



Universidade Federal de Ouro Preto
Escola de Minas
Engenharia de Controle e Automação



Filipe Rodrigues Araújo

APRENDIZADO DE MÁQUINA NÃO SUPERVISIONADO PARA A DETECÇÃO DE ANOMALIAS APLICADA À MANUTENÇÃO PREDITIVA DE BOMBAS DE POLPA

Ouro Preto
2026

Filipe Rodrigues Araújo

**APRENDIZADO DE MÁQUINA NÃO
SUPERVISIONADO PARA A DETECÇÃO DE
ANOMALIAS APLICADA À MANUTENÇÃO
PREDITIVA DE BOMBAS DE POLPA**

Trabalho apresentado ao Colegiado do Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheiro(a) de Controle e Automação.

Orientador: Profa. Dra. Adrielle C. Santana

Ouro Preto
2026



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA CONTROLE E
AUTOMACAO



FOLHA DE APROVAÇÃO

Filipe Rodrigues Araújo

Aprendizado de Máquina não Supervisionado para a Detecção de Anomalias Aplicada a Manutenção Preditiva de Bombas de Polpa

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Controle e Automação

Aprovada em 20 de julho de 2026

Membros da banca

Dra. Adrielle de Carvalho Santana - Orientadora - Universidade Federal de Ouro Preto
Dra. Andrea Gomes Campos - Convidada - Universidade Federal de Ouro Preto
Me. Bruno Moreira da Silva - Convidado - Instituto Tecnológico Vale

Adrielle de Carvalho Santana, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 20/07/2026



Documento assinado eletronicamente por **Adrielle de Carvalho Santana, PROFESSOR DE MAGISTERIO SUPERIOR**, em 20/07/2026, às 11:57, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1141797** e o código CRC **53F4F6E0**.

Resumo

Este trabalho apresenta o desenvolvimento de um pipeline preditivo baseado em aprendizado de máquina não supervisionado para identificar padrões anômalos e antecipar falhas operacionais em sistemas de bombeamento de polpa mineral. Na rotina da mineração, quebras inesperadas desses ativos geram severos impactos financeiros e riscos operacionais, como o entupimento de tubulações. Para solucionar esse problema, estruturou-se um pipeline de dados em Python utilizando séries temporais reais de variáveis elétricas e hidráulicas de uma bomba industrial. Por meio da engenharia de atributos e filtragem estatística, o algoritmo *Isolation Forest* foi calibrado com uma taxa de contaminação de 1,0%, superando a abordagem tradicional do *Local Outlier Factor* (LOF). O modelo alcançou 100% de sensibilidade na interceptação das paradas registradas no histórico do ativo. Quando integrado a uma proposta de interface operacional associada a um Filtro de Persistência Temporal de 15 minutos, o sistema demonstrou capacidade de antecipar anomalias críticas com um horizonte estratégico de até 6 horas. Esse intervalo confere à equipe de manutenção a janela útil necessária para coordenar manobras preventivas e procedimentos de lavagem de linha, eliminando o estresse operacional.

Palavras-chave: Manutenção Preditiva. Detecção de Anomalias. Isolation Forest. Local Outlier Factor. Bomba de Polpa. Automação Industrial. Confiabilidade.

Abstract

This work presents the development of a predictive pipeline based on unsupervised machine learning to identify anomalous patterns and anticipate operational failures in slurry pumping systems. In mining routines, unexpected breakdowns of these assets generate severe financial impacts and operational risks, such as pipeline plugging. To address this issue, a data pipeline was structured in Python using real-world time-series data of electrical and hydraulic variables from an industrial pump. Through feature engineering and statistical filtering, the *Isolation Forest* algorithm was calibrated with a 1.0% contamination rate, outperforming the traditional *Local Outlier Factor* (LOF) approach. The model achieved 100% sensitivity in intercepting the failures recorded in the asset's history. When integrated into a proposed operational interface associated with a 15-minute Temporal Persistence Filter, the system demonstrated the capability to anticipate critical anomalies with a strategic horizon of up to 6 hours. This interval provides the maintenance team with the necessary window of opportunity to coordinate preventive maneuvers and line flushing procedures, thereby eliminating operational stress.

Keywords: Predictive Maintenance. Anomaly Detection. Isolation Forest. Local Outlier Factor. Slurry Pump. Industrial Automation. Reliability.

Lista de figuras

Figura 1 – Bomba de Polpa. Fonte: Souza <i>et al.</i> (2017)	22
Figura 2 – Componentes de uma bomba de polpa. Fonte: Carvalho (1999)	23
Figura 3 – <i>Isolation Forest</i> através das árvores. Fonte: Cavalcante, Medeiros e Gonçalves (2023)	28
Figura 4 – Uma propriedade de LOF. Fonte: Santos (2025)	29
Figura 5 – Fluxograma sequencial das etapas metodológicas do pipeline analítico. Fonte: O autor (2026).	33
Figura 6 – Diagrama de dispersão entre a velocidade de operação e a corrente do motor mapeado pelo estado dinâmico das válvulas de intertravamento. Fonte: O autor (2026).	35
Figura 7 – Gráfico de dispersão espacial das variáveis indexadas pelo Isolation Forest (1,0%). Fonte: o autor (2026).	50
Figura 8 – Interface de monitoramento e conferência de dados de processo na bomba. Fonte: O autor (2026).	51
Figura 9 – Janela operacional de alarmes dinâmicos e persistência temporal. Fonte: O autor (2026).	52

Lista de tabelas

Tabela 1 – Os principais tipos de manutenção e suas descrições.	21
Tabela 2 – Dicionário de variáveis operacionais e <i>tags</i> do PIMS utilizadas na modelagem.	33
Tabela 3 – Sensibilidade e cobertura de paradas macro sob variações de parâmetros.	46
Tabela 4 – Matriz de avaliação de desempenho operacional dos modelos.	47
Tabela 5 – Horizonte de antecipação preditiva por evento de parada macro.	49

Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas [cite: 63]
AED	Análise Exploratória de Dados [cite: 414, 415]
CEFET-MG	Centro Federal de Educação Tecnológica de Minas Gerais [cite: 64]
CLP	Controlador Lógico Programável [cite: 66]
COI	Centro de Operações Integradas [cite: 68]
EWMA	<i>Exponentially Weighted Moving Average</i> (Média Móvel Exponencialmente Ponderada) [cite: 518]
IA	Inteligência Artificial [cite: 70]
IBRAM	Instituto Brasileiro de Mineração [cite: 72]
IEC	<i>International Electrotechnical Commission</i> [cite: 839]
IQR	<i>Interquartile Range</i> (Intervalo Interquartil) [cite: 471]
ISA	<i>International Society of Automation</i> [cite: 840]
k-NN	<i>K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos) [cite: 571, 655]
LOF	<i>Local Outlier Factor</i> (Fator de Outlier Local) [cite: 74]
MTTR	<i>Mean Time To Repair</i> (Tempo Médio de Reparo) [cite: 221]
NaN	<i>Not a Number</i> (Valor Nulo / Ausente) [cite: 463]
PCM	Planejamento e Controle de Manutenção [cite: 76]
PID	Proporcional, Integral e Derivativo [cite: 78]
PIMS	<i>Process Information Management System</i> (Sistema de Gerenciamento de Informações de Processo) [cite: 152, 405]
RNA	Rede Neural Artificial [cite: 80]
SCADA	<i>Supervisory Control and Data Acquisition</i> (Sistema Supervisório) [cite: 82]
SEI	Sistema Eletrônico de Informações [cite: 34, 38]
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte) [cite: 84]

TCC	Trabalho de Conclusão de Curso [cite: 86]
UFOP	Universidade Federal de Ouro Preto [cite: 88]

Lista de símbolos

$\%$	Percentual
ν	Parâmetro de contaminação ou taxa de <i>outliers</i> esperada no modelo
σ	Desvio padrão amostral de uma variável de processo
σ^2	Variância amostral de uma série temporal
σ_Q	Volatilidade dinâmica local da vazão de polpa
σ_I	Volatilidade dinâmica local da corrente elétrica do motor
Δt	Intervalo ou janela de tempo operacional
t	Instante de tempo discreto de amostragem dos sensores
x_i	Valor instantâneo de uma variável de processo medida em campo
$\bar{x}$	Média aritmética de uma variável dentro de uma janela móvel
D_{bomba}	Conjunto de dados limitado ao regime de operação ativa do ativo
I_{motor}	Corrente elétrica total consolidada consumida pelo motor
I_R, I_S, I_T	Correntes elétricas instantâneas das fases elétricas brutas
I_{inv}	Corrente de saída analógica modulada pelo inversor de frequência
N_{ref}	Velocidade de referência demandada pelo sistema de automação
w	Velocidade operacional síncrona do ativo
Q	Vazão instantânea de escoamento da polpa mineral na tubulação
ρ	Densidade real da polpa de minério
L_{sup}, L_{inf}	Fronteiras superior e inferior de dispersão estatística para <i>capping</i>
α	Fator matemático de decaimento e suavização exponencial do filtro EWMA
$h(x)$	Comprimento do caminho de isolamento de uma instância na árvore binária
$c(n)$	Comprimento médio de busca de uma árvore binária malsucedida atuando como fator de escala

$s(x, n)$	Escore de anomalia normalizado final do algoritmo <i>Isolation Forest</i>
$lrd_k(p)$	Densidade de alcançabilidade local de uma coordenada espacial de atributos
$S_A(t)$	Indicador de acumulação temporal de falhas consecutivas
τ_{alarme}	Limiar de tolerância mínima para persistência de alarmes

Sumário

1	INTRODUÇÃO	13
1.1	Justificativas e Relevância	14
1.2	Objetivos	14
1.2.1	Geral	14
1.2.2	Específicos	15
1.3	Metodologia	15
1.3.1	Estudo do Processo e Compreensão do Problema:	15
1.3.2	Levantamento e Organização da Base de Dados	15
1.3.3	Pesquisa e Seleção dos Algoritmos:	16
1.3.4	Tratamento e Preparação dos Dados:	17
1.3.5	Treinamento, Validação e Teste dos Modelos	17
1.3.6	Execução e Análise dos Resultados	18
1.4	Organização e estrutura	18
2	REVISÃO DE LITERATURA	20
2.1	Manutenção	20
2.1.1	Tipos de Manutenção	20
2.2	Bombas de Polpa Mineral	22
2.3	Machine Learning	24
2.4	Algoritmos de Detecção de Anomalias Não Supervisionados	27
2.4.1	<i>Isolation Forest</i>	28
2.4.2	<i>Local Outlier Factor - LOF</i>	29
2.5	Trabalhos Relacionados	30
3	DESENVOLVIMENTO	32
3.1	Caracterização da Base de Dados Industriais	32
3.2	Análise Exploratória dos Dados (AED)	34
3.3	Pipeline de Condicionamento de Sinais	35
3.3.1	Saneamento Contextual e Filtro de Bomba Ativa	35
3.3.2	Tratamento de <i>Outliers</i> via <i>Capping</i> Geométrico	36
3.4	Escalonamento e Engenharia de Atributos	37
3.4.1	Indicador de Desbalanceamento Elétrico de Fases	37
3.4.2	Volatilidade Dinâmica de Curto Termo	38
3.4.3	Injeção de Memória Temporal via Filtro EWMA	38
3.4.4	Escalonamento Normalizado Robusto	39
3.5	Modelagem Matemática Não Supervisionada	39

3.5.1	Formalização Matemática do <i>Isolation Forest</i>	39
3.5.2	Formalização Matemática do <i>Local Outlier Factor</i> (LOF)	41
3.6	Pipeline de Pós-Processamento e Lógica de Alarmística	43
3.6.1	Filtro de Persistência Temporal	43
3.6.2	Arquitetura de Estados de Alarme	43
3.7	Métricas	44
3.8	Disponibilidade do Script	45
4	RESULTADOS	46
4.1	Análise Comparativa de Hiperparametrização e Sensibilidade	46
4.2	Avaliação de Desempenho Computacional e Métricas Operacionais	47
4.2.1	Análise de Latência Operacional e Horizonte de Antecipação	48
4.2.2	Análise de Dispersão Espacial e Envelope Operacional	49
4.3	Integração de Telas e Interface de Operação Prática	50
5	CONSIDERAÇÕES FINAIS	53
	Referências	55
	ANEXO A – DECLARAÇÃO DE USO DE IAGEN NO PROCESSO DE ESCRITA CIENTÍFICA	59

1 Introdução

No Brasil, a mineração é uma das principais atividades econômicas, se destacando a produção de minério de ferro, no qual a exploração consolidou o país com um dos maiores exportadores mundiais. Em 2023 a exportação atingiu 392,3 milhões de toneladas de minério, ocasionando uma receita de 42,9 bilhões de dólares, em que 71% correspondiam ao ferro, reforçando a sua importância para o comércio brasileiro ([IBRAM, 2024](#)).

Para viabilizar o processamento e a disposição desses volumes nas usinas de beneficiamento, o transporte de polpa mineral por tubulações internas consolida-se como a solução técnica mais eficiente e economicamente sustentável. Esse sistema hidráulico contínuo permite a movimentação de grandes volumes de minério e rejeitos em suspensão aquosa entre as etapas de moagem, flotação e espessamento, reduzindo drasticamente custos operacionais e impactos ambientais no complexo minerário. Contudo, a operação interna dessas redes impõe desafios severos relacionados à dinâmica de escoamentos multifásicos, onde pequenas oscilações de fluxo podem acarretar o desgaste prematuro de componentes e a sedimentação de sólidos na linha.

Nesse cenário, as bombas de polpa são os ativos críticos que sustentam a energia hidráulica de todo o sistema. Falhas mecânicas nesses equipamentos ou desvios de variáveis operacionais — como a alteração na densidade da polpa ou problemas de recirculação e operação em seco — podem causar obstruções imediatas nas tubulações e paradas catastróficas em equipamentos subsequentes, a exemplo dos espessadores de rejeitos. Conforme adverte [Abulnaga \(2002\)](#), uma única interrupção não planejada nesse circuito interrompe o fluxo de produção da planta, gerando graves prejuízos financeiros e criando gargalos severos no tratamento de efluentes da usina.

Nesse contexto, uma grande mineradora brasileira adota práticas de planejamento e controle da manutenção para mitigar falhas inesperadas e evitar impactos na produção. No entanto, alguns equipamentos apresentam desafios específicos na programação de sua manutenção, como a bomba de polpa mineral, cuja complexidade operacional frequentemente resulta em paradas não planejadas, dificultando o planejamento eficaz de suas intervenções. Esse cenário reforça a diretriz clássica de [Kardec e Nascif \(2013\)](#) sobre a necessidade da adoção de estratégias avançadas de manutenção, como a preditiva, para garantir a confiabilidade desse ativo e minimizar prejuízos decorrentes da indisponibilidade do equipamento.

A criticidade desse ativo se torna ainda mais evidente quando se considera que, devido à necessidade de isolamento elétrico, bloqueio operacional, drenagem de longos trechos de tubulação e substituição de componentes sujeitos a desgaste severo (como

rotores e volutas) após episódios de cavitação ou estresse por funcionamento em seco, as paradas não programadas interrompem imediatamente o fluxo de transporte de sólidos da usina, impactando significativamente a produção planejada (DIAS, 2020). Dessa forma, acredita-se que a implementação de técnicas avançadas de inteligência computacional contribuirá para a excelência operacional, reduzindo paradas inesperadas e otimizando o desempenho da planta industrial

1.1 Justificativas e Relevância

No processo de beneficiamento mineral, o transporte hidráulico de polpa é uma das etapas fundamentais da infraestrutura de suporte da planta, sendo responsável pela movimentação contínua de concentrados ou rejeitos espessados entre os diferentes circuitos de tratamento, pelletização ou barragens de descarte. A interrupção desse fluxo, majoritariamente decorrente de falhas mecânicas ou hidráulicas em sistemas de bombeamento centrífugo, gera impactos diretos e imediatos na eficiência produtiva global da usina, paralisando as operações a montante e tornando imprescindível manter a disponibilidade operacional e a confiabilidade das bombas de polpa (LUZ; SAMPAIO; FRANÇA, 2010).

Neste contexto, o presente trabalho visa contribuir significativamente para melhorar o processo decisório das equipes de manutenção e engenharia de confiabilidade. As paradas não planejadas desses sistemas de bombeamento geram custos operacionais elevados referente à perda de produção, o que reforça a necessidade de prever com antecedência as possíveis falhas e estados de estresse crônico (KARDEC; NASCIF, 2013). Caso seja viável monitorar continuamente as condições de funcionamento do ativo por meio de variáveis elétricas e de processo nativas da instrumentação industrial, tornando possível antecipar desvios com antecedência adequada, a manutenção pode ser executada preditivamente (SIQUEIRA, 2005). Essa abordagem, frequentemente denominada manutenção proativa ou baseada na condição, otimiza a vida útil dos componentes rotativos, reduz o índice de paradas não programadas e mitiga riscos de perdas operacionais no ambiente industrial (SELEME, 2015).

1.2 Objetivos

1.2.1 Geral

Desenvolver um pipeline preditivo baseado em técnicas de aprendizado de máquina não supervisionado para identificar padrões anômalos e antecipar falhas operacionais em sistemas de bombeamento de polpa mineral, contribuindo para o planejamento mais eficiente da manutenção, aumento da assertividade preditiva e maximização da disponibilidade operacional do equipamento.

1.2.2 Específicos

- Compreender e desenvolver a expertise na predição de falhas em ativos industriais, com foco na aplicação em bombas de polpa mineral.
- Desenvolver e testar um modelo de aprendizado de máquina para prever falhas na bomba de polpa, visando reduzir as paradas não programadas e otimizar a manutenção.
- Aplicar o modelo em um caso real de uma grande mineradora brasileira, avaliando sua eficácia na minimização das paradas de produção causadas por falhas na bomba de polpa mineral.

1.3 Metodologia

Esta seção apresenta os procedimentos adotados para o desenvolvimento deste estudo, cuja finalidade principal é construir um modelo preditivo baseado em técnicas de *machine learning*, capaz de antecipar falhas relacionadas ao bombeamento de polpa mineral. A metodologia foi estruturada de forma a atender aos objetivos específicos definidos neste trabalho.

1.3.1 Estudo do Processo e Compreensão do Problema:

Inicialmente, realizou-se uma análise detalhada do funcionamento hidráulico e elétrico do sistema de bombeamento de polpa mineral utilizado na planta industrial estudada. Este levantamento foi realizado por meio de entrevistas e reuniões técnicas com o engenheiro de automação responsável pela operação do equipamento, além da consulta a documentação técnica e registros históricos de manutenção do ativo. Essa etapa visou compreender os aspectos operacionais do equipamento, as variáveis de campo monitoradas e as principais causas associadas a falhas.

1.3.2 Levantamento e Organização da Base de Dados

Com base no estudo do processo, procedeu-se à extração de dados históricos contínuos do ativo diretamente do sistema PIMS (*Process Information Management System*). Os dados contemplaram um período de mais de 60 dias de monitoramento, estruturados em uma régua temporal com taxa de amostragem minuto a minuto, totalizando uma matriz estável de 87.820 registros históricos. As variáveis coletadas foram:

- Correntes elétricas instantâneas das fases A, B e C (I_A, I_B, I_C);
- Corrente elétrica total consolidada consumida pelo motor (I_{motor});

- Velocidade operacional síncrona do ativo (ω);
- Vazão instantânea de escoamento da polpa mineral (Q);
- Densidade real da polpa de minério (ρ);
- *Status* categórico de posicionamento da válvula de controle (Válvula-Estado);

1.3.3 Pesquisa e Seleção dos Algoritmos:

Em paralelo à organização dos dados, realizou-se uma revisão bibliográfica para contornar o severo desbalanceamento de classes que afeta bancos de dados industriais reais, onde quebras catastróficas são eventos escassos e altamente indesejados. Conforme apontado por [Géron \(2019\)](#), a aplicação de classificadores supervisionados tradicionais torna-se inviável nesses cenários, pois a extrema assimetria na proporção das amostras com e sem ocorrência de falha impede que os algoritmos aprendam as fronteiras de decisão da classe minoritária, resultando em modelos negligentes com alto viés de falso-negativo. Complementarmente, [Tan, Steinbach e Kumar \(2005a\)](#) afirmam que, sob escassez crônica de rótulos de falha, o paradigma computacional deve migrar para a Detecção de Anomalias Não Supervisionada, modelando puramente o contorno estatístico do processo estável. Como resultado dessa pesquisa, selecionou-se dois algoritmos consagrados na literatura para a condução de um *benchmarking* comparativo neste estudo:

- *Isolation Forest*: Proposto originalmente por [Liu, Ting e Zhou \(2008\)](#), fundamenta-se no princípio de que as anomalias são mais fáceis de serem isoladas no espaço geométrico do que as instâncias normais. O algoritmo executa partições aleatórias de atributos por meio de florestas de árvores binárias, identificando os desvios operacionais com base em caminhos de busca sensivelmente mais curtos na estrutura das árvores;
- *Local Outlier Factor (LOF)*: Desenvolvido por [Breunig et al. \(2000\)](#), o algoritmo fundamenta-se na determinação da densidade estatística local de uma instância em comparação à de seus k -vizinhos mais próximos no espaço multidimensional de atributos. No monitoramento do sistema de bombeamento, essa abordagem baseada em densidade permite que o pipeline identifique pontos cujas coordenadas operacionais apresentem um isolamento local anômalo em relação ao regime estável.

O uso do *Isolation Forest* e do *Local Outlier Factor* fundamenta-se na capacidade de ambos os métodos em detectar anomalias sem a necessidade de dados previamente etiquetados. A literatura da área, conforme apontam [Breunig et al. \(2000\)](#) e [Liu, Ting e Zhou \(2008\)](#) indica que essas ferramentas são eficientes para lidar com o desbalanceamento extremo que caracteriza os processos de fluxo contínuo na indústria. Desse modo, o estudo

comparativo e a parametrização dessas técnicas viabilizam o cumprimento do segundo objetivo específico desta pesquisa.

1.3.4 Tratamento e Preparação dos Dados:

Nesta etapa, os dados históricos passam por um processamento estruturado para mitigar ruídos de instrumentação e condicionar as variáveis antes da modelagem computacional. Esta fase subdivide-se nos seguintes procedimentos:

- Filtragem de Estados Inoperantes: Remoção dos registros em que o ativo encontrava-se desligado ou em período de repouso legítimo, isolando a base de dados exclusivamente para o regime de funcionamento ativo;
- Tratamento de *outliers*: Etapa de pré-processamento voltada à identificação e ao tratamento de valores atípicos presentes nos dados dos sensores, como picos isolados e variações bruscas causadas por falhas momentâneas de leitura, com o objetivo de reduzir ruídos e melhorar a qualidade das informações utilizadas na análise.;
- Normalização das variáveis: Escalonamento e equalização das grandezas físicas heterogêneas (como vazão, corrente elétrica e densidade), garantindo que as diferentes escalas dos sensores não distorçam a sensibilidade do modelo.;
- Definição do Limiar de Anomalia: Estabelecimento de um parâmetro estatístico de contaminação para delimitar a fronteira do regime estável do ativo, dispensando o uso de rótulos prévios de falha;
- Engenharia de Recursos Sintéticos: Criação de novas variáveis matemáticas baseadas em princípios físicos e temporais, incluindo indicadores de assimetria elétrica e acumuladores ponderados de estresse, conferindo memória operacional ao *pipeline*.

1.3.5 Treinamento, Validação e Teste dos Modelos

Nesta fase, os algoritmos são calibrados e testados para garantir que o sistema identifique corretamente desvios operacionais sem gerar instabilidade na leitura dos dados. Este procedimento envolve as seguintes ações:

- Particionamento dos Dados: Divisão da base histórica em conjuntos independentes, isolando os dados necessários para o ajuste inicial do comportamento estável e aqueles destinados aos testes de validação;
- Otimização de Hiperparâmetros: Parametrização das variáveis internas de configuração de cada algoritmo para refinar a sensibilidade de detecção e evitar o sobreajuste (*overfitting*) dos modelos;

- Seleção de Métricas de Desempenho: Avaliação do desempenho dos algoritmos com foco na capacidade de cobertura e detecção de falhas reais (sensibilidade) e no controle estatístico de falsos alertas;
- Integração de Filtros de Persistência: Aplicação de regras temporais pós-inferência para amortecer oscilações rápidas, garantindo que apenas anomalias contínuas e representativas consolidem um estado de alarme.

1.3.6 Execução e Análise dos Resultados

Nesta etapa final, os modelos calibrados são aplicados aos dados de teste para avaliar o seu desempenho em cenários práticos. Esse procedimento é composto pelas seguintes ações:

- Simulação de Monitoramento: Execução do pipeline sobre a base de testes para reproduzir o comportamento dos algoritmos em um cenário de fluxo contínuo;
- Confronto com Históricos Reais: Comparação direta dos alarmes gerados pelos modelos com os registros de macroparas e intervenções mecânicas reais catalogadas pelas equipes de campo;
- Avaliação de Impacto Operacional: Análise do tempo de antecedência e da assertividade das ferramentas na identificação das falhas, validando a viabilidade prática do sistema como ferramenta de suporte à decisão.

1.4 Organização e estrutura

O Capítulo 1 deste trabalho apresenta a introdução ao tema, trazendo a contextualização do problema na indústria mineral, as justificativas socioeconômicas e operacionais, as delimitações do escopo e a definição dos objetivos geral e específicos do projeto. O Capítulo 2 é dedicado ao referencial teórico, reunindo a revisão da literatura especializada sobre engenharia de confiabilidade, gestão de manutenção preditiva e o embasamento matemático dos algoritmos não supervisionados de detecção de anomalias aplicados a processos contínuos. Já o Capítulo 3 aborda a metodologia e a preparação de dados, detalhando o mapeamento e o tratamento estatístico das séries temporais da bomba industrial, bem como as etapas de engenharia de atributos para a construção das variáveis de volatilidade em Python. O Capítulo 4 detalha toda a análise e discussão dos resultados, apresentando os testes de validação, o procedimento de *grid search*, o desempenho operacional comparativo frente ao LOF, a dispersão física das anomalias e a integração das regras de persistência temporal na interface sinótica. Por fim, o Capítulo 5 traz a conclusão do

trabalho, discutindo de forma integrada os resultados obtidos, o cumprimento de cada objetivo específico e as sugestões de frentes técnicas para trabalhos futuros.

2 Revisão de literatura

2.1 Manutenção

De acordo com [De Almeida e Rocca \(2015\)](#), a palavra “manutenção” origina-se do latim *manu tenere*, cujo significado remete a “segurar com a mão” ou “manter o que se possui”, traduzindo-se no ambiente industrial como o conjunto de ações técnicas destinadas a garantir a regularidade operacional e o reparo de ativos. Os autores enfatizam que a engenharia de manutenção deve exercer influência direta ainda na fase de projeto, de modo a contemplar a manutenibilidade do equipamento por meio do acesso facilitado aos componentes mecânicos e da padronização dimensional de peças sobressalentes. Esses conceitos e os termos técnicos correlatos à confiabilidade são normatizados nacionalmente pela NBR 5462 da Associação Brasileira de Normas Técnicas (ABNT).

Complementarmente, [Seleme \(2015\)](#) descreve que a aplicação eficiente dessas práticas resulta em ganhos diretos de confiabilidade e segurança devido à redução de quebras inesperadas. O autor também elenca vantagens estratégicas como a estabilidade qualitativa dos processos, a otimização de custos operacionais, a extensão da vida útil dos maquinários e a preservação do valor patrimonial dos ativos industriais.

2.1.1 Tipos de Manutenção

Conforme descrito por [Siqueira \(2005\)](#), as estratégias de manutenção são categorizadas de acordo com a postura adotada pela equipe de engenharia diante da iminência ou da ocorrência de falhas nos ativos. Nesse contexto, o gerenciamento de ativos industriais divide-se em três linhas principais de atuação, conforme demonstrado na Tabela 1.

Ao permitir o diagnóstico de anomalias operacionais sem a necessidade de paralisar o fluxo produtivo, a abordagem preditiva viabiliza o planejamento antecipado das ações de reparo. A adoção dessa metodologia proporciona vantagens diretas à planta, tais como:

- Incremento da segurança operacional e da disponibilidade física dos equipamentos, mitigando paradas inesperadas e riscos de acidentes de trabalho;
- Eliminação da substituição prematura de peças e componentes mecânicos que ainda preservam vida útil remanescente;
- Otimização de custos logísticos e redução do tempo médio de reparo (MTTR) devido à previsibilidade das intervenções;

- Preservação das condições gerais de instalação do ativo, reduzindo o estresse mecânico secundário sobre os demais componentes do conjunto.

Tabela 1 – Os principais tipos de manutenção e suas descrições.

Tipos de Manutenção	Descrição
Corretiva	Caracteriza-se pela atuação puramente reativa, onde os procedimentos de reparo ou substituição de componentes são executados apenas após a manifestação física da quebra;
Preventiva	Baseia-se no planejamento cronológico ou por tempo de operação, realizando intervenções programadas de forma sistemática para mitigar a probabilidade de falhas e estender a vida útil operacional do maquinário;
Preditiva	Atua por meio do monitoramento contínuo de parâmetros operacionais e variáveis de estado (como vibração, temperatura e corrente elétrica) para identificar o desenvolvimento de falhas em estágio incipiente, permitindo o agendamento da manutenção antes que o dano se torne crítico.
Detectiva	Consiste na atuação em sistemas de proteção, comando ou intertravamentos, realizando testes e checagens funcionais periódicas para identificar falhas ocultas que não são perceptíveis pela equipe de operação, garantindo a atuação das defesas em emergências.
Engenharia de Manutenção	Concentra-se na aplicação de engenharia de confiabilidade para modificar o projeto, alterar componentes ou readequar processos operacionais, buscando eliminar a causa raiz de falhas repetitivas e aumentar a manutenibilidade do ativo.

Fonte: O autor, 2026.

Adicionalmente, o registro e a análise estatística das variáveis históricas coletadas por essa modalidade de manutenção geram inteligência de dados para a engenharia de confiabilidade. Esse histórico viabiliza a identificação de modos de falha recorrentes (maquinários crônicos) e o direcionamento assertivo de modificações de projeto. Da mesma forma, os dados permitem auditar a eficácia das intervenções realizadas pelas equipes de campo, fornecendo subsídios para programas de capacitação técnica e orientando a implementação de novas tecnologias de automação (SIQUEIRA, 2005).

De maneira geral, conforme apontado por Mobley (2002), a consolidação da manutenção preditiva em indústrias de processo após o período inicial de investimento em instrumentação e infraestrutura de dados é capaz de reduzir em aproximadamente dois terços (2/3) as perdas financeiras associadas a interrupções catastróficas na linha de produção, além de contrair em cerca de um terço (1/3) os custos totais de manutenção a médio e longo prazo.

2.2 Bombas de Polpa Mineral

O bombeamento de polpa é o método mais simples, rápido e econômico de se transportar sólidos que são usados em plantas de beneficiamento mineral, em que as polpas de minério são bombeadas entre os processos de concentração, quando o arranjo mecânico não possibilita o transporte da polpa por gravidade ou quando é necessária uma pressão residual predeterminada (SILVA, V. D., 2011).

Para transportar de acordo com as normas ambientais e com custo-benefício adequado, é importante realizar o dimensionamento das tubulações e selecionar as bombas de modo apropriado, para que obtenha uma mistura de água e resíduos minerais com as características reológicas adequadas para serem bombeadas de modo eficaz (BARCELOS; ARAÚJO; ALONSO, 2015).

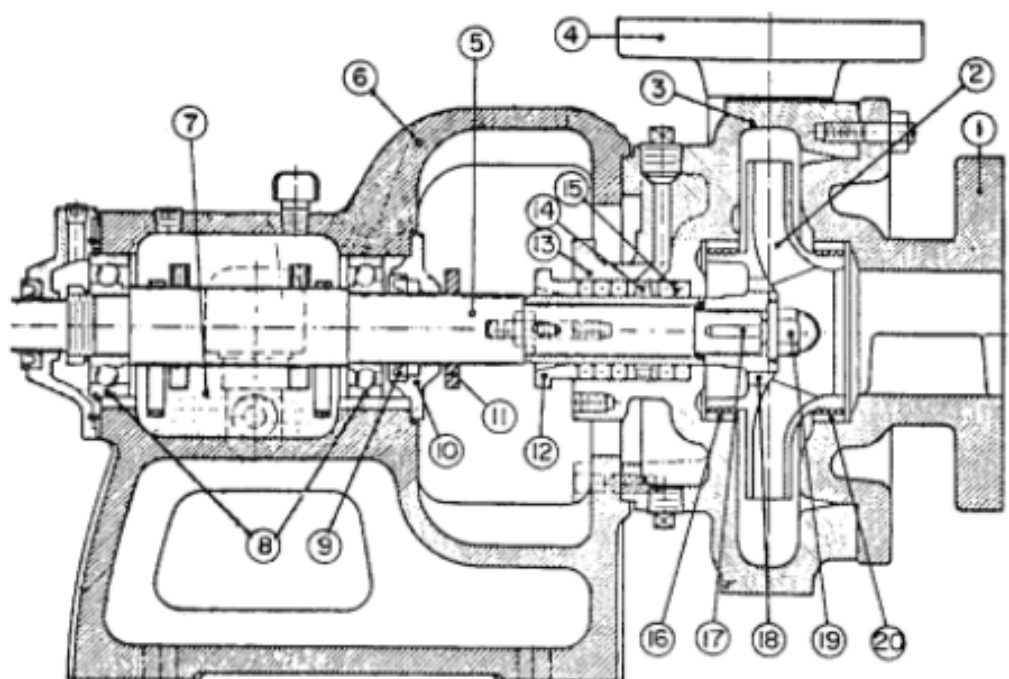
A bomba de polpa na Figura 1 é composta por uma carcaça estacionária e na sua parte interna possui um rotor montado sobre um eixo rotativo. Suas manutenções impactam na disponibilidade do equipamento, afetando a capacidade de produção do beneficiamento do minério de ferro e, conseqüentemente, o custo do produto (FERREIRA, 2020).



Figura 1 – Bomba de Polpa. Fonte: Souza *et al.* (2017)

As bombas de polpa mineral são majoritariamente do tipo centrífugo, no qual a energia cinética transferida pelo rotor ao fluido é convertida em pressão ao longo da voluta, impulsionando a mistura de sólidos e água através da tubulação V. D. Silva (2011). A Figura 2 apresenta os componentes de uma bomba de polpa.

Além do modelo centrífugo, também são empregadas bombas volumétricas (ou de deslocamento positivo) em aplicações que exigem pressões mais elevadas ou vazões mais constantes, especialmente quando a polpa apresenta viscosidade ou concentração de sólidos



ITEM	NOME DA PEÇA
01	Flange de sucção
02	Rotor
03	Carcça ou caixa espiral
04	Flange de descarga
05	Eixo
06	Cavalete
07	Caixa de óleo
08	Rolamentos
09	Retentor
10	Tampa da caixa de óleo
11	Defletor
12	Sobreposta ou aperta-gaxetas
13	Estojo de gaxetas
14	Cadeado hidráulico
15	Gaxetas
16	Anel de desgaste traseiro
17	Chaveta
18	Furos de compensação
19	Porca do rotor
20	Anel de desgaste dianteiro

Figura 2 – Componentes de uma bomba de polpa. Fonte: [Carvalho \(1999\)](#)

mais elevada (BARCELOS; ARAÚJO; ALONSO, 2015).

O dimensionamento adequado de uma bomba de polpa depende diretamente das propriedades reológicas da mistura transportada, como a concentração de sólidos, a distribuição granulométrica e a viscosidade aparente, uma vez que essas variáveis influenciam diretamente a perda de carga ao longo da tubulação e o desempenho hidráulico do sistema (FONSECA, 2024).

Um dos parâmetros centrais nesse dimensionamento é a velocidade crítica de deposição, isto é, a velocidade mínima de escoamento abaixo da qual as partículas sólidas tendem a se sedimentar no fundo da tubulação, favorecendo entupimentos, aumento da perda de carga e desgaste localizado (FERREIRA, 2020). Operar a bomba de forma consistente acima dessa velocidade crítica é essencial para preservar a vida útil do equipamento e evitar paradas não programadas.

Devido ao contato contínuo com partículas sólidas abrasivas em suspensão, os principais componentes internos da bomba, como rotor, voluta e selos mecânicos, estão sujeitos a desgaste erosivo progressivo, que reduz gradualmente a eficiência hidráulica do equipamento e pode culminar em falhas catastróficas se não for identificado a tempo (DIAS, 2020). Além do desgaste abrasivo, a operação em condições inadequadas acelera a deterioração dos componentes rotativos e amplia o risco de paradas não programadas, tornando o monitoramento contínuo das condições operacionais uma medida essencial para a gestão da confiabilidade desses ativos.

Nesse sentido, a importância do acompanhamento sistemático de indicadores de manutenção para avaliar a efetividade das estratégias adotadas e orientar decisões de intervenção. A consolidação desses indicadores, associada ao monitoramento contínuo de variáveis de processo, permite não apenas quantificar a criticidade do ativo, mas também identificar oportunidades de melhoria na programação da manutenção, reforçando a relevância de abordagens preditivas baseadas em dados para esse tipo de equipamento (SOUZA *et al.*, 2017).

2.3 *Machine Learning*

O aprendizado de máquina, do inglês *machine learning*, constitui um dos ramos mais expressivos da inteligência artificial contemporânea, dedicado ao desenvolvimento de algoritmos capazes de identificar padrões, extrair conhecimento e realizar previsões a partir da análise de conjuntos de dados, sem que suas regras de decisão precisem ser explicitamente programadas (GÉRON, 2019). Diferentemente da programação tradicional, na qual o comportamento do sistema é definido previamente por regras fixas estabelecidas pelo desenvolvedor, os algoritmos de aprendizado de máquina ajustam seus próprios parâmetros internos a partir da exposição a exemplos, construindo modelos matemáticos

capazes de generalizar o comportamento observado para situações ainda não vivenciadas.

Essa capacidade de generalização é o que confere ao aprendizado de máquina sua principal utilidade prática em contextos industriais, nos quais a complexidade dos sistemas físicos e a multiplicidade de variáveis envolvidas tornam inviável a modelagem puramente analítica ou baseada em regras determinísticas (MORAES *et al.*, 2022). Ao integrar conceitos oriundos da estatística, da teoria da probabilidade, da otimização matemática e da ciência da computação, esses algoritmos permitem que o comportamento de ativos industriais complexos seja aprendido diretamente a partir de seus próprios registros históricos, dispensando a necessidade de um modelo físico completo e explícito do sistema monitorado.

De maneira geral, os algoritmos de aprendizado de máquina podem ser classificados em dois grandes paradigmas: o aprendizado supervisionado e o aprendizado não supervisionado (DIAS, 2020). No aprendizado supervisionado, o modelo é treinado a partir de um conjunto de dados previamente rotulado, no qual cada observação está associada a uma classe ou valor de saída conhecido, permitindo que o algoritmo aprenda, por meio de exemplos, a associar determinadas características de entrada a determinados resultados. Essa abordagem exige, contudo, a disponibilidade de rótulos confiáveis e representativos de todas as classes de interesse, condição que nem sempre é atendida em ambientes industriais reais.

Já o aprendizado não supervisionado prescinde da existência de rótulos, concentrando-se exclusivamente na identificação de estruturas, agrupamentos ou desvios estatísticos inerentes à própria distribuição dos dados (GAIGHER, 2023). Por não depender de exemplos previamente classificados, essa abordagem se mostra particularmente adequada para cenários em que a ocorrência de falhas é rara, os registros históricos de eventos anômalos são escassos ou inexistentes, e a obtenção de rótulos confiáveis demandaria um esforço manual desproporcional em relação ao benefício obtido, situação recorrente em ativos industriais de operação contínua, como é o caso das bombas de polpa mineral.

Dentro do paradigma não supervisionado, destaca-se a detecção de anomalias como uma das aplicações mais relevantes para o monitoramento de sistemas físicos complexos (GAIGHER, 2023). A detecção de anomalias consiste em identificar, dentro de um conjunto de dados, observações que se desviam significativamente do padrão considerado normal ou esperado, sem que seja necessário conhecer previamente exemplos rotulados de falha. Ao modelar exclusivamente o comportamento estatístico do regime estável de operação, esses algoritmos tornam-se capazes de sinalizar desvios sutis que podem representar o estágio inicial de um processo de degradação mecânica, ainda antes que a falha se manifeste de forma catastrófica.

Independentemente do paradigma adotado, o desenvolvimento de um modelo de aprendizado de máquina segue, de forma geral, um fluxo estruturado de etapas que

compreende a coleta e o pré-processamento dos dados, a extração e seleção de atributos relevantes, o treinamento do modelo propriamente dito, sua validação e, por fim, a avaliação de desempenho em dados de teste (MORAES *et al.*, 2022). Essa sequência metodológica garante que o modelo seja avaliado em condições que se aproximem do cenário real de aplicação, reduzindo o risco de conclusões enviesadas por um ajuste excessivo às particularidades do conjunto utilizado no treinamento.

A qualidade dos dados empregados nesse processo é determinante para o desempenho final do modelo, no qual ruídos de instrumentação, valores ausentes, leituras espúrias e variáveis registradas em escalas heterogêneas podem distorcer significativamente a capacidade do algoritmo de identificar padrões relevantes, tornando indispensável um processo criterioso de tratamento, normalização e filtragem dos dados antes da etapa de modelagem (DIAS, 2020). Complementarmente, é possível incorporar ao modelo informações físicas, temporais e operacionais que não estão diretamente disponíveis nos dados brutos, ampliando sua capacidade de discriminar entre comportamento normal e anômalo.

Um dos principais desafios enfrentados no treinamento de modelos de aprendizado de máquina é o equilíbrio entre a capacidade de generalização e o ajuste aos dados de treinamento. Quando o modelo se ajusta excessivamente às particularidades do conjunto utilizado em seu treinamento, ocorre o fenômeno conhecido como sobreajuste, ou *overfitting*, que compromete sua capacidade de generalizar corretamente para dados novos e ainda não observados (GÉRON, 2019).

Em sentido oposto, quando o modelo não é capaz de capturar adequadamente os padrões subjacentes aos dados, ocorre o subajuste, ou *underfitting*, resultando em um desempenho insatisfatório mesmo sobre o próprio conjunto de treinamento. Encontrar esse ponto de equilíbrio é especialmente relevante em contextos industriais, nos quais o comportamento do processo pode apresentar variações sutis e legítimas ao longo do tempo, sem que isso represente necessariamente uma condição de falha (GÉRON, 2019).

A avaliação do desempenho de modelos de aprendizado de máquina aplicados a problemas com forte desbalanceamento de classes, como é o caso típico da detecção de falhas industriais, exige métricas que vão além da simples acurácia, uma vez que essa métrica pode apresentar valores elevados mesmo quando o modelo ignora completamente a classe minoritária de interesse, isto é, os eventos de falha (GAIGHER, 2023).

Nesse contexto, métricas como a precisão, definida pela proporção de alarmes corretos entre todos os classificados como anomalia, a sensibilidade, ou *recall*, definida pela proporção de falhas reais corretamente identificadas, e o F1-score, calculado como a média harmônica entre as duas anteriores, mostram-se consideravelmente mais informativas para mensurar a real capacidade de um modelo de detectar eventos raros sem gerar um volume excessivo de falsos alertas operacionais (GAIGHER, 2023).

2.4 Algoritmos de Detecção de Anomalias Não Supervisionados

A escolha dos algoritmos de detecção de anomalias não supervisionados se deve ao fato de eles se ajustarem perfeitamente às características dos dados industriais deste trabalho. Como essa classe de algoritmos não precisa de um histórico de falhas já marcadas ou rotuladas, o que é um evento raro no dia a dia de uma fábrica, ela foca em aprender diretamente como o equipamento se comporta quando está operando de forma normal e estável. Com isso, o sistema consegue identificar o início de pequenas alterações no funcionamento da máquina bem antes de acontecer uma quebra física, baseando-se puramente na identificação de padrões excentricamente distantes do regime comum.

A detecção de anomalias consiste em identificar, dentro de um conjunto de dados, observações que se afastam de maneira significativa do comportamento considerado normal, sem que seja necessário conhecer previamente exemplos rotulados dessas ocorrências atípicas. Essa característica torna a abordagem especialmente adequada a ativos industriais de operação contínua, nos quais os registros de falha costumam ser raros, descontínuos ou simplesmente inexistentes. Assim, os algoritmos de detecção de anomalias não supervisionada permitem que o próprio contorno estatístico do regime estável de operação seja utilizado como referência para a identificação de desvios (BALDISSERA; TRENTIN, 2025).

De maneira geral, as técnicas de detecção de anomalias não supervisionada podem ser organizadas em diferentes estratégias conceituais, dentre as quais se destacam as abordagens estruturais, que avaliam a facilidade com que uma observação pode ser isolada dentro do espaço de atributos, e as abordagens baseadas em densidade, que avaliam o grau de isolamento de uma observação em relação à concentração local de seus vizinhos mais próximos (SILVA, J. P. S., 2025). Cada uma dessas estratégias tende a capturar tipos distintos de comportamento anômalo, de modo que sua aplicação conjunta amplia a capacidade de um pipeline preditivo de identificar diferentes manifestações de desvio operacional.

Diante disso, neste trabalho foram selecionados dois algoritmos representativos dessas duas estratégias, o *Isolation Forest* e o *Local Outlier Factor*. Do ponto de vista computacional, ambos se mostram adequados ao processamento de bases de dados extensas e de alta dimensionalidade, como as séries históricas de variáveis de processo industrial coletadas em alta frequência de amostragem, ainda que apresentem particularidades distintas quanto ao custo de execução, o que deve ser considerado na etapa de otimização de hiperparâmetros descrita na metodologia deste trabalho. Os fundamentos teóricos e os mecanismos de funcionamento de cada uma dessas técnicas são apresentados nas seções seguintes.

2.4.1 Isolation Forest

O *Isolation Forest* parte do princípio de que observações anômalas são, por definição, raras e estatisticamente diferentes das observações normais, características que as tornam mais fáceis de isolar dentro de um espaço de atributos multidimensional ([CAVALCANTE; MEDEIROS; GONÇALVES, 2023](#)). Para explorar essa propriedade, o algoritmo constrói um conjunto de árvores binárias, denominadas árvores de isolamento, ilustrado na Figura 3, por meio da seleção aleatória sucessiva de um atributo e de um valor de corte dentro do intervalo observado desse atributo, particionando recursivamente o espaço de dados até que cada observação fique isolada em um nó folha individual.

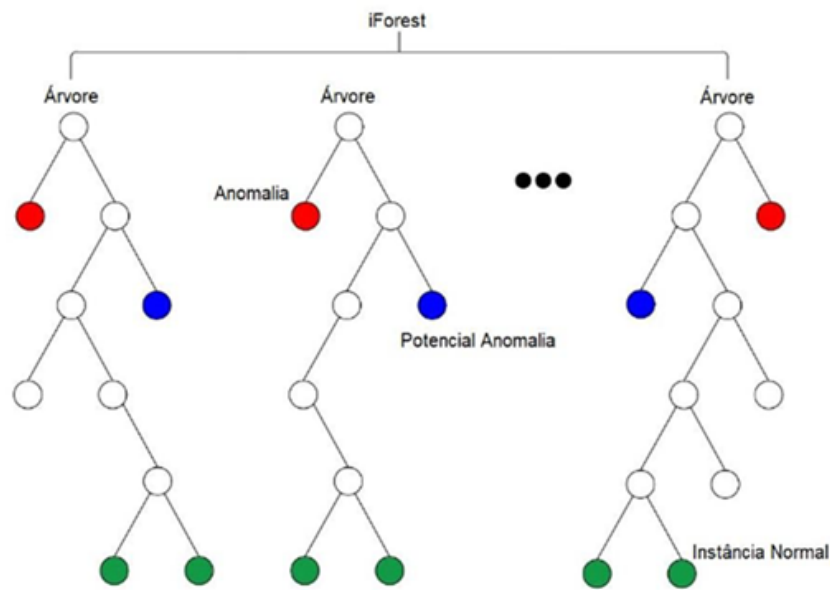


Figura 3 – *Isolation Forest* através das árvores. Fonte: [Cavalcante, Medeiros e Gonçalves \(2023\)](#)

Observações anômalas tendem a requerer, em média, menos partições para serem isoladas, resultando em caminhos mais curtos entre a raiz e o nó folha correspondente dentro de cada árvore. Observações normais, por estarem posicionadas em regiões mais densas e homogêneas do espaço de atributos, exigem um número maior de partições sucessivas até que sejam completamente isoladas, produzindo caminhos consideravelmente mais longos ([FREITAS; VIEIRA, 2026](#)). A partir do comprimento médio desses caminhos, calculado ao longo de todas as árvores que compõem a floresta, o algoritmo atribui a cada observação um escore de anomalia normalizado, no qual valores mais próximos da unidade indicam maior probabilidade de comportamento anômalo, enquanto valores próximos de zero indicam maior conformidade com o regime considerado normal.

Entre suas principais vantagens está a baixa complexidade computacional quando comparado a métodos que dependem do cálculo de distâncias entre todos os pares de observações, característica que o torna atrativo para bases de dados de grande volume, além

de não exigir suposições prévias sobre a distribuição estatística dos dados (CAVALCANTE; MEDEIROS; GONÇALVES, 2023). Como limitação, o algoritmo tende a apresentar desempenho reduzido em situações nas quais as anomalias não se distinguem estruturalmente das observações normais, manifestando-se apenas por sutis variações de densidade local, cenário em que abordagens baseadas em densidade tendem a se sobressair.

2.4.2 Local Outlier Factor - LOF

Em contraste com a abordagem estrutural do *Isolation Forest*, o *Local Outlier Factor* adota uma perspectiva baseada em densidade, avaliando o grau de isolamento de cada observação em relação à densidade local de sua vizinhança, em vez de considerar apenas sua distância a um centro global de referência, como visto na Figura 4. Essa diferença conceitual confere ao algoritmo a capacidade de identificar anomalias locais, isto é, observações que se destacam de maneira significativa em relação à vizinhança imediata que as circunda, mesmo quando sua distância em relação ao conjunto de dados como um todo não é excepcionalmente elevada, situação em que técnicas baseadas exclusivamente em distância global tendem a apresentar desempenho limitado (SANTOS, 2025).

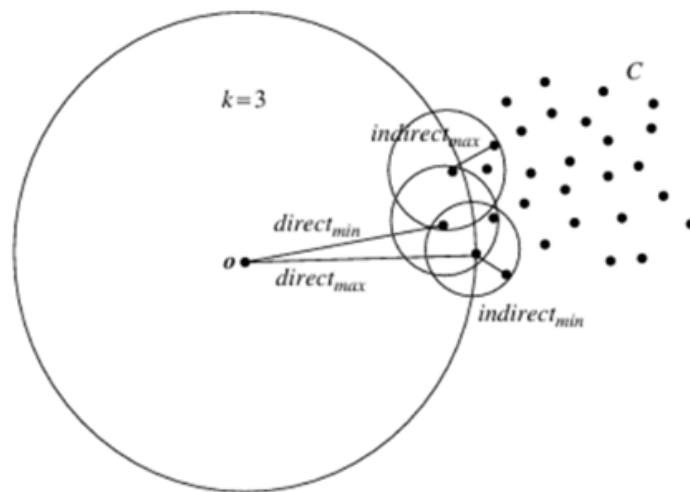


Figura 4 – Uma propriedade de LOF. Fonte: Santos (2025)

O cálculo do *Local Outlier Factor* parte da determinação da densidade de alcançabilidade local de cada observação, obtida a partir da distância média entre essa observação e seus k vizinhos mais próximos no espaço multidimensional de atributos. Em seguida, essa densidade local é comparada com a densidade média observada entre os próprios vizinhos, resultando em uma razão que expressa o grau relativo de isolamento da observação avaliada (LIMA *et al.*, 2024). Quando essa razão se aproxima da unidade, entende-se que a observação possui densidade semelhante à de sua vizinhança, sendo classificada como normal. Quando o valor obtido é significativamente superior a um, a ob-

servação apresenta densidade local muito inferior à de seus vizinhos, condição interpretada como um forte indicativo de comportamento anômalo.

A aplicação do *Local Outlier Factor* é especialmente relevante para identificar desvios operacionais sutis em ativos como bombas de polpa mineral, que podem não se isolar completamente no espaço de atributos, mas que rompem com o padrão de densidade característico do regime estável de operação. Como limitação, seu desempenho é fortemente dependente da escolha adequada do número de vizinhos considerado no cálculo da densidade local, parâmetro cuja definição inadequada pode comprometer a capacidade do algoritmo de distinguir entre variações legítimas do processo e desvios efetivamente anômalos, além de o algoritmo demandar maior esforço computacional à medida que o volume de dados cresce, por depender do cálculo de distâncias entre observações (SILVA, J. P. S., 2025).

2.5 Trabalhos Relacionados

A aplicação de técnicas de inteligência computacional para predição de falhas em ativos industriais tem sido tema recorrente na literatura, principalmente em equipamentos críticos para o processo produtivo. Nesta seção, são apresentados estudos que abordam o uso de algoritmos de aprendizado de máquina para diagnóstico e previsão de falhas, servindo de base para o desenvolvimento deste trabalho. Destaca-se que, enquanto o monitoramento de bombas comuns de água já é muito conhecido, as bombas que transportam polpa de minério lidam com uma mistura pesada e abrasiva que muda constantemente de comportamento. Por causa dessa variação constante do processo, esses ativos exigem modelos preditivos mais robustos e ajustados à realidade prática da planta.

No trabalho desenvolvido por Mendes, Rodrigues e Lima (2018), os autores propuseram um modelo de detecção de anomalias e diagnóstico de falhas em bombas centrífugas industriais utilizando técnicas de aprendizado de máquina baseadas em árvores de decisão e florestas aleatórias. Para isso, foram monitorados parâmetros de processo como vazão, pressão de descarga e corrente elétrica do motor. Os resultados indicaram que a abordagem preditiva foi capaz de identificar desvios incipientes de cavitação e desgaste de rotor com antecedência substancial, alcançando taxas de acerto superiores a 94%, evidenciando a eficácia do monitoramento de variáveis hidráulicas e elétricas combinadas.

No estudo realizado por Silva e Santos (2021), investigou-se a aplicação de algoritmos de aprendizado não supervisionado, com foco na detecção de anomalias em turbomáquinas e sistemas de bombeamento hidráulico de grande porte. Utilizando dados de séries temporais de sensores de processo, os autores avaliaram o desempenho de modelos geométricos de isolamento na identificação de regimes anômalos de operação, como o funcionamento em seco e o refluxo de linha. Os resultados demonstraram que abordagens não supervisionadas

apresentam alta aderência ao chão de fábrica devido à capacidade de modelar o envelope de normalidade sem a necessidade de histórico prévio de falhas catastróficas rotuladas.

No estudo conduzido por [Kumar, Santhosh e Vinayak \(2019\)](#), os autores exploraram a fusão de dados e algoritmos de detecção de anomalias, comparando o comportamento de modelos de densidade espacial e isolamento geométrico em bombas de palhetas e deslocamento positivo. Inicialmente, foram extraídas características dinâmicas dos sinais sob diferentes condições operacionais de carga e estrangulamento de válvulas. Os resultados obtidos evidenciaram que, embora ambos os modelos apresentassem alta precisão em ambientes controlados, os algoritmos baseados no *isolation forest* mostraram-se significativamente mais robustos frente aos ruídos de transientes operacionais normais e ao severo desbalanceamento de classes típico de séries temporais industriais.

3 Desenvolvimento

Para executar o projeto com sucesso, é fundamental compreender o conceito de detecção de anomalias e aplicar o método do *pipeline* de engenharia de dados, que consiste nas etapas de condicionamento de sinais, engenharia de atributos e modelagem não supervisionada. Para isso, utiliza-se como referência os materiais de [Montgomery \(2009\)](#) e [Géron \(2019\)](#), que situam o aprendizado de máquina no contexto de processos industriais e fornecem percepções sobre a identificação e o tratamento de desvios operacionais. A partir desse pipeline estatístico e computacional, é possível localizar e entender as instabilidades no contexto do bombeamento de polpa e propor alertas adequados para apoiar a gestão da manutenção.

O desenvolvimento do trabalho foi dividido em quatro etapas principais: caracterização da base de dados, condicionamento dos sinais brutos, engenharia de recursos com memória temporal e aplicação dos algoritmos de detecção.

A estrutura sequencial do *pipeline* analítico desenvolvido e suas respectivas interconexões estão ilustradas na Figura 5.

Para isso, utilizou-se como referência o fluxo de análise sequencial de séries temporais industriais, porém sem incluir ações de controle corretivo em malha fechada, visto que o objetivo do trabalho é propor um sistema preditivo de monitoramento de integridade do ativo, e não a regulação dinâmica das malhas automatizadas do inversor de frequência.

3.1 Caracterização da Base de Dados Industriais

O estudo de caso baseia-se em dados históricos reais extraídos do sistema *Process Information Management System* (PIMS) de uma unidade mineradora, cobrindo o período contínuo de 02 de abril de 2026 a 30 de maio de 2026. A taxa de amostragem configurada no servidor de dados é de uma amostra por minuto, fornecendo uma série temporal densa que reflete os regimes permanentes, transientes de carga e eventos de estresse hidráulico na bomba.

A matriz de dados de entrada é composta por oito variáveis operacionais analógicas e digitais, que monitoram simultaneamente o comportamento elétrico do motor e a dinâmica do fluido no circuito. As variáveis coletadas via *tags* do PIMS são:

Esse conjunto de variáveis constitui o espaço vetorial heterogêneo que passará pelas etapas subsequentes de higienização, engenharia de atributos e modelagem preditiva.



Figura 5 – Fluxograma sequencial das etapas metodológicas do pipeline analítico. Fonte: O autor (2026).

Tabela 2 – Dicionário de variáveis operacionais e *tags* do PIMS utilizadas na modelagem.

Variável / Tag	Unidade	Descrição Operacional
I_R, I_S, I_T	A	Correntes elétricas trifásicas do motor de indução.
I_{inv}	A	Corrente de saída analógica modulada pelo inversor.
N_{ref}	rpm	Velocidade de referência demandada pelo sistema de automação.
Q	m^3/h	Vazão volumétrica de polpa na tubulação.
ρ	g/L	Densidade da polpa.
$St_{válv}$	Booleano	Estado discreto da válvula (aberta/fechada).

Fonte: O autor (2026).

3.2 Análise Exploratória dos Dados (AED)

Antes da aplicação de qualquer algoritmo de modelagem ou rotina de higienização, realizou-se a Análise Exploratória de Dados (AED). Conforme preconizado por [Tukey \(1977\)](#), a AED é uma etapa indispensável para compreender a estrutura subjacente das séries temporais, identificar o comportamento estatístico das variáveis e mapear visualmente as perturbações decorrentes da dinâmica do processo industrial. No contexto deste trabalho, a exploração focou em correlacionar os sinais elétricos do acionamento com as variáveis hidráulicas do circuito da bomba.

A análise inicial das distribuições de frequência revelou um comportamento multimodal e altamente assimétrico na maioria das variáveis analógicas. A corrente de saída do inversor (I_{inv}) e as correntes das fases do motor (I_R, I_S, I_T), por exemplo, apresentaram uma densidade de pontos massiva concentrada próxima a zero. Esse fenômeno estatístico, denominado “bloco morto de dados”, reflete os períodos em que o ativo **BP2055VG01** encontrava-se desenergizado ou em estado de repouso reserva. Sob a ótica de [Fitzgerald, Kingsley e Umans \(2003\)](#), o comportamento de regime em vazio ou desenergizado difere fundamentalmente das características eletromecânicas do motor sob carga operacional. Dessa forma, manter esses registros de repouso na base distorce as medidas de tendência central e induz o modelo computacional ao viés de interpretar a inoperação como um estado de estabilidade em regime de carga.

Complementarmente, mapeou-se o acoplamento eletromecânico e hidráulico do sistema por meio do diagrama de dispersão entre a velocidade de operação e a corrente do motor, estratificado pelas transições lógicas da válvula de controle, conforme ilustrado na [Figura 6](#).

A inspeção visual da [Figura 6](#) revela características fundamentais do processo. Abaixo de 50% de velocidade, a escassez de pontos confirma que a região funciona estritamente como rampa transiente de aceleração. Acima desse limiar, observa-se o comportamento quadrático clássico da corrente em relação à velocidade, característico de cargas de fluxo centrífugo. O mapeamento do estado da válvula explicita a influência fluido-mecânica no motor: os registros em regime de fechamento (*FECHANDO*) impõem perda de carga e elevam a corrente em faixas intermediárias de rotação, ao passo que a condição de restrição crítica (*FECHADA*) concentra-se em altas velocidades com correntes severas (próximas a 120 A), indicando recirculação total e estresse térmico.

Por fim, identificam-se *outliers* isolados no quadrante superior esquerdo, incompatíveis com a física do acionamento, que denotam ruídos espúrios de instrumentação e justificam os critérios de contenção estatística implementados nas etapas subsequentes.

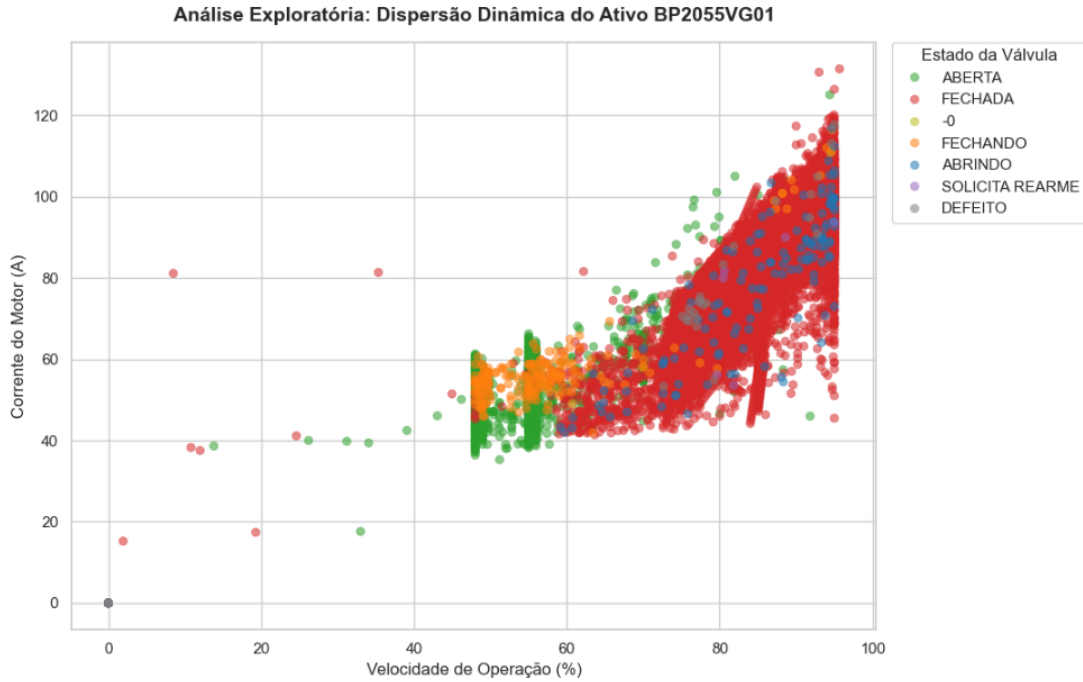


Figura 6 – Diagrama de dispersão entre a velocidade de operação e a corrente do motor mapeado pelo estado dinâmico das válvulas de intertravamento. Fonte: O autor (2026).

3.3 Pipeline de Condicionamento de Sinais

Após a identificação das perturbações estatísticas e operacionais na etapa exploratória, estruturou-se um *pipeline* sequencial de condicionamento e higienização de sinais. Esta etapa é crítica para eliminar registros que não representam o processo dinâmico real e para atenuar ruídos de alta frequência da instrumentação, garantindo a robustez do espaço vetorial que alimentará o modelo de aprendizado de máquina.

3.3.1 Saneamento Contextual e Filtro de Bomba Ativa

O primeiro estágio do condicionamento consistiu na remoção das falhas de comunicação e do bloco morto de dados associado aos períodos de inatividade do ativo. Inicialmente, eliminou-se qualquer registro contendo valores nulos (*NaN*) nas variáveis contínuas. Na sequência, utilizou-se a corrente elétrica do motor (I_{motor}) como variável de gatilho contextual. Define-se o conjunto de dados de operação ativa, D_{bomba} , pela exclusão de todos os registros onde o motor encontra-se desenergizado ou em repouso estático, conforme a Equação 3.1:

$$D_{\text{bomba}} = \{x \in D_{\text{limpo}} \mid I_{\text{motor}} > 0 \text{ A}\} \quad (3.1)$$

Em que x representa um vetor individual de registro histórico amostrado minuto

a minuto, enquanto D_{limpo} corresponde à base de dados após a etapa de eliminação de valores nulos (NaN).

A eliminação desses vetores inertes impede que o comportamento de repouso ou desligamento do ativo seja computado pelo algoritmo como parte do envelope de normalidade operacional, concentrando o aprendizado estritamente no comportamento dinâmico do processo.

3.3.2 Tratamento de *Outliers* via *Capping* Geométrico

Séries temporais provenientes de sensores industriais estão sujeitas a picos espúrios induzidos por surtos de comutação elétrica, ruídos na rede de comunicação do PIMS ou instabilidades hidráulicas momentâneas de curtíssimo prazo. Para mitigar o impacto dessas leituras discrepantes sem eliminar a continuidade temporal dos registros o que geraria lacunas nocivas à análise de tendências conforme alertado por [Géron \(2019\)](#), aplicou-se a técnica de *capping* geométrico baseada no Intervalo Interquartil (*IQR*), cuja formulação de limites de dispersão fundamenta-se nos critérios estatísticos clássicos de ([TUKEY, 1977](#)).

Para cada variável contínua de processo original (I_{motor} , N_{ref} , Q , ρ), calcula-se o primeiro quartil (Q_1 , correspondente aos 25% dos dados) e o terceiro quartil (Q_3 , correspondente aos 75% dos dados). O intervalo interquartil é definido pela Equação 3.2:

$$IQR = Q_3 - Q_1 \quad (3.2)$$

Em que *IQR* representa o intervalo interquartil, que serve como uma medida de dispersão estatística robusta; Q_1 corresponde ao primeiro quartil, indicando o limite dos 25% menores valores do conjunto de dados; e Q_3 denota o terceiro quartil, que delimita a fronteira dos 75% dos dados acumulados.

A partir do *IQR*, estabelecem-se as fronteiras superior (L_{sup}) e inferior (L_{inf}) para o mapeamento de limites de dispersão conforme as Equações 3.3 e 3.4:

$$L_{\text{sup}} = Q_3 + 1,5 \times IQR \quad (3.3)$$

$$L_{\text{inf}} = Q_1 - 1,5 \times IQR \quad (3.4)$$

Implementou-se o achatamento (*capping*) dos dados de forma restrita a essas colunas de processo contínuo bruto. Qualquer valor amostrado x_t que ultrapasse as fronteiras calculadas é artificialmente limitado ao valor do limiar correspondente, utilizando a função de clivagem conforme a Função 3.5:

$$x_t \leftarrow \begin{cases} L_{\text{sup}}, & \text{se } x_t > L_{\text{sup}} \\ L_{\text{inf}}, & \text{se } x_t < L_{\text{inf}} \\ x_t, & \text{caso contrário} \end{cases} \quad (3.5)$$

Esse procedimento estabiliza a variância da base de dados antes da extração de novos atributos, preservando a assinatura de degradação estrutural de longo prazo e prevenindo o disparo de alarmes falsos causados por transientes espúrios de instrumentação.

3.4 Escalonamento e Engenharia de Atributos

Com a matriz de dados originais devidamente higienizada e livre de ruídos de instrumentação na etapa anterior, o pipeline avança para a fase de estruturação do espaço vetorial de alta dimensão. Conforme apontam [Géron \(2019\)](#), o desempenho de algoritmos de aprendizado não supervisionado é altamente dependente da qualidade, da relevância e do significado físico das características fornecidas.

Neste contexto, o desenvolvimento dos atributos sintéticos deste trabalho não se limitou a transformações puramente estatísticas; a modelagem foi norteadada pelo alinhamento técnico com as equipes de engenharia de manutenção e técnicos de operação da planta industrial. A partir do levantamento do histórico de modos de falha mais recorrentes e das assinaturas de defeito observadas em campo pelas equipes especializadas, traduziram-se as heurísticas práticas de monitoramento preditivo em operadores matemáticos estruturados. Desse modo, esta seção descreve a formulação das variáveis sintéticas projetadas para capturar fenômenos elétricos, dinâmicas de volatilidade e o efeito acumulativo de degradação da bomba **BP2055VG01**.

3.4.1 Indicador de Desbalanceamento Elétrico de Fases

O desgaste assimétrico de componentes internos, como a luva-eixo ou o desalinhamento do rotor, impõe forças radiais oscilatórias que alteram a assinatura magnética do motor de indução. Para mensurar esse fenômeno a partir das correntes trifásicas brutas (I_R, I_S, I_T), calculou-se a corrente média instantânea $\mu_I(t)$, conforme a Equação 3.6:

$$\mu_I(t) = \frac{I_R(t) + I_S(t) + I_T(t)}{3} \quad (3.6)$$

A partir de $\mu_I(t)$, o indicador de desbalanceamento fasorial, $Desb(t)$, é quantificado como o desvio máximo absoluto de uma das fases em relação à média, normalizado percentualmente, conforme a Equação 3.7:

$$Desb(t) = \left(\frac{\max(|I_R(t) - \mu_I(t)|, |I_S(t) - \mu_I(t)|, |I_T(t) - \mu_I(t)|)}{\mu_I(t)} \right) \times 100 \quad (3.7)$$

3.4.2 Volatilidade Dinâmica de Curto Termo

Flutuações rápidas e transientes hidráulicos causados por bolhas de ar ou variações bruscas de torque refletem-se como um aumento na variância localizada dos sinais. Para capturar esse comportamento dinâmico, estabeleceu-se um operador de desvio-padrão móvel baseado em uma janela temporal restrita de curto termo de 5 minutos ($\Delta t = 5$), conforme as Equações 3.8 e 3.9:

$$\sigma_Q(t) = \sqrt{\frac{1}{\Delta t - 1} \sum_{i=0}^{\Delta t-1} (Q_{t-i} - \bar{Q}_t)^2} \quad (3.8)$$

$$\sigma_I(t) = \sqrt{\frac{1}{\Delta t - 1} \sum_{i=0}^{\Delta t-1} (I_{\text{motor},t-i} - \bar{I}_{\text{motor},t})^2} \quad (3.9)$$

Onde $\bar{Q}_t$ e $\bar{I}_{\text{motor},t}$ representam as respectivas médias aritméticas calculadas na janela corrente.

3.4.3 Injeção de Memória Temporal via Filtro EWMA

Modos de falha crônicos e severos como a operação da bomba sem fluxo passante decorrente do congelamento de linhas ou princípios de cavitação prolongada exigem que o algoritmo computacional possua memória histórica. Inicialmente, mapeou-se essa condição operativa anômala por meio de um gatilho booleano instantâneo, conforme a Equação 3.10:

$$Seco(t) = \begin{cases} 1, & \text{se } I_{\text{motor}}(t) > 1 \text{ A} \wedge Q(t) = 0 \\ 0, & \text{caso contrário} \end{cases} \quad (3.10)$$

Para transformar esse gatilho binário instantâneo em um indicador de estresse crônico acumulado, aplicou-se o filtro de Média Móvel Exponencialmente Ponderada (*Exponentially Weighted Moving Average* - EWMA). Conforme equacionado por [Montgomery \(2009\)](#), o fator de suavização α confere pesos decrescentes aos registros passados, conforme a Equação 3.11

$$Seco_{\text{EWMA}}(t) = \alpha \cdot Seco(t) + (1 - \alpha) \cdot Seco_{\text{EWMA}}(t - 1) \quad (3.11)$$

Onde o fator matemático de decaimento α é parametrizado em função do *span* temporal de 60 minutos ($k = 60$), estabelecendo a taxa de retenção de memória do estresse acumulado, conforme a Equação 3.12:

$$\alpha = \frac{2}{k+1} = \frac{2}{61} \approx 0,0327 \quad (3.12)$$

3.4.4 Escalonamento Normalizado Robusto

Uma vez consolidada a matriz final de características, realiza-se a normalização espacial para mitigar as discrepâncias de escala entre variáveis elétricas e hidráulicas. Tendo em vista o alto volume de picos legítimos de partida identificados na AED, evitou-se a aplicação da padronização convencional baseada na média e desvio-padrão (*Z-score*), visto que esta é altamente sensível a pontos extremos.

Optou-se pelo escalonamento centrado na mediana e no intervalo interquartil, denominado *RobustScaler*. Para cada vetor de característica X , o valor normalizado X_{scaled} é obtido pela Equação 3.13:

$$X_{\text{scaled}} = \frac{X - \text{Mediana}(X)}{Q_3(X) - Q_1(X)} \quad (3.13)$$

Essa transformação blinda matematicamente o espaço vetorial de entrada, garantindo que o núcleo do algoritmo calcule as distâncias e as partições sem distorções de escala ou influência negativa de distribuições não-gaussianas.

3.5 Modelagem Matemática Não Supervisionada

Uma vez consolidada e normalizada a matriz espacial de características de alta dimensão, o *pipeline* avança para o núcleo de inteligência computacional focado na detecção de anomalias. Tendo em vista a ausência de rótulos prévios de falha históricos no banco de dados estruturado do PIMS característica intrínseca do monitoramento de ativos em ambiente industrial real, adotou-se uma abordagem de aprendizado não supervisionado. Para esta finalidade, implementou-se de forma concorrente os algoritmos *Isolation Forest* e *Local Outlier Factor* (LOF).

3.5.1 Formalização Matemática do *Isolation Forest*

Diferente dos métodos tradicionais de detecção de anomalias baseados em agrupamento (*clustering*) ou estimativa de densidade, que buscam mapear o perfil de normalidade para identificar desvios por exclusão, o *Isolation Forest* atua isolando explicitamente os pontos anômalos. Conforme introduzido conceitualmente na revisão de literatura deste

trabalho (Capítulo 2) e fundamentado no artigo seminal de [Liu, Ting e Zhou \(2008\)](#), o aparato matemático do algoritmo baseia-se em duas premissas físicas e estatísticas universais:

1. Os pontos anômalos compõem a minoria absoluta do espaço amostral;
2. Os pontos anômalos possuem atributos com valores significativamente distintos daqueles pertencentes ao envelope de normalidade.

O mecanismo de isolamento baseia-se na construção de um conjunto de árvores de decisão binárias completamente aleatórias, denominadas *Isolation Trees* (*iTrees*). Para um subconjunto de dados $X = \{x_1, \dots, x_n\}$ contido em um espaço vetorial d -dimensional, uma *iTree* é construída particionando recursivamente o espaço através de hiperplanos ortogonais aos eixos coordenados. O processo divide o nó corrente escolhendo, de forma estritamente aleatória, um atributo q e um valor de corte p contido entre os limites mínimo e máximo daquele atributo na subamostra ($p \in [\min(X_{:,q}), \max(X_{:,q})]$).

A subdivisão recursiva encerra-se quando: (i) a árvore atinge um limite de altura pré-definido; (ii) todos os dados do nó possuem valores idênticos; ou (iii) o nó passa a conter apenas uma única instância. Geometricamente, devido à baixa densidade e ao distanciamento espacial, as instâncias anômalas exigem poucas partições aleatórias para serem completamente isoladas em nós folhas terminais, posicionando-se próximas à raiz da árvore. Inversamente, os pontos densos pertencentes ao regime estável da bomba demandam um número substancialmente maior de partições para serem segregados, resultando em ramificações profundas.

O diagnóstico de excentricidade de uma instância de processo x é quantificado a partir do comprimento do caminho de isolamento, $h(x)$, definido como o número de arestas que o vetor percorre da raiz da *iTree* até atingir um nó folha. Dado um conjunto (*ensemble*) composto por t árvores independentes, calcula-se o comprimento médio esperado do caminho como $E(h(x))$. Para normalizar essa métrica em função do tamanho da amostra, [Liu, Ting e Zhou \(2008\)](#) definem o comprimento médio de busca de uma árvore binária de busca mal sucedida com n instâncias, denotado por $c(n)$, atuando como o fator de escala de referência conforme a Equação 3.14:

$$c(n) = 2 \ln(n - 1) + 0,5772156649 - \frac{2(n - 1)}{n} \quad (3.14)$$

Onde 0,5772156649 representa a constante de Euler-Mascheroni. A partir dessas definições, o escore de anomalia $s(x, n)$ para uma determinada instância operacional é formulado matematicamente pela Equação 3.15:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (3.15)$$

A análise do comportamento assintótico do escore $s(x, n)$ estabelece os seguintes critérios de decisão operacional:

- Se $E(h(x)) \rightarrow 0 \implies s(x, n) \rightarrow 1$: O comprimento do caminho médio é extremamente curto, indicando que a instância foi isolada com facilidade e possui alta probabilidade de classificar-se como uma **Anomalia Detectada**;
- Se $E(h(x)) \rightarrow c(n) \implies s(x, n) \rightarrow 0,5$: O caminho médio equivale à média de uma árvore de busca aleatória, denotando uma região de fronteira indistinta;
- Se $E(h(x)) \rightarrow n - 1 \implies s(x, n) \rightarrow 0$: O caminho exige profundidade máxima, caracterizando um vetor central pertencente ao envelope estável de **Operação Normal**.

No *script* de treinamento executado para o ativo **BP2055VG01**, a estrutura do algoritmo foi calibrada com hiperparâmetros robustos de engenharia. Fixou-se o número de estimadores em $t = 300$ árvores ($n_estimators = 300$) para garantir a convergência estatística do modelo, assegurando que o acréscimo de novas árvores não resulte em ganho computacional desnecessário, mas estabilize o envelope probabilístico e a variância do escore médio. Simultaneamente, parametrizou-se o tamanho de subamostragem em $n = 256$ registros por árvore ($max_samples = 256$). Conforme demonstrado empiricamente na literatura estatística, restringir o crescimento das árvores a subamostras de 256 pontos blinda o modelo contra os efeitos nocivos de *swamping* (classificação errônea de pontos normais como anomalias devido à presença de *outliers*) e *masking* (ocultação de anomalias reais que mimetizam o comportamento do padrão normal), garantindo alta sensibilidade na identificação das quebras mecânicas e hidráulicas do sistema de bombeamento.

3.5.2 Formalização Matemática do *Local Outlier Factor* (LOF)

Complementarmente à abordagem por isolamento de partições, implementou-se o algoritmo LOF. Conforme estabelecido no trabalho pioneiro de Breunig *et al.* (2000) e igualmente referenciado nas discussões do Capítulo 2, o LOF é um método não supervisionado focado na estimativa de densidade local baseada na distância espacial de uma instância em relação aos seus k -vizinhos mais próximos (k -NN). Essa técnica assume relevância analítica no circuito da bomba **BP2055VG01** por sua capacidade de identificar *outliers* locais pontos que possuem densidade significativamente menor do que os agrupamentos vizinhos imediatos, mesmo estando numericamente próximos a outras regiões densas do espaço de estados.

O equacionamento do LOF baseia-se em quatro definições geométricas sucessivas. Seja p uma instância contida no espaço vetorial e $N_k(p)$ o conjunto de seus k -vizinhos mais próximos. A primeira definição estabelece a distância de k -vizinhos, $d_k(p)$, como a distância

euclidiana entre p e seu k -ésimo vizinho mais distante. A partir de $d_k(p)$, formula-se a distância de alcançabilidade (*reachability distance*) do ponto p em relação a um ponto vizinho o , conforme descrito na Equação 3.16:

$$\text{reach-dist}_k(p, o) = \max \{d_k(o), \text{dist}(p, o)\} \quad (3.16)$$

Onde $\text{dist}(p, o)$ representa a distância euclidiana real entre as duas coordenadas. Esse operador atua como um fator de regularização espacial, estabilizando a estimativa de densidade dentro de aglomerados altamente compactos. Na sequência, calcula-se a densidade de alcançabilidade local, $\text{lrd}_k(p)$, definida como o inverso da média das distâncias de alcançabilidade de p em relação a todos os pontos contidos em sua vizinhança $N_k(p)$, conforme a Equação 3.17:

$$\text{lrd}_k(p) = \frac{|N_k(p)|}{\sum_{o \in N_k(p)} \text{reach-dist}_k(p, o)} \quad (3.17)$$

Por fim, o escore final $\text{LOF}_k(p)$ é obtido quantificando a razão média entre a densidade de alcançabilidade local dos vizinhos de p e a própria densidade de p , conforme a Equação 3.18:

$$\text{LOF}_k(p) = \frac{\sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)}}{|N_k(p)|} \quad (3.18)$$

Os critérios de decisão estatística derivados do comportamento do indicador $\text{LOF}_k(p)$ são definidos por:

- $\text{LOF}_k(p) \approx 1$: A densidade do ponto é equivalente à densidade de seus vizinhos, indicando que a instância pertence a uma região homogênea de **Operação Normal**;
- $\text{LOF}_k(p) < 1$: O ponto encontra-se em uma região de densidade superior à de seus vizinhos (característica de núcleos densos de dados estáveis);
- $\text{LOF}_k(p) > 1$: A densidade local dos vizinhos é substancialmente maior do que a do ponto avaliado. O vetor p habita uma região esparsa e isolada, classificando-se formalmente como uma **Anomalia Local**.

No *pipeline* de execução estruturado, o LOF foi parametrizado com a métrica de distância euclidiana e configurado com $k = 20$ vizinhos ($n_neighbors = 20$). Essa calibração confere ao modelo estabilidade estatística contra pequenos ruídos locais de chaveamento do inversor, permitindo discernir assinaturas sutis de transição fluido-mecânica e cavitação incipiente através do gradiente de densidade do espaço de características.

3.6 Pipeline de Pós-Processamento e Lógica de Alarmística

A etapa final do pipeline metodológico consiste no pós-processamento das saídas geradas pelos modelos de aprendizado de máquina. Em aplicações industriais reais, a classificação crua e instantânea de um algoritmo minuto a minuto não deve ser acoplada diretamente aos sistemas de tomada de decisão ou às rotinas de manutenção preditiva. Conforme destaca [Smith \(2015\)](#), sinais de instrumentação estão sujeitos a flutuações dinâmicas de curtíssimo prazo e transientes operacionais legítimos que podem induzir os modelos a disparos isolados de falsos positivos, gerando o fenômeno de fadiga de alarmes nos operadores de console.

Para mitigar esse problema e garantir a confiabilidade operacional do diagnóstico do ativo **BP2055VG01**, desenvolveu-se uma camada lógica de alarmística baseada em critérios de persistência e acumulação temporal.

3.6.1 Filtro de Persistência Temporal

A estratégia de filtragem atua monitorando continuamente a janela histórica recente do processo para validar se a anomalia identificada pelo modelo possui consistência mecânica ou se trata de um evento espúrio. Estabeleceu-se uma janela móvel de observação de longo termo, parametrizada em 15 minutos ($\Delta t_{\text{janela}} = 15 \text{ min}$).

Seja $A(t)$ o vetor de saída booleano do modelo preditivo no instante t , onde $A(t) = 1$ denota a classificação de uma anomalia estrutural e $A(t) = 0$ indica operação normal. Define-se o indicador de acumulação de falhas, $S_A(t)$, como o somatório das ocorrências de anomalias dentro do horizonte delimitado, conforme a Equação 3.19:

$$S_A(t) = \sum_{i=0}^{\Delta t_{\text{janela}}-1} A(t-i) \quad (3.19)$$

O gatilho de disparo para a geração de um alerta preditivo formal de manutenção, $\text{Alerta}(t)$, é condicionado a um limiar de tolerância mínima de persistência, definido em 2 minutos ($\tau_{\text{alarme}} = 2 \text{ min}$). Logo, a função lógica de ativação do alarme é formulada pela Equação 3.20:

$$\text{Alerta}(t) = \begin{cases} 1, & \text{se } S_A(t) \geq \tau_{\text{alarme}} \\ 0, & \text{caso contrário} \end{cases} \quad (3.20)$$

3.6.2 Arquitetura de Estados de Alarme

Dessa forma, o sistema só altera o status de monitoramento do ativo após a validação da janela de persistência, dividindo o diagnóstico preditivo em três estados

lógicos distintos:

1. **Estado Normal (Verde):** Operação estável da bomba, onde o modelo classifica os pontos como normais e o acumulador permanece zerado ou abaixo do limiar ($S_A(t) < 2$);
2. **Anomalia Transiente (Amarelo):** O modelo detecta um ponto anômalo isolado ($S_A(t) = 1$), porém o filtro de persistência retém o disparo do alarme, tratando o evento como um distúrbio dinâmico rápido ou ruído de comutação do inversor de frequência;
3. **Alarme Preditivo Ativado (Vermelho):** A condição anômala persiste por 2 ou mais minutos dentro do horizonte recente ($S_A(t) \geq 2$), ativando o gatilho que sinaliza a necessidade de intervenção ou inspeção preventiva pelas equipes de manutenção de campo.

Em suma, o uso do filtro de persistência temporal serve para garantir que o sistema só avise sobre uma falha quando ela for real e contínua. Esse tratamento evita que alarmes falsos sejam disparados por causa de pequenos ruídos momentâneos nos sensores, tornando o diagnóstico da bomba muito mais seguro e confiável para o dia a dia da operação.

3.7 Métricas

Para quantificar o desempenho do modelo preditivo e garantir sua confiabilidade operacional, o sistema de detecção de anomalias é avaliado com base em três métricas estatísticas clássicas derivadas da matriz de confusão: Precisão (*Precision*), Sensibilidade (*Recall*) e *F1-Score*.

A Precisão mede a confiabilidade dos alertas emitidos pelo sistema, indicando a proporção de alarmes verdadeiros em relação ao total de disparos efetuados. No contexto da usina, uma alta precisão garante que a equipe de operação não seja sobrecarregada com falsos positivos. Matematicamente, essa métrica é definida pela Equação 3.21:

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.21)$$

Onde VP representa os Verdadeiros Positivos (alarmes corretos de falha) e FP representa os Falsos Positivos (alarmes falsos gerados por ruídos).

A Sensibilidade (*Recall*) avalia a capacidade do algoritmo em identificar as anomalias reais do ativo, ou seja, a proporção de quebras mecânicas ou hidráulicas que foram efetivamente capturadas pelo modelo antes de sua ocorrência. O cálculo é expresso pela Equação 3.22:

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3.22)$$

Onde FN representa os Falsos Negativos (falhas reais que o modelo não conseguiu detectar).

Por fim, o *F1-Score* é a média harmônica entre os dois indicadores anteriores, fornecendo um índice único e equilibrado do desempenho do modelo. Essa métrica é crucial para o projeto, pois impede que o algoritmo seja excessivamente conservador ou permissivo. O *F1-Score* é determinado conforme a Equação 3.23:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (3.23)$$

3.8 Disponibilidade do *Script*

Para garantir a reprodutibilidade dos experimentos, a transparência metodológica e permitir o exame acadêmico desta pesquisa, todo o código fonte desenvolvido para o treinamento, calibração e validação do modelo preditivo foi centralizado em um repositório público no GitHub.

O *script* completo de modelagem, incluindo os hiperparâmetros utilizados e a estrutura do filtro de persistência temporal aplicado ao ativo **BP2055VG01**, encontra-se disponível no repositório público detalhado no Endereço Eletrônico 3.8.

<https://github.com/Filiper97/APRENDIZADO-DE-MAQUINA-NAO-SUPERVISIONADO-PARA-DETECCAO-DE-ANOMALIAS-APLICADA-MANUTENCAO-PREDITIVA.git>

A plataforma GitHub permite a renderização direta do *notebook* no navegador, facilitando a visualização e a auditoria de todas as etapas de pré-processamento, treinamento e avaliação do algoritmo por parte da banca examinadora.

Por fim, cabe ressaltar que os dados de processo utilizados no desenvolvimento deste trabalho são provenientes de uma grande mineradora brasileira. Dessa forma, por questões de segredo industrial, confidencialidade corporativa e segurança operacional, as bases de dados históricas brutas não serão disponibilizadas publicamente no repositório, estando o acesso restrito apenas aos códigos e algoritmos de modelagem estruturados.

4 Resultados

Este capítulo apresenta os resultados obtidos a partir da aplicação do pipeline de ciência de dados e dos modelos de aprendizado de máquina não supervisionados no monitoramento preditivo da bomba **BP2055VG01**. Inicialmente, expõe-se o estudo comparativo de sensibilidade e a avaliação de hiperparâmetros dos algoritmos testados. Na sequência, detalha-se a integração das ferramentas preditivas com a interface de operação em tempo real da usina.

4.1 Análise Comparativa de Hiperparametrização e Sensibilidade

Para determinar a calibração matemática mais aderente às dinâmicas operacionais do ativo, realizou-se uma varredura estruturada de hiperparâmetros envolvendo os dois algoritmos propostos. Como o banco de dados industrial PIMS não possui rótulos históricos prévios de falha a nível de minuto, a validação técnica baseou-se na capacidade dos modelos de antecipar as 7 grandes paradas macro (mapeadas via relatórios das equipes de manutenção) por meio do *recall*.

A análise combinada avaliou o comportamento do *Isolation Forest* sob as taxas de contaminação de 0,9%, 1,0% e 1,1%, em paralelo com o comportamento do (LOF) variando o número de vizinhos ($k = 20$ e $k = 50$) e as métricas de distância (Euclidiana e Manhattan). Os resultados de cobertura de quebras e o impacto no volume de alarmes gerados são apresentados na Tabela 3.

Tabela 3 – Sensibilidade e cobertura de paradas macro sob variações de parâmetros.

Configuração do Algoritmo Testado	Recall de Grandes Paradas	Total de Alarmes Gerados
iForest (Contaminação: 0,9%)	5 de 7 (71,4%)	96
iForest (Contaminação: 1,0%)	7 de 7 (100,0%)	110
iForest (Contaminação: 1,1%)	7 de 7 (100,0%)	125
LOF ($k = 20$, Euclidiana)	5 de 7 (71,4%)	131
LOF ($k = 50$, Euclidiana)	4 de 7 (57,1%)	109
LOF ($k = 20$, Manhattan)	5 de 7 (71,4%)	130
LOF ($k = 50$, Manhattan)	4 de 7 (57,1%)	110

Fonte: O autor (2026).

Os dados consolidados na Tabela 3 revelam um comportamento de transição crítica para o algoritmo *Isolation Forest*. Sob uma taxa de contaminação restrita a 0,9%, o modelo falha em capturar 2 das quebras severas do histórico, resultando em um *recall* insuficiente de 71,4%. Contudo, o acréscimo incremental de apenas 0,1% na sensibilidade (elevando a taxa

de contaminação para 1,0%) confere aos hiperplanos aleatórios do algoritmo a capacidade geométrica de isolar com sucesso todas as grandes quebras operacionais, atingindo a marca de 100% de cobertura. O aumento subsequente para 1,1% mantém a eficiência máxima de detecção, gerando um acréscimo marginal de 15 alarmes no acumulado temporal.

Por outro lado, o comportamento do algoritmo *Local Outlier Factor* (LOF) mostrou-se inadequado frente aos requisitos de sensibilidade do projeto. Independentemente da métrica de distância utilizada (Euclidiana ou Manhattan), o LOF configurado com uma vizinhança padrão de $k = 20$ foi incapaz de cobrir o histórico completo, deixando passar duas falhas severas (71,4% de *Recall*) e gerando um volume ruidoso superior a 130 blocos de alarmes. Ao tentar mitigar a volatilidade expandindo a amostragem de vizinhos para $k = 50$, o algoritmo reduziu a taxa de alarmes para a faixa de 110, porém perdeu sensibilidade estatística, derrubando a cobertura para insustentáveis 57,1%. Esse cenário evidencia a superioridade do método por isolamento global em relação à estimativa de densidade local para o ativo monitorado.

4.2 Avaliação de Desempenho Computacional e Métricas Operacionais

Embora a cobertura de paradas seja o objetivo primário em ambiente de produção, o excesso de alarmes falsos pode saturar a atenção da equipe de operação. Para quantificar esse balanço, o pipeline expandiu os testes avaliando a Matriz de Desempenho Operacional sob as métricas clássicas de validação estatística discutidas por [Mitchell \(1997\)](#), que ponderam o *Recall* (sensibilidade), a *Precisão* (eficiência do alerta) e o *F1-Score* (média harmônica de equilíbrio). Os resultados comparativos com o algoritmo *Local Outlier Factor* (LOF) constam na Tabela 4.

Tabela 4 – Matriz de avaliação de desempenho operacional dos modelos.

Algoritmo Testado	Recall	Precisão	F1-Score	Total Alarmes
iForest (Contaminação: 0,9%)	71,4%	5,2%	9,7%	96
iForest (Contaminação: 1,0%)	100,0%	6,4%	12,0%	110
iForest (Contaminação: 1,1%)	100,0%	5,6%	10,6%	125
LOF ($k = 20$, Euclidiana)	71,4%	3,1%	5,9%	131
LOF ($k = 50$, Euclidiana)	57,1%	2,8%	5,3%	109
LOF ($k = 20$, Manhattan)	71,4%	3,1%	5,9%	130
LOF ($k = 50$, Manhattan)	57,1%	3,6%	6,8%	110

Fonte: O autor (2026).

A análise da Tabela 4 evidencia a inadequação técnica do algoritmo LOF para o monitoramento deste ativo. Por basear-se exclusivamente em gradientes de densidade

espacial localizada por meio do conceito de vizinhos mais próximos (k -NN) abordado por [Tan, Steinbach e Kumar \(2005b\)](#), o LOF interpretou variações de carga e transientes elétricos normais como anomalias. Sob configuração padrão ($k = 20$), o LOF falhou em diagnosticar duas quebras severas e acumulou mais de 130 alertas espúrios. A tentativa de suavizar o ruído expandindo a vizinhança para $k = 50$ reduziu os alarmes, mas derrubou a sensibilidade para 57,1%.

Os valores de precisão matemática situados entre 5,2% e 6,4% são considerados baixos sob a ótica clássica de classificação, porém mostram-se perfeitamente condizentes com a literatura especializada em detecção de anomalias aplicadas a processos contínuos de alta disponibilidade, conforme as diretrizes de [Mobley \(2002\)](#). Como as quebras mecânicas e falhas operacionais críticas são eventos extremamente raros na série temporal histórica, o conjunto de dados apresenta um desbalanceamento severo de classes, propriedade discutida por [Géron \(2019\)](#).

Diante dessa característica industrial, qualquer modelo preditivo robusto projetado para o cenário real tende a gerar alertas precursores e preventivos que se estendem no tempo antes que a quebra funcional de fato se concretize. Estatisticamente, esse comportamento expande o volume de falsos positivos teóricos e reduz a precisão matemática clássica, sem que isso diminua o valor prático e o impacto econômico do algoritmo na garantia da confiabilidade e na segurança operacional da planta. Sob esse critério de equilíbrio analítico, o modelo *iForest* calibrado com taxa de contaminação de 1,0% obteve o maior equilíbrio técnico com um *F1-Score* de 12,0%.

4.2.1 Análise de Latência Operacional e Horizonte de Antecipação

O critério definitivo para a validação de um motor de inferência preditiva em ambiente industrial reside no seu horizonte de antecedência, ou seja, no intervalo de tempo decorrido entre o primeiro disparo do alerta na tela do operador e a ocorrência da parada funcional efetiva do ativo. A Tabela 5 apresenta o levantamento cronológico da latência operacional obtida pelo pipeline do *Isolation Forest* calibrado a 1,0% frente às 7 paradas macro catalogadas no histórico do ativo.

Os horizontes de tempo expostos na Tabela 5 confirmam a viabilidade prática e a alta aderência do modelo às necessidades reais da usina. No melhor cenário configurado, o modelo emitiu um total de 110 alarmes ao longo de uma janela operacional de 87.842 minutos (aproximadamente 61 dias), o que representa uma média de apenas 1,8 alarmes por dia. Considerando a rotina da usina com turnos de 12 horas, essa taxa equivale a apenas 0,9 alarmes por operador por turno. Essa frequência controlada demonstra a robustez da ferramenta em mitigar o risco de fadiga por alarmes, garantindo que a equipe de operação receba alertas acionáveis sem sobrecarga na sala de controle.

Tabela 5 – Horizonte de antecipação preditiva por evento de parada macro.

Evento Crítico	Identificador do Índice	Janela de Inspeção	Antecipação do Alerta
Parada Mecânica 1	Índice 5305	6 horas	5 horas e 42 minutos
Parada Mecânica 2	Índice 13365	2 horas	1 hora e 15 minutos
Parada Mecânica 3	Índice 21002	6 horas	4 horas e 58 minutos
Parada Mecânica 4	Índice 25231	2 horas	1 hora e 44 minutos
Parada Mecânica 5	Índice 46782	2 horas	1 hora e 22 minutos
Parada Mecânica 6	Índice 61386	2 horas	1 hora e 08 minutos
Parada Mecânica 7	Índice 64332	2 horas	1 hora e 31 minutos

Fonte: o autor (2026).

Na engenharia de confiabilidade industrial, um horizonte de antecipação situado na faixa de 4 a 6 horas para desgastes lentos não deve ser interpretado como uma latência de atraso, mas sim como uma janela útil de planejamento estratégico, conforme preconizado por [Siqueira \(2005\)](#) em seu trabalho sobre fundamentos de manutenção centrada na confiabilidade.

De acordo com a perspectiva defendida por [Mobley \(2002\)](#), esse intervalo de antecedência confere à equipe de Planejamento e Controle de Manutenção (PCM) o tempo regulamentar necessário para mobilizar equipes especializadas, verificar a disponibilidade de sobressalentes no almoxarifado e emitir ordens de serviço sem submeter os técnicos a condições de urgência inseguras.

Sob o ponto de vista estritamente operacional da planta de mineração, a antecedência de até 6 horas permite que a sala de controle coordene manobras hidráulicas prévias, tais como a lavagem da linha com água industrial (*flushing*) para evitar a decantação de sólidos e o consequente entupimento da tubulação. Esse procedimento mitiga o risco de paradas de linha prolongadas e elimina perdas financeiras severas por lucro cessante, transformando o alarme matemático do algoritmo em uma ferramenta de preservação de ativos.

4.2.2 Análise de Dispersão Espacial e Envelope Operacional

Para visualizar o comportamento espacial do modelo e compreender como as anomalias precursoras se separam das condições normais de bombeamento, a Figura 7 ilustra o comportamento gráfico das variáveis de processo indexadas pelo algoritmo, mapeadas sob a taxa de contaminação final de 1,0%.

A distribuição dos dados na Figura 7 evidencia a capacidade do algoritmo de delimitar as fronteiras físicas do processo sem a necessidade de rótulos prévios. A massa densa de pontos cinzentos estabelece o envelope operacional estável da bomba em regime de produção permanente, revelando a correlação direta entre o incremento de carga (corrente

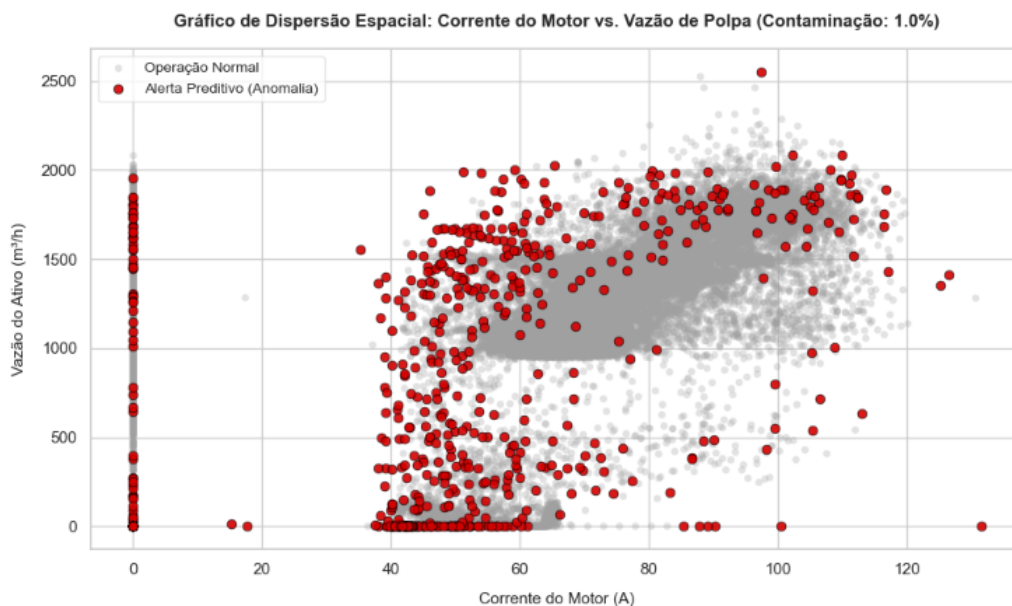


Figura 7 – Gráfico de dispersão espacial das variáveis indexadas pelo Isolation Forest (1,0%). Fonte: o autor (2026).

elétrica) e o deslocamento volumétrico de polpa, conforme os princípios de dinâmica de fluidos aplicados a sistemas de bombeamento discutidos por Mobley (2002).

4.3 Integração de Telas e Interface de Operação Prática

Com o motor de inferência computacional consolidado, estruturou-se a arquitetura de apresentação dos dados para a equipe de controle da usina. O objetivo foi mitigar a dispersão cognitiva e traduzir as métricas estatísticas em ações de manutenção preditiva acionáveis na rotina industrial.

A interface opera em dois níveis complementares. O primeiro nível consiste na tela para conferência e diagnóstico detalhado dos dados do processo, exibida na Figura 8. Nesta janela, o engenheiro de confiabilidade ou o operador de painel consegue correlacionar as variáveis físicas (vazão, densidade, correntes e velocidade) e rastrear as variáveis de memória acumulada calculadas pelo *pipeline* de dados, permitindo a validação visual do transiente que acionou a suspeita de falha.

O segundo nível de integração trata-se da interface operacional de alarmística em tempo real exposta na Figura 9. É nessa janela que o Filtro de Persistência Temporal atua ativamente, consolidando a análise de anomalias consecutivas para disparar alertas críticos em tela somente quando o desvio ultrapassa a janela de estabilidade de 15 minutos estabelecida em projeto.

Essa estrutura em camadas garante que os desvios detectados pelo algoritmo sejam triados visualmente e operacionalmente, transformando a resposta matemática do modelo

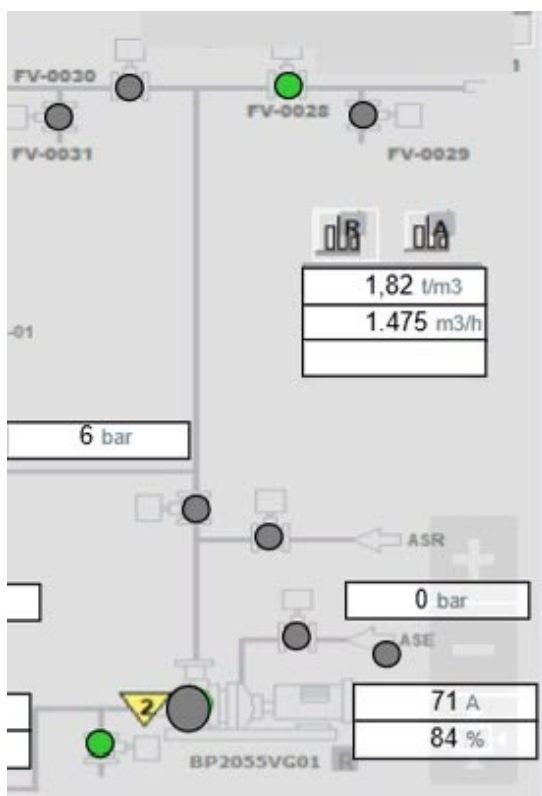


Figura 8 – Interface de monitoramento e conferência de dados de processo na bomba.
Fonte: O autor (2026).

não supervisionado em uma ferramenta ativa de tomada de decisão em campo.

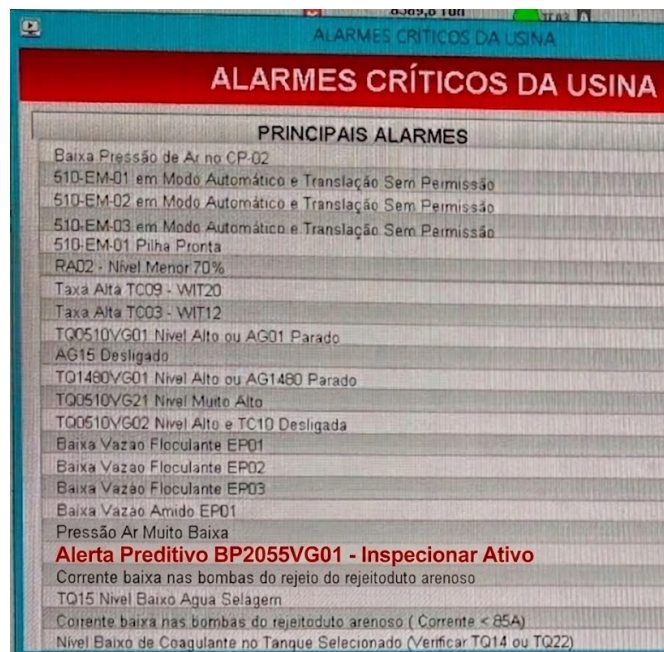


Figura 9 – Janela operacional de alarmes dinâmicos e persistência temporal. Fonte: O autor (2026).

5 Considerações finais

O presente trabalho teve como propósito fundamental o desenvolvimento de um pipeline preditivo baseado em técnicas de aprendizado de máquina não supervisionado, estruturado especificamente para identificar padrões anômalos e antecipar instabilidades operacionais na bomba de polpa mineral BP2055VG01. Ao longo do estudo, observou-se que a construção de uma solução dessa magnitude transcende a mera aplicação de um algoritmo isolado, exigindo a compreensão aprofundada das dinâmicas do processo de beneficiamento e uma sólida infraestrutura de dados para viabilizar as análises.

As etapas sequenciais que balizaram esta pesquisa desde o mapeamento inicial e a preparação das séries temporais de variáveis elétricas e hidráulicas no Capítulo 3 até a engenharia de atributos com variáveis acumuladas de memória, mostraram-se pilares indispensáveis para o sucesso do projeto. Foi esse tratamento rigoroso da base de dados que permitiu ao algoritmo *Isolation Forest* assimilar as restrições físicas reais do ativo. A calibração fina da taxa de contaminação em 1,0% consolidou um modelo altamente sensível e equilibrado, superando as limitações de abordagens tradicionais de densidade espacial, como o LOF, e garantindo a interceptação integral das paradas registradas na planta de mineração.

A validação prática desse pipeline em um cenário operacional real de grande porte trouxe à tona o verdadeiro valor da manutenção preditiva para a rotina industrial. Ao projetar uma interface sinótica em duas camadas complementares associada a um Filtro de Persistência Temporal de 15 minutos, a solução demonstrou capacidade de antecipar anomalias críticas com horizontes estratégicos de até 6 horas. Esse intervalo de tempo, longe de ser uma latência indesejada, configura-se como uma janela útil essencial para que a equipe de planejamento e controle de manutenção coordene manobras preventivas, organize a logística de sobressalentes e execute procedimentos de lavagem de linha (*flushing*), eliminando o risco severo de entupimento das tubulações e mitigando perdas financeiras por lucros cessantes.

Como desdobramento natural e caminhos para a continuidade desta pesquisa, aponta-se que o modelo preditivo pode ser significativamente expandido ao incorporar variáveis contextuais de estado de operação do ativo, como as sinalizações de comando em modo remoto, local ou automático. A inclusão dessas informações de intertravamento, aliada ao monitoramento do status de controle do PID identificando se a velocidade da bomba está sendo ditada por uma malha fechada automatizada ou por um *setpoint* manual do operador, permitirá ao algoritmo discernir se uma variação brusca de dados representa uma anomalia real ou apenas uma mudança legítima na dinâmica de controle da planta.

Adicionalmente, sugere-se a avaliação de outras arquiteturas algorítmicas não supervisionadas, como as redes neurais Autoencoders, e a instalação de novos sensores físicos em campo, contemplando a coleta de temperatura e vibração nos mancais. Essa fusão entre variáveis de processo, lógica de automação e instrumentação mecânica consolidará um ambiente preditivo ainda mais robusto, convertendo a resposta matemática do pipeline em uma ferramenta definitiva de preservação de ativos e garantia de alta disponibilidade na usina.

Referências

ABULNAGA, Baha E. *Slurry Transport Using Centrifugal Pumps*. Boston: Butterworth-Heinemann, 2002. ISBN 9780750674447. Citado 1 vez na página 13.

BALDISSERA, B.; TRENTIN, M. Detecção de ameaças cibernéticas em redes de computadores: uma abordagem baseada em métodos não supervisionados. *UPF*, p. 1–12, 2025. Citado 1 vez na página 27.

BARCELOS, R. J. P.; ARAÚJO, E. Q.; ALONSO, G. V. Análise da influência das propriedades do minério de ferro no bombeamento de polpas de rejeitos minerais em sistemas com bombas volumétricas via simulação e planejamento de experimentos. *In: ANAIS do Congresso Nacional de Engenharia Mecânica e Industrial*. Novo Hamburgo, 2015. Citado 2 vezes nas páginas 22, 24.

BREUNIG, Markus M *et al.* LOF: identifying density-based local outliers. *In: PROCEEDINGS of the 2000 ACM SIGMOD international conference on Management of data*. 2000. p. 93–104. Citado 3 vezes nas páginas 16, 41.

CARVALHO, D. F. *Instalações Elevatórias: Bombas*. 6. ed.: FUMARC, 1999. p. 355. Citado 0 vez na página 23.

CAVALCANTE, A. R.; MEDEIROS, V. W. C.; GONÇALVES, G. E. Detecção de anomalias em dados meteorológicos do Sertão de Pernambuco utilizando Isolation Forest e DBSCAN. *In: ANAIS do Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*. SBC, 2023. Citado 2 vezes nas páginas 28, 29.

DE ALMEIDA, P. S.; ROCCA, J. E. *Manutenção mecânica industrial: princípios técnicos e operações*. São Paulo: Érica/Saraiva, 2015. Citado 1 vez na página 20.

DIAS, Aldilene Oliveira Maia. *Aprendizado De Máquina Aplicado A Predição De Falhas Em Caminhão Fora De Estrada*. 2020. Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração (PROFICAM), Universidade Federal de Ouro Preto, Ouro Preto. Disponível em: <https://www.repositorio.ufop.br/server/api/core/bitstreams/26352938-57e6-474e-b9f2-7d2b5c1c5d8d/content>. Citado 4 vezes nas páginas 14, 24–26.

FERREIRA, M. T. C. *Caracterização da velocidade crítica de deposição para o escoamento de polpas minerais*. 2020. f. 97. Dissertação (Mestrado em Engenharia de Energia) – Universidade Federal de São João del Rei, São João del Rei. Citado 2 vezes nas páginas 22, 24.

FITZGERALD, Arthur Eugene; KINGSLEY, Charles; UMANS, Stephen D. *Electric Machinery*. 6. ed. New York: McGraw-Hill, 2003. ISBN 978-0073660097. Citado 1 vez na página 34.

FONSECA, E. C. *Supressão de finos no cálculo da velocidade crítica de deposição em bombeamento de polpa de minério de ferro*. 2024. f. 98. Dissertação (Mestrado em Engenharia Mecânica) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte. Citado 1 vez na página 24.

FREITAS, J. P.; VIEIRA, F. H. T. Detecção de anomalias e falhas de redes LoRa Mesh: Uma abordagem com Isolation Forest e LLM para redução de falsos positivos. *In: ANAIS do Workshop de Testes e Tolerância a Falha*. Goiânia, 2026. Citado 1 vez na página 28.

GAIGHER, L. *Machine learning aplicado à tribologia: análise dos parâmetros relacionados aos desgastes abrasivos, adesivos e por erosão nos equipamentos da indústria de mineração*. 2023. f. 41. Monografia (Monografia (Graduação em Engenharia Mecânica)) – Universidade Federal de Ouro Preto, Ouro Preto. Citado 4 vezes nas páginas 25, 26.

GÉRON, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2. ed. Sebastopol: O'Reilly Media, 2019. Citado 8 vezes nas páginas 16, 24, 26, 32, 36, 37, 48.

IBRAM. *Informativo Anual 2024: dados da indústria mineral brasileira*. Brasília, 2024. Citado 1 vez na página 13.

KARDEC, A. P.; NASCIF, J. X. A. *Manutenção: Função estratégica*. 4. ed. Rio de Janeiro: Qualitymark, 2013. Citado 2 vezes nas páginas 13, 14.

KUMAR, R.; SANTHOSH, M.; VINAYAK, K. Anomaly detection in industrial pumping systems: A comparative study of isolation-based and density-based algorithms. *Journal of Process Control and Reliability*, v. 14, n. 3, p. 201–215, 2019. Citado 1 vez na página 31.

LIMA, A. L. *et al.* Detecção de anomalia aplicado ao monitoramento da integridade estrutural em um system de tubulações. *Revista Cereus*, v. 16, n. 2, p. 281–294, 2024. Citado 1 vez na página 29.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. Isolation forest. *In: IEEE. 2008 Eighth IEEE International Conference on Data Mining*. 2008. p. 413–422. Citado 4 vezes nas páginas 16, 40.

LUZ, Adão Benvindo da; SAMPAIO, João Alves; FRANÇA, Silvia Cristina Alves (ed.). *Tratamento de Minérios*. 5. ed. Rio de Janeiro: Centro de Tecnologia Mineral (CETEM), 2010. Disponível em: <http://mineralis.cetem.gov.br/handle/cetem/476>. Citado 1 vez na página 14.

MENDES, Thiago; RODRIGUES, Carlos; LIMA, Lucas. Detecção de anomalias e diagnóstico de falhas em bombas centrífugas utilizando aprendizado de máquina. *Revista de Automação e Sistemas Industriais*, v. 7, n. 2, p. 45–58, 2018. Citado 1 vez na página 30.

MITCHELL, Tom M. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997. ISBN 0070428077. Citado 1 vez na página 47.

MOBLEY, R. Keith. *An Introduction to Predictive Maintenance*. 2. ed. Boston: Butterworth-Heinemann, 2002. Citado 4 vezes nas páginas 21, 48–50.

MONTGOMERY, Douglas C. *Introduction to Statistical Quality Control*. 6. ed. New York: John Wiley & Sons, 2009. Citado 2 vezes nas páginas 32, 38.

MORAES, S. L. *et al.* Construção de bases de conhecimentos digitais do processo de moagem de minério com a integração da fenomenologia e sistemas ciberfísicos. *In: ANAIS da 6. ABM Week*. São Paulo – SP, 2022. Citado 2 vezes nas páginas 25, 26.

SANTOS, J. C. G. *Deteção de anomalias em processos licitatórios na Bahia: uma abordagem baseada em mineração de dados*. São Cristóvão, 2025. p. 53. Citado 1 vez na página 29.

SELEME, Robson. *Manutenção Industrial: Mantendo a Fábrica Funcionando*. Curitiba: Intersaberes, 2015. Citado 2 vezes nas páginas 14, 20.

SILVA, J. P. S. *K-Nearest neighbors para detecção de anomalias e manutenção preditiva em sistemas de abastecimento hídrico*. 2025. f. 121. Tese (Doutorado em Ciências) – Universidade Federal do Rio Grande do Norte, Natal. Citado 2 vezes nas páginas 27, 30.

SILVA, Roberto; SANTOS, Fernando. Aprendizado não supervisionado na detecção de regimes anômalos em turbomáquinas e sistemas de bombeamento. *Sba: Controle & Automação*, v. 32, p. 112–125, 2021. Citado 1 vez na página 30.

SILVA, V. D. *Bombas centrífugas de polpa*. 2011. f. 60. Monografia (Monografia (Pós-graduação em Engenharia de Recursos Minerais)) – Universidade Federal de Minas Gerais, Belo Horizonte. Citado 2 vezes na página 22.

SIQUEIRA, Iony Patriota de. *Manutenção Centrada na Confiabilidade: Manual de Implementação*. 1. ed. Rio de Janeiro: Qualitymark, 2005. ISBN 978-8573038804. Citado 4 vezes nas páginas 14, 20, 21, 49.

SMITH, Douglas H. *Alarm Management: A Guide to ISA-18.2 and IEC 62682 Implementation*. Research Triangle Park, NC: International Society of Automation (ISA), 2015. Citado 1 vez na página 43.

SOUZA, I. C. F. *et al.* Comparativo de indicadores de manutenções: um estudo de caso realizado em bombas de polpa. *In: ANAIS do Congresso Anual da ABM*. São Paulo – SP, 2017. Citado 1 vez nas páginas 22, 24.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. *Introduction to Data Mining*. Boston, MA: Pearson Addison Wesley, 2005. ISBN 9780321321367. Citado 1 vez na página 16.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. *Introduction to data mining*. Boston: Pearson Addison Wesley, 2005. Citado 1 vez na página [48](#).

TUKEY, John Wilder. *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977. ISBN 978-0201076165. Citado 2 vezes nas páginas [34](#), [36](#).

ANEXO A – Declaração de uso de IAGen no processo de escrita científica

Durante a preparação deste documento, eu, Filipe Rodrigues Araújo, estudante de graduação do curso de Engenharia de Controle e Automação, declaro o uso da ferramenta Gemini (Google), baseada em modelos de linguagem de grande escala (LLM), para apoio à revisão textual, reescrita acadêmica, organização estrutural de capítulos, ajuste de coerência entre seções, tradução de trechos para o inglês e refinamento de resumos, objetivos e conclusões.

Após o uso desta ferramenta/modelo/serviço, a pessoa autora revisou e editou o conteúdo em conformidade com os princípios éticos para uso de IAGen e com os acordos estabelecidos com a pessoa orientadora da pesquisa. Dessa forma, assumo total responsabilidade pelo conteúdo da publicação.

Etapas em que a IAGen foi utilizada e exemplos de prompts adotados:

- Apoio à formulação, revisão e reescrita de elementos textuais da dissertação incluindo resumo, palavras-chave e organização do trabalho com foco em clareza, concisão e fluidez de leitura. Prompts representativos: “mude este final de parágrafo para contextualizar melhor com o meu trabalho operando com a bomba de polpa”; “Revise o resumo e as palavras-chave”.
- Organização estrutural de capítulos e seções incluindo o ajuste de transições, formatação estruturada de tabelas e listas de elementos pré-textuais. Prompts representativos: “quero uma tabela fechada com linhas pretas verticais e horizontais em todas as células (estilo grade estruturada) com o texto centralizado”; “com base nos meus objetivos, o que eu ainda não contextualizei nos resultados”.
- Apoio à sintaxe do código Latex e a estrutura do texto quanto as normas ABNT auxílio na verificação e formatação de citações e referências bibliográficas de acordo com o padrão normativo. Prompts representativos: “Como citar este autor de forma indireta neste parágrafo?”.
- Apoio à leitura, interpretação e aproveitamento de trechos de artigos, livros e demais referências. Prompts representativos: “O que o autor quis dizer exatamente nesta parte?”; “me explique melhor como o autor chegou neste resultado”.