REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

# DETECÇÃO DE FALHAS MECÂNICAS EM SISTEMAS ROTATIVOS UTILIZANDO INTELIGÊNCIA ARTIFICIAL: UMA ABORDAGEM BASEADA EM ÁRVORES DE DECISÃO

## *DETECTION OF MECHANICAL FAULTS IN ROTATING SYSTEMS USING ARTIFICIAL INTELLIGENCE: A DECISION TREE–BASED APPROACH*

## *DETECCIÓN DE FALLAS MECÁNICAS EN SISTEMAS ROTATIVOS UTILIZANDO INTELIGENCIA ARTIFICIAL: UN ENFOQUE BASADO EN ÁRBOLES DE DECISIÓN*

Vinícius Galhardo[1], Hygor Santiago Lara[2], Jorge Nei Brito[3]

e747372
https://doi.org/10.47820/recima21.v7i4.7372

PUBLICADO: 04/2026

**RESUMO**

Detecção precoce de falhas incipientes em máquinas rotativas é fundamental para a manutenção preditiva e a redução de custos operacionais. Este estudo avalia a aplicação de inteligência artificial, especificamente o classificador baseado em árvore de decisão, na identificação de falhas incipientes de desbalanceamento e desalinhamento a partir de assinaturas de vibração, considerando também a condição normal de operação e assumindo a ausência de outros tipos de falha dentre os analisados. Para isso, foi utilizado um conjunto de dados simulado, com respostas previamente conhecidas, destinado ao treinamento de um modelo de aprendizado supervisionado implementado na linguagem Python. O modelo foi treinado para reconhecer frequências determinísticas e seus harmônicos associados a cada condição de falha. As previsões obtidas foram comparadas com as respostas conhecidas, permitindo avaliar o desempenho do classificador, que apresentou acurácia de 93,67% na classificação das condições analisadas. Os resultados demonstram a eficácia da abordagem e evidenciam o potencial da técnica de árvores de decisão para aplicações em monitoramento de condição e diagnóstico de falhas em máquinas rotativas.

**PALAVRAS-CHAVE:** Inteligência artificial. Árvore de decisão. Diagnóstico de falhas. Máquinas rotativas. Manutenção preditiva.

***ABSTRACT***

*Early detection of incipient faults in rotating machinery is essential for predictive maintenance and reducing operational costs. This study evaluates the use of artificial intelligence, specifically a decision tree classifier, to identify incipient unbalance and misalignment faults based on vibration signatures and normal operating conditions, assuming no other fault types among those analyzed. A simulated dataset with known responses was used to train a supervised learning model implemented in Python. The model was trained to recognize deterministic frequencies and their harmonics associated with each fault condition. Model predictions were compared with the known responses to evaluate classification performance, resulting in an accuracy of 93.67%. The results demonstrate the effectiveness of the approach and highlight the potential of decision tree models for condition monitoring and fault diagnosis in rotating machinery.*

***KEYWORDS:*** *Artificial intelligence. Decision tree. Failure diagnosis. Rotating machinery. Predictive maintenance.*

[1] Bacharel em Engenharia Mecânica, Universidade Federal de São João del Rei (UFSJ), Taubaté, São Paulo, Brasil.
[2] Doutor em Engenharia Mecânica, Universidade de Campinas (Unicamp), São Tiago, Minas Gerais, Brasil.
[3] Doutor, Universidade Federal de São João del Rei, São João del Rei, Minas Gerais, Brasil.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



***RESUMEN***
*La detección temprana de fallas incipientes en maquinaria rotativa es esencial para el mantenimiento predictivo y la reducción de costos operativos. Este estudio evalúa el uso de inteligencia artificial, específicamente un clasificador basado en árboles de decisión, para identificar fallas incipientes de desbalanceo y desalineación a partir de firmas de vibración y condiciones normales de operación, asumiendo la ausencia de otros tipos de fallas entre las analizadas. Se utilizó un conjunto de datos simulado con respuestas previamente conocidas para entrenar un modelo de aprendizaje supervisado implementado en Python. El modelo fue entrenado para reconocer frecuencias determinísticas y sus armónicos asociados a cada condición de falla. Las predicciones del modelo se compararon con las respuestas conocidas para evaluar el desempeño de clasificación, obteniéndose una precisión del 93,67%. Los resultados demuestran la eficacia del enfoque y destacan el potencial de los modelos de árboles de decisión para aplicaciones en el monitoreo de condición y el diagnóstico de fallas en maquinaria rotativa*

***PALABRAS CLAVE****: Inteligencia artificial. Árbol de decisión. Diagnóstico de fallas. Maquinaria rotativa. Mantenimiento predictivo.*

## INTRODUÇÃO

Sistemas rotativos são amplamente utilizados em diversos setores industriais, como manufatura, geração de energia e transporte, sendo essenciais para o funcionamento contínuo das operações. A detecção precoce de possíveis falhas mecânicas em máquinas rotativas é um fator crítico para a manutenção preditiva e a maximização da confiabilidade operacional, (Kardec, 2011). No entanto, falhas nesses equipamentos podem resultar em custos elevados de manutenção, redução da eficiência produtiva e, em casos extremos, danos irreversíveis às máquinas e ao ambiente ao redor (Mozammil, 2020).

Dentre as falhas mais frequentes, o desalinhamento e o desbalanceamento se destacam por sua influência direta no desempenho e na longevidade dos componentes mecânicos. Se não forem detectadas e corrigidas a tempo, essas falhas podem causar aumento nas forças dinâmicas, intensificação do desgaste e até mesmo, em casos mais graves, falhas catastróficas (Bently; Hatch, 2003; Randall, 2011; Zhang, 2021). Métodos tradicionais de monitoramento de condição incluem a análise de vibração, inspeções periódicas e além de aplicação de outras técnicas baseadas na resposta em frequência.

Embora essas abordagens sejam amplamente utilizadas na indústria, elas frequentemente requerem um alto nível de especialização para a interpretação dos dados e podem ser limitadas na identificação de padrões complexos de falha. Além disso, Análises baseadas em experiência humana e sem um tratamento adequado dos dados estão sujeitas a erros de interpretação e podem se tornar inviáveis quando se lida com grandes volumes de dados coletados continuamente por sensores industriais (Randall, 2011; Carvalho, 2020; Zhao, 2022).

Com os avanços na inteligência artificial (IA), novas estratégias vêm sendo exploradas para aprimorar a eficiência, precisão e automação da detecção de falhas.


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



Entre as técnicas de aprendizado de máquina, o método de árvore de decisão se destaca por sua capacidade de classificar diferentes estados operacionais das máquinas de forma interpretável e eficiente, apesar de não ser uma técnica relativamente nova, ela se adequa muito bem para estudo proposto. Esse modelo se baseia em um conjunto de regras derivadas dos dados de entrada para diferenciar entre condições normais e falhas específicas, permitindo a identificação automática do estado do equipamento (Bishop, 2006; Quinlan, 2014; Carvalho, 2020).

No contexto da análise de vibração, o modelo pode utilizar harmônicos característicos das falhas para reconhecer padrões distintos entre desalinhamento, desbalanceamento e condições normais (Randall, 2011; Bently; Hatch, 2003; Zhang, 2021). A principal vantagem da aplicação da IA nesse contexto é a capacidade de processar grandes volumes de dados em tempo real, detectando padrões sutis que poderiam passar despercebidos em uma análise convencional. Além disso, a automação reduz a necessidade de intervenção humana, tornando o processo de manutenção preditiva mais ágil, confiável e acessível.

A aplicação de modelos baseados em aprendizado de máquina pode, portanto, aumentar a eficiência operacional e reduzir custos associados a falhas inesperadas e intervenções corretivas (Carvalho, 2020; Zhao, 2022; Nguyen; Medjaher, 2019). Neste estudo, propõe-se a utilização de um modelo baseado em árvore de decisão para classificar a condição operacional de um sistema rotativo, com base em dados simulados de vibração. O objetivo é avaliar a eficácia do método na detecção de desalinhamento, desbalanceamento e ausência de defeitos, verificando sua viabilidade como ferramenta de monitoramento industrial. O modelo foi treinado e testado com um conjunto de dados gerado artificialmente, contendo frequências de rotação e seus respectivos harmônicos, para determinar o desempenho da técnica na classificação dessas falhas. Os resultados obtidos foram analisados por meio de métricas de desempenho, como acurácia, precisão, recall e F1-score, garantindo uma avaliação quantitativa da eficácia do método proposto.

## 1. REFERENCIAL TEÓRICO

A fundamentação teórica deste estudo concentra-se na integração entre técnicas de processamento de sinais de vibração e métodos de aprendizado de máquina supervisionado, amplamente empregados no diagnóstico de falhas em máquinas rotativas. A análise de vibração é reconhecida como uma das principais ferramentas de monitoramento de condição, pois permite extrair informações relevantes sobre o estado mecânico do sistema a partir de grandezas como aceleração, velocidade e deslocamento. Conforme Randall (2011), a análise no domínio do tempo e, principalmente, no domínio da frequência, por meio de técnicas como a Transformada Rápida de Fourier (FFT), possibilita identificar componentes espectrais associados a falhas específicas. Bently e Hatch (2003) complementam que a correta interpretação dessas assinaturas é essencial para a identificação de anomalias como desbalanceamento e desalinhamento.


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

No processo de aquisição de dados, aspectos como posicionamento de sensores, taxa de amostragem e resolução espectral são determinantes para a qualidade das análises. Em aplicações industriais, os sinais estão sujeitos a ruídos, interferências externas e variações operacionais, o que exige o uso de técnicas de pré-processamento para garantir a confiabilidade dos dados. Segundo Goodfellow, Bengio e Courville (2016), a qualidade e a consistência dos dados são fatores críticos para o desempenho de modelos de aprendizado de máquina. Nesse contexto, procedimentos como filtragem digital, normalização, remoção de outliers e interpolação de dados faltantes são amplamente utilizados para reduzir ruídos e melhorar a representatividade dos sinais (Carvalho, 2020).

A etapa de extração de características *(feature extraction)* é fundamental para a aplicação de algoritmos de classificação. Em sistemas rotativos, características baseadas em frequências determinísticas e seus harmônicos são frequentemente utilizadas para representar o comportamento dinâmico do sistema. De acordo com Randall (2011) e Mozammil, Panigrahi e Israr (2020), parâmetros como amplitude de componentes espectrais, energia em bandas específicas e indicadores estatísticos no domínio do tempo permitem distinguir diferentes condições operacionais. Trabalhos recentes também destacam o uso de atributos derivados de transformadas no domínio do tempo-frequência e estatísticas descritivas como entradas para modelos de aprendizado de máquina (Zhang, 2021; Das; Birant, 2023).

Após a extração das características, os dados são organizados em conjuntos destinados ao treinamento, validação e teste dos modelos preditivos. Essa divisão é essencial para avaliar a capacidade de generalização dos algoritmos e evitar problemas como *overfitting*. Bishop (2006) destaca que a separação adequada dos dados permite ajustar os parâmetros do modelo e validar seu desempenho de forma confiável. Em aplicações de manutenção preditiva, Carvalho (2020) ressaltam que a utilização de dados rotulados, provenientes de condições conhecidas de operação, é fundamental para o desenvolvimento de modelos supervisionados eficazes.

Entre os algoritmos de classificação, as árvores de decisão são amplamente utilizadas devido à sua interpretabilidade e facilidade de implementação. O método consiste na construção de uma estrutura hierárquica baseada em regras de decisão extraídas dos dados, permitindo classificar diferentes estados operacionais. Conforme Breiman (1984), o processo de divisão dos nós é orientado por métricas de impureza, como o índice de Gini, que busca maximizar a separação entre classes. Quinlan (2014) destaca que esse tipo de modelo é particularmente eficiente em problemas onde as relações entre variáveis podem ser expressas por regras lógicas simples.

A aplicação de árvores de decisão no diagnóstico de falhas em máquinas rotativas tem sido amplamente investigada na literatura recente. Estudos indicam que, quando combinadas com um conjunto adequado de características extraídas de sinais de vibração, essas técnicas podem apresentar desempenho competitivo em relação a métodos mais complexos.


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Breiman (2001) argumenta que a qualidade dos dados e das variáveis utilizadas exerce maior influência no desempenho do modelo do que sua complexidade estrutural. Nesse sentido, abordagens interpretáveis são particularmente vantajosas em ambientes industriais, onde a compreensão do processo de decisão é essencial (Rudin, 2019).

Além disso, a implementação computacional desses modelos é facilitada pelo uso de bibliotecas consolidadas na linguagem Python, como o NumPy para manipulação de dados (Harris, 2020) e o Scikit-learn para construção e avaliação de modelos de aprendizado de máquina (Pedregosa, 2011). Essas ferramentas permitem a aplicação eficiente de técnicas de classificação, validação cruzada e análise de desempenho por meio de métricas como acurácia, precisão, recall e F1-score, amplamente utilizadas em estudos de diagnóstico de falhas (Carvalho, 2020; Zhao, 2019).

Por fim, a literatura recente reforça que a combinação de processamento de sinais, extração de características relevantes e modelos de aprendizado de máquina constitui uma abordagem robusta para o diagnóstico de falhas em máquinas rotativas. Trabalhos como os de Zhang (2021), Zhao (2019) e Neupane (2025) demonstram que essas técnicas permitem identificar padrões complexos em grandes volumes de dados, contribuindo para o desenvolvimento de sistemas de manutenção preditiva mais eficientes.

## 2. MÉTODOS

Este estudo propõe a aplicação de um modelo de inteligência artificial baseado no método de árvore de decisão para a detecção de falhas mecânicas em máquinas rotativas. Para isso, foi desenvolvido um algoritmo gerador de dados sintéticos baseado em conceitos consolidados na área de vibrações mecânicas (Randall, 2011). Esse algoritmo simula um conjunto de valores representativos de um ensaio de vibração, fornecendo informações sobre a frequência de rotação do sistema e seus respectivos harmônicos (primeiro, segundo terceiro e quarto), além da condição associada do elemento rotativo (desalinhamento, desbalanceamento ou normal). A metodologia adotada pode ser dividida nas seguintes etapas principais:

### 2.1. Geração dos dados sintéticos

A geração dos dados sintéticos seguiu um modelo determinístico no qual os sinais de vibração são representados por seus componentes harmônicos. Considerando um sistema rotativo operando a uma determinada frequência de rotação (fr), os quatro primeiros harmônicos foram calculados como múltiplos dessa frequência. Condição Normal (Sem Falha): Quando o sistema opera normalmente, os harmônicos apresentam amplitudes dentro de um intervalo de operação típico, sem oscilações excessivas. A distribuição da energia vibracional ocorre de forma equilibrada entre os componentes espectrais, sem picos anômalos.


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

O que seria a condição normal de operação da máquina. Desalinhamento: O desalinhamento ocorre quando os eixos de acoplamento de um sistema rotativo não estão corretamente alinhados, gerando forças radiais e axiais adicionais. Essa falha se manifesta por um aumento significativo na amplitude dos harmônicos 1 e 2, na Figura 1 temos uma representação gráfica de como deve ser o espectro nessa condição, especialmente em eixos flexíveis. Em casos mais graves, pode haver também excitação de frequências superiores, criando padrões de vibração não lineares e aumentando o consumo energético do sistema.

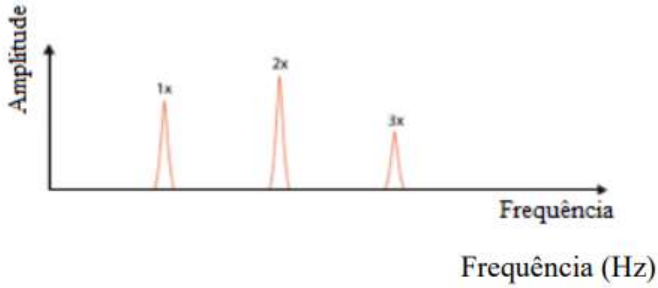


Figura 1-Representação do espectro de vibração de um eixo desalinhado. *(Dewesoft, 2023)*

Desbalanceamento: O desbalanceamento ocorre quando há uma distribuição desigual de massa ao redor do eixo de rotação, gerando forças centrífugas que resultam em vibração excessiva. Essa condição é caracterizada por uma amplificação anômala do primeiro harmônico, como pode ser observado na Figura 2, pois a força gerada é proporcional ao quadrado da velocidade angular. Em casos mais severos, o terceiro harmônico também pode ser afetado, especialmente quando há ressonância estrutural ou efeitos giroscópicos.




*Figura 2- Representação do espectro de vibração de eixo desbalanceado.(Dewesoft, 2023)*


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A qualidade e a representatividade dos dados desempenham um papel fundamental no sucesso da aplicação de quaisquer métodos de *Machine Learning,* especialmente no contexto da detecção de falhas mecânicas em máquinas rotativas (Goodfellow; Bengio; Courville, 2016; Carvalho, 2020; Zhang, 2021).

Em sistemas baseados em aprendizado de máquina, os dados são a base para o treinamento, validação e teste dos modelos preditivos. Dessa forma, qualquer limitação na coleta, tratamento ou estruturação dos dados pode comprometer diretamente a eficácia e a generalização dos algoritmos utilizados (Bishop, 2006; Goodfellow; Bengio; Courville, 2016; Carvalho, 2020). Os dados são, em geral, obtidos por meio de sensores de vibração, corrente elétrica, entre outros dispositivos de instrumentação. A escolha adequada dos sensores, sua calibração e posicionamento são determinantes para garantir a fidelidade dos sinais coletados. Erros nessa fase podem resultar em dados com baixa representatividade do fenômeno físico monitorado, comprometendo a acurácia dos modelos de detecção de falhas.

Além da correta aquisição, a definição dos parâmetros monitorados deve ser tecnicamente fundamentada. Em máquinas rotativas, falhas como desalinhamento e desbalanceamento geram assinaturas específicas nos domínios do tempo e da frequência (Randall, 2011; Bently; Hatch, 2003; Zhang, 2021). Parâmetros como Aceleração, velocidade de vibração e componentes harmônicas do sinal são particularmente sensíveis a essas anomalias. A seleção inadequada das variáveis de entrada pode levar à perda de informações relevantes ou à introdução de ruídos desnecessários nos dados.

A qualidade dos dados também é influenciada por fatores externos, falhas nos sensores, variações ambientais e problemas de sincronização temporal. Dados corrompidos ou inconsistentes podem introduzir viés nos modelos, além de aumentar o risco de *overfitting*. Para mitigar esses efeitos, torna-se essencial a aplicação de técnicas de pré-processamento, incluindo filtragem, normalização, remoção de outliers e interpolação de lacunas (Randall, 2011; Bently; Hatch, 2003; Carvalho, 2020; Goodfellow; Bengio; Courville, 2016). Neste trabalho foram utilizados dados sintéticos para validação do modelo, todavia como exposto anteriormente, para aplicação com dados reais deve-se atentar a esses pontos que podem influenciar nos resultados, e consequentemente na validação do modelo final.

### 2.2. Desenvolvimento do modelo de árvore de decisão

O modelo de árvore de decisão foi implementado utilizando a biblioteca *scikit-learn* em *Python*. Esse método foi escolhido devido à sua alta interpretabilidade e capacidade de realizar classificações baseadas em regras de decisão extraídas dos dados. O algoritmo de aprendizado seguiu a abordagem *CART (Classification and Regression Tree)*. Para garantir a robustez do modelo e evitar *overfitting*, foi aplicada a técnica de validação cruzada do tipo *kfold*, com k = 5.


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



O conjunto de dados foi inicialmente dividido em 70% para treinamento e 30% para teste. A validação cruzada foi empregada exclusivamente no subconjunto de treinamento, permitindo que o conjunto de teste fosse utilizado apenas para avaliação final, assegurando imparcialidade nos resultados. Na prática, o conjunto de treinamento foi dividido em cinco partes (*folds*) aproximadamente iguais. Em cada uma das cinco iterações, quatro partes foram utilizadas para o treinamento do modelo e uma parte distinta foi usada para validação. Esse processo foi repetido até que todas as partições fossem utilizadas como conjunto de validação exatamente uma vez. A acurácia média entre os cinco ciclos serviu como critério para avaliação do desempenho de cada configuração de modelo. Combinado à validação cruzada, utilizou-se a técnica de *Grid Search* por meio da função *GridSearchCV* da biblioteca *Scikit-learn*, com o objetivo de otimizar os parâmetros do classificador de árvore de decisão. Essa técnica realiza uma busca exaustiva por todas as combinações possíveis de parâmetros dentro de uma grade definida pelo pesquisador, avaliando cada uma delas com validação cruzada. No presente trabalho, foram avaliados os seguintes parâmetros:

- Critério de Divisão.
  Valores testados: *'gini'* e *'entropy'*
- Profundidade máxima da Árvore.
  Valores testados: 3, 4, 5, 6, e *None* (sem limite)
- Mínimo de amostras para dividir um nó.
  Valores testados: 2, 5, e 10
- Mínimo de amostras em cada folha.
  Valores testados: 1, 2, e 4

A combinação dessas opções resultou em um total de 90 configurações únicas de modelo (2×5×3×3), cada uma avaliada com validação cruzada 5-*fold*. O modelo com a melhor média de acurácia entre os *folds* foi selecionado como o modelo final, posteriormente validado no conjunto de teste. Essa abordagem sistemática de otimização permitiu a seleção de um modelo mais generalizável, alinhada às boas práticas de aprendizado de máquina para diagnóstico de falhas mecânicas, conforme demonstrado em trabalhos anteriores na área (Oguzhan Das, 2023; Dhiraj Neupane, 2025; Mozammil, 2020).


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



| Critério de divisão | índice de Gini |
|---|---|
| Profundidade máxima da árvore | 4 |
| Semente aleatória para reprodutibilidade | 42 |
| Estratégia de divisão | Melhor |
| Mínimo de amostras para dividir um nó | 2 |
| Mínimo de amostras em cada folha | 1 |
| Fração mínima de peso em cada folha | 0.0 |
| Número máximo de recursos considerados em cada divisão | Nenhum |
| Número máximo de nós folha | Nenhum |
| Redução mínima na impureza para uma divisão | 0.0 |
| Peso das classes | Nenhum |
| Parâmetro de poda | 0.0 |

Critério de divisão: O índice de Gini é um critério adequado para medir a impureza dos nós e tomar decisões de divisão durante a construção da árvore de decisão (Breiman, 1984; Hastie; Tibshirani; Friedman, 2009; Carvalho, 2020).

Profundidade máxima da árvore: Limitar a profundidade, ajuda a evitar *overfitting*. Um valor 4 é um bom ponto de partida, mas pode ser ajustado.

Semente aleatória para reprodutibilidade: A adoção de uma semente aleatória permite que os resultados obtidos possam ser replicados exatamente nas mesmas condições por outros pesquisadores ou em execuções futuras do experimento. Neste trabalho, foi utilizada a semente 42.

Estratégia de divisão: A estratégia “melhor” seleciona a melhor divisão possível com base no critério escolhido (gini).

Mínimo de amostras para dividir um nó: Essa é a configuração padrão e permite que qualquer nó com pelo menos 2 amostras possa ser dividido.

ISSN: 2675-6218 - RECIMA21
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



Mínimo de amostras em cada folha: Com 1, cada folha pode conter apenas uma amostra. Esse valor pode ser adequado, mas se houver muito ruído nos dados, pode ser interessante aumentá-lo.

Fração mínima de peso em cada folha: Como não há pesos específicos atribuídos às amostras, este valor faz sentido.

Número máximo de recursos considerados em cada divisão: Quando Nenhum, todos os recursos disponíveis são considerados em cada divisão. Isso é coerente, pois queremos utilizar todas as variáveis disponíveis.

Número máximo de nós folha: Sem limite para o número de nós folha, permitindo que a árvore cresça dentro das restrições de profundidade.

Redução mínima na impureza para uma divisão: Significa que qualquer redução na impureza pode justificar uma divisão. Faz sentido, mas pode ser ajustado se quisermos podar a árvore automaticamente.

Peso das classes: Como as classes de defeito parecem estar balanceadas, “Nenhum” é aceitável. Se houver desbalanceamento nas classes, poderíamos definir “Balanceado” para ajustar os pesos automaticamente.

Parâmetro de poda: Sem poda automática. Pode ser ajustado para evitar *overfitting* se a árvore estiver muito complexa. Os parâmetros podem ser modificados, dependendo dos dados que se deseja estudar ou das classificações, a fim de obter melhores resultados.

### 2.3. Avaliação do modelo

Para avaliar a performance da árvore de decisão, foram calculadas métricas de desempenho, incluindo: Acurácia: Proporção de classificações corretas sobre o total de amostras. Precisão (*Precision*): Taxa de classificações corretas para cada classe positiva. É calculado através da fórmula (1).

$$Precisão = \frac{Verdadeiro\ Positivos}{Verdadeiros\ Positivo + Falsos\ Positivos} \quad (1)$$

Revocação (*Recall*): Proporção de amostras corretamente classificadas dentro de cada classe (2).

$$Revocação = \frac{Verdadeiro\ Positivos}{Verdadeiros\ Positivo + Falsos\ Negativos} \quad (2)$$

*F1-score*: Média harmônica entre precisão e *recall*. (3)

ISSN: 2675-6218 - RECIMA21

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



$$F1score = 2\left(\frac{Precisão.Revocação}{Precisão+Revocação}\right) \quad (3)$$

As métricas foram computadas utilizando a matriz de confusão gerada pelo modelo, permitindo uma análise detalhada da eficácia na classificação das condições da máquina.

### 2.4. Comparação com os dados de referência

Os resultados obtidos pelo modelo foram comparados com os resultados sintéticos gerados pelo algoritmo gerador de dados, a fim de verificar a precisão na identificação das falhas. Essa comparação permitiu avaliar a capacidade da árvore de decisão em reconhecer padrões nos sinais de vibração e distinguir corretamente entre desalinhamento, desbalanceamento e condição normal. Isso contribui para validação do modelo uma vez que dados reais forem inseridos, há maior probabilidade de acertos.

**Figura 3**. Forma de apresentação dos Dados

```
Entrada: {'freq_rot': 1.573, 'harm_1': 0.4671, 'harm_2': 0.4848, 'harm_3': 0.4608, 'harm_4': 0.5553} -> Previsto: desbalanceamento | Real: desbalanceamento
Entrada: {'freq_rotacao': 0.9260, 'harm_1': 0.6681, 'harm_2': 0.6741, 'harm_3': 0.8055, 'harm_4': 0.6563} -> Previsto: desalinhamento | Real: desalinhamento
Entrada: {'freq_rotacao': 1.0853, 'harm_1': 0.5194, 'harm_2': 0.5401, 'harm_3': 0.4872, 'harm_4': 0.4934} -> Previsto: normal | Real: normal
Entrada: {'freq_rotacao': 0.8157, 'harm_1': 0.4202, 'harm_2': 0.4896, 'harm_3': 0.4122, 'harm_4': 0.5455} -> Previsto: normal | Real: normal
Entrada: {'freq_rotacao': 0.9583, 'harm_1': 0.6944, 'harm_2': 0.4669, 'harm_3': 0.5785, 'harm_4': 0.5164} -> Previsto: normal | Real: desalinhamento
```

## 3. RESULTADOS E DISCUSSÃO

A avaliação do modelo de inteligência artificial baseado em árvore de decisão demonstrou um desempenho sólido na classificação das condições operacionais de sistemas rotativos. A matriz de confusão apresentada na Tabela 1 resume os resultados obtidos. A avaliação do desempenho do modelo otimizado de árvore de decisão evidenciou resultados consistentes e promissores na tarefa de classificação automática de falhas mecânicas em máquinas rotativas. A classe desalinhamento apresentou uma precisão de 96% e recall de 98%, evidenciando a elevada capacidade do modelo em identificar corretamente essa condição com baixa taxa de falsos positivos. O *F1-score* de 97% complementa esse resultado, reforçando a robustez da abordagem adotada.

ISSN: 2675-6218 - RECIMA21
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 1- Matriz de Confusão

| Real \ Previsto | Desal. | Desbal. | Normal |
|---|---|---|---|
| Desal. | 98 | 0 | 2 |
| Desbal. | 0 | 97 | 3 |
| Normal | 4 | 4 | 92 |

A classe desbalanceamento obteve desempenho semelhante, com precisão de 96% e *recall* de 97%, o que indica uma alta taxa de detecção correta das ocorrências reais, com poucos equívocos de classificação. O *F1- score*, também de 97%, reforça o equilíbrio entre sensibilidade e especificidade. A classe normal, representando o funcionamento livre de falhas, apresentou precisão de 95% e recall de 92%. Esse resultado sugere que aproximadamente 8% das amostras normais foram erroneamente classificadas como defeituosas, o que pode ser atribuído à semelhança espectral entre sinais normais e sinais de falhas incipientes, um desafio comum em sistemas de diagnóstico baseados em vibração (Randall, 2011; Mozammil, 2020).

De forma geral, o modelo alcançou uma acurácia global de 93,67% e um *F1-score* médio de 96%, confirmando a eficácia da estratégia baseada em árvore de decisão com validação cruzada e otimização por *Grid Search* (Pedregosa, 2011). Diversos estudos na literatura de aprendizado de máquina indicam que modelos relativamente simples e interpretáveis podem apresentar desempenho semelhante ao de algoritmos mais complexos, especialmente quando os dados utilizados são representativos e as características extraídas capturam adequadamente o comportamento do sistema analisado.

De acordo com Breiman (2001), a eficácia de um modelo preditivo depende fortemente da qualidade das variáveis utilizadas e da relação destas com o fenômeno estudado, não sendo necessariamente proporcional à complexidade do algoritmo empregado. Nesse contexto, modelos interpretáveis, como árvores de decisão, podem alcançar níveis de desempenho comparáveis aos de métodos mais sofisticados quando as informações relevantes estão adequadamente representadas nos dados. Além disso, Rudin (2019) argumenta que, em muitos problemas práticos, especialmente em aplicações de engenharia e sistemas industriais, modelos interpretáveis podem atingir resultados equivalentes aos de modelos de “caixa-preta”, oferecendo ainda a vantagem


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

adicional de transparência e facilidade de interpretação. No contexto da manutenção preditiva e do diagnóstico de falhas em máquinas rotativas, revisões recentes também apontam que algoritmos de classificação relativamente simples podem apresentar desempenho competitivo, desde que os dados sejam adequadamente pré-processados e as características selecionadas sejam relevantes para o fenômeno monitorado (Carvalho, 2020). Na Tabela 2 temos o relatório de classificação.

**Tabela 2. Relatório de classificação**

| Classe | Precisão | Recall | F1-Score |
|---|---|---|---|
| Desalinhamento | 96% | 98% | 97% |
| Desbalanceamento | 96% | 97% | 97% |
| Normal | 95% | 92% | 93% |

Além disso, o uso de bibliotecas consolidadas em ciência de dados e aprendizado de máquina, como *NumPy* (Harris, 2020), *Pandas* (McKinney, 2010) e *Scikitlearn* (Pedregosa, 2011), foi essencial para a implementação eficiente da solução, corroborando a importância dessas ferramentas no desenvolvimento de aplicações industriais de manutenção preditiva. Uma vez que são ferramentas que permitem um desenvolvimento mais rápido, são menos propensas a erros, são mais robustas e possuem grande documentação. O desenvolvimento do código sem essas ferramentas poderia dar mais flexibilidade e controle ao algoritmo, mas poderia trazer alguns riscos, como o de erros no resultado e de desempenho. Em consonância com estudos recentes que destacam o papel da inteligência artificial no diagnóstico preditivo de falhas (Oguzhan Das, 2023; Dhiraj Neupane, 2025), os resultados obtidos reforçam o potencial da combinação entre técnicas supervisionadas de classificação e análise espectral para o monitoramento automatizado de condições operacionais em sistemas rotativos.

## 4. CONSIDERAÇÕES

Os resultados obtidos indicam que a técnica de árvore de decisão é eficiente para a detecção automática de falhas mecânicas, alcançando alta acurácia e boas métricas de classificação para as três condições avaliadas (desalinhamento, desbalanceamento e ausência de falhas).


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A identificação de falhas críticas com precisão e rapidez possibilita a implementação de estratégias eficazes de manutenção preditiva, reduzindo o risco de falhas inesperadas e otimizando a confiabilidade dos equipamentos.

No entanto, alguns desafios ainda permanecem. A análise da matriz de confusão sugere que o modelo pode estar propenso a superestimar falhas em determinadas situações, o que pode levar a diagnósticos conservadores. Além disso, destaca-se como limitação do estudo o uso de dados sintéticos para o treinamento e validação do modelo. Embora esse tipo de abordagem permita controle das variáveis e conhecimento prévio das respostas, os dados simulados podem não representar integralmente a complexidade dos sinais reais, que estão sujeitos a ruídos, variações operacionais, interferências ambientais e não linearidades do sistema. Dessa forma, o desempenho observado pode ser superior ao que seria obtido em condições reais, limitando a capacidade de generalização do modelo.

Melhorias podem ser realizadas por meio de ajustes nos parâmetros da árvore de decisão, balanceamento do conjunto de dados e inclusão de novas variáveis de entrada para refinar a detecção. Como trabalhos futuros, sugere-se a comparação do desempenho da árvore de decisão com outras abordagens de aprendizado de máquina, como florestas aleatórias (Random Forest), redes neurais e métodos híbridos, que podem oferecer maior robustez na classificação e minimizar possíveis vieses do modelo. Além disso, a validação do modelo com dados reais provenientes de sensores industriais é fundamental para avaliar sua aplicabilidade prática no monitoramento de sistemas rotativos e confirmar sua capacidade de generalização em ambientes operacionais reais.

## REFERÊNCIAS

BENTLY, Donald E.; HATCH, Charles T. **Fundamentals of rotating machinery diagnostics**. 2. ed. Minden: Bently Pressurized Bearing Press, 2003.

BISHOP, Christopher M. **Pattern recognition and machine learning**. New York: Springer, 2006.

BREIMAN, Leo. Statistical modeling: the two cultures. **Statistical Science**, v. 16, n. 3, p. 199–231, 2001.

BREIMAN, Leo; FRIEDMAN, Jerome H.; OLSHEN, Richard A.; STONE, Charles J. **Classification and regression trees**. New York: Chapman & Hall/CRC, 1984.

CARVALHO, Tiago P. *et al.* A systematic literature review of machine learning methods applied to predictive maintenance. **Computers & Industrial Engineering**, v. 137, p. 106024, 2020. DOI: https://doi.org/10.1016/j.cie.2019.106024

DAS, Oguzhan; DAS, Duygu Bagci; BIRANT, Derya. Machine learning for fault analysis in rotating machinery: a comprehensive review. **Heliyon**, v. 9, n. 6, e17584, 2023. DOI: https://doi.org/10.1016/j.heliyon.2023.e17584


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

DEWESOFT. **Gráfico de vibração do eixo** [imagem]. 2023. Disponível em: https://dewesoft.com/pt/blog/como-interpretar-dados-de-monitoramento-de-condicao Acesso em: 17 mar. 2026.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning. Cambridge: MIT Press, 2016.

HAO, X.; ZHANG, Guigang; SHANG, M. Deep learning. **International Journal of Semantic Computing**, v. 10, n. 3, p. 417–439, 2016. DOI: https://doi.org/10.1142/S1793351X16500045

HARRIS, Charles R. *et al.* Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, 2020. DOI: https://doi.org/10.1038/s41586-020-2649-2

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning**: data mining, inference, and prediction. 2. ed. New York: Springer, 2009.

KARDEC, Allan; NASCIF, Júlio. **Manutenção**: função estratégica. 3. ed. Rio de Janeiro: Qualitymark, 2009.

KOUTNÍK, Jan; GREFF, Klaus; FAUSTINO, G.; SCHMIDHUBER, Jürgen. A clockwork RNN. *In:* INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 31., 2014, Beijing. **Proceedings** […] PMLR, 2014. p. 1863–1871.

MCKINNEY, Wes. Data structures for statistical computing in Python. *In:* PYTHON IN SCIENCE CONFERENCE, 9., 2010, Austin. **Proceedings** [...] SciPy, 2010. p. 51–56.

MOZAMMIL, S.; PANIGRAHI, S.; ISRAR, M. Vibration detection and analysis for fault prediction of electric rotating machines. **Dogo Rangsang Research Journal**, v. 10, n. 12, p. 442–446, 2020.

NEUPANE, Dhiraj; BOUADJENEK, Mohamed Reda; DAZELEY, Richard; ARYAL, Sunil. Data-driven machinery fault diagnosis: a comprehensive review. **Neurocomputing**, v. 627, p. 129588, 2025. DOI: https://doi.org/10.1016/j.neucom.2025.129588

NGUYEN, Kien; MEDJAHER, Kamal. A new dynamic predictive maintenance framework using machine learning for failure prognostics. **Reliability Engineering & System Safety**, v. 188, p. 251–262, 2019. DOI: https://doi.org/10.1016/j.ress.2019.03.018

PEDREGOSA, F. *et al.* Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

QUINLAN, J. Ross. **C4.5**: programs for machine learning. Burlington: Morgan Kaufmann, 2014.

RANDALL, Robert B. **Vibration-based condition monitoring**: industrial, aerospace and automotive applications. Chichester: John Wiley & Sons, 2011.

RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, v. 1, p. 206–215, 2019. DOI: https://doi.org/10.1038/s42256-019-0048-x

ZHANG, Wei et al. A review on machinery fault diagnosis using machine learning for rotating equipment. **Mechanical Systems and Signal Processing**, v. 148, p. 107247, 2021. DOI: https://doi.org/10.1016/j.ymssp.2020.107247


Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218



ZHAO, Rui et al. Deep learning and its applications to machine health monitoring: a survey. **Mechanical Systems and Signal Processing**, v. 115, p. 213–237, 2019. DOI: https://doi.org/10.1016/j.ymssp.2018.05.050

**ISSN: 2675-6218 - RECIMA21**

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.