

IA PARA PREDIÇÃO DE REFUGO: Avaliação de modelos na indústria de papel reciclado

ROBERTA VANESSA ARAGÃO FELIX DA SILVA
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)

TACIANA DE BARROS JERÔNIMO
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)

JOSÉ CLAUDIO VILAR GOMES JÚNIOR
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)

FERNANDO GOMES DE PAIVA JÚNIOR
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)

JARDIEL DE MOURA GOMES
UNIVERSIDADE FEDERAL DE PERNAMBUCO (UFPE)

Agradecimento à órgão de fomento:

Este trabalho foi realizado com apoio da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

IA PARA PREDIÇÃO DE REFUGO: Avaliação de modelos na indústria de papel reciclado

INTRODUÇÃO

A indústria de manufatura tem passado por mudanças estruturais marcadas pela incorporação de tecnologias voltadas à automação e à inteligência artificial (IA), à eficiência dos processos e à sustentabilidade ambiental (Cheng; Yu, 2021; Merlin; Coleman; Padgett, 2024; Ma; Yin; Li, 2024). Em específico, na indústria de papel e celulose, o Brasil se destaca, liderando as exportações mundiais de celulose e ocupando a segunda posição em produção global, atrás apenas dos Estados Unidos, com movimentação comercial de US\$ 12,7 bilhões em 2023 (IBÁ, 2024). No entanto, o segmento enfrenta desafios operacionais, especialmente no controle em tempo real de variáveis críticas, como gramatura e umidade, que impactam diretamente a qualidade do produto final e os custos industriais (He *et al.*, 2022; Charles *et al.*, 2024).

No caso do papel reciclado, a heterogeneidade da matéria-prima eleva a instabilidade do processo, dificultando a padronização da produção e contribuindo para uma taxa de refugo estimada em 1,7% (Devi; Priyanka; Sakthivel, 2023). Esse percentual, embora aparentemente baixo, representa perdas expressivas em ambientes de produção contínua e reforça a necessidade de soluções que antecipem falhas com base em variáveis monitoradas na linha de produção (Devi; Priyanka; Sakthivel, 2023).

A consolidação da Indústria 4.0 introduziu novas possibilidades para o tratamento automatizado de dados em tempo real (Frank *et al.*, 2019; Soori *et al.*, 2024; Piron *et al.*, 2025). A integração entre sistemas ciberfísicos, computação em nuvem e IA permite que decisões operacionais baseadas em análise preditiva, otimizando recursos e reduzindo desperdícios (Soori *et al.*, 2024; Nessari *et al.*, 2024; Piron *et al.*, 2025).

No setor papeleiro, o uso de algoritmos de aprendizado de máquina supervisionado tem ganhado espaço por sua capacidade de prever variáveis críticas antes da finalização dos lotes, contribuindo para a minimização de perdas e a racionalização dos processos (Soni *et al.*, 2022). Entretanto, a aplicação desses modelos ainda é limitada, e a literatura concentra-se em abordagens restritas ao monitoramento isolado de variáveis ou classificações binárias de qualidade (Schroth *et al.*, 2022; Keskar, 2021). Faltam propostas que integrem análise preditiva em tempo real, especialmente em ambientes onde a variabilidade da matéria-prima compromete a previsibilidade dos resultados.

Além das limitações no controle de variáveis físicas internas ao processo, há outro aspecto crítico relacionado à qualidade: a detecção de defeitos superficiais no produto final. A manufatura moderna exige mecanismos que assegurem não apenas a conformidade dimensional, mas também a integridade visual e estrutural da produção. Estudos indicam que a inspeção de qualidade pode representar até 20% do custo total da manufatura (Matt *et al.*, 2020), e que falhas afetam a confiabilidade do produto (Cheng; Yu, 2021). Métodos tradicionais de inspeção visual, realizados manualmente, são demorados, sujeitos a erros humanos e apresentam baixa escalabilidade frente à produção contínua (Lu *et al.*, 2022; Ma; Yin; Li, 2024).

Nesse contexto, tem-se intensificado o uso de métodos automatizados baseados em modelos tradicionais de aprendizado de máquina para detectar e classificar diferentes tipos de defeitos com maior precisão e eficiência (Matt *et al.*, 2020; Merlin; Coleman; Padgett, 2024).

Nesse sentido, técnicas baseadas em aprendizado profundo, como Redes Neurais Convolucionais (CNNs), têm sido empregadas para reconhecer padrões complexos em imagens industriais, permitindo a detecção automatizada de defeitos com maior precisão e rapidez (Han; Kamber; Pei, 2011; James *et al.*, 2021). A partir desse cenário, este estudo propõe avaliar a aplicabilidade de três algoritmos de classificação supervisionada: *Random Forest*, *Gradient*

Boosting e *MLPClassifier*, como ferramenta de predição da qualidade em bobinas de papel reciclado, com ênfase na redução do refugo.

Os modelos foram selecionados por sua capacidade de lidar com conjuntos de dados estruturados, variabilidade nas entradas e comportamento não linear. O *Random Forest* destaca-se pelo uso de múltiplas árvores de decisão que reduzem o risco de sobreajuste (Han; Kamber; Pei, 2011). O *Gradient Boosting* é apropriado para melhorar progressivamente a acurácia ao minimizar os erros residuais ao longo de iterações (James *et al.*, 2021). O *MLPClassifier*, por sua vez, aplica redes neurais multicamadas para aprender relações complexas entre variáveis como gramatura e umidade e os resultados dos testes físicos. A base de dados utilizada contempla 6.613 lotes de produção de uma planta industrial localizada em Pernambuco, com medições transversais de gramatura e umidade realizadas a cada 105 segundos.

O presente artigo, portanto, tem como objetivo avaliar o desempenho preditivo de três algoritmos supervisionados na classificação de bobinas de papel reciclado como conformes ou refugadas, com foco na capacidade de identificar corretamente a classe minoritária (refugo), com o objetivo de apoiar a tomada de decisão industrial.

FUNDAMENTAÇÃO TEÓRICA

IA na indústria 4.0

A Indústria 4.0 é um modelo de produção baseado na integração entre tecnologias digitais e processos industriais. A digitalização da manufatura facilita o monitoramento contínuo da produção, ampliando a capacidade de resposta a eventos operacionais (Lee *et al.*, 2018; Santos, 2021; Douimia *et al.*, 2025). O uso de sensores conectados e redes industriais possibilita a coleta de dados em tempo real (Souza; Silva, 2024).

A Indústria 4.0 favorece práticas produtivas mais rápidas e fundamentadas em análise de dados (Janiesch; Zschech; Heinrich, 2021; Douimia *et al.*, 2025) com fins de predição, análise e controle (Zhang *et al.*, 2023). Essa estrutura tecnológica permite adequações mais dinâmicas às exigências do mercado, sem interromper o fluxo operacional. Kagermann, Wahlster e Helbig (2013) destacam que a digitalização completa das operações industriais transforma a lógica produtiva, não apenas automatizando tarefas, mas introduzindo inteligência nos processos. Assim, a cadeia produtiva torna-se capaz de antecipar desvios, otimizar recursos e aumentar a disponibilidade de equipamentos com menor intervenção direta de operadores (Zhang *et al.*, 2023).

Sua função consiste em apoiar o processamento de grandes volumes de dados e permitir que sistemas computacionais identifiquem padrões relevantes para decisões operacionais (Zhang *et al.*, 2023). O aprendizado de máquina (*Machine Learning* – ML) e o aprendizado profundo (*Deep Learning* – DL) são duas das técnicas mais empregadas na produção. Ambas operam com base em algoritmos capazes de extrair relações a partir de variáveis contínuas e discretas (Rai *et al.*, 2021).

No setor de papel e celulose, por exemplo, técnicas de ML vêm sendo aplicadas para controlar parâmetros críticos como gramatura e umidade, que impactam diretamente na qualidade do produto final (Wan *et al.*, 2022). Liu *et al.* (2020) mostraram que o uso de *Extreme Learning Machines* (ELM) contribuiu para a melhora da acurácia preditiva em processos industriais. Houve um ganho de 27,5% na previsão da qualidade de efluentes, sugerindo que modelos de IA podem apoiar decisões com maior sensibilidade aos dados de processo, ajustando variáveis conforme padrões detectados automaticamente.

Outra tecnologia incorporada ao ambiente fabril é o gêmeo digital (*Digital Twin*), que permite simular cenários com base em réplicas virtuais de equipamentos e processos reais. Essa abordagem tem sido aplicada na indústria papeleira como suporte à tomada de decisão em

tempo real, integrando dados operacionais com modelos computacionais para ajustes preventivos (Zhang *et al.*, 2023).

A simulação contínua favorece o diagnóstico de anomalias sem a necessidade de interrupção da linha de produção, promovendo intervenções mais rápidas com base em condições detectadas virtualmente. Liu *et al.* (2020) apontam que a utilização de gêmeos digitais auxilia na análise da variabilidade dos processos, o que amplia o controle sobre parâmetros críticos de qualidade. Esse tipo de recurso depende da capacidade de integração entre sensores, algoritmos e infraestrutura de dados, o que reforça o papel da IA na consolidação de um sistema produtivo adaptativo e monitorado em tempo real.

A redução de refugo na cadeia de produção é uma das aplicações mais relevantes da IA em contextos industriais. He *et al.* (2022) analisaram o uso de sensores e algoritmos preditivos na indústria de papel reciclado e constataram redução de até 17,3% no refugo, após a implementação de modelos baseados em IA. A identificação precoce de anomalias em variáveis como gramatura e umidade evita que materiais não conformes avancem nas etapas produtivas.

O uso de modelos supervisionados de ML, treinados com dados históricos, permite que o sistema reconheça condições que antecedem falhas de qualidade. A cada ciclo, os modelos podem ser ajustados com novos dados, promovendo aprendizado contínuo. Wan *et al.* (2022) observaram que modelos treinados para padrões específicos apresentam maior assertividade na detecção de desvios sutis. Isso se torna relevante em processos de alta variabilidade, onde o controle manual apresenta limitações.

Essa capacidade de atualização dos modelos é essencial para manter a eficiência em ambientes industriais sujeitos a mudanças na matéria-prima ou no processo produtivo, como ocorre frequentemente na produção com papel reciclado.

Algoritmos de aprendizado de máquina

Os algoritmos de aprendizado de máquina podem ser classificados de acordo com a natureza do problema abordado e do conjunto de dados disponíveis. Além da distinção entre os tipos de aprendizado de máquina, é importante considerar fatores que afetam o desempenho dos modelos (Murphy, 2012; Alpaydin, 2020). Em problemas supervisionados, modelos são treinados com dados rotulados para aprender a mapear entradas (X) a saídas (Y), sendo aplicados em tarefas de classificação ou regressão (James *et al.*, 2021).

Já o aprendizado não supervisionado, opera sobre dados não rotulados, buscando descobrir padrões ocultos ou agrupamentos naturais, por meio de técnicas como *clustering*, regras de associação ou redução de dimensionalidade (Goodfellow; Bengio; Courville, 2016). O aprendizado supervisionado constitui uma das abordagens do campo do aprendizado de máquina, sendo empregado, por exemplo, em tarefas de classificação e regressão (James *et al.*, 2021).

Essa técnica fundamenta-se no treinamento de algoritmos a partir de um conjunto de dados rotulados. Assim, o modelo aprende a reconhecer padrões entre as variáveis de entrada e a saída correspondente, tornando-se capaz de realizar previsões sobre novos dados com base nesse histórico (James *et al.*, 2021). Géron (2019) defende que esse tipo de aprendizado é particularmente eficaz em contextos em que há disponibilidade de dados estruturados e com resultados previamente definidos, como nos ambientes industriais monitorados por sensores.

No contexto da manufatura inteligente, o aprendizado supervisionado tem sido aplicado para prever falhas em equipamentos, identificar produtos fora de especificação e otimizar parâmetros operacionais (Goodfellow; Bengio; Courville, 2016; He *et al.*, 2022). As variáveis coletadas durante o processo produtivo, como temperatura, pressão, gramatura e umidade, são utilizadas como insumos para os algoritmos, que, por sua vez, aprendem a associar tais variáveis a categorias específicas de saída, como "aprovado" ou "refugado" (James *et al.*, 2021). Essa

abordagem permite antecipar eventos que comprometeriam a eficiência produtiva, viabiliza ações corretivas mais rápidas e precisas.

Contudo, a aplicação do aprendizado supervisionado em dados industriais apresenta desafios (Merlin; Caleman; Padgett, 2024; Ma; Yin; Li, 2024). Um dos principais refere-se ao desequilíbrio de classes, situação em que há grande disparidade na quantidade de exemplos de cada categoria. Em processos industriais, é comum que a maioria dos lotes produzidos seja aprovada, enquanto os casos de refugo representam uma parcela minoritária. Isso pode levar os modelos a desenvolverem um viés em favor da classe majoritária, comprometendo a capacidade de detectar adequadamente os casos menos frequentes. Segundo Chawla *et al.* (2002), esse desequilíbrio reduz a efetividade de métricas tradicionais, como a acurácia, sendo necessário empregar indicadores como precisão, revocação e *F1-score* para avaliar o desempenho com maior robustez (James *et al.*, 2021; Grinsztajn; Oyallon; Varoquaux, 2022).

A precisão indica a proporção de acertos entre todas as previsões positivas realizadas pelo modelo, enquanto a revocação mede a capacidade do modelo em identificar corretamente todos os casos positivos existentes. Já o *F1-score* corresponde à média harmônica entre precisão e revocação, oferecendo uma medida equilibrada do desempenho, especialmente útil quando há assimetria entre as classes. Essas métricas são mais robustas do que a acurácia em cenários nos quais a classe de interesse (como produtos com defeito) representa uma minoria significativa (Grinsztajn; Oyallon; Varoquaux, 2022).

A variabilidade dos processos e dos ruídos nos dados dificultam a tarefa de classificação, exigindo algoritmos mais robustos e estratégias eficazes de pré-processamento. Por isso, é essencial escolher métodos que lidem com incertezas e façam a generalização com dados novos.

Algoritmos supervisionados de classificação

A construção de modelos também depende da qualidade e do equilíbrio dos dados: em conjuntos desbalanceados o algoritmo tende a privilegiar a classe majoritária, exigindo técnicas de reamostragem ou ajustes de custo (Lemaître *et al.*, 2017).

Outra questão refere-se à representação: estatísticas agregadas reduzem esforço computacional, mas perdem granularidade, o que pode ocultar variações relevantes (Géron, 2019). Por isso, recomenda-se utilizar dados brutos ou sequenciais, sobretudo em processos industriais onde pequenas oscilações sinalizam desvios significativos. Diante desse cenário, o presente estudo aplica três classificadores supervisionados, *Gradient Boosting*, *Random Forest* e *MLPClassifier*, cujos resultados se destacam na predição de variáveis críticas, como gramatura e umidade, em ambientes fabris (Grinsztajn *et al.*, 2022).

O *Gradient Boosting (GBoost)* combina aditivamente modelos fracos (pequenas árvores) em um processo iterativo de minimização de perda pelo gradiente (Friedman, 2001). A predição é dada por $F(x) = \sum(\beta_m h(x, a_m))$, onde cada função base $h(x, a_m)$ é ajustada na direção do gradiente negativo da perda $L(y, F(x))$, com relação à predição corrente. Para evitar *overfitting* e melhorar a capacidade de generalização, Friedman (2001) também propôs o *shrinkage*, no qual cada novo termo adicionado ao modelo é multiplicado por uma taxa de aprendizado $v \in (0, 1]$, controlando sua contribuição.

O *Random Forest* forma um conjunto de K árvores $h(x, \theta_k)$, cada qual treinada em amostras *bootstrap* e subconjuntos aleatórios de atributos; a classificação final resulta da moda das saídas individuais, $\hat{y} = \text{mode}\{h(x, \theta_1), h(x, \theta_2), \dots, h(x, \theta_k)\}$ (Breiman, 2001). A margem é definida por $mr(X, Y) = P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X, \theta) = j)$. Essa equação define a margem de confiança do classificador para a classe correta em relação à maior probabilidade dada a qualquer classe incorreta. Isso expressa o "grau de certeza" do modelo. A força do

classificador é a expectativa dessa margem: $s = E_{X,Y}[mr(X,Y)]$.

A expectativa da função de margem representa a força média da floresta. É um indicador do “quão boas são”, em média, as árvores para classificar corretamente os dados. A correlação média entre as árvores é representada por \underline{p} , e a taxa de erro generalizada (erro de generalização) é limitada superiormente por: $PE^* \leq \frac{p(1-s^2)}{s^2}$, em que PE^* é o erro esperado de generalização; s é a força média (*strength*) dos classificadores individuais; e \underline{p} é a correlação média entre os classificadores da floresta. Quanto maior s e menor \underline{p} , melhor o desempenho.

Por fim, o *MLPClassifier* utiliza uma rede *feed-forward* com ao menos três camadas; a transformação de entrada é $h = \phi(W^{(1)}x + b^{(1)})$, $\hat{y} = \psi(W^{(2)}h + b^{(2)})$. O treinamento ocorre por retropropagação, com otimização estocástica sobre a perda (Rumelhart *et al.*, 1986; Haykin, 1998). A generalização depende da arquitetura, taxa de aprendizado e regularização (dropout, L2, ajuste de hiperparâmetros). Apesar de exigir maior capacidade computacional, o MLP captura relações não lineares complexas, sendo aplicado à detecção de defeitos em bobinas de papel; em He *et al.* (2022), a rede reduziu falhas ao permitir ajustes automáticos de processo.

METODOLOGIA

Este estudo caracteriza-se como uma pesquisa quantitativa, aplicada e descritiva-comparativa. A abordagem quantitativa possibilita mensurar e analisar dados operacionais obtidos em ambiente industrial. O caráter aplicado justifica-se pela busca de soluções concretas para o problema do refugo de bobinas de papel reciclado, enquanto o caráter descritivo-comparativo se evidencia na avaliação do desempenho de três diferentes algoritmos de classificação supervisionada com base em métricas estatísticas (Gil, 2017).

O estudo foi desenvolvido a partir de um estudo de caso em uma única empresa, estratégia metodológica adequada quando se pretende realizar uma investigação profunda sobre um fenômeno contemporâneo inserido em um contexto real (Yin, 2015). A pesquisa foi realizada em uma empresa localizada no estado de Pernambuco, considerada a maior recicladora de papéis do Brasil, com capacidade anual de produção de aproximadamente 400 mil toneladas. A base de dados contemplou 6.613 lotes de bobinas produzidas ao longo de 11 meses consecutivos (janeiro a novembro de 2024). Cada lote foi composto por 254 medições transversais de gramatura e umidade, realizadas automaticamente por um scanner industrial, com intervalo de 105 segundos entre as coletas.

Os dados brutos foram inicialmente tratados no *software Microsoft Excel®*, sendo removidos registros incompletos, inconsistentes ou com presença de outliers, conforme o método do intervalo interquartil (IQR), recomendado por Han, Pei e Kamber (2011) para limpeza de dados em ambientes industriais. Lotes com número inferior a 250 medições foram descartados com base na recomendação de que conjuntos de dados com número reduzido de observações podem gerar estimativas estatísticas instáveis e não representativas da variabilidade real do processo (Han; Pei; Kamber, 2011). Os dados foram organizados em estrutura de colunas, viabilizando sua importação para os modelos de aprendizado de máquina.

Como variável de saída para rotular os dados como “aprovado” ou “refugado”, foram utilizados os resultados de dois testes físicos: *Ring Crush Test* (RCT) e teste de ruptura (Mullen). O RCT mede a resistência à compressão anular, sendo crucial para avaliar a capacidade do papel de suportar cargas verticais, aspecto essencial na produção de embalagens cilíndricas. Já o teste Mullen avalia a resistência à ruptura sob pressão, sendo indicativo da integridade estrutural do papel. Ambos os testes são amplamente reconhecidos como critérios objetivos de qualidade na indústria papeleira, impactando diretamente a aprovação ou o refugo dos lotes (Srinivasan *et al.*, 2024).

Dada a grande quantidade de dados por lote e as limitações computacionais, adotou-se a estratégia de representação estatística agregada das variáveis. Para cada lote, foram extraídas estatísticas descritivas como média, mediana, desvio padrão, quartis, valores mínimo e máximo de gramatura e umidade. Essa abordagem é indicada por Zhang *et al.* (2023) como eficaz para transformar séries temporais industriais em vetores de entrada para modelos supervisionados.

A etapa seguinte consistiu na análise exploratória dos dados (AED) para compreender a distribuição das variáveis e verificar a presença de desequilíbrios entre as classes “aprovado” e “refugado”. Em seguida, os dados foram submetidos à modelagem preditiva com os algoritmos *Gradient Boosting*, *Random Forest* e *MLPClassifier*, implementados na linguagem *Python*, por meio das bibliotecas *scikit-learn* e *pandas* (Pedregosa *et al.*, 2011).

Cada parâmetro foi selecionado com o objetivo de equilibrar desempenho, estabilidade e capacidade de generalização, especialmente diante de um conjunto de dados desbalanceado e com variáveis estatísticas agregadas. Essa sistematização permitiu padronizar os critérios de comparação entre os algoritmos, assegurando maior robustez à análise preditiva conduzida neste estudo.

Para a avaliação dos modelos, foram utilizados os seguintes indicadores de desempenho: acurácia, precisão, revocação (*recall*) e *F1-score*, conforme recomendado por James *et al.* (2021), especialmente em cenários com classes desbalanceadas. A divisão dos dados em conjunto de treinamento e teste seguiu a proporção de 80%-20%, com validação cruzada *K-Fold* ($k=5$), prática considerada robusta para avaliação de modelos supervisionados.

Adicionalmente, foram analisadas as matrizes de confusão, a fim de compreender a distribuição dos acertos e erros, contribuindo para a avaliação prática da aplicabilidade dos modelos no contexto industrial. Para melhor compreensão das decisões metodológicas no ajuste dos modelos supervisionados, o Quadro 1 apresenta os hiperparâmetros definidos, os modelos em que foram aplicados e a justificativa técnica para cada configuração adotada.

Quadro 1 – Parâmetros utilizados nos modelos

Parâmetro	Significado e função
n_estimators	Número de árvores que o modelo irá construir. Quanto mais árvores, maior a robustez e capacidade preditiva, até certo ponto. Usado no <i>Random Forest</i> e no <i>Gradient Boosting</i> .
max_depth	Profundidade máxima de cada árvore. Se for muito alta, o modelo pode memorizar os dados (overfitting). Em <i>Random Forest</i> foi deixado como None (crescimento livre), e em GB limitado a 3 (mais controle).
learning_rate	Apenas no <i>Gradient Boosting</i> . Controla o quanto cada nova árvore contribui para a previsão final. Valores mais baixos tornam o modelo mais conservador e evitam sobreajuste.
random_state	Define uma "semente" para o gerador de números aleatórios. Isso garante que os resultados sejam os mesmos toda vez que o código for executado.
hidden_layer_sizes	Define o número de camadas ocultas e neurônios em cada uma para o <i>MLPClassifier</i> .
activation	Função usada nos neurônios da rede. 'relu' (Unidade Linear Retificada) é a mais comum para modelos com muitas entradas contínuas, ajuda a aprender padrões complexos.
solver	Algoritmo de otimização da rede neural. 'adam' é eficiente para bases de dados pequenas ou médias e converge rapidamente.
max_iter	Número máximo de iterações que o modelo pode executar para tentar convergir (parar de errar). Usado para evitar loops infinitos ou travamentos no treinamento.

Fonte: dados da pesquisa (2025).

Os hiperparâmetros dos modelos foram definidos com base nas recomendações da biblioteca *Scikit-learn* e ajustados empiricamente por meio de validações sucessivas. No *Gradient Boosting*, utilizaram-se: n_estimators=100, learning_rate=0.1, max_depth=3 e random_state=42, buscando equilíbrio entre desempenho e prevenção de sobreajuste. Para o *Random Forests*, os parâmetros foram n_estimators=100, max_depth=None e random_state=42, permitindo que as árvores crescessem livremente. Já no *MLPClassifier*,

adotaram-se: `hidden_layer_sizes=(50,50)`, `activation='relu'`, `solver='adam'`, `max_iter=500` e `random_state=42`, otimizando a eficiência e estabilidade do modelo.

ANÁLISE DOS RESULTADOS

Produção de papel reciclado

A produção de papel reciclado inicia-se na central de aparas, onde os resíduos de papel são triados, classificados e preparados para remoção de contaminantes conforme os critérios da NBR 15483:2009. A matéria-prima pode ser de pré ou pós-consumo e é desagregada mecanicamente em água no equipamento *intensa pulper*, passando por processos de limpeza e depuração para remoção de impurezas físicas, como metais e plásticos (Alvarenga; Reis, 2018; Campos, 2011).

Após a preparação da massa, segue-se a etapa de refinação, cujo objetivo é modificar mecanicamente as fibras para melhorar sua capacidade de entrelaçamento, aumentando a resistência, porosidade e absorção do papel (Campos, 2011; Alvarenga; Reis, 2018). Na máquina de papel, a produção ocorre em etapas sequenciais: formação da folha sobre uma tela drenante; prensagem para remoção da água e aumento da densidade; pré-secagem por cilindros aquecidos; aplicação de amido (quando aplicável) para reforço superficial; pós-secagem, que reduz a umidade final do papel para 7–9%; e, por fim, o enrolamento e rebobinamento, que formam as bobinas finais conforme as especificações de uso (Campos, 2011).

O processo exige controle rigoroso de variáveis como gramatura e umidade, que afetam diretamente os resultados dos testes laboratoriais de qualidade, como o *Ring Crush Test* (RCT) e o Teste Mullen, verificam a resistência à compressão e à ruptura, respectivamente (He *et al.*, 2022; Devi *et al.*, 2023).

Parâmetros de qualidade no papel reciclado

A qualidade do papel reciclado depende de parâmetros técnicos como: gramatura, umidade, *Ring Crush Test* (RCT) e Mullen, os quais foram utilizados como variáveis na modelagem preditiva deste estudo. A gramatura, expressa em g/m^2 , influencia a rigidez, resistência e impressão do papel, sendo essencial manter sua uniformidade (Cardoso *et al.*, 2012). Já a umidade, calculada pela diferença entre o peso inicial e seco da amostra, afeta a estabilidade e resistência mecânica do produto. Flutuações excessivas comprometem o desempenho e podem gerar falhas (Sousa *et al.*, 2016).

O RCT avalia a resistência do papel à compressão anular, importante para aplicações em embalagens. O ensaio, realizado conforme a ABNT NBR NM-ISO 12192:2013, utiliza uma amostra moldada em anel e submetida à compressão até sua ruptura (Borba, 2012; Devi *et al.*, 2023). Já o teste de Mullen, normatizado pela ABNT NBR 7755:1983, mede a resistência à ruptura por pressão hidrostática, verifica a robustez do papel frente a impactos (Sousa *et al.*, 2016; Devi *et al.*, 2023). As Figuras 1 e 2 ilustram, respectivamente, os equipamentos utilizados nos testes de RCT e Mullen.

Figura 1 - Máquina de teste de compressão (RCT)



Fonte: dados da pesquisa (2025).

Figura 2 - Mullen Tester



Fonte: dados da pesquisa (2025).

Desempenho geral dos modelos

Após a realização das análises preditivas, a classificação final dos lotes de papel reciclado pode ser: “Classificado” (conforme) ou “Desclassificado” (refugo). Os algoritmos testados foram: *Gradient Boosting*, *Random Forest* e *MLPClassifier*. A Tabela 1 apresenta as métricas de desempenho obtidas pelos modelos, considerando acurácia, precisão e revocação (*recall*). Observa-se que os três algoritmos apresentaram desempenho elevado, com acurácia superior a 97%, demonstrando forte capacidade preditiva para o problema proposto.

Tabela 1 – Resultados das Métricas dos Modelos de Classificação Previsão da Classificação Final dos Lotes

Modelo	Acurácia	Precisão	Revocação
<i>Gradient Boosting</i>	0,9743	0,9772	0,9969
<i>Random Forest</i>	0,9751	0,9758	0,9992
<i>MLPClassifier</i>	0,9766	0,9773	0,9992

Fonte: Dados da pesquisa (2025).

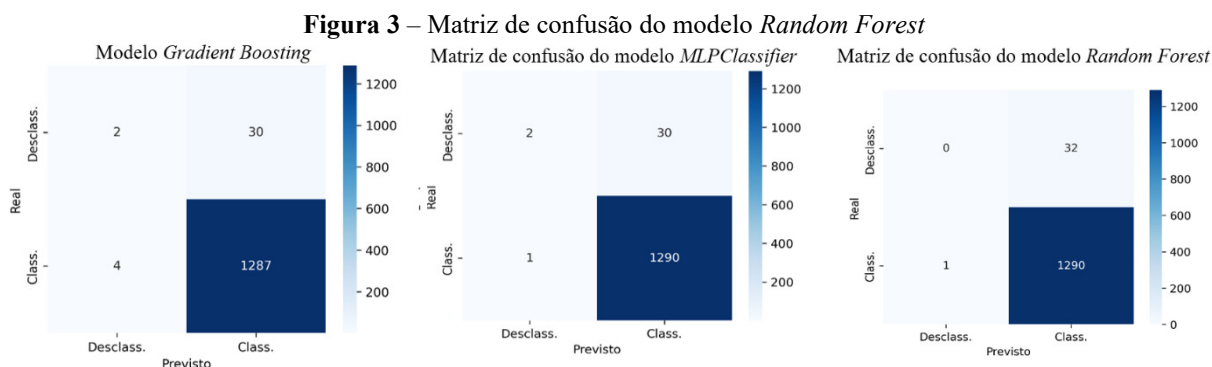
Entre os modelos avaliados, o *MLPClassifier* obteve ligeira vantagem nos três indicadores: acurácia de 97,66%, precisão de 97,73% e revocação de 99,92%. Em seguida, o *Gradient Boosting* alcançou precisão de 97,72% e revocação de 99,69%, enquanto o *Random Forest* obteve os mesmos 99,92% de revocação do MLP, porém com menor precisão (97,58%).

Apesar da leve variação numérica entre os modelos, todos se mostraram eficazes na tarefa de classificação dos lotes. Como a diferença de desempenho é estatisticamente pequena, a escolha do modelo mais adequado pode considerar critérios adicionais, como tempo de treinamento, interpretabilidade e facilidade de atualização no ambiente fabril. A alta taxa de revocação para a classe minoritária reforça a confiabilidade do processo, ao impedir que produtos com desvios avancem para as etapas subsequentes da cadeia.

Análise da matriz de confusão

As matrizes de confusão, apresentadas nas Figuras 3 ilustra os padrões de erro dos modelos testados. Todos mantiveram elevada taxa de acerto para a classe majoritária (bobinas conformes), com destaque para o *MLPClassifier*, que apresentou apenas 1 falso negativo e 30 falsos positivos. O *Gradient Boosting* teve desempenho semelhante, com 4 falsos negativos e

30 falsos positivos, enquanto o *Random Forest* registrou 1 falso negativo, mas nenhum verdadeiro negativo, resultando em 32 falsos positivos.



Fonte: Dados da pesquisa (2025).

Esses resultados evidenciam um desafio importante: a dificuldade dos modelos em identificar corretamente os lotes refugados. A baixa quantidade de verdadeiros negativos, apenas dois nos modelos *MLPClassifier* e *Gradient Boosting*, e nenhum no *Random Forest*, reflete a limitação na detecção de padrões associados à não conformidade, aspecto crítico para a triagem automatizada de bobinas com risco de refugo.

Do ponto de vista operacional, tanto os falsos negativos (bobinas refugáveis classificadas como conformes) quanto os falsos positivos (bobinas conformes classificadas como refugadas) acarretam impactos financeiros significativos. Enquanto os primeiros comprometem a qualidade final e podem gerar retrabalho ou insatisfação do cliente, os segundos implicam desperdício de matéria-prima e queda na eficiência produtiva. A Tabela 2 resume esses erros de classificação, reforçando a importância de considerar a sensibilidade à classe minoritária como critério estratégico na escolha do classificador mais adequado.

Tabela 2 - Matriz de Confusão Simplificada dos Modelos de Classificação

Modelo	Falsos Positivos (FP)	Falsos Negativos (FN)	Verdadeiros Positivos (TP)	Verdadeiros Negativos (TN)
<i>Gradient Boosting</i>	27	4	1050	2
<i>Random Forest</i>	32	0	1054	0
<i>MLPClassifier</i>	30	1	1053	2

Fonte: Dados da pesquisa (2025).

Diante disso, a taxa de erro dos modelos deve ser analisada em conjunto com os custos industriais associados. Mesmo pequenas quantidades de falsos positivos podem resultar em perdas financeiras acumuladas ao longo do tempo, o que justifica a priorização de classificadores com maior equilíbrio entre precisão e revocação. Nesse sentido, o *MLPClassifier* demonstrou desempenho superior, conciliando alta sensibilidade com baixos índices de erro, além de maior estabilidade na detecção de padrões críticos.

Os achados do estudo confirmam a efetividade dos modelos supervisionados, especialmente do *MLPClassifier*, na classificação de bobinas recicladas quanto à conformidade. Esse resultado dialoga com pesquisas anteriores, como a de Jauhar *et al.* (2024), que destacam a importância da estabilidade preditiva e da detecção precisa de padrões operacionais no setor papelero. Ainda assim, persistem limitações técnicas, como o desbalanceamento da base de dados e a utilização de estatísticas agregadas em vez de séries temporais brutas — fatores que reduzem a capacidade de captura de variações sutis. Estudos como os de He *et al.* (2022) e Devi *et al.* (2023) indicam que dados contínuos oriundos de sensores inteligentes ampliam significativamente a acurácia preditiva.

Observa-se um claro *trade-off* entre profundidade analítica e viabilidade computacional. O uso de estatísticas descritivas, como média e desvio padrão, viabilizou a modelagem com recursos limitados e o tratamento de grandes volumes de dados, mas implicou perda de granularidade, o que pode ter comprometido a detecção de variações pontuais relevantes, especialmente em casos de não conformidade.

Ainda assim, tais medidas mostraram-se eficazes na identificação de padrões relacionados à aprovação e ao refugo. Lotes com menor variabilidade interna apresentaram maior probabilidade de aprovação, indicando maior estabilidade do processo. No entanto, o uso exclusivo dessas estatísticas agregadas revelou limitações gerenciais, sobretudo diante do desbalanceamento da base, que dificulta a detecção de anomalias raras, como os lotes refugados.

Dessa forma, destaca-se a necessidade de complementar essas medidas com dados sequenciais, bem como de empregar técnicas específicas de reamostragem ou penalização de classes para aprimorar a capacidade preditiva em contextos industriais. Apesar dessas limitações, os modelos aplicados, especialmente o *MLPClassifier*, demonstraram potencial para apoiar decisões preventivas no controle de qualidade, com impacto direto na redução de não conformidades. Sua integração a sistemas de apoio à decisão pode racionalizar o uso de insumos, minimizar perdas e padronizar critérios técnicos. Com a coleta contínua de dados e ajustes incrementais, tais ferramentas tendem a evoluir para sistemas adaptativos, mais alinhados à lógica da Indústria 4.0.

Por fim, o Quadro 2, abaixo, sintetiza as principais características dos três algoritmos avaliados neste estudo no que tange a comparação quanto ao tipo de aprendizado, sensibilidade à revocação, robustez e adequação a diferentes contextos produtivos:

Quadro 2 - Comparativo dos algoritmos aplicados à decisão industrial

Algoritmo	Aprendizado	Análise	Precisão e Revocação	Desvantagens	Vantagens
<i>MLPClassifier</i>	Utilizado em tarefas de aprendizado supervisionado. Pode lidar com dados complexos, como imagens e textos.	Voltado para análise de desempenho global médio. Útil com dados altamente não lineares.	Depende fortemente dos hiperparâmetros e da qualidade dos dados de treinamento.	Requer ajustes extensivos e pode ser menos interpretável.	Capaz de capturar padrões complexos; indicado para dados não estruturados.
<i>Random Forest</i>	Aplica múltiplas árvores de decisão para reduzir erros individuais. Útil para classificação e regressão.	Focado na análise de dispersão. Indicado para grandes volumes de dados.	Alta sensibilidade à classe minoritária, ideal para bases desbalanceadas.	Não aprende sequencialmente; pode ser menos preciso que algoritmos mais modernos.	Robusto contra <i>overfitting</i> ; bom desempenho com grandes volumes de dados.
<i>Gradient Boosting</i>	Combina modelos fracos de forma sequencial para formar um modelo forte.	Otimiza erros residuais iterativamente. Indicado para padrões complexos e não lineares.	A revocação pode ser comprometida se houver desequilíbrio nas classes.	Sensível a hiperparâmetros e mais lento no treinamento.	Alta precisão; aprende com os erros anteriores; excelente para dados tabulares.

Fonte: Dados da pesquisa (2025).

Assim, a escolha dos algoritmos a serem utilizados deve considerar não apenas o desempenho estatístico, mas também a aderência às particularidades do processo produtivo e aos objetivos gerenciais. Cada modelo oferece vantagens específicas que podem ser mais ou menos relevantes conforme o contexto da aplicação, reforçando a importância de decisões técnicas alinhadas às necessidades operacionais da indústria.

CONCLUSÃO / CONTRIBUIÇÃO

A presente pesquisa evidenciou o potencial da aplicação de modelos de IA, especialmente os baseados em aprendizado supervisionado. Ao testar diferentes algoritmos com variáveis estatísticas agregadas de umidade e gramatura, o estudo demonstrou que o classificador MLP apresentou o melhor desempenho preditivo na identificação de bobinas com maior probabilidade de serem refugadas.

Essa abordagem se mostra particularmente relevante diante das limitações práticas da indústria, que frequentemente não dispõe de recursos para medições contínuas ou de testes físicos em tempo real. Ao utilizar dados de sensores já existentes na planta, como o scanner de gramatura e umidade, os modelos desenvolvidos contribuem para uma triagem automatizada dos lotes, antecipando falhas de qualidade e evitando o avanço de materiais fora de especificação para as etapas subsequentes do processo produtivo.

A ênfase na classe minoritária (refugo) ao longo da avaliação dos modelos também contribui metodologicamente para estudos que lidam com dados desbalanceados, ao reconhecer a importância de minimizar erros de classificação justamente nos casos mais críticos para o desempenho fabril.

Em termos de contribuições para a indústria e para a literatura, este estudo destaca-se por propor uma metodologia replicável em outras plantas industriais que operam com processos contínuos e grande volume de dados. A integração de modelos de classificação ao fluxo produtivo representa um passo em direção à Indústria 4.0, especialmente para empresas que desejam evoluir do monitoramento reativo para a prevenção inteligente de defeitos. A replicabilidade da abordagem, aliada à sua capacidade de gerar ganhos em produtividade, qualidade e sustentabilidade, consolida o papel da IA como aliada estratégica na modernização dos processos industriais.

Para pesquisas futuras, recomenda-se expandir a base de dados com registros contínuos, incorporar variáveis financeiras que permitam mensurar o impacto econômico dos erros de classificação, e realizar comparações entre modelos classificatórios e algoritmos de regressão, especialmente em casos em que os desfechos possam ser representados em escalas contínuas.

Também se destaca o potencial de integrar essas soluções a sistemas industriais de controle, como plataformas MES (*Manufacturing Execution Systems*), favorecendo a retroalimentação automática dos modelos e a implementação de estratégias de aprendizado contínuo alinhadas aos princípios da Indústria 4.0.

REFERÊNCIAS

ALPAYDIN, Ethem. **Introduction to Machine Learning**. 4. ed. Cambridge: MIT Press, 2020.

ALVARENGA, Davi Lopes; REIS, Rodrigo Eustáquio dos. **Análise do processo de fabricação do papel utilizando matéria-prima reciclada**. 2018. 34 f. Monografia (Graduação em Engenharia Química) – Universidade Uberaba, Uberaba, 2018.

Disponível em:

<<https://dspace.uniube.br:8443/bitstream/123456789/1092/1/ANALISE%20DO%20PROCESSO%20DE%20FABRICA%C3%87%C3%83O%20DO%20PAPEL%20UTILIZ>

ANDO%20MAT%C3%89RIA%20PRIMA%20RECICLADA%20-%20revisado.pdf>.

BREIMAN, Leo. **Random forests**. Machine Learning, v. 45, n. 1, p. 5–32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>.

CAMPOS, E. S. **Curso básico de fabricação de celulose e papel**. Eucalyptus Online Book & Newsletter, 2011.

CARDOSO, Marco Túlio et al. **Propriedades físicas e mecânicas de papéis reciclados utilizados para fabricação de tubetes**. Ciência Florestal, v. 22, n. 2, p. 403-411, 2012.

CHARLES, Molly et al. **The role of the pulp and paper industry in achieving net zero U.S. CO₂ emissions in 2050**. Energy and Climate Change, v. 5, p. 100160, 2024. <https://doi.org/10.1016/j.egycc.2024.100160>.

CHAWLA, Nitesh V.; BOWYER, Kevin W.; HALL, Lawrence O.; KEGELMEYER, William P. **SMOTE: Synthetic Minority Over-sampling Technique**. Journal of Artificial Intelligence Research, v. 16, p. 321–357, 2002. DOI: <https://doi.org/10.1613/jair.953>.

CHENG, X.; YU, J. **RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection**. IEEE Transactions on Instrumentation and Measurement, v. 70, p. 1–11, 2021. DOI: <https://doi.org/10.1109/TIM.2020.3040485>.

DEVI, T. K.; PRIYANKA, E. B.; SAKTHIVEL, P. **Paper quality enhancement and model prediction using machine learning techniques**. Results in Engineering, v. 17, 2023. Art. 100950. DOI: 10.1016/j.rineng.2023.100950.

DOUIMIA, Soufiane et al. **Machine learning and deep learning applications in the automotive manufacturing industry: A systematic literature review and industry insights**. Robotics and Computer-Integrated Manufacturing, [S. l.], v. 84, 2025. DOI: <https://doi.org/10.1016/j.rcim.2025.103034>.

FRANK, A. G.; DALENOGARE, L. S.; AYALA, N. F. **Industry 4.0 technologies: implementation patterns in manufacturing companies**. International Journal of Production Economics, v. 210, p. 15–26, 2019. DOI: <https://doi.org/10.1016/j.ijpe.2019.01.004>.

FRIEDMAN, Jerome. **Greedy Function Approximation: A Gradient Boosting Machine**. Annals of Statistics, v. 29, n. 5, p. 1189-1232, 2001.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2017.

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

GRINSZTAJN, Léo; OYALLON, Edouard; VAROQUAUX, Gaël. **Why do tree-based models still outperform deep learning on tabular data?**. Journal of Machine Learning Research, v. 23, p. 1-27, 2022.

LEMAÎTRE, Guillaume; NOGUEIRA, Fernando; ARIDAS, Christos K. **Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning**.

Journal of Machine Learning Research, v. 18, n. 17, p. 1-5, 2017. Disponível em: <https://www.jmlr.org/papers/volume18/16-365/16-365.pdf>.

LEE, J.; DAVARI, H.; SINGH, J.; PANDHARE, V. **Industrial artificial intelligence for industry 4.0-based manufacturing systems**. Manufacturing Letters, [S. l.], v. 18, p. 20–23, 2018. DOI: <http://dx.doi.org/10.1016/j.mfglet.2018.09.002>.

LIU, Hongbin; ZHANG, Yuchen; ZHANG, Hao. **Prediction of effluent quality in papermaking wastewater treatment processes using dynamic kernel-based extreme learning machine**. Process Biochemistry, v. 97, p. 72–79, 2020. DOI: <https://doi.org/10.1016/j.procbio.2020.06.020>.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data mining: concepts and techniques**. 3rd ed. San Francisco: Morgan Kaufmann, 2011.

HAYKIN, Simon. **Neural Networks: A Comprehensive Foundation**. 2. ed. Upper Saddle River: Prentice Hall, 1998.

HE, Z. et al. **Data-driven soft sensors of papermaking process and its application to cleaner production with multi-objective optimization**. Journal of Cleaner Production, v. 372, 2022. Art. 133803. DOI: 10.1016/j.jclepro.2022.133803.

HEO, Chang Hoe; CHO, Hyunjun; YEO, Yeong-Koo. **Dynamic modeling of paper drying processes**. Korean Journal of Chemical Engineering, v. 28, n. 8, p. 1651–1657, 2011.

INDÚSTRIA BRASILEIRA DE ÁRVORES – IBÁ. **Relatório Anual 2024**. São Paulo: IBÁ, 2024. Disponível em: <https://iba.org/publicacoes/relatorios/>.

JAMES, Gareth et al. **An Introduction to Statistical Learning: with Applications in R**. 2. ed. New York: Springer, 2021.

JAUHAR, S. K. et al. **A deep learning-based approach for performance assessment and prediction: a case study of pulp and paper industries**. Annals of Operations Research, v. 332, p. 405–431, 2024. DOI: 10.1007/s10479-022-04528-3.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. **Machine learning and deep learning**. Electronic Markets, [S. l.], v. 31, n. 3, p. 685–695, 2021. DOI: <https://doi.org/10.1007/s12525-021-00475-2>.

KAGERMANN, H.; WAHLSTER, W.; HELBIG, J. **Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry**. Final report of the Industrie 4.0 Working Group. Frankfurt: National Academy of Science and Engineering, 2013. Disponível em: <https://www.din.org/en/innovation-and-research/industrie-4-0>.

KESKAR, Ankush. **AI and Machine Learning-Driven Manufacturing: Pioneering Best Practices for Intelligent, Scalable, and Sustainable Industrial Operations**. International Journal of All Research Education and Scientific Methods, v. 9, n. 4, p. 3038-3041, 2021. Disponível em: <https://www.researchgate.net/publication/388189976>.

LU, P.; ZHAO, T. S.; WANG, J. et al. **Review on damage identification and health monitoring of steel structures based on computer vision**. *Industrial Construction*, v. 52, n. 10, p. 22–27, 2022. DOI: <https://doi.org/10.13204/j.gyjzg22071401>.

MA, Yuxin; YIN, Jiaying; HUANG, Feng; LI, Qipeng. **Surface defect inspection of industrial products with object detection deep networks: a systematic review**. *Artificial Intelligence Review*, v. 57, p. 333–390, 2024. DOI: <https://doi.org/10.1007/s10462-024-10956-3>.

MATT, D. T.; MODRÁK, V.; ZSIFKOVITS, H. **Industry 4.0 for SMEs: Challenges, opportunities and requirements**. Cham: Springer, 2020. DOI: <https://doi.org/10.1007/978-3-030-25425-4>.

MERLIN, Simone; CALEMAN, Silvia Morais; PADGETT, Marilene. **Technological innovation and competitiveness in the pulp and paper industry: An integrated analysis**. *Technological Forecasting and Social Change*, v. 196, 2024.

MURPHY, Kevin P. **Machine Learning: A Probabilistic Perspective**. Cambridge: MIT Press, 2012.

NESSARI, Saman et al. **A data-driven decision-making approach for evaluating the projects according to resilience, circular economy and industry 4.0 dimension**. *Engineering Applications of Artificial Intelligence*, v. 134, p. 108608, 2024. DOI: <https://doi.org/10.1016/j.engappai.2024.108608>.

PIRON, Mirco; BOVO, Enrico; LUCCHETTA, Giovanni; MANZARDO, Alessandro. **Soft-sensors to drive manufacturing toward clean production: LCA based on Digital Twin**. *Journal of Cleaner Production*, v. 498, 2025. DOI: <https://doi.org/10.1016/j.jclepro.2025.145192>.

RAI, R.; TIWARI, M. K.; IVANOV, D.; DOLGUI, A. **Machine learning in manufacturing and industry 4.0 applications**. *International Journal of Production Research*, [S. l.], v. 59, n. 16, p. 4773–4778, 2021. DOI: <http://dx.doi.org/10.1080/00207543.2021.1956675>.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. **Learning representations by back-propagating errors**. *Nature*, v. 323, n. 6088, p. 533–536, 1986. DOI: <https://doi.org/10.1038/323533a0>.

SANTOS, Frederico de Oliveira. **Inovação, tecnologia e o uso potencial do aprendizado de máquina na fabricação inteligente para detecção de anomalias nos processos de produção**. 2021. 158 f. Dissertação (Mestrado em Desenvolvimento e Sociedade) – Universidade do Alto Vale do Rio do Peixe – Uniarp, Caçador, 2021.

SCHROTH, Moritz et al. **Optimization paper production through digitalization by developing an assistance system for machine operators including quality forecast: a concept**. Preprint, 2022. DOI: <https://arxiv.org/abs/2206.11581>.

SONI, Gunjan et al. **A decision-making framework for Industry 4.0 technology implementation: The case of FinTech and sustainable supply chain finance for SMEs**. *Technological Forecasting and Social Change*, v. 180, p. 121686, 2022. DOI: <https://doi.org/10.1016/j.techfore.2022.121686>.

SOORI, Mohsen et al. **AI-Based Decision Support Systems in Industry 4.0: A Review**. Journal of Economy and Technology, 2024. DOI: <https://doi.org/10.1016/j.ject.2024.08>.

SOUSA, Derlicio Carlos Goes; MATOS, Leandro Lisboa; ARAUJO, Myllane Kelry Sa; LIMA, Elon Vieira. **A importância da reciclagem do papel na melhoria da qualidade do meio ambiente**. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO (ENEGEP), 36., 2016, João Pessoa. Anais [...]. João Pessoa: ABEPRO, 2016.

SOUZA, Alan Christian Almeida; SILVA, Paula Renatha Nunes da. **Aplicação da inteligência artificial na engenharia de confiabilidade e manutenção preditiva: um estudo de caso na indústria de mineração**. Revista Ibero-Americana de Humanidades, Ciências e Educação, São Paulo, v. 10, n. 10, p. 3646-3658, out. 2024. ISSN 2675-3375. DOI: <https://doi.org/10.51891/rease.v10i10.16252>.

SRINIVASAN, S.; MURUGAN, S.; SWAGTHY, R.; OLIVEIRA, C.; SRIVIDHYA, V.; MUTHULEKSHMI, M. **Data-driven recycling transformation to enhance the efficiency of paper and cardboard bins using IoT and Random Forest**. In: AMATHE 2024: International Conference on Advances in Modern Technology for Health Sciences and Engineering, Anais [...], 2024. DOI: 10.1109/AMATHE61652.2024.10582250. Disponível em: <https://ieeexplore.ieee.org/document/10582250>.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 5. ed. Porto Alegre: Bookman, 2015.

WAN, Xin et al. **Water quality prediction model using Gaussian process regression based on deep learning for carbon neutrality in papermaking wastewater treatment system**. Environmental Research, v. 211, p. 112942, 2022. DOI: <https://doi.org/10.1016/j.envres.2022.112942>.

ZHANG, Jiwei et al. **An intelligent digital twin system for paper manufacturing in the paper industry**. Expert Systems with Applications, v. 230, p. 120614, 2023. DOI: <https://doi.org/10.1016/j.eswa.2023.120614>.