

## ANÁLISE EXPLORATÓRIA DOS DADOS E EXTRAÇÃO DE CARACTERÍSTICAS UTILIZADOS PARA CLASSIFICAÇÃO DE FALHAS EM SISTEMAS ELETROPNEUMÁTICOS<sup>1</sup>

YAMAMOTO, Érick Toshio<sup>2</sup>  
SUYAMA, Ricardo<sup>3</sup>

### RESUMO

Este estudo realizou uma análise exploratória de dados e extração de características para melhorar a manutenção preditiva de equipamentos industriais. Utilizando dados simulados gerados pelo software FluidSIM, foram extraídas 200 características dos domínios temporais, estatísticos e espectrais. As bibliotecas TSFEL e PyCaret foram usadas para a extração de características e desenvolvimento dos modelos de aprendizado de máquina, respectivamente. Random Forest e Decision Tree superaram significativamente o MLP em todas as métricas de avaliação, com acurácia de 96,83% no treino e acima de 90% no teste. Embora a Random Forest tenha demandado mais tempo de treinamento, apresentou melhor desempenho em AUC e F1-Score. Esses resultados enfatizam a importância de escolher o modelo certo para a manutenção preditiva, equilibrando precisão e eficiência computacional. A manutenção preditiva bem-sucedida otimiza a performance dos equipamentos, assegura a continuidade da produção e mantém a competitividade das empresas.

**Palavras-chave:** Manutenção preditiva, Análise de dados, Aprendizado de máquina, Extração de características, Sistema Eletropneumática.

### INTRODUÇÃO

No cenário competitivo atual, otimizar a eficiência dos equipamentos é crucial para garantir uma produção contínua e manter a competitividade das empresas. Interrupções inesperadas podem resultar em inatividade dispendiosa e comprometer a qualidade dos produtos. Por isso, cresce o interesse pela manutenção preditiva para prever esses eventos e minimizar impactos negativos.

A análise exploratória dos dados e a extração de características são abordagens promissoras, pois permitem entender detalhadamente as condições operacionais dos equipamentos, identificando padrões e anomalias que indicam falhas. A análise exploratória oferece insights sobre o desempenho dos sistemas, enquanto a extração de características destaca variáveis relevantes, facilitando a construção de modelos preditivos robustos.

A manutenção preditiva bem-sucedida requer tecnologias como IoT e Machine Learning e uma compreensão das limitações que podem afetar a integridade dos equipamentos. Essa abordagem representa um avanço significativo, permitindo estratégias que previnem falhas e otimizam a performance dos equipamentos, assegurando a continuidade da produção e a competitividade das empresas.

---

<sup>1</sup> Trabalho Vinculado a projeto de pesquisa de doutorado no programa de pós graduação em engenharia da informação da Universidade Federal do ABC (UFABC).

<sup>2</sup>MSc.; Fundação Universidade Federal do ABC (UFABC); Santo André; SP; erick.yamamoto@ufabc.edu.br; **YAMAMOTO, Érick Toshio.**

<sup>3</sup>Prof.Dr; Fundação Universidade Federal do ABC (UFABC); Santo André; SP; ricardo.suyama@ufabc.edu.br; **SUYAMA, Ricardo.**

## OBJETIVO DO ESTUDO

O principal objetivo deste estudo é realizar uma análise exploratória de dados e a extração de características para aprimorar a compreensão das condições operacionais dos equipamentos industriais, visando a implementação de estratégias de manutenção preditiva eficazes. Este estudo busca identificar padrões e anomalias nos dados coletados, facilitando a construção de modelos preditivos robustos que possam antecipar falhas e otimizar a performance dos sistemas. Os objetivos específicos incluem a coleta e organização dos dados, visualização dos dados, extração de características, desenvolvimento de modelos preditivos e avaliação da eficiência dos modelos.

## FUNDAMENTAÇÃO TEÓRICA

As técnicas de aprendizado de máquina são cruciais para a manutenção preditiva, facilitando a identificação de padrões e a previsão de necessidades futuras de manutenção, como destacado por Di Maggio (2022) e Li, Zhang e Ding (2022). A análise exploratória de dados é vital para compreender as condições operacionais dos equipamentos, destacando tendências, outliers e anomalias (Li, Zhang e Ding, 2022), e utilizando métodos avançados como a transformada de Fourier e análise de Wavelet (Di Maggio 2022). No presente trabalho, consideramos que o processo de manutenção preditiva envolve monitoramento contínuo através de sensores que registram vazão e pressão, permitindo o uso de classificadores como Árvore de Decisão (DT), Random Forest (RF) e Perceptron de Múltiplas Camadas (MLP) para prever falhas de forma eficiente. Essas técnicas permitem que equipes de manutenção atuem proativamente, otimizando recursos e minimizando inatividades, apoiadas por métodos consolidados na literatura (Li, Zhang e Ding, 2022; Zhao et al., 2022; Fernandes, Corchado e Marreiros, 2022, Sharma, Mittal e Soni, 2022; Assim, Obeidat e Hammad, 2020; Lee, Lee e Kim, 2019; Wang e Wang, 2018).

## DESENVOLVIMENTO

Este estudo desenvolve um sistema de predição com dados simulados gerados pelo FluidSIM, um software renomado da Art-Systems (2024) para simulação de sistemas pneumáticos, eletropneumáticos, hidráulicos e automação industrial. O FluidSIM permite a construção, simulação e análise de circuitos com um editor intuitivo e recursos visuais detalhados. A Figura 1 ilustra o sistema estudado, destacando sua aplicação prática.

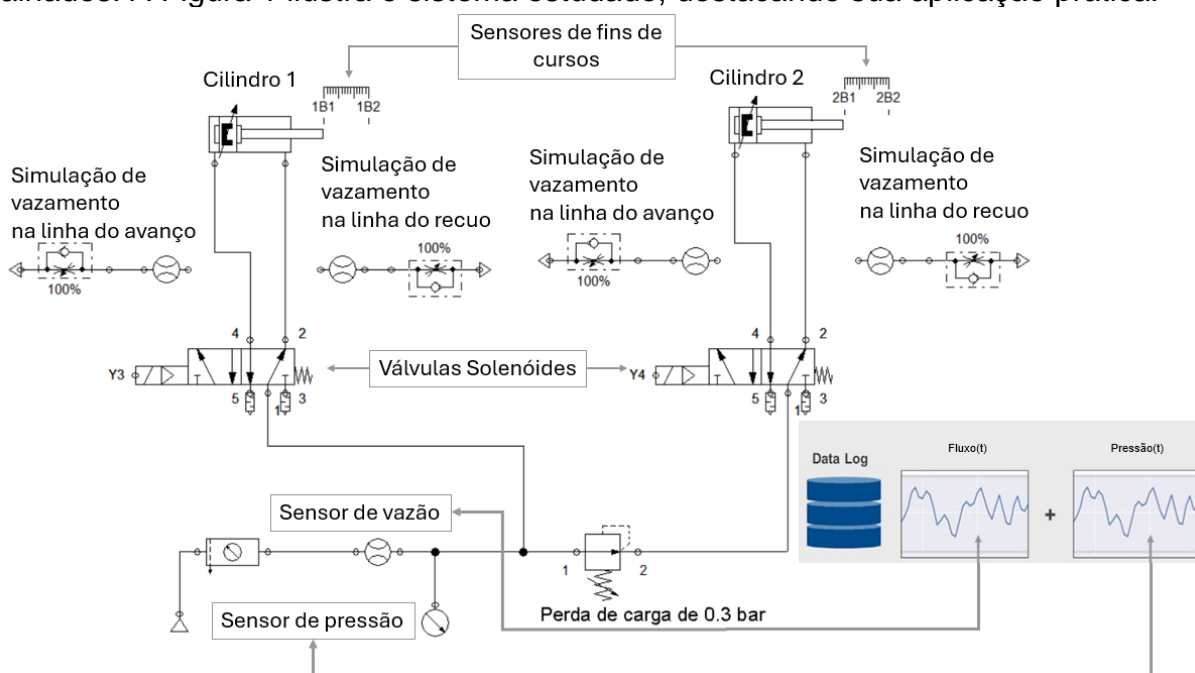


Figura 1 - Circuito Eletropneumático construído no FluidSIM.  
Fonte: Elaborado pelo autor (2024).

O sistema simulado permitiu a coleta de dados de pressão (em bar) e vazão (em l/min) em uma faixa de pressão entre 5,4 e 6,6 bar, comumente utilizado em sistemas pneumáticos (Foit, Banas e Cwikla, 2018), esta variação é devido a característica do sistema ser ideal e sem perdas.

Para desenvolver o modelo preditivo, utilizou-se a biblioteca TSFEL (Barandas et al., 2020) para a extração de 200 características nos domínios temporais, estatísticos e espectrais dos dados coletados. O objetivo foi classificar dados em sete categorias: Funcionamento Normal, Vazamento no Avanço/Recuo da Válvula 1, Vazamento Interno do Cilindro 1, Vazamento no Avanço/Recuo da Válvula 2, e Vazamento Interno do Cilindro 2.

Os classificadores foram implementados com a biblioteca PyCaret (Ali, 2020), utilizando 80% dos dados para treinamento e 20% para teste. A técnica GridSearchCV (2024), da biblioteca scikit-learn, foi empregada para otimizar hiperparâmetros.

Após o treinamento, a função *evaluate\_model* facilita a análise do desempenho do modelo. A técnica de *feature importance* (Brownlee, 2020), Figura 2(a) e Figura 2(b), foi aplicada para identificar as características mais influentes, e a *permutation importance*, Figura 2(c), da scikit-learn foi utilizada para a rede neural MLP.

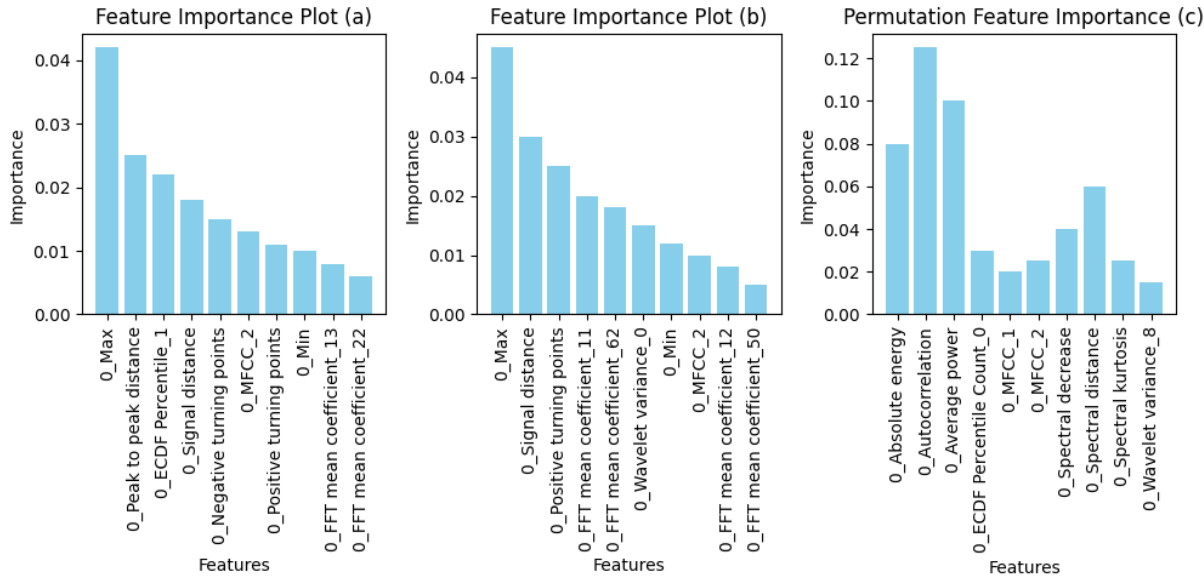


Figura 2 - Feature Importance: (a) Árvore de Decisão; (b) Random Forest; (c) MLP.  
Fonte: Elaborado pelo autor (2024).

As características extraídas ajudam a analisar e classificar defeitos comuns em sistemas eletropneumáticos. As características que foram vistas de forma expressiva na Figura 2 estão apresentadas na Tabela 1.

**Tabela 1 – FEATURE IMPORTANCE X POSSÍVEIS DEFEITOS.**

Feature	Descritivo da Importância	Possíveis Defeitos
0_Max	Indicador crucial de picos de pressão ou vazão que podem sinalizar condições críticas no sistema.	Sobrecarga, vazamentos, obstrução ou resistência no sistema.
0_Autocorrelation	Fundamental para identificar padrões repetitivos e anomalias cíclicas que podem levar a falhas.	Falhas cíclicas ou persistentes.
0_Average power	Essencial para avaliar o consumo energético e a eficiência, indicando possíveis sobrecargas.	Desgaste, sobrecarga, consumo energético e eficiência do sistema.
0_Peak to peak distance	Importante para detectar flutuações na pressão ou fluxo, sugerindo instabilidades operacionais.	Instabilidade de controle, vibrações excessivas, problemas de controle de pressão.

Fonte: Elaborado pelo autor (2024)

## RESULTADOS E DISCUSSÃO

Os resultados do estudo foram divididos em duas partes: a fase de treinamento e a fase de teste dos modelos. A seguir, apresentamos uma análise detalhada do desempenho dos modelos Árvore de Decisão, Random Forest e Multi-Layer Perceptron (MLP) em termos de acurácia, AUC (Área Sob a Curva), revocação, precisão, F1-Score e tempo de execução. Essas métricas são cruciais para avaliar a eficácia dos modelos em prever falhas e otimizar a manutenção dos equipamentos industriais. As tabelas a seguir mostram os resultados detalhados para cada modelo nas fases de treino e teste, considerando todas as características (Tabela 2) e as 10 características mais importantes para cada modelo (Tabela 3).

**Tabela 2 – MÉTRICAS DE TODAS AS CARACTERÍSTICAS PARA CADA MODELO**  
**TREINO: MÉTRICAS DE DESEMPENHO E TEMPO**

Modelo	Acurácia	AUC	Revocação	Precisão	F1-Score	Tempo (s)
Decision Tree Classifier	96,83%	0,9990	96,83%	97,08%	96,83%	24,02
Random Forest Classifier	96,83%	0,9987	96,83%	96,99%	96,83%	102,42
MLP Classifier	34,00%	0,6840	34,00%	53,09%	29,95%	40,30
TESTE: MÉTRICAS DE DESEMPENHO E TEMPO						
Modelo	Acurácia	AUC	Revocação	Precisão	F1-Score	Tempo (s)
Decision Tree Classifier	90,88%	0,9574	90,88%	90,98%	90,80%	1,75
Random Forest Classifier	92,55%	0,9919	92,55%	92,65%	92,88%	8,03
MLP Classifier	33,16%	0,6743	33,16%	51,28%	28,84%	2,61

Fonte: Elaborado pelo autor (2024).

**Tabela 3 - MÉTRICAS DAS 10 CARACTERÍSTICAS MAIS IMPORTANTES DE CADA MODELO**  
**TREINO: MÉTRICAS DE DESEMPENHO E TEMPO**

Modelo	Acurácia	AUC	Revocação	Precisão	F1-Score	Tempo (s)
Decision Tree Classifier	98,61%	0,9998	98,61%	98,72%	98,60%	5,19
Random Forest Classifier	98,61%	0,9998	98,61%	98,61%	98,61%	20,81
MLP Classifier	23,34%	0,5820	23,34%	37,83%	17,20%	21,84
TESTE: MÉTRICAS DE DESEMPENHO E TEMPO						
Modelo	Acurácia	AUC	Revocação	Precisão	F1-Score	Tempo (s)
Decision Tree Classifier	92,30%	0,9563	92,30%	92,31%	92,24%	0,18
Random Forest Classifier	94,06%	0,9952	94,06%	94,09%	94,02%	1,33
MLP Classifier	23,06%	0,5793	23,06%	31,40%	16,62%	1,37

Fonte: Elaborado pelo autor (2024).

## CONSIDERAÇÕES FINAIS/CONCLUSÃO

Os resultados deste estudo mostram que os modelos Árvore de Decisão e *Random Forest* superaram o MLP em todas as métricas de avaliação. Árvore de Decisão e *Random Forest* alcançaram 96,83% de acurácia no treino e 90,88% e 92,55% no teste, respectivamente, enquanto o MLP obteve 34,00% no treino e 33,16% no teste. A análise exploratória dos dados e a extração de 200 características foram cruciais para identificar padrões e anomalias. A importância das características foi avaliada, destacando variáveis chave para a manutenção preditiva. Esses achados ressaltam a importância de escolher o

modelo certo, equilibrando precisão e eficiência computacional, para otimizar a performance dos equipamentos e manter a competitividade.

## REFERÊNCIAS

- Di Maggio, L. G. "Intelligent Fault Diagnosis of Industrial Bearings Using Transfer Learning and CNNs Pre-Trained for Audio Classification." *Sensors* 23(1): 211. 2022. <https://doi.org/10.3390/s23010211>.
- Li, X., W. Zhang, e Q. Ding. "Multi-Stage Convolutional Neural Network for Automatic Bearing Fault Diagnosis." *IEEE Transactions on Industrial Informatics* 18(6): 3767-3776. 2022. <https://doi.org/10.1109/TII.2021.3091658>.
- Zhao, R., R. Yan, J. Wang, K. Mao, F. Shen, e X. Wang. "Machine Learning-Based Autonomous Diagnosis of Rotating Machinery Faults: A Comprehensive Review and Future Research Direction." *Journal of Manufacturing Systems* 52: 227-243. 2022. <https://doi.org/10.1016/j.jmsy.2019.06.012>.
- Sharma, J., Mittal, M. L., e Soni, G., "Condition-based maintenance using machine learning and role of interpretability: a review". *International Journal of System Assurance Engineering and Management*. Vol.15, Issue 4, pp.1345-1360, 2022. <https://doi.org/10.1007/s13198-022-01843-7>.
- Fernandes, M., Corchado, J.M., e Marreiros, G., "Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review". *Applied Intelligence*. Vol. 52, Issue 12, pp.14246-14280, 2022. <https://doi.org/10.1007/s10489-022-03344-3>.
- Foit, K., Banas, W. e Cwikła, G. "The pneumatic and electropneumatic systems in the context of 4th industrial revolution", *IOP Conference Series: Materials Science and Engineering*, ModTech, 2018.
- Wang, J. e Wang, Q. "Intelligent explicit model predictive control based on machine learning for microbial desalination cells", *SAGE Publishing*, 2018. Disponível em: <https://doi.org/10.1177/0959651818816845>.
- Lee, S.M., Lee, D. e Kim, Y.S. "The quality management ecosystem for predictive maintenance in the Industry 4.0 era". *Int J Qual Innov* 5, Springer Nature, 2019. Disponível em: <https://doi.org/10.1186/s40887-019-0029-5>.
- Gohel, P., Singh, P., e Mohanty, M. "Explainable AI: current status and future directions", *IEEE ACCESS*, 2021.
- Assim, M., Obeidat, Q., Hammad, M. "Software Defects Prediction using Machine Learning Algorithms", 2020 *International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Sakheer, Bahrain, pp. 1-6, 2020.
- Art-Systems. *FluidSIM*. 2024. Disponível em: *Art-Systems FluidSIM*.
- Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T. e Gamboa, H., *TSFEL: Time Series Feature Extraction Library*, SoftwareX, 2020. Disponível em: <https://doi.org/10.1016/j.softx.2020.100387>.
- Ali, M., *PyCaret: An open source, low-code machine learning library in Python*, 2020. Disponível em: <https://pycaret.gitbook.io/docs>.
- GridSearchCV. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html). Acesso em 05 de Jul. de 2024.
- Brownlee, J., *How to Calculate Feature Importance With Python*, 2020. Disponível em: <https://machinelearningmastery.com/calculate-feature-importance-with-python/>. Acesso em 05 de Jul. de 2024.