

Aplicação de Algoritmos de Aprendizado de Máquina na Manutenção Preditiva: Uma Análise das Técnicas de IA para Predição de Falhas[★]

Victor F. Magalhães* Arthur F. Foureaux** Carlos A. Silva***
Daniel B. F. Conrado**** Bruno N. Gomes† Diego O. Miranda‡

* Área de Informática, Instituto Federal de Minas Gerais, Sabará, MG,
(e-mail: victorfernandesmagalhaes@gmail.com)

** Área de Informática, Instituto Federal de Minas Gerais, Sabará,
MG, (e-mail: arthurffoureaux@gmail.com)

*** Área de Informática, Instituto Federal de Minas Gerais, Sabará,
MG, (e-mail: carlos.silva@ifmg.edu.br)

**** Área de Informática, Instituto Federal de Minas Gerais, Sabará,
MG, (e-mail: daniel.conrado@ifmg.edu.br)

† Área de Informática, Instituto Federal de Minas Gerais, Sabará, MG,
(e-mail: bruno.nonato@ifmg.edu.br)

‡ Área de Controle e Processos Industriais, Instituto Federal de Minas
Gerais, Sabará, MG, (e-mail: diego.miranda@ifmg.edu.br)

1

Abstract: With the significant increase in the volume of operational data collected, machine learning algorithms have stood out in identifying the need for predictive maintenance in equipment, especially in scenarios where the partial or total inoperability of a piece of equipment is extremely costly. This study analyzed the feasibility of three algorithms, training them and comparing their performance using synthetic data that simulates real-world scenarios. The results point to strong potential for the use of the Random Forest algorithm, which achieved an accuracy of 0,99 in the tests conducted, with a training time of 0,22 seconds on the test set. Furthermore, this algorithm outperformed the others in all evaluation metrics considered, establishing itself as the most effective approach among those analyzed. Thus, the results obtained can aid in the selection of more efficient algorithms for predictive maintenance in industrial environments with limited data.

Resumo: Com o aumento expressivo do volume de dados operacionais coletados, algoritmos de aprendizado de máquina têm se destacado na identificação da necessidade de manutenção preditiva em equipamentos, principalmente em cenários nos quais a inoperância parcial ou total de algum equipamento é extremamente custosa. Este estudo analisou a viabilidade de três algoritmos, treinando-os e comparando seus desempenhos com dados fictícios que simulam cenários reais. Os resultados apontam um forte potencial para o uso da Floresta Randômica, que alcançou acurácia de 0,99 nos testes realizados, com tempo de treinamento de 0,22 segundo no conjunto de testes. Além disso, esse algoritmo superou os demais em todas as métricas de avaliação consideradas, consolidando-se como a abordagem mais eficaz entre as analisadas. Dessa forma, os resultados obtidos podem auxiliar na escolha de algoritmos mais eficientes para manutenção preditiva em ambientes industriais com dados limitados.

Keywords: Predictive Maintenance; Machine Learning; Random Forest; Failure Prediction; Industry 4.0.

Palavras-chaves: Manutenção Preditiva; Aprendizado de Máquina; Floresta Randômica; Previsão de Falhas; Indústria 4.0.

¹ Este trabalho foi publicado no XVII Simpósio Brasileiro de Automação Inteligente (SBAI), em São João Del Rei - MG, ago. 2025.

1. INTRODUÇÃO

A transformação digital da indústria tem impulsionado a coleta e o armazenamento massivo de dados operacionais, o que abre novas possibilidades para a análise preditiva da integridade de equipamentos. No contexto da manutenção industrial, esses dados contêm informações relevantes sobre o comportamento dos ativos e os fatores associados a diferentes modos de falha. Quando devidamente processados, podem revelar padrões ocultos que permitem prever falhas - que neste estudo se caracterizam como situações de inoperância parcial ou total do equipamento - com antecedência, contribuindo para o aumento da confiabilidade dos sistemas e a extensão da vida útil dos equipamentos, como afirma Carvalho et al. (2019).

Além disso, abordagens analíticas orientadas por dados possibilitam a extração de *insights* interpretáveis, apoiando decisões estratégicas relacionadas à manutenção. Entre os principais benefícios observados estão a redução de custos operacionais, a diminuição de falhas inesperadas, o aumento da segurança dos operadores e a melhora da eficiência geral do processo produtivo. Tais benefícios estão diretamente ligados à adoção de estratégias avançadas, como a manutenção preditiva, que se baseia na previsão de falhas para programar intervenções de forma otimizada. Conforme discutido por Carvalho et al. (2019) e reforçado por Susto et al. (2015), essas estratégias podem ser consideravelmente mais eficazes quando complementadas por técnicas de aprendizado de máquina (*Machine Learning - ML*), que otimizam a identificação de padrões e o processo decisório com base em dados históricos.

Estudos recentes, como o de Alves et al. (2023), apontam que a implementação de estratégias de manutenção preditiva pode reduzir os custos de manutenção em até 30%, aumentar a disponibilidade de equipamentos em até 20% e diminuir a ocorrência de falhas em até 25%. Esses dados evidenciam o potencial significativo dessa abordagem para transformar a gestão da manutenção industrial.

Com o avanço das técnicas de aprendizado de máquina, novas ferramentas vêm sendo utilizadas para aprimorar a capacidade de previsão de falhas. Neste trabalho, investigamos a aplicação de algoritmos de aprendizado de máquina para prever falhas em diferentes equipamentos, com base em um conjunto de dados sintético que simula o comportamento de um ambiente industrial real. O foco está na comparação de três algoritmos amplamente utilizados na literatura: Máquinas de Vetores de Suporte (*Support Vector Machines - SVM*), Redes Neurais Artificiais (*Artificial Neural Networks - ANN*) e Florestas Randômicas (*Random Forests - RF*).

A principal contribuição deste estudo está na análise comparativa dessas técnicas, considerando métricas de desempenho como acurácia, precisão, sensibilidade e *F1-score*. Além disso, destaca-se o uso de técnicas modernas de pré-processamento, como a superamostragem com SMOTE,

* Este trabalho foi financiado pelo Polo de Inovação IFMG, um centro de Pesquisa, Desenvolvimento e Inovação (PD&I) credenciado pela EMBRAPPII (Associação Brasileira de Pesquisa e Inovação Industrial).

visando lidar com o desbalanceamento presente nos dados. Diferentemente de outros trabalhos focados em apenas um tipo de máquina ou falha específica, este estudo aplica os algoritmos a uma base que simula diferentes tipos de compartimentos industriais (como motores, redutores e transmissões), o que permite uma avaliação mais generalista das técnicas. Assim, embora não haja foco em um único tipo de falha, a abordagem visa explorar a adaptabilidade dos modelos a diferentes condições e perfis operacionais. Por fim, reforça-se que a literatura aponta benefícios concretos na integração entre manutenção preditiva e aprendizado de máquina, destacando que essa combinação favorece a antecipação de falhas e a tomada de decisão baseada em dados.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta uma revisão de trabalhos relacionados ao uso de aprendizado de máquina na manutenção preditiva. Na Seção 3, são descritos os algoritmos selecionados, o processo de preparação dos dados e a configuração dos modelos. A Seção 4 apresenta e discute os resultados obtidos. Por fim, a Seção 5 apresenta as conclusões e as principais contribuições do trabalho.

2. TRABALHOS RELACIONADOS

A manutenção preditiva é essencial para reduzir custos e evitar falhas inesperadas em equipamentos industriais, conforme aponta Relkar (2021). Além disso, para ser mais eficiente que abordagens corretivas e planejadas, essa pode ser aprimorada com aprendizado de máquina para prever falhas em equipamentos, como evidenciado por Susto et al. (2012, 2015).

Com base nisso, diversos estudos na literatura têm se dedicado à realização de análises preditivas de modos e efeitos de falha em diferentes tipos de equipamentos, valendo-se de distintas metodologias e métricas de avaliação para mensurar o desempenho dos modelos utilizados. O trabalho de Álvaro Huertas García et al. (2024), aplica múltiplos algoritmos de aprendizado de máquina no contexto de manutenção preditiva e cibersegurança em pequenas e médias empresas, incluindo Floresta Randômica, Redes Neurais e SVM. No entanto, esse estudo não realiza uma comparação sistemática e aprofundada entre os algoritmos utilizados. Por sua vez, Teramoto et al. (2024) conduz análises preditivas baseadas na composição de óleos lubrificantes, mas sem empregar técnicas de aprendizado de máquina. Ressalta-se também o trabalho de Junior et al. (2023), que examina diversos estudos aplicados à indústria, com a intenção de identificar um algoritmo de destaque para manutenção preditiva — objetivo que não foi alcançado. Já o estudo de Veloso et al. (2022) tem como foco a criação de um *dataset benchmark* para manutenção preditiva em sistemas de transporte público, mas não explora a aplicação de algoritmos preditivos. Complementarmente, Maheshwari et al. (2024) e Vago et al. (2024) propõem o uso de aprendizado de máquina em contextos industriais semelhantes ao deste trabalho, mas adotam modelos baseados em redes neurais recorrentes, notadamente o *Long Short-Term Memory* (LSTM), que operam com dados sequenciais ou séries temporais. O presente estudo diferencia-se por não

utilizar dados temporais, concentrando-se em atributos estáticos multivariados. As principais distinções entre os trabalhos analisados estão apresentadas na Tabela 1.

Tabela 1. Comparação entre este artigo e estudos relacionados.

Estudo	Objetivo principal	ML	Compara	Contexto / dados	Séries
Este artigo	Prever falhas com ML e comparar algoritmos	Sim	Sim	Compartimentos industriais diversos	Não
García et al. (2024)	Manut. preditiva e cibersegurança em PMEs	Sim	Não	Sensores em tempo real	Sim
Teramoto et al. (2024)	Análise preditiva com óleos lubrificantes	Não	Não	Dados físico-químicos de lubrificantes	Não
Júnior et al. (2023)	Análise de estudos sobre manutenção com ML	Não	Não	Grandezas elétricas industriais	Não
Veloso et al. (2022)	Criação de <i>dataset</i> para manutenção preditiva	Não	Não	Dados de sensores em trens urbanos	Não
Maheshwari et al. (2024)	Comparar técnicas de ML para manutenção preditiva	Sim	Sim	Dados de falhas e desempenho de máquinas	Sim
Vago et al. (2024)	Avaliar janelas de leitura/-previsão em modelos de falha	Sim	Sim	Três máquinas industriais	Sim

À luz do que foi exposto, o presente trabalho visa aplicar algoritmos de aprendizado de máquina na gestão da manutenção de equipamentos, utilizando características presentes na base de dados e focando em prever possíveis falhas. Após revisar a literatura, constatou-se que embora existam trabalhos que realizam comparações entre algoritmos de aprendizado de máquina e que fazem previsões

considerando características físico-químicas de óleos, este estudo se destaca ao treinar e comparar tais algoritmos de forma mais aprofundada, empregando uma base de dados diversificada, com múltiplas amostras de diferentes tipos de equipamentos e variáveis, com o propósito de identificar, de forma precoce, a ocorrência potencial de uma falha em novas amostras.

3. METODOLOGIA

A metodologia deste trabalho é estruturada com o objetivo de criar uma base de dados utilizável para responder à hipótese de que a Floresta Randômica apresenta desempenho superior na predição de falhas mecânicas em comparação com os demais modelos analisados, devido à sua robustez frente a ruídos e desbalanceamento nos dados. Para isso, esta seção é dividida em quatro pontos: base de dados, pré-processamento e tratamento dos dados, escolha dos modelos de aprendizado de máquina e treinamento dos modelos, conforme descrito a seguir.

3.1 Base de dados

A base de dados foi criada a partir de registros reais, coletados por uma empresa especializada na prestação de serviços de análises de efeitos e modos de falhas em diferentes equipamentos a partir da análise de óleos lubrificantes, preservando sua estrutura e distribuição estatística, mas com dados sintéticos devido a restrições de confidencialidade. O conjunto original possui 100 colunas de *features* e 51871 amostras de diferentes equipamentos.

3.2 Pré-processamento e tratamento dos dados

A base de dados criada e utilizada para este trabalho segue o processo ilustrado na Figura 1, que apresenta o fluxo de geração dos dados sintéticos, desde a amostragem dos compartimentos até as etapas de tratamento e balanceamento. Foi feita a coleta das 50 primeiras amostras de cada tipo de compartimento de óleo presente nos equipamentos (como motor, transmissão, redutor, entre outros) presente na base original, preservando características como estrutura, tipos e faixas de valores. Nesse sentido, *outliers* numéricos presentes na base original foram removidos utilizando o método estatístico de intervalos interquartis (IQR). Posteriormente, os limites de valores foram estabelecidos com base nos valores mínimo e máximo de cada coluna da base original, seguindo uma distribuição uniforme contínua. Para as colunas textuais, os valores foram gerados aleatoriamente com a mesma distribuição observada na base de dados original, garantindo que a proporção de ambas as classes fosse preservada. Essa abordagem assegura que o balanceamento original dos dados seja mantido na análise preditiva. Posteriormente, alguns valores textuais foram reescritos como valores booleanos, preservando sua respectiva significância. Adicionalmente, colunas não relevantes ou com dados insuficientes para impactar no resultado foram removidas. Por fim, os dados faltantes nas colunas numéricas e categóricas foram preenchidos, respectivamente, com a média e a moda da coluna correspondente. Após essas alterações, o conjunto resultante contém 61 colunas de *features* e 301 amostras de equipamentos.

Com a base criada, foi realizado o tratamento necessário para viabilizar o treinamento dos modelos de aprendizado de máquina. Esse processo envolveu a codificação (*encoding*) das variáveis categóricas por meio do método conhecido como *Label Encoding*, no qual cada valor distinto é convertido em um número inteiro. Por exemplo, para a variável ‘Avaliação do Óleo’, contendo os valores ‘Normal’ e ‘Anormal’, a técnica atribuiria os valores 0 e 1, respectivamente. A escolha por esse tipo de codificação, em detrimento de alternativas como o *One Hot Encoding*, deve-se à alta cardinalidade presente, já que o *Label Encoding* preserva a estrutura original da base sem aumentar significativamente a quantidade de colunas. Esse crescimento, comum em representações *one-hot*, poderia comprometer o desempenho da Rede Neural, pois aumentaria a esparsidade da matriz de entrada e exigiria mais memória, processamento e tempo de execução. Além disso, para lidar com a escassez de amostras em alguns tipos de compartimento, aplicou-se a técnica de superamostragem SMOTE, que gera exemplos sintéticos para as classes minoritárias, promovendo o balanceamento do conjunto. Neste estudo, utilizou-se a configuração padrão da implementação disponível na biblioteca *scikit-learn*, na qual todas as classes passam a conter o mesmo número de instâncias da classe mais frequente.

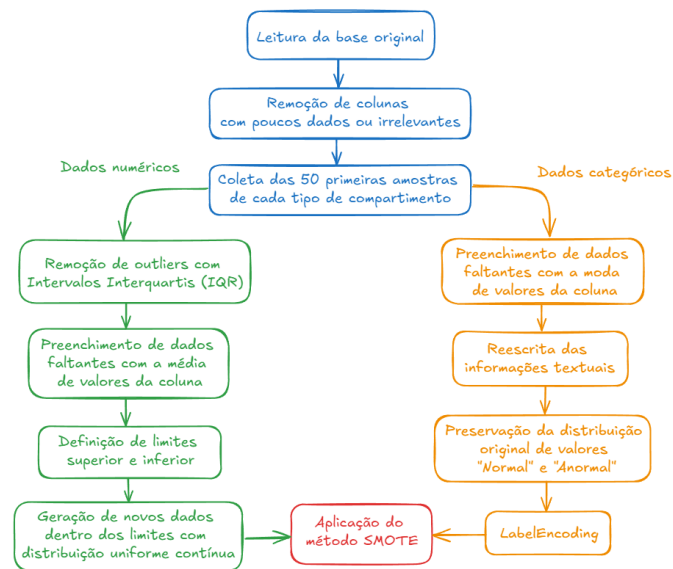


Figura 1. Diagrama de fluxo da montagem da base de dados sintética.

3.3 Escolha dos modelos de aprendizado de máquina

Atualmente, diversos modelos de aprendizado de máquina são utilizados para reconhecimento de padrões e classificação. Neste trabalho, foram selecionados três algoritmos para avaliar e comparar sua viabilidade nesse contexto: Floresta Randômica, Rede Neural Artificial e Máquina de Vetores de Suporte. A seguir, são apresentados e discutidos os principais aspectos de cada um desses modelos.

Rede Neural Artificial. Proposta por McCulloch and Pitts (1943), a Rede Neural Artificial é um modelo computacional inspirado no cérebro humano, composto por neurônios artificiais organizados em camadas. Cada neurônio recebe entradas, aplica uma função de ativação — uma função

matemática que transforma o valor resultante da soma ponderada das entradas e é responsável por introduzir não linearidades ao modelo e permitir que ele aprenda padrões complexos — e transmite a saída para a próxima camada. As redes possuem camadas de entrada, ocultas e de saída, ajustando os pesos das conexões durante o treinamento para minimizar o erro.

Floresta Randômica. Introduzida por Breiman (2001), a Floresta Randômica é um modelo de aprendizado de máquina baseado na construção de múltiplas Árvores de Decisão aleatórias. Cada árvore é treinada com uma amostra *bootstrap* dos dados — amostra obtida ao selecionar registros aleatoriamente com repetição a partir do conjunto original — e, em cada nó, apenas um subconjunto aleatório de características é considerado para divisão. Esse processo reduz o sobreajuste (*overfitting*) e melhora a generalização. A previsão final é obtida por votação majoritária (classificação) ou média (regressão), tornando o modelo mais robusto e preciso do que uma única árvore de decisão.

Máquina de Vetores de Suporte - SVM. Introduzido por Vapnik (1995), o SVM é um modelo amplamente utilizado para classificação, baseado na separação de dados por um hiperplano ótimo. Esse hiperplano maximiza a margem entre as classes, garantindo alta precisão e generalização. Novos dados são classificados conforme sua posição em relação a essa fronteira de decisão.

A Floresta Randômica foi selecionada devido à sua capacidade de generalização aprimorada, resultante da combinação de várias árvores de decisão, o que a torna menos propensa ao sobreajuste e mais robusta em comparação a uma única árvore. Por outro lado, a Rede Neural Artificial foi escolhida por sua flexibilidade e capacidade de modelar relações complexas entre os dados, especialmente em cenários onde os padrões são não lineares e difíceis de capturar por métodos tradicionais. Finalmente, a Máquina de Vetores de Suporte foi selecionada por sua capacidade de aprender padrões de classificação de dados com precisão equilibrada e reprodutibilidade. Essa combinação de modelos permite uma análise comparativa rica, onde a simplicidade interpretativa das árvores de decisão que compõem as florestas randômicas contrasta com a potência e a adaptabilidade das redes neurais, enquanto as Máquinas de Vetores de Suporte agregam robustez na separação eficiente de classes, especialmente em espaços de alta dimensionalidade. Dessa forma, a abordagem combinada proporciona uma visão abrangente da eficácia de diferentes técnicas no reconhecimento de padrões e classificação, equilibrando interpretabilidade, flexibilidade e precisão.

3.4 Treinamento dos modelos

O conjunto de dados foi dividido em duas partes, destinando-se 70% do total para o treinamento dos modelos e 30% para a validação e avaliação de seu desempenho. Os melhores hiperparâmetros — conjunto de configurações definidas antes do treinamento de um modelo de aprendizado de máquina que controlam o comportamento do algoritmo — para cada modelo da biblioteca *scikit-learn* foram definidos através do uso de *Grid Search*, um algoritmo que explora combinações entre hiperparâmetros definidos pelo usuário e define a melhor delas. Este estudo utiliza

a classe *GridSearchCV*, fornecido pela mesma biblioteca para facilitar esse processo.

A classe *GridSearchCV* permite a configuração de diversos parâmetros, dentre os quais, aqueles utilizados nos testes deste trabalho estão especificados na Tabela 2.

Tabela 2. Parâmetros e suas funções no *Grid Search*.

Parâmetro	Função	Valores atribuídos
model	Define o modelo de aprendizado de máquina a ser avaliado.	RandomForestClassifier, MLPClassifier, SVC.
param_grid	Define os hiperparâmetros a serem combinados e testados.	Dicionário de valores específico para cada modelo.
cv	Define a quantidade de divisões dos dados em treinamento e validação.	5
scoring	Define a métrica de avaliação do modelo.	Acurácia.
n_jobs	Controla a quantidade de núcleos da CPU usados.	-1 (todos os núcleos disponíveis).

Os hiperparâmetros foram ajustados via *Grid Search*, de acordo com valores frequentemente encontrados na literatura acadêmica, com variações em *n_estimators* (50-1000), *max_depth* (10-100) e *min_samples_split* (2-10) para RF; *hidden_layer_sizes* (32-128), *activation (relu, tanh)* e *learning_rate (constant, adaptive)* para ANN; e *C* (0.10-100) e *kernel (linear, rbf)* para SVM.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os testes foram conduzidos em um computador com processador Intel Core i3 10100f a 3,60 GHz e 16 GB de RAM. Para cada modelo, os hiperparâmetros foram ajustados via *Grid Search*, buscando a combinação que maximizasse a acurácia, utilizando *random_state = 42* para garantir reprodutibilidade. A melhor configuração foi então avaliada no conjunto de testes, também com *random_state = 42*. As métricas consideradas foram: acurácia — proporção de classificações corretas; *F1-score* — média harmônica entre precisão e *recall*; AUC-ROC — capacidade de discriminação entre classes pela taxa de verdadeiros e falsos positivos; e Matriz de Correlação de *Matthews* (MCC) — métrica equilibrada que considera todos os tipos de acertos e erros.

A melhor combinação para o modelo RF foi: *max_depth=10*, *min_samples_split=5* e *n_estimators=100*. O modelo RF apresentou bom desempenho, com acurácia próxima a 1,0 no treinamento e 0,99 no conjunto de teste. As métricas *F1-score*, AUC-ROC e MCC foram aproximadamente 0,98, 0,99 e 0,98, respectivamente. O treinamento levou cerca de 0,22 segundos.

Para a ANN, a melhor configuração obtida foi: *activation='tanh'*, *alpha=0,0001*, *batch_size=64*, *hidden_layer_sizes=(64,)*, *learning_rate='constant'*, *max_iter=1000* e *solver='adam'*. O modelo apresentou desempenho inferior, com acurácia de aproximadamente 0,68 no treinamento e 0,65 no teste, e métricas *F1-score*, AUC-ROC e MCC em torno de 0,05, 0,58 e -0,02, respectivamente. O treinamento durou cerca de 0,15 segundos. Embora os valores de MCC e

F1-score para a ANN tenham sido baixos, esses índices são consistentes com resultados de modelos não ajustados em bases industriais ruidosas, conforme observado por Susto et al. (2015).

Para o modelo SVM, a melhor combinação foi *C=0,10* e *kernel='linear'*, visto que o parâmetro *degree* só faz sentido caso o *kernel* seja polinomial. Além disso, para calcular algumas métricas para o modelo SVM, é necessário definir o parâmetro adicional de configuração 'probability' como 'True'. Dessa forma, o modelo realiza um ajuste adicional usando validação cruzada (*cross-validation*) — uma técnica comum na avaliação de modelos de aprendizado de máquina, que consiste em particionar os dados em subconjuntos para treinar e testar o modelo de forma alternada. Ressaltamos que não foi utilizada correlação cruzada (*cross-correlation*) neste trabalho, uma vez que os dados não são séries temporais. A metodologia adotada visa otimizar o desempenho preditivo dos modelos sem a análise de dependência entre variáveis ao longo do tempo. O modelo SVM obteve acurácia de aproximadamente 0,77 no treinamento e 0,75 no teste, com métricas *F1-score*, AUC-ROC e MCC em cerca de 0,58, 0,82 e 0,42, respectivamente. O treinamento foi o mais lento, durando em torno de 15 minutos e 29 segundos.

Após a obtenção das métricas de desempenho de cada modelo, a Tabela 3 apresenta um resumo comparativo dos valores de acurácia e tempo de treinamento, permitindo uma visão consolidada da performance dos métodos avaliados.

Tabela 3. Comparação da acurácia e tempo de treinamento dos modelos de aprendizado de máquina.

Método ML	Acurácia		Tempo de Treinamento (seg.)
	Treino	Teste	
ANN	0,68	0,65	0,15
RF	1,00	0,99	0,22
SVM	0,77	0,75	929,00

Realizou-se uma análise estatística ANOVA, que verifica se existem diferenças significativas entre as médias dos grupos. Antes disso, foram avaliados os pressupostos do ANOVA: a normalidade dos resíduos, testada pelo *Shapiro-Wilk*, e a homogeneidade das variâncias, avaliada pelo teste de *Levene*. Ambos indicaram que os dados atendem aos requisitos do método. O resultado do ANOVA mostrou diferença estatisticamente significativa entre os modelos ($p < 0,05$ e aproximando-se muito de 0), e o teste *post-hoc* de *Tukey* identificou que a Floresta Randômica teve desempenho superior aos demais.

O desempenho inferior dos outros modelos pode ser atribuído a limitações na modelagem dos dados e à otimização limitada dos hiperparâmetros, que foi feita com base em valores comumente encontrados na literatura, sem uma busca exaustiva. O modelo *Random Forest*, por usar várias árvores de decisão, é menos sensível a ajustes finos e mais robusto para generalizar em novos dados, capturando padrões complexos com menor risco de *overfitting*. Por isso, ele é mais adequado para o ambiente industrial estudado, que geralmente envolve dados variados e ruidosos.

5. CONCLUSÕES E TRABALHOS FUTUROS

A manutenção preditiva de equipamentos se destaca como uma estratégia para redução de custos em organizações. Por isso, este estudo busca avaliar a viabilidade do uso de algoritmos de aprendizado de máquina no contexto de previsão de falhas em diferentes equipamentos a partir da análise de características físico-químicas destes. Para o treinamento dos modelos de aprendizado de máquina, foi criada uma base de dados sintética baseada em uma base de dados real, fornecida por uma empresa especializada na prestação de serviços de análises de efeitos e modos de falhas em diferentes equipamentos.

A Floresta Randômica se mostrou a alternativa mais eficiente para predição de falhas, com alta acurácia e baixo tempo de treinamento. Entretanto, os modelos SVM e ANN podem se beneficiar de ajustes adicionais, como *Feature Selection* para RF, *Dropout* para ANN e otimização do *kernel* para SVM. Além disso, destaca-se as limitações deste estudo, como a necessidade do uso de uma base sintética, restrição ao cenário industrial simulado e o número limitado de modelos analisados. Nesse sentido, estudos futuros irão prioritariamente realizar uma busca exaustiva por hiperparâmetros mais adequados e, posteriormente, incluir a validação dos resultados com dados reais. Também será considerada a ampliação do escopo com modelos baseados em séries temporais e o uso de técnicas modernas de interpretabilidade dos modelos.

Além disso, destaca-se uma limitação deste estudo: a base de dados utilizada não apresenta marcações temporais entre os eventos, o que impossibilita a análise do tempo decorrido entre a ocorrência de uma falha e sua detecção pelo modelo. Este é um parâmetro relevante em aplicações industriais reais, pois está diretamente relacionado à capacidade de resposta do sistema. Dessa forma, como trabalho futuro, propomos a utilização de bases com registros temporais ou em tempo real, permitindo quantificar essa latência de detecção e avaliar a efetividade dos modelos nesse aspecto crítico da manutenção preditiva.

AGRADECIMENTOS

Os autores gostariam de agradecer ao Polo de Inovação do IFMG, um centro de Pesquisa, Desenvolvimento e Inovação (PD&I) credenciado pela EMBRAPPII (Associação Brasileira de Pesquisa e Inovação Industrial), pelo financiamento e suporte ao desenvolvimento deste trabalho. Agradecemos também ao IFMG campus Sabará pela infraestrutura e pelo ambiente propício à realização desta pesquisa.

REFERÊNCIAS

Alves, A.S., Silva, E.S., and Garcia, L.F.C. (2023). Análise de falhas e manutenção preditiva na indústria 4.0. *Revista FT*, 27(128), 25–35. doi:10.5281/zenodo.10205935. URL <https://revistaft.com.br/analise-de-falhas-e-manutencao-preditiva-na-industria-4-0/>.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., da P. Francisco, R., Basto, J.P., and Alcalá, S.G.S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. doi:<https://doi.org/10.1016/j.cie.2019.106024>. URL <https://www.sciencedirect.com/science/article/pii/S0360835219304838>.

Junior, M.C.S.P., Araújo, M.L.V., de Oliveira, E.A.R., Guimarães, V.B., Saba, H., Jorge, E.M.d.F., Nascimento Filho, A.S., and Ferreira, J.S. (2023). Estudo da aplicação de algoritmos de machine learning na manutenção preditiva de motores elétricos. *Revista de Gestão e Secretariado*, 14(10), 16542–16561. doi:10.7769/gesec.v14i10.2754. URL <https://ojs.revistagesec.org.br/secretariado/article/view/2754>.

Maheshwari, S., Tiwari, S., Rai, S., and Singh, S.V.D.P. (2024). Comprehensive study of predictive maintenance in industries using classification models and lstm model. URL <https://arxiv.org/abs/2403.10259>.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.

Relkar, A.S. (2021). Risk analysis of equipment failure through failure mode and effect analysis and fault tree analysis. *Journal of Failure Analysis and Prevention*, 21, 793–805.

Susto, G.A., Beghi, A., and Luca, C.D. (2012). A predictive maintenance system for epitaxy processes based on filtering and prediction techniques. *IEEE Transactions on Semiconductor Manufacturing*, 25, 638–649. URL <https://api.semanticscholar.org/CorpusID:46612877>.

Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., and Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3), 812–820. doi:10.1109/TII.2014.2349359.

Teramoto, A.M., dos Santos, L.L.C., and Daguam, W.G. (2024). Estruturação e aplicação de análise preditiva de óleo em plataformas de petróleo. *Unisanta Science and Technology*, 13(2), 77–83.

Vago, N.O.P., Forbicini, F., and Fraternali, P. (2024). Predicting machine failures from multivariate time series: An industrial case study. *Machines*, 12(6), 357.

Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York, NY.

Veloso, B., Gama, J., Ribeiro, R.P., and Pereira, P.M. (2022). A benchmark dataset for predictive maintenance. URL <https://arxiv.org/abs/2207.05466>.

Álvaro Huertas García, Muñoz, J., Ambite, E.D.M., Camarmas, M.A., and Ovejero, J.F. (2024). Detecta 2.0: Research into non-intrusive methodologies supported by industry 4.0 enabling technologies for predictive and cyber-secure maintenance in smes. URL <https://arxiv.org/abs/2405.15832>.