Anais do **Simpósio Acadêmico de Engenharia de Produção (SAEPRO)** da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Análise Estatística Aplicada à Validação de Modelos Preditivos para Reparabilidade de PCBA

Julyana Nazario Alves

Centro Paula Souza (julyana.alves@cpspos.sp.gov.br)

Vinicius Querencia

Centro Paula Souza (vinicius.querencia@cpspos.sp.gov.br)

Antonio Gualhardi

Centro Paula Souza (antonio.gualhardi@cpspos.gov.br)

Resumo

A crescente complexidade dos processos produtivos impõe desafios às empresas de serviços técnicos, especialmente às que atuam no reparo de equipamentos eletrônicos. A variabilidade dos defeitos, a diversidade de modelos e a subjetividade dos diagnósticos tornam o controle da qualidade e o planejamento operacional mais difíceis. Neste contexto, a aplicação de técnicas de Inteligência Artificial (IA) e aprendizado de máquina surge como uma alternativa promissora para apoiar decisões e reduzir incertezas no processo de reparo. Este estudo tem como objetivo validar e comparar três algoritmos supervisionados *K-Nearest Neighbors* (KNN), *Random Forest* e *Multilayer Perceptron (MLP)* na previsão da reparabilidade de placas-mãe em uma operação de serviço de uma empresa multinacional do setor eletrônico. Foram analisados 2.309 registros reais de reparo, sendo 77,3% de placas recuperáveis e 22,7% devolvidas ao cliente. Todos os modelos apresentaram associação significativa com os resultados reais ($p < 0,001$). Os resultados indicam que os modelos de aprendizado de máquina, especialmente *Random Forest* e MLP, são ferramentas eficazes para previsão de

reparabilidade, permitindo padronizar diagnósticos, reduzir retrabalhos e apoiar a gestão da qualidade em ambientes de alta variabilidade.

Palavras-chave: Inteligência Artificial; Aprendizado de Máquina; Processos de Reparo; Predição de Falhas; Machine Learning.

Abstract

The growing complexity of production processes poses challenges for technical service companies, especially those involved in the repair of electronic equipment. The variability of defects, the diversity of models, and the subjectivity of diagnoses make quality control and operational planning more difficult. In this context, the application of Artificial Intelligence (AI) and machine learning techniques emerges as a promising alternative to support decision-making and reduce uncertainties in the repair process. This study aims to validate and compare three supervised algorithms K-Nearest Neighbors (KNN), Random Forest, and Multilayer Perceptron in predicting the repairability of motherboards in the service operation of a multinational company in the electronics sector. A total of 2,309 real repair records were analyzed, with 77.3% being recoverable boards and 22.7% returned to the customer. All models showed a significant association with the actual outcomes ($p < 0.001$). The results indicate that the machine learning models, especially Random Forest and MLP, are effective tools for predicting repairability, enabling the standardization of diagnoses, reducing rework, and supporting quality management in high-variability environments.

Keywords: Artificial Intelligence; Machine Learning; Repair Processes; Failure Prediction; Predictive Maintenance.

1. INTRODUÇÃO

A crescente complexidade dos processos produtivos impõe desafios às organizações, especialmente às que operam em setores de tecnologia e prestação de serviço. O setor de serviços é responsável por mais de dois terços da atividade econômica global, destacando-

se como o principal gerador de valor em economias desenvolvidas e emergentes (Fitzsimmons & Fitzsimmons, 2014). No Brasil, mais de 52% das empresas estão voltadas para a prestação de serviços, conforme dados da ABDI (2024), reforçando o setor terciário como o motor econômico do país. Neste cenário, as empresas de reparo de equipamentos eletrônicos desempenham um papel estratégico, pois suas soluções prolongam a vida útil dos produtos, reduzem custos de substituição e mitigam impactos ambientais (Slack et al., 2019).

Entretanto, as operações de reparo enfrentam desafios significativos devido à variabilidade dos processos, heterogeneidade dos componentes e influência direta dos clientes nos critérios de qualidade e tempo de resposta (Johnston & Clark, 2008). Essa variabilidade dificulta a padronização de diagnósticos e a consolidação de controles operacionais, especialmente em linhas de reparo de placas-mãe de notebooks e desktops, onde a diversidade de modelos e tipos de falhas gera incertezas que afetam o planejamento da capacidade, alocação de mão de obra e cumprimento de acordos comerciais. Assim, a gestão eficiente dessas operações requer ferramentas de apoio à decisão que funcionem bem em ambientes complexos e incertos.

A Inteligência Artificial (IA) surge como um recurso estratégico para o setor de serviços. Avanços em aprendizado de máquina e mineração de dados permitem a análise de grandes volumes de informações históricas, identificando padrões ocultos e prevendo falhas antes da intervenção humana (Russell & Norvig, 2021; Jordan & Mitchell, 2015). Classificadores supervisionados e técnicas de ensemble learning têm se mostrado eficazes em diagnósticos e manutenção preditiva (Zhang & Ma, 2012; Lee et al., 2014). Ao minimizar a subjetividade do julgamento humano e oferecer previsões baseadas em evidências, essas ferramentas aumentam a confiabilidade dos diagnósticos, otimizam o tempo de reparo e reduzem o retrabalho.

Este estudo justifica-se pela necessidade de introduzir decisões baseadas em dados empíricos em um ambiente de serviços técnicos que, tradicionalmente, depende fortemente do conhecimento tácito e da experiência individual de técnicos. A alta

variabilidade e complexidade dos defeitos em placas-mãe criam um cenário propício a inconsistências diagnósticas, retrabalho, custos elevados e insatisfação do cliente.

Este estudo tem como objetivo principal validar e comparar diferentes classificadores de aprendizado de máquina para prever a efetividade de reparos em placas-mãe. A validação será feita mediante a comparação direta das previsões de cada modelo com os dados reais de uma operação de uma empresa multinacional do setor eletrônico. A análise estatística do desempenho de cada classificador, frente aos resultados observados, buscará identificar a abordagem mais precisa para o diagnóstico de falhas, demonstrando seu potencial para a padronização de processos e a redução de incertezas em cenários de alta variabilidade operacional.

2. FUNDAMENTAÇÃO TEÓRICA

A gestão de serviços distingue-se da manufatura pela maior intensidade da participação do cliente e pela intrínseca variabilidade dos processos (Fitzsimmons & Fitzsimmons, 2014). Esse cenário é particularmente complexo em linhas de reparo de equipamentos eletrônicos, onde a diversidade de falhas, modelos e gerações de dispositivos acentua essa variabilidade, dificultando a padronização dos diagnósticos e a previsibilidade do tempo de execução. Conforme destacado por Slack, Brandon-Jones e Burgess (2019), a variabilidade no fluxo operacional impacta diretamente custos, produtividade e a qualidade percebida pelo cliente, tornando imperativo o desenvolvimento de mecanismos de controle e previsão.

Em modelos de negócio *business-to-business* (B2B), a previsibilidade do processo de reparo é estratégica, transcendendo o mero cumprimento de prazos Service *Level Agreement* (SLA) para se tornar a base de uma vantagem competitiva pautada em confiabilidade e agilidade (Johnston & Clark, 2008). Nesse contexto, a capacidade de prever a reparabilidade de placas eletrônicas antes da intervenção manual completa emerge como uma solução pivotal, capaz de reduzir desperdícios, otimizar a alocação da mão de obra especializada e promover o uso racional dos recursos.

Diante dessa conjuntura de incerteza e complexidade, a inteligência artificial (IA) e, mais especificamente, o aprendizado de máquina, surgem como abordagens promissoras para o suporte à gestão operacional. Essas tecnologias permitem o processamento de grandes volumes de dados, identificando padrões complexos e realizando predições em contextos de alta variabilidade (Russell & Norvig, 2021). A aplicação de técnicas de IA em serviços incrementa a precisão decisória e reduz a dependência de análises puramente subjetivas, contribuindo para a padronização e a otimização de recursos (Van Klompenburg et al., 2022).

- **k-NN (*k-Nearest Neighbour*):** Este classificador baseia-se na formação de *clusters* que calculam a distância entre objetos vizinhos para classificar amostras semelhantes. Como um método de aprendizagem não paramétrico, é eficaz com dados ruidosos e de fácil implementação (Acharya et al., 2020). Suas aplicações incluem a reconstituição de dados ausentes e a previsão de rendimentos em processos (Shahbeik et al., 2023; İrik et al., 2023).
- ***Random Forest* (RF):** Fundamentado em um conjunto de árvores de decisão, o RF realiza predições por meio de divisões sucessivas nos nós, baseadas nos atributos dos dados. A predição final é definida pela votação da maioria das árvores (comitê) (Oshiro, Perez e Baranauskas, 2012). Este modelo apresenta alto potencial para prever rendimento (İrik et al., 2023) e evita *overfitting* devido à sua variabilidade interna (Liu et al., 2020).
- ***Multilayer Perceptron* (MLP):** O MLP é uma rede neural artificial composta por múltiplas camadas (entrada, ocultas e saída), com neurônios totalmente conectados. Este modelo é capaz de aprender padrões não-lineares e complexos, embora sua performance possa ser sensível a ruídos e *outliers* nos dados (Rezaie-Balf, M. et al., 2020).

O jamovi é um software de análise de dados desenvolvido para ser gratuito, intuitivo e sempre atualizado com os métodos estatísticos mais recentes. Sua filosofia central é ser

uma plataforma feita pela comunidade, onde qualquer pesquisador pode criar e compartilhar novas formas de análise, tornando-as disponíveis para todos os usuários.

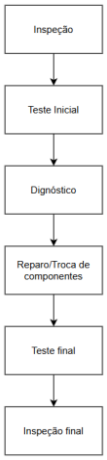
Uma característica fundamental do jamovi é sua neutralidade científica: ele não favorece nenhuma escola de pensamento estatístico específica, mas sim cria um ambiente onde diferentes abordagens podem coexistir e ser igualmente valorizadas (jamovi, 2025).

3. MÉTODO

A pesquisa foi realizada em uma empresa multinacional especializada em reparo de equipamentos eletrônicos. Para o desenvolvimento do estudo, foi utilizado a operação de reparo de notebooks e desktops de um cliente com grande relevância para o mercado de equipamentos computacionais.

O escopo da operação adotado para pesquisa se limita a uma linha de reparo de PCBA (figura 1). Toda a mão de obra envolvida no processo recebe treinamento específico para a função que desempenha. No entanto, as etapas de diagnóstico e reparo exigem profissionais altamente especializados, em razão da elevada densidade de componentes, da presença de múltiplas camadas e de vias internas nas PCBA. Além disso, a grande variabilidade de modelos e o reparo simultâneo de placas com diferentes níveis de complexidade tornam o processo de difícil gerenciamento. Nesse contexto, a análise de previsibilidade de reparo por meio de modelos preditivos, mostra-se uma ferramenta promissora para apoiar a tomada de decisão e otimizar o uso dos recursos disponíveis. Para validar a aplicabilidade dessa abordagem, os resultados preditos pelos modelos serão comparados com os dados reais de reparo, utilizando-se métodos estatísticos que permitam avaliar a acurácia e a correlação entre as previsões e os resultados observados.

Figura 1 – Linha de reparo de PCBA.



Fonte: Elaborado pelos autores (2025).

3.1 Modelos preditivos

3.1.1 Coleta de dados

Os dados utilizados são dados secundários e qualitativos, foram obtidos através do sistema ERP da empresa, do período de fevereiro até junho de 2025. Os metadados utilizados pelos classificadores k-NN, *Randon Forest* e MLT estão apresentados no quadro 1:

Quadro 1 - Metadados do *dataset*.

Coluna	Descrição	Tipo da variável
Product	Código do produto	entrada (*feature)*
PN	*Partnumber* do produto	entrada (*feature)*
Model	Modelo do produto	entrada (*feature)*
Category	Tipo do produto	entrada (*feature)*
Part Type	Equipamento com defeito	entrada (*feature)*
Result	Resultado dos testes após o reparo	saída (*target*)

Fonte: Elaborado pelos autores (2025)

Os dados coletados correlacionam o diagnóstico técnico com o componente responsável pela falha, bem como com as variações de modelo, *Part Number* (PN) e demais características da *Printed Circuit Board Assembly* (PCBA) que influenciam a eficácia do reparo.

3.1.2 Treinamento dos modelos preditivos

Para realizar a predição, as colunas de entrada (features) selecionadas para o treinamento dos modelos foram: 'PN', 'Category', 'Model' e 'Part Type'. O pré-processamento dos dados seguiu as seguintes etapas:

a. Carregamento dos dados: O arquivo de planilha eletrônica foi lido utilizando a biblioteca pandas.

b. Codificação de variáveis categóricas: As variáveis categóricas presentes nas features ('PN', 'Category', 'Model', 'Part Type') foram convertidas em representações numéricas utilizando LabelEncoder da biblioteca sklearn.preprocessing. Este passo foi necessário, pois a maioria dos algoritmos de aprendizagem de máquina requer entradas numéricas.

c. Separação de *features* e *target*: As colunas de entrada (features) foram separadas da coluna de saída (target), que é a variável a ser prevista ('Result').

d. Normalização dos dados: As *features* numéricas (após a codificação das categóricas) foram normalizadas utilizando StandardScaler da biblioteca sklearn.preprocessing. A normalização foi realizada para garantir que todas as *features* contribuam igualmente para o treinamento do modelo, evitando que features com escalas maiores dominem o processo de aprendizagem.

e. Divisão em conjuntos de treino e teste: O conjunto de dados pré-processado foi dividido em conjuntos de treino e teste, com 70% dos dados destinados ao treinamento dos modelos e 30% para a avaliação de seu desempenho. A divisão foi realizada de forma aleatória, garantindo a reprodutibilidade dos resultados por meio da definição de uma semente.

Para avaliação dos resultados de três modelos preditivos distintos, foram selecionados três algoritmos supervisionados que apresentaram complementares características de aprendizagem:

1. k-Nearest Neighbors (k-NN): Um classificador não paramétrico que classifica uma instância com base na maioria das classes de seus k vizinhos mais próximos no espaço de características. Para este estudo, foi configurado com 4 vizinhos.

2. *MultiLayer Perceptron* (MLP): Uma rede neural artificial *feedforward*, treinada com o algoritmo de retropropagação. Para esta aplicação foi configurada com duas camadas ocultas compostas por 10 e 5 neurônios, função de ativação ReLU, otimizador Adam e máximo de 10.000 iterações.

3. *Random Forest*: Um método de ensemble baseado em árvores de decisão. Ele constrói múltiplas árvores de decisão durante o treinamento e produz a classe que é a moda das classes (classificação) ou a previsão média (regressão) das árvores individuais. Para este trabalho, foi utilizado com 500 árvores na floresta. Cada um desses modelos foi treinado independentemente utilizando o conjunto de dados de treino e, posteriormente, utilizado para fazer previsões no conjunto de teste.

3.2 Análise estática

O jamovi foi empregado como ferramenta de análise estatística, e para seu uso foi necessário a tabulação dos dados obtidos dos classificados e dos dados do sistema ERP que permitissem a correlação dos resultados dos modelos preditivos e os dados reais. Os dados foram classificados de acordo com o Quadro 2:

Quadro 2 – Classificação dos dados

Coluna	Descrição	Tipo de Variaiveis
Result	Resultado dos testes após reparo	Variavel dependente
Repair Result	Resultado do reparo após troca do componente	Variavel independente
Prediçao KNN	Resultado da predição através do classificador KNN	Variavel independente
Prediçao Random Forest	Resultado da predição através do classificador Random Forest	Variavel independente
Prediçao MLP	Resultado da predição através do classificador MLP	Variavel independente

Fonte: Elaborado pelos autores (2025)

No contexto de classificadores preditivos aplicados a dados reais, o teste Qui-quadrático (X^2) é o modelo estatístico apropriado porque permite verificar se existe dependência significativa entre as classes previstas pelo modelo e os resultados observados na prática.

Fórmula Qui-quadratico:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Onde:

- O_i= valor observado,
- E_i= valor esperado,
- e a soma é feita sobre todas as categorias analisadas.

Fórmula df - Graus de liberdade

$$df = (r-1)(c-1)$$

onde:

- r = número de linhas da tabela de contingência

- c = número de colunas

Valor de significância (p-value)

Probabilidade de obter um valor de χ^2 igual ou maior que o observado, assumindo que a hipótese nula (H_0) de independência entre as variáveis é verdadeira.

4. RESULTADOS

Após o treinamento os três classificadores apresentaram desempenhos distintos devido às diferenças na forma como processam os dados e aprendem os padrões. O k-NN **(***K-Nearest Neighbors)* realiza a classificação com base na similaridade entre amostras, o que o torna sensível à distribuição e à escala dos dados, e pode ter desempenho inferior quando há ruído ou muitas variáveis. O *Random Forest*, por outro lado, combina vários modelos de árvore de decisão para reduzir o risco de *overfitting* e tende a apresentar maior estabilidade em bases heterogêneas e com variáveis categóricas. Já o MLP (*Multilayer Perceptron*), uma rede neural artificial, é capaz de capturar relações não lineares mais complexas, mas depende fortemente do ajuste de parâmetros e da quantidade de dados disponíveis para treinamento. As taxas de acerto por classificadores podem ser observadas no Tabela 1:

Tabela 1 – Taxa de acerto

Classificador	Taxa de acerto
k-NN	95,35 %
Random Forest	99,12%
MLP	98,75%

Fonte: Elaborado pelos autores (2025)

Os dados da tabela 1 relacionados a equação Qui-quadrado apresentaram o resultado a seguir:

Tabela 2 – Resultado Qui-quadratico k-NN

k-Nearest Neighbors

	Repair Result		
	P - Pass	T - Return to Customer	Total
P - Pass			
Observado	1753	92	1845
% em relação á linha	95,00%	5,00%	100,00%
T - Return to Customer			
Observado	31	433	464
% em relação á linha	6,70%	93,30%	100,00%
Total			
Observado	1784	525	2309
% em relação á linha	77,30%	22,70%	100,00%

	Valor	**df**	**p**
χ^2	1647	1	< ,001
N	2309		

Fonte: Elaborado pelos autores com uso do software Jamovi (versão 2.4).

Tabela 3 – Resultado Qui-quadratico *Randon Forest*

Random Forest

	Repair Result		
	P - Pass	T - Return to Customer	Total
P - Pass			
Observado	1782	25	1807
% em relação á linha	98,6%	1,4%	100,00%
T - Return to Customer			
Observado	2	500	502
% em relação á linha	0,40%	99,6%	100,00%
Total			
Observado	1784	525	2309
% em relação á linha	77,3%	22,70%	100,00%

	Valor	**df**	**p**
χ^2	2157	1	< ,001
N	2309		

Fonte: Elaborado pelos autores com uso do software Jamovi (versão 2.4).

Tabela 4 – Resultado Qui-quadratico MLP

Multilayer Perceptron

	Repair Result		
	P - Pass	T - Return to Customer	Total
P - Pass			
Observado	1784	38	1822
% em relação á linha	97,9%	2,10%	100,00 %
T - Return to Customer			
Observado	0	487	487
% em relação á linha	0,00%	100,00 %	100,00 %
Total			
Observado	1784	525	2309
% em relação á linha	77,30%	22,70%	100,00 %

	Valor	**df**	**p**
χ^2	2097	1	< ,001
N	2309		

Fonte: Elaborado pelos autores com uso do software Jamovi (versão 2.4).

5. DISCUSSÕES

Os três algoritmos (k-NN, *Random Fores*t e MLP) apresentaram relação significativa com os resultados reais do reparo ($p < 0,001$), mostrando que suas previsões não ocorreram por acaso. O k-NN obteve bons resultados, com 95% de acerto para placas “Pass” e 93,3% para “Return to Customer”. Mesmo assim, apresentou mais erros que os outros modelos (5% de falsos positivos e 6,7%de falsos negativos). Esse comportamento é comum no k-NN, pois ele é sensível a variações e ruídos nos dados.

O *Random Forest* teve o melhor desempenho geral, com 98,6% de acerto para “Pass” e 99,6% para “Return to Customer”. Apresentou o maior valor de qui-quadrado ($\chi^2 = 2157$), indicando forte associação com os dados reais. Esse resultado mostra que o modelo aprendeu os padrões do processo com alta confiabilidade.

O MLP apresentou bom desempenho, com 97,9% de acerto para “Pass” e 100% para “Return to Customer”. Se mostrando assertivo em todas as previsões de irreparabilidade

de placas, sendo possível reduzir retrabalhos. No entanto, teve 2,1% de falsos positivos, apresentando-se menos eficiente quando comparado *Random Forest*.

Tabela 5 – Comparação modelos

Algoritmo	χ^2	Acurácia "P - Pass"	Acurácia "T - Return"	Destaque principal
KNN	1647	95,0%	93,3%	Simples, mas sensível a ruídos
Random Forest	2157	98,6%	99,6%	Maior precisão geral
MLP	2097	97,9%	100,0%	Melhor na detecção de "não reparáveis"

Fonte: Elaborado pelos autores (2025).

Os dados reais de reparo refletem 77,3% de placas recuperáveis "Pass" e 22,7% de "Return to Customer", indicando um desequilíbrio de classes. Mesmo com essa desproporção, todos os modelos mantiveram boa capacidade preditiva, especialmente *Random Forest* e MLP, que se mostraram robustos contra o viés da classe majoritária.

Isso demonstra que ambos os modelos conseguem aprender padrões de falha e sucesso de forma eficiente, sendo adequados para suporte preditivo ao processo de qualidade e diagnóstico técnico.

6. CONSIDERAÇÕES FINAIS

Desta forma, o presente trabalho evidencia a importância da interface entre estatística, aprendizado de máquina e gestão de serviços industriais. A análise estatística não foi apenas uma etapa metodológica, mas o alicerce para a validação da confiabilidade preditiva dos modelos em um ambiente operacional real. Através de testes de significância ($p < 0,001$) e métricas de desempenho validadas contra dados reais, o estudo demonstrou, com robustez quantitativa, a viabilidade e o impacto do *Machine Learning* aplicadas a operações complexas e suscetíveis à variabilidade. Este rigor estatístico é fundamental para transformar outputs de algoritmos em insights

acionáveis para a gestão, fundamentando a tomada de decisão em evidências sólidas e não em correlações espúrias.

Por sua vez, a gestão operacional pode se beneficiar do desenvolvimento de uma ferramenta que, lastreada por essa análise estatística, facilita a padronização dos diagnósticos e reduz a variabilidade associada a avaliações subjetivas. Isso pode implicar em redução de custos operacionais, diminuição de retrabalho e melhoria na satisfação do cliente, fortalecendo a competitividade de empresas do setor de serviço.

REFERÊNCIAS

ACHARYA, S. et al. A long short term memory deep learning network for the classification of negative emotions using EEG signals. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2020. Anais. 2020. DOI: https://doi.org/10.1109/IJCNN48605.2020.9207280.

AGÊNCIA BRASILEIRA DE DESENVOLVIMENTO INDUSTRIAL (ABDI). PIB do 2º Trimestre de 2024: indústria e serviços garantem recuperação, 2024. Disponível em: https://www.abdi.com.br. Acesso em: 23 ago. 2025.

ÇETIN, N.; SAĞLAM, C. Rapid detection of total phenolics, antioxidant activity and ascorbic acid of dried apples by chemometric algorithms. *Food Bioscience*, v. 47, p. 101670, 2022. DOI: https://doi.org/10.1016/j.fbio.2022.101670.

FITZSIMMONS, J. A.; FITZSIMMONS, M. J. Administração de Serviços: operações, estratégia e tecnologia da informação. 7. ed. Porto Alegre: McGraw-Hill, 2014.

JAMOVI. Sobre o jamovi. 2025. Disponível em: https://www.jamovi.org/about.html. Acesso em: 20 outubro 2025

JOHNSTON, R.; CLARK, G. Service Operations Management: Improving Service Delivery. 3. ed. Harlow: Pearson, 2008.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, v. 349, n. 6245, p. 255-260, 2015.

LEE, J.; BAGHERI, B.; KAO, H. A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, v. 3, p. 18-23, 2014.

LIU, D. *et al*. Random Forest regression evaluation model of regional flood disaster resilience based on the whale optimization algorithm. *Journal of Cleaner Production*, v. 250, art. 119468, 2020.

MINISTÉRIO DA INDÚSTRIA, COMÉRCIO E SERVIÇOS (Brasil). Mapa de Empresas: novembro de 2024. 2024. Disponível em: https://www.gov.br/mapa-de-empresas. Acesso em: 22 ago. 2025.

OSHIRO, T. M.; PEREZ, P. S.; BARANAUSKAS, J. A. How many trees in a random forest? In: INTERNATIONAL WORKSHOP ON MACHINE LEARNING AND DATA MINING IN PATTERN RECOGNITION, 2012, Berlin. *Proceedings...* Berlin: Springer, 2012. p. 154-168.

REZAIE-BALF, M. *et al*. Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: comparative assessment of a noise suppression hybridization approach. *Journal of Cleaner Production*, v. 271, art. 122576, 2020.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 4. ed. Harlow: Pearson, 2021.

SLACK, N.; BRANDON-JONES, A.; BURGESS, N. Operations Management. 8. ed. Harlow: Pearson, 2019.

TASAN, S. et al. Estimation of eggplant yield with machine learning methods using spectral vegetation indices. *Computers and Electronics in Agriculture*, v. 202, p. 107367, 2022. DOI: https://doi.org/10.1016/j.compag.2022.107367.

VAN KLOMPENBURG, T. et al. Artificial Intelligence in Service Operations: A Review and Research Agenda. *International Journal of Operations & Production Management*, v. 42, n. 5, p. 631-653, 2022.

Zhang, C.; Ma, Y. Ensemble Machine Learning: Methods and Applications. Springer, 2012.