20 de junho de 2024. Disponível em: <<https://tractian.com/en/vibration-sensor>>. Citado 3 vezes nas páginas 26, 45 e 49.

TSFRESH DEVELOPMENT TEAM. *tsfresh Documentation*. 2024. Acessado em 19 de junho de 2024. Disponível em: <<https://tsfresh.readthedocs.io/en/latest/>>. Citado na página 39.

XIA, Z. et al. Spectral regression based fault feature extraction for bearing accelerometer sensor signals. *Sensors*, v. 12, n. 10, p. 13694–13719, 2012. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/12/10/13694>>. Citado na página 36.