UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
FACULDADE DE ENGENHARIA ELÉTRICA

JADE MELO DA CONCEIÇÃO

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE ÓLEO NA MANUTENÇÃO PREDITIVA DE EQUIPAMENTOS MÓVEIS NA INDÚSTRIA DE MINERAÇÃO

TUCURUÍ-PA
2026

JADE MELO DA CONCEIÇÃO

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE ÓLEO NA MANUTENÇÃO PREDITIVA DE EQUIPAMENTOS MÓVEIS NA INDÚSTRIA DE MINERAÇÃO

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia Elétrica, na Universidade Federal do Pará - Campus Universitário de Tucuruí, como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica.

Orientador(a): Prof. Dr(a). Raphael Barros Teixeira.

TUCURUÍ-PA
2026

JADE MELO DA CONCEIÇÃO

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE ÓLEO NA MANUTENÇÃO PREDITIVA DE EQUIPAMENTOS MÓVEIS NA INDÚSTRIA DE MINERAÇÃO

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia Elétrica, na Universidade Federal do Pará - Campus Universitário de Tucuruí, como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica.

Data da aprovação: ____/____/_____

Conceito: _____

BANCA EXAMINADORA

Documento assinado digitalmente
gov.br RAPHAEL BARROS TEIXEIRA
Data: 05/08/2026 14:15:17-0300
Verifique em https://validar.iti.gov.br

_____ _____

Orientador: Prof. Dr. Raphael Barros Teixeira (UFPA – CAMTUC)

Documento assinado digitalmente
gov.br DANIEL DA CONCEICAO PINHEIRO
Data: 04/08/2026 21:13:14-0300
Verifique em https://validar.iti.gov.br

_____ _____

Membro interno: Dr. Daniel da Conceição Pinheiro (UFPA - CAMTUC)

Documento assinado digitalmente
gov.br RAPHAEL SARAIVA DE SOUSA
Data: 03/08/2026 20:11:28-0300
Verifique em https://validar.iti.gov.br

_____ _____

Membro externo: Prof. Msc. Raphael Saraiva de Souza (IFPA)

RESUMO

A análise de óleo lubrificante é uma ferramenta importante para identificar falhas prematuras em equipamentos, prolongar a vida útil dos componentes e reduzir custos de manutenção. Nesse contexto, torna-se necessário desenvolver métodos que contribuam para a melhoria da qualidade dos diagnósticos e para a antecipação de possíveis falhas. Dessa forma, este trabalho propõe a aplicação de técnicas de aprendizagem de máquina para classificar os resultados de análises de óleo em equipamentos móveis da indústria de mineração, identificando a condição do óleo com base em parâmetros estabelecidos pela gestão de ativos. Para a realização do estudo, foram utilizados dados de diagnósticos coletados com amostragem diária ao longo de dois anos. O conjunto de dados foi submetido a etapas de pré-processamento e dividido em subconjuntos de treino (80%) e validação (20%), permitindo treinamento e avaliação de diferentes algoritmos de classificação, incluindo Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Árvore de Decisão e Multilayer Perceptron (MLP). Os modelos foram aplicados com o objetivo de identificar padrões nos dados e classificar o estado do óleo de acordo com os parâmetros definidos nas classes normal, monitorar e crítico. Os algoritmos avaliados apresentaram índices competitivos de assertividade, com destaque para o Random Forest, que obteve o melhor desempenho global ao atingir 83,25% de acurácia e 82,02% de *F1-score* Macro e elevada estabilidade na curva *Precision-Recall* (*Average Precision* de 0,84 para a classe crítico). Desse modo, o estudo valida a eficácia do aprendizado de máquina como uma ferramenta robusta de apoio à tomada de decisão em estratégias de manutenção preditiva.

Palavras-chave: Análise de óleo. Aprendizado de máquina. Classificação de dados. Manutenção preditiva. Equipamentos móveis.

ABSTRACT

Lubricating oil analysis is an important tool for identifying premature equipment failures, extending component lifespan, and reducing maintenance costs. In this context, it is necessary to develop methods that contribute to the improvement of diagnostic quality and the anticipation of potential failures. Thus, this work proposes the application of machine learning techniques to classify the results of oil analyses in mobile equipment within the mining industry, identifying the oil condition based on parameters established by asset management. For this study, diagnostic data collected through daily sampling over two years were used. The dataset underwent preprocessing stages and was divided into training (80%) and validation (20%) subsets, allowing for the training and evaluation of different classification algorithms, including Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, and Multilayer Perceptron (MLP). The models were applied with the objective of identifying data patterns and classifying the oil state according to the defined parameters into *normal*, *monitor*, and *critical* classes. The evaluated algorithms presented competitive predictive performance, with emphasis on the Random Forest, which achieved the best overall performance, reaching an accuracy of 83.25%, a Macro F1-score of 82.02%, and high stability in the Precision-Recall curve (Average Precision of 0.84 for the critical class). Thus, the study validates the effectiveness of machine learning as a robust decision-support tool in predictive maintenance strategies.

Keywords: Oil Analysis. Machine Learning. Data Classification. Predictive Maintenance. Mobile Equipment.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A lubrificação é um fator crucial nos processos produtivos da indústria de mineração, uma vez que muitos equipamentos, especialmente os móveis, operam sob elevadas cargas de trabalho. Entre os principais ativos que compõem essas frotas, destacam-se os caminhões fora de estrada, as escavadeiras hidráulicas, os tratores de esteira e os mineradores de superfície, conforme ilustrado na Figura 1. Nesse contexto, a análise de óleo desempenha um papel fundamental no monitoramento das condições dos fluidos lubrificantes e dos equipamentos, permitindo maximizar o desempenho e a confiabilidade dos ativos. Além disso, essa técnica possibilita a detecção precoce de falhas por meio de métodos não invasivos e não destrutivos (RODRIGUES et al., 2019).

Figura 1 – Equipamentos móveis de grande porte: (a) Escavadeira Hidráulica; (b) Minerador de superfície; (c) Trator de esteira; (d) Caminhão fora de estrada.

Fonte: (A, C, D) Elaborado pelo autor (2026); (B) Adaptado de Molland/Hydro (2016).

Na indústria de mineração, essas máquinas são constantemente expostas a ambientes severos, caracterizados pela presença de poeira, umidade e partículas contaminantes, que comprometem o desempenho dos lubrificantes e aceleram o desgaste dos componentes mecânicos. A contaminação por partículas sólidas é considerada um fator crítico nesse contexto,

sendo responsável por aproximadamente 70% das falhas operacionais, devido à deterioração acelerada de vedações e componentes internos de bombas hidráulicas (KHAYAL, 2025).

Diante dessas condições operacionais severas, diversos fatores influenciam diretamente a vida útil do óleo lubrificante, tornando a análise laboratorial uma ferramenta preditiva indispensável. A análise de óleo permite identificar os primeiros sinais de desgaste de um componente por meio do estudo da quantidade de partículas presentes no lubrificante, bem como de seu tamanho, forma e composição. Essas partículas são geradas pelo atrito dinâmico entre as superfícies em contato e fornecem informações precisas sobre as condições internas do equipamento, sem a necessidade de desmontagem do conjunto mecânico (GONÇALVES; SILVA, 2011).

Além da avaliação das partículas de desgaste, o monitoramento do lubrificante envolve a análise de diversos parâmetros físico-químicos, como viscosidade, contaminação e propriedades químicas do óleo, correlacionando essas informações ao modelo do equipamento e ao lubrificante específico de cada componente. Dessa forma, o monitoramento contínuo contribui para a redução de falhas catastróficas, paradas não planejadas e custos elevados de manutenção.

A análise de óleo envolve a avaliação simultânea de diversas variáveis, tornando o processo manual complexo e, muitas vezes, demorado. Além disso, as empresas enfrentam uma elevada demanda diária de amostras, dificultando a gestão eficiente dos diagnósticos. Nesse contexto, o uso de técnicas de aprendizado de máquina surge como uma solução promissora para otimizar esse processo, permitindo a análise automatizada e a classificação do estado das amostras de óleo nas classes normal, monitorar e crítica, contribuindo para diagnósticos mais ágeis e eficientes. O aprendizado de máquina é o processo de indução de uma hipótese ou aproximação de função a partir da experiência passada (FACELI et al., 2011).

Devido à sua capacidade de identificar padrões em grandes volumes de dados, o aprendizado de máquina tem sido amplamente aplicado em diferentes áreas. Na saúde, modelos supervisionados são utilizados em tarefas como diagnóstico de doenças, análise de imagens médicas e previsão de resultados clínicos, auxiliando profissionais na detecção de anomalias e no suporte à tomada de decisões (LUDERMIR, 2021). No setor financeiro, essas técnicas são empregadas na previsão de tendências de mercado, detecção de fraudes e análise de risco de crédito, possibilitando decisões mais seguras e estratégicas (FAÇANHA, 2019). Os resultados obtidos nesses contextos evidenciam o potencial dessas abordagens para aplicações industriais, como a análise de óleo lubrificante, na qual a identificação de padrões pode contribuir para a

automatização dos diagnósticos e para o aumento da confiabilidade das decisões de manutenção.

Diante desse cenário, este trabalho propõe a aplicação de técnicas de aprendizado de máquina supervisionado para a classificação de um banco de dados composto por 11.853 amostras de óleo lubrificante. Esses registros são provenientes da análise de óleo na indústria de mineração, refletindo as condições severas de campo por meio das análises de viscosidade, partículas de desgaste, visual e o ensaio de crepitação. Para essa finalidade, são avaliados os algoritmos *Random Forest*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), Árvore de Decisão e *Multi-Layer Perceptron* (MLP). As amostras são classificadas nas categorias normal, monitorar e crítica, com base nos resultados laboratoriais. O desempenho dos modelos é avaliado por meio da matriz de confusão e das métricas de precisão, *recall* e *F1-score*, permitindo comparar a capacidade de classificação de cada algoritmo e identificar a abordagem mais adequada para o problema estudado.

1.1 Objetivo

Desenvolver e avaliar modelos de aprendizado de máquina para a classificação do estado de contaminação de óleo lubrificante em equipamentos móveis da indústria de mineração, identificando o algoritmo com melhor desempenho preditivo para apoiar a detecção antecipada de falhas por meio da análise de dados físico-químicos e inspeções visuais. A partir dessa abordagem, são definidos os seguintes objetivos específicos:

- Analisar os fundamentos da análise de óleo lubrificante aplicada à estratégia de manutenção preditiva em frotas de equipamentos móveis;
- Estruturar o banco de dados histórico contendo as variáveis físico-químicas e os registros analíticos do ativo;
- Realizar o pré-processamento dos dados por meio do tratamento de valores nulos, codificação de variáveis categóricas via *One-Hot Encoding* e padronização de escala das variáveis numéricas;
- Implementar os classificadores supervisionados baseados nos algoritmos *Random Forest*, SVM, MLP, Árvore de Decisão e KNN;
- Otimizar os hiperparâmetros dos modelos utilizando a estratégia de busca em grade (*Grid Search*) sob validação cruzada e divisão estratificada dos dados;

- Avaliar o desempenho global dos modelos com base nas métricas de acurácia, precisão, *recall* e *F1-score*, utilizando a agregação *Macro Average* devido ao desbalanceamento das classes;
- Diagnosticar o comportamento dinâmico dos erros interclasses por meio de Matrizes de Confusão, avaliando a capacidade dos modelos em mitigar Falsos Negativos na classe Crítico;
- Comparar e selecionar o modelo preditivo mais robusto para atuar como ferramenta de apoio à engenharia de confiabilidade na classificação das amostras em Normal, Monitorar e Crítico.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Lubrificação e Análise de Óleos

Os óleos lubrificantes são constituídos por uma mistura de óleos básicos obtidos através de processos de refinação. Essa mistura deve ser realizada em proporções específicas, com o objetivo de atingir viscosidades adequadas e atender às exigências operacionais. Para isso, são adicionados aditivos ou realizados tratamentos específicos (ARAÚJO et al., 2019).

A monitoramento do estado das máquinas e equipamentos é crucial neste contexto, e a análise de óleo desempenha um papel fundamental para alcançar esse objetivo. Ao examinar a qualidade do óleo lubrificante, falhas precoces podem ser detectadas, resultando em redução de custos e no prolongamento da vida útil dos componentes. A análise do óleo é uma manutenção de caráter preditivo de diagnóstico, no intuito de acompanhar e examinar as condições dos fluídos e equipamentos. Esse processo permite melhorar o desempenho e a confiabilidade dos ativos, uma vez que possibilita a identificação de problemas antes que se tornem falhas. Essa ferramenta fornece objetividade e garantia na tomada de decisões dos gestores, desta maneira, poupando tempo e economizando custos com manutenção (SANTOS; FERREIRA, 2016).

Amostras são coletadas do equipamento para realizar análises de óleo lubrificante, que são então categorizadas em vários grupos, incluindo análise de partículas de desgaste, análise de viscosidade, análise visual e crepitação.

2.1.1 Coleta de Óleo

A execução de cada técnica começa com a coleta de amostras, que devem ser obtidas enquanto o equipamento estiver em funcionamento. Durante um determinado período de operação do equipamento, amostras devem ser cuidadosamente retiradas de locais específicos do ponto de coleta, conforme a Figura 2. A frequência da coleta das amostras varia de acordo com as especificações do fabricante, das condições e criticidade do serviço.

Figura 2 – Coleta de óleo.

Fonte: Próprio Autor.

A entrega das amostras no laboratório deve ocorrer em até 48 horas após a coleta, sendo as próprias amostras coletadas em recipiente PET transparente com capacidade para até 100 ml de óleo lubrificante e lacrado com tampa de rosca branca, como mostrado na Figura 3. Os recipientes devem estar devidamente etiquetados com as seguintes informações:

Figura 3 – Recipiente PET.

Fonte: Próprio Autor.

- **Equipamento:** Identificado com a TAG do equipamento em qual ativo a amostra foi coletada;
- **Data da coleta:** Data da coleta da amostra;
- **Hora da coleta:** Horário em que foi realizado a coleta da amostra;

- **Lubrificante:** Identificar o tipo de óleo usado atualmente no componente;
- **Horímetro:** Identificar as horas de operação do ativo. Os dados têm que estar de acordo com a leitura que foi realizada no momento da coleta;
- **Drenagem do equipamento:** Identificar se óleo foi drenado após a coleta;
- **Tipo de coleta:** Identificar se a coleta foi proveniente de uma programação preventiva, demanda corretiva, ocorrência em campo ou solicitação de contraprova;
- **Componente:** Identificar o componente onde foi coletada a amostra;
- **Executante:** Responsável pela coleta.

Para iniciar o processo de análise no laboratório as amostras são cadastradas no banco de dados de acordo com as informações das etiquetas, conforme Tabela 1.

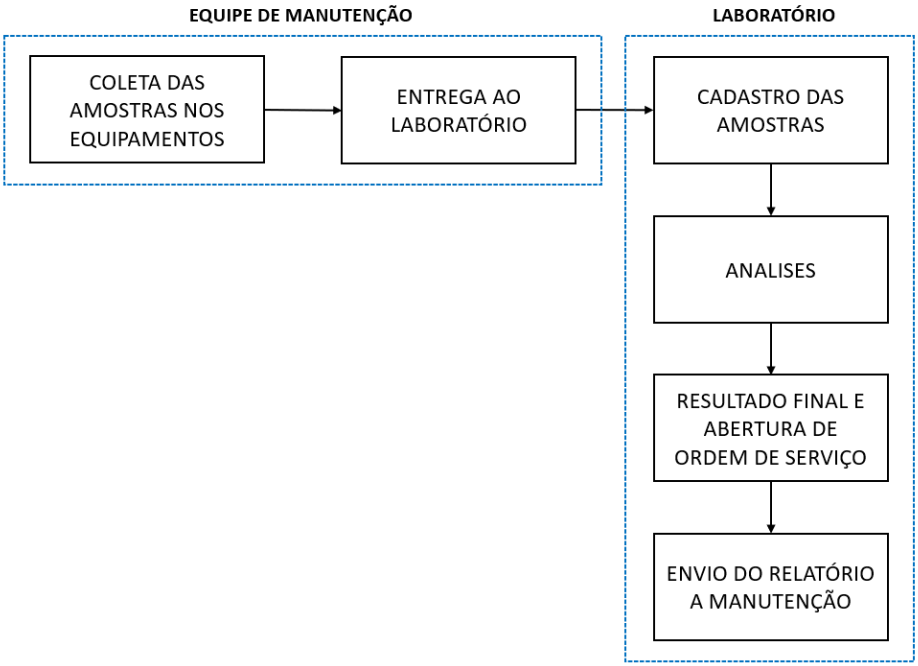
Tabela 1 – Dados do cadastro inicial.

CADASTRO INICIAL													
Nº DE AMOSTRA	SUPERVISÃO	HORA	DATA DA COLETA	RECEBIMENTO DATA	ANÁLISE DATA	ENTREGUE	ATIVO	FROTA	COMPARTIMENTO	DRENADO	HORÍMETRO	H/T DO ÓLEO	TIPO DE FLUIDO
90694	CAMINHÕES	22:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	TOP_6447	4844K	MOTOR 4844K	SIM	23870	673	15W40
90695	CAMINHÕES	22:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	TOP_6447	4844K	HIDRÁULICO 4844K	NÃO	23870	1751	15W40
90696	CAMINHÕES	22:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	TOP_6447	4844K	CX. DE MARCHA 4844K	SIM	23870	1751	80W
90697	CAMINHÕES	22:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	TOP_6447	4844K	1º DIFERENCIAL 4844K	SIM	23870	673	MB90
90698	CAMINHÕES	22:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	TOP_6447	4844K	2º DIFERENCIAL 4844K	SIM	23870	673	MB90
90699	CAMINHÕES	10:00:00	31-out-21	3-out-21	4-out-21	ENTREGUE	CB_6431	G440	MOTOR G440	SIM	27838	509	15W40
90700	CAMINHÕES	10:00:00	31-out-21	3-out-21	4-out-21	ENTREGUE	CB_6431	G440	HIDRÁULICO G440	NÃO	27838	4463	15W40
90701	CAMINHÕES	10:00:00	31-out-21	3-out-21	4-out-21	ENTREGUE	CB_6431	G440	CX. DE MARCHA G440	SIM	27838	509	85W140
90702	CAMINHÕES	10:00:00	31-out-21	3-out-21	4-out-21	ENTREGUE	CB_6431	G440	1º DIFERENCIAL G440	SIM	27838	509	85W140
90703	CAMINHÕES	10:00:00	31-out-21	3-out-21	4-out-21	ENTREGUE	CB_6431	G440	2º DIFERENCIAL G440	SIM	27838	509	85W140
90704	CAMINHÕES	10:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	CB_6462	4844K	MOTOR 4844K	SIM	23139	707	15W40
90705	CAMINHÕES	10:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	CB_6462	4844K	HIDRÁULICO 4844K	NÃO	23139	16953	15W40
90706	CAMINHÕES	10:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	CB_6462	4844K	CX. DE MARCHA 4844K	NÃO	23139	707	80W
90707	CAMINHÕES	10:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	CB_6462	4844K	1º DIFERENCIAL 4844K	NÃO	23139	707	MB90
90708	CAMINHÕES	10:00:00	31-out-21	3-nov-21	4-nov-21	ENTREGUE	CB_6462	4844K	2º DIFERENCIAL 4844K	NÃO	23139	707	MB90
90709	DECAPEAMENTO	10:00:00	2-nov-21	3-nov-21	4-nov-21	ENTREGUE	TE_8033	D11 AMA	MOTOR D11 AMA	SIM	17863	159	15W40
90710	DECAPEAMENTO	10:00:00	2-nov-21	3-nov-21	4-nov-21	ENTREGUE	TE_8033	D11 AMA	HIDRÁULICO D11 AMA	NÃO	17863	528	15W40
90711	DECAPEAMENTO	10:00:00	2-nov-21	3-nov-21	4-nov-21	ENTREGUE	TE_8033	D11 AMA	TRANSMISSÃO D11 AMA	NÃO	17863	528	SAE50
90712	DECAPEAMENTO	10:00:00	2-nov-21	3-nov-21	4-nov-21	ENTREGUE	TE_8033	D11 AMA	COMANDO FINAL DIR. D11 AMA	SIM	17863	281	SAE60
90713	DECAPEAMENTO	10:00:00	2-nov-21	3-nov-21	4-nov-21	ENTREGUE	TE_8033	D11 AMA	COMANDO FINAL ESQ. D11 AMA	SIM	17863	528	SAE60

Fonte: Próprio Autor.

O laboratório recebe em média 60 amostras por dia, que são analisadas e computadas na base de dados. O fluxo completo desse processo de análise de óleo é apresentado na Figura 4. Após as análises, os resultados são divulgados à manutenção junto com a ordem de serviço e o relatório final, contendo as devidas ações indicadas, como ilustrado na Figura 5.

Figura 4 – Fluxo do processo de análise de óleo.



Fonte: Próprio Autor.

Figura 5 – Relatório de análise de óleo.

Relatório de Análise de óleo

ATIVO	COMPARTIMENTO	Nº DA AMOSTRA	SUPERVISÃO	RESULTADO	NOTA LUBRIFICAÇÃO	ANÁLISE DATA
MS_1006	RLTE HEAVY DUTY	93053	CARREGAMENTO LAVRA	CRITICO		18/01/2022

PAUTA	HORÍMETRO	H/T DO ÓLEO	DRENADO	DIAGNÓSTICO/RECOMENDAÇÃO	RESULTADO	DRENAGEM EMERGENCIAL	DATA DA COLETA
125	16535	177	SIM	A ANÁLISE DE ÓLEO APRESENTOU FORTE ESCURECIMENTO E CONTAMINAÇÃO POR TRAÇOS DE ÁGUA. QUE PODE INDICAR DEGRADAÇÃO FISICA OU QUIMICA DO LUBRIFICANTE E ENTRADA DE CONTAMINAÇÃO EXTERNA. VERIFICAR TEMPERATURA DE TRABALHO E DANOS EM VEDAÇÕES. DRENADO.	CRITICO	RELATÓRIO	13/01/22
2000	16358	206	SIM	A ANÁLISE DE ÓLEO APRESENTOU CONTAMINAÇÃO POR TRAÇOS DE ÁGUA. VERIFICAR DANOS NAS VEDAÇÕES.	MONITORAR	-	03/01/22
125	16152	195	SIM	A ANÁLISE DE ÓLEO APRESENTOU ALTERAÇÕES NOS TEORES DE ALUMÍNIO E SILÍCIO, QUE PODE INDICAR ENTRADA DE CONTAMINANTES EXTERNOS. VERIFICAR DANOS NAS VEDAÇÕES.	MONITORAR	-	24/12/21
250	15957	198	SIM	A AMOSTRA APRESENTOU CONDIÇÃO NORMAL.	NORMAL	-	13/12/21

Retorno da Manutenção

DATA DA COLETA	FEEDBACK
13/12/21	-
24/12/21	-
03/01/22	-
13/01/22	-

Análises: Visual, Viscosidade e Contagem de partículas

DATA DA COLETA	ÁGUA	IMPUREZA	LIMALHAS	SÍLICA	VIS 40°C	4µ	6µ	14µ
13/01/22	T	N	N	N	215			
03/01/22	T	N	N	N	210			
24/12/21	T	N	N	N	206			
13/12/21	N	N	N	N	211			

Análise FTIR

DATA DA COLETA	OXID	NIT	SUL	TBN	FULIGEM	COMB
13/01/22						
03/01/22						
24/12/21						
13/12/21						

Spectrometria

DATA DA COLETA	Fe	Cu	Cr	Pb	Sn	Ni	Ag	Al	Si	Na	Ca	Mo	P	Zn	Mg	B	Ba
13/01/22	445,00	4,00	6,00	8,00	0,00	0,00	0,00	7,00	24,00	5,00	2.558,00	1,00	767,00	1.158,00	11,00	2,00	9,00
03/01/22	49,00	0,00	0,00	1,00	0,00	0,00	0,00	16,00	23,00	3,00	2.291,00	2,00	722,00	1.110,00	11,00	1,00	0,00

Fonte: Próprio Autor.

2.1.2 Análise de partículas de desgaste

O estudo e a quantificação do desgaste ao qual um sistema mecânico está submetido é fundamental para diferentes aplicações. Esse estudo fundamenta-se na tribologia, ciência e tecnologia que estuda a interação entre superfícies em movimento relativo, abrangendo os fenômenos de atrito, desgaste e lubrificação (STACHOWIAK; BATCHELOR, 2013). Suas aplicações estendem-se por diversas áreas, desde o setor industrial, focado nos parâmetros de controle de qualidade de fabricação, até os processos de acompanhamento e manutenção preventiva de maquinários (CUNHA, 2005).

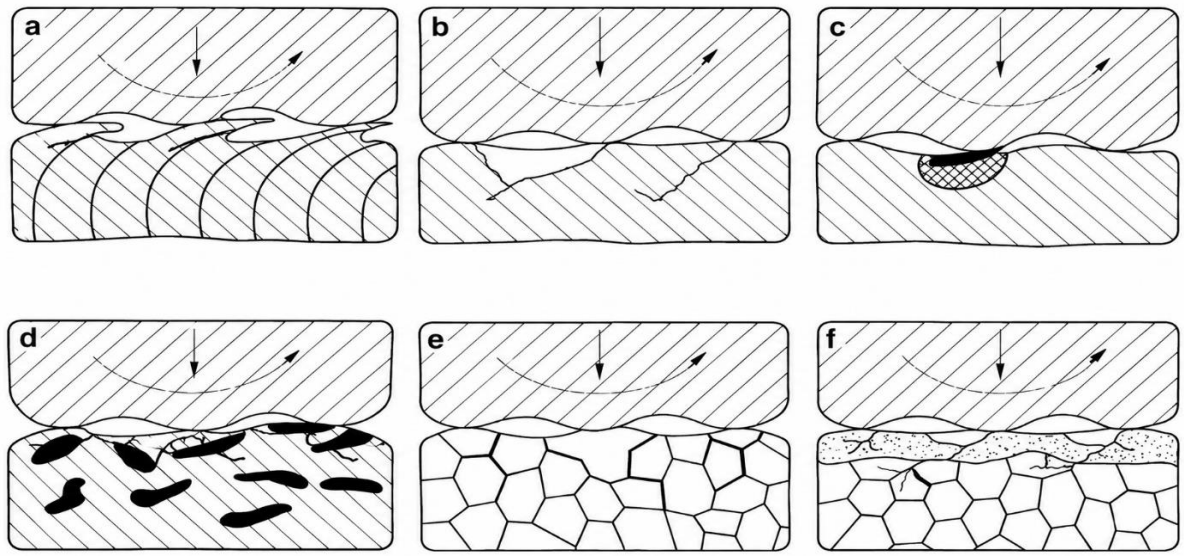
A ocorrência de defeitos resultantes de processos de deterioração pode comprometer a continuidade e a qualidade do serviço de qualquer sistema mecânico ou equipamento. Isto significa que uma quebra não prevista se traduz em uma parada abrupta, gerando grandes prejuízos e a perda de tempo de produção (LAGO; GONÇALVES, 2006).

A análise de partículas de desgaste surge como um forte indicador da interação tribológica. As informações acerca da quantidade de partículas, tamanho, forma e composição são necessárias para verificar as condições das superfícies em movimento, sem a necessidade de se desmontar o conjunto, a qual estas partes pertencem (MOREIRA, 2022).

Além disso, de acordo com Osmari, Santos e Herzog (2022), os fenômenos de fadiga superficial, abrasão severa e degradação química, cada um destes aspectos está relacionado, a uma maneira diferente em que ocorre a carga e a ação de deslizamento necessários ao desgaste. Na maioria dos casos, todas as formas de desgaste, podem resultar da introdução de energia mecânica no sistema, haja vista que se pode ter desgaste por erosão, abrasão por erosão e assim por diante, todas geradas dentro do mesmo processo erosivo. Levando esses fatores em consideração, ainda podem ocorrer trincas superficiais, o que pode ocorrer em componentes como engrenagens, devido a combinação de rolamento e deslizamento que promove a ocorrência dessas trincas.

Constata-se que em certas condições de laminação a lubrificação pode ser inexistente ou deficiente, conduzindo ao contato direto entre os pares sob rolamento-deslizamento e, consequentemente, causando o desgaste. Os diferentes danos e tipos de falhas superficiais originados durante o contato rolamento-deslizamento nessas condições de má lubrificação são apresentados de forma esquemática na Figura 6 (SILVA, 2006).

Figura 6 – Representação esquemática dos danos causados à superfície durante o contato rolamento-deslizamento com lubrificação deficiente.

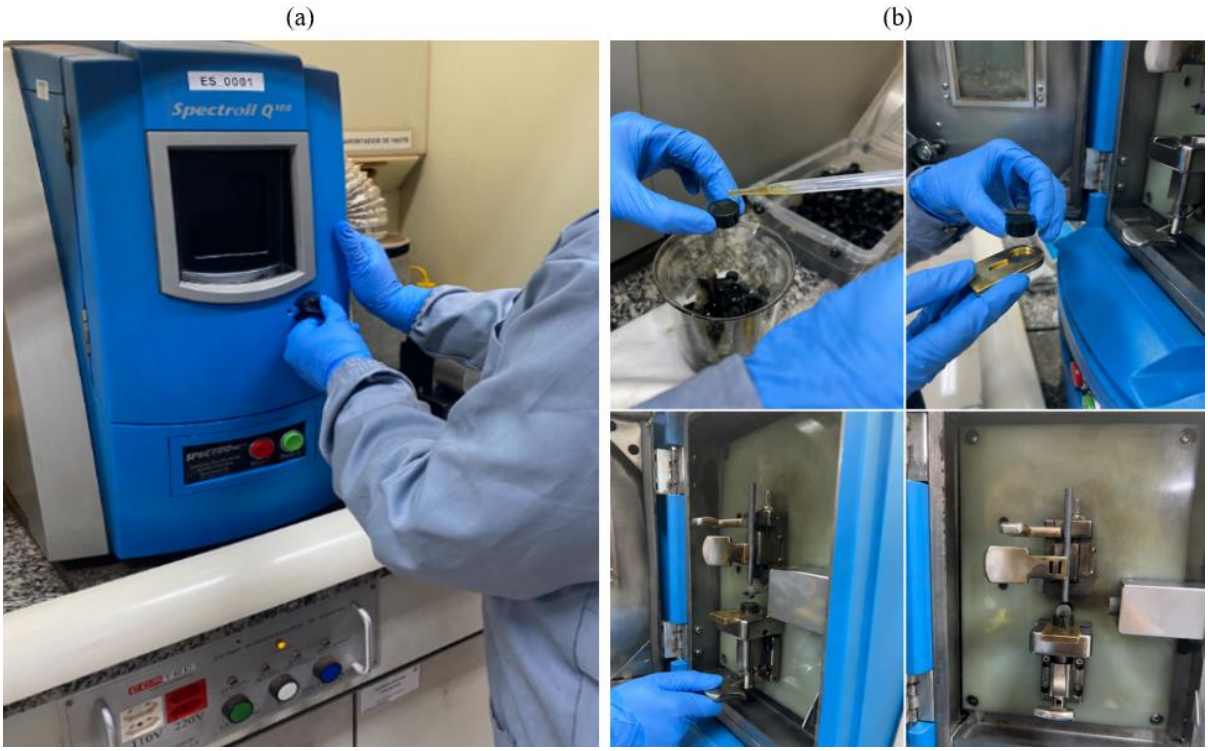


(a) Deformação plástica e cisalhamento; (b) Trincas originadas na superfície; (c) Impressões causadas por partículas duras; (d, e) Trincas devidas a elementos microestruturais como precipitados ou contornos de grão fragilizados; (f) Trincas das camadas com reações químicas induzidas por atrito.

Fonte: Silva (2006).

A análise do desgaste é realizada por espectrometria de emissão óptica por eletrodo rotativo, permitindo a avaliação das concentrações de elementos, incluindo ferro, cobre, alumínio, silício, estanho, sódio, potássio, níquel, cromo, chumbo e prata, medidas em partes por milhão (ppm) para amostras de óleo lubrificante. A Figura 7 apresenta o equipamento de espectrometria.

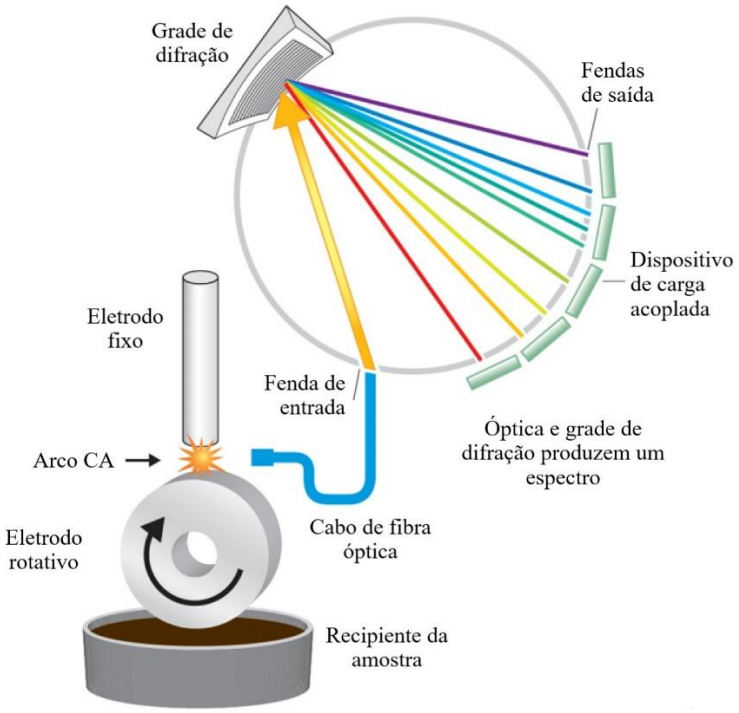
Figura 7 – (a) Espectrômetro de emissão óptica SpectrOil Q100 utilizado nas análises; (b) preparação e posicionamento da amostra para análise espectrométrica.



Fonte: Autor próprio.

A Figura 8 apresenta os principais componentes de um espectrômetro de análise de óleo equipado com um sistema óptico. Conforme apresentado por Spectro Scientific (2014), durante a análise, a luz gerada no processo de excitação por arco elétrico da amostra é transmitida por uma fibra óptica, atravessa a fenda de entrada e é focalizada sobre uma rede de difração por meio de uma lente. A fenda de entrada recebe a radiação emitida pelos elementos presentes na amostra e define a forma das linhas espectrais produzidas após a difração. A rede de difração tem como função separar a radiação em seus diferentes comprimentos de onda, permitindo a identificação dos elementos químicos presentes. As linhas espectrais resultantes podem ser registradas fotograficamente ou quantificadas eletronicamente por detectores, como tubos fotomultiplicadores ou dispositivos de carga acoplada.

Figura 8 – Esquema do espectrômetro de absorção atômica.



Fonte: Adaptado de Spectro Scientific (2025).

O software do espectrômetro processa os sinais elétricos e gera a interface de resultados diretamente na tela do computador, conforme ilustrado na Figura 9. Esses dados coletados são, posteriormente, organizados no banco de dados com a finalidade de avaliá-los comparativamente em relação aos parâmetros de referência definidos pela empresa, conforme exemplificado na Tabela 2.

Figura 9 – Tela de exibição de dados do software de emissão óptica.

Operations Databases Tools Help

Setup Detection:7060

Ba	Ca	Cd	Cr	Cu	Fe	K	Li	MG	Mn	Mo
12.67	12.17	11.80	10.19	11.90	11.23	10.74	52.83	12.26	11.55	12.65
10.95	11.49	11.07	9.76	11.34	10.74	10.63	9.79	11.24	10.98	10.64
11.03	11.43	10.81	10.26	11.70	10.46	11.28	9.96	3.75	11.49	10.27
9.32	11.00	11.14	9.90	11.01	10.39	10.38	9.77	7.24	10.63	11.63
9.94	11.97	11.95	9.76	11.07	10.90	10.81	10.94	10.20	10.93	10.58
9.76	10.89	10.17	9.91	10.90	10.26	11.08	10.55	10.64	10.40	9.18
10.30	11.12	9.88	9.45	11.04	9.49	11.50	11.20	9.95	9.69	10.46
10.03	9.60	11.32	9.23	10.80	9.49	10.71	9.31	7.65	10.01	10.33
11.68	10.35	10.11	9.81	10.97	10.07	11.28	10.82	5.29	9.63	10.27
11.44	9.50	9.11	8.75	10.23	9.36	10.63	9.70	2.71	9.46	9.25
12.57	10.28	9.10	9.16	10.39	10.10	11.23	11.04	10.73	9.27	10.64

Fonte: Adaptado de Spectro Scientific (2025).

Tabela 2 – Guia de parâmetros de desgaste.

RESPONSÁVEL	INFORMAÇÕES DO EQUIPAMENTO				PARÂMETROS DE ANÁLISE DE DESGASTE										
SUPERVISÃO	SÉRIE	COMPONENTES	TIPO DE FLUIDO	TAG do Óleo	Fe	Cu	Cr	Pb	Sn	Ni	Ag	Al	Si	Na	K
CARREGAMENTO	964 LIEBHERR	MOTOR 964LB	15W40	ESC-15W40-M	50	60	5	20	6	2	2	15	15	25	25
CARREGAMENTO	964 LIEBHERR	HIDRÁULICO 964LB	15W40	ESC-15W40-H	40	40	5	4	5	2	2	16	26	25	25
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO DIR. 964LB	SAE90	ESC-SAE90	400	50	4	4	4	2	2	20	30	25	25
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO ESQ. 964LB	SAE90	ESC-SAE90	400	50	4	4	4	2	2	20	30	25	25
CARREGAMENTO	964 LIEBHERR	REDUTOR DE GIRO 964LB	SAE90	ESC-SAE90	400	150	4	4	4	2	2	5	15	25	25
CARREGAMENTO	964 LIEBHERR	P.T.O 964LB	SAE90	ESC-SAE90	100	10	5	6	5	2	2	11	22	25	25
CARREGAMENTO	R9250 LIEBHERR	MOTOR R9250	15W40	ESC-15W40-M	50	60	5	20	6	2	2	15	15	20	20
CARREGAMENTO	R9250 LIEBHERR	HIDRÁULICO R9250	15W40	ESC-15W40-HR9250	13	40	5	4	5	2	2	15	26	25	25
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO DIR. R9250	VG460	ESC-VG460	400	50	4	4	4	2	2	20	30	25	25
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO ESQ. R9250	VG460	ESC-VG460	400	50	4	4	4	2	2	20	30	25	25
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO DIR. R9250	VG460	ESC-VG460	400	150	4	4	4	2	2	5	15	25	25
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO ESQ. R9250	VG460	ESC-VG460	400	150	4	4	4	2	2	5	15	25	25
CARREGAMENTO	R9250 LIEBHERR	P.T.O R9250	SAE90	ESC-SAE90	100	10	5	6	5	2	2	11	20	25	25
CARREGAMENTO LAVRA	CAT 390FL	MOTOR 390FL	15W40	LAV-15W40-M390	51	12	4	5	12	2	2	5	20	25	25
CARREGAMENTO LAVRA	CAT 390FL	HIDRÁULICO 390FL	15W40	LAV-15W40-H390	132	26	4	4	15	2	2	8	20	25	25
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO DIR. 390FL	SAE 50	LAV-SAE50-390	63	14	4	5	12	2	2	5	20	25	25
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO ESQ. 390FL	SAE 50	LAV-SAE50-390	63	14	4	5	12	2	2	5	20	25	25
CARREGAMENTO LAVRA	CAT 390FL	GIRO DIANT. 390FL	SAE 50	LAV-SAE50-390	87	9	4	4	19	2	2	5	30	25	25
CARREGAMENTO LAVRA	CAT 390FL	GIRO TRAS. 390FL	SAE 50	LAV-SAE50-390	87	9	4	4	19	2	2	5	30	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	MOTOR SM2500	15W40	LAV-15W40-MMS	50	60	5	20	6	2	2	15	151	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	HIDRÁULICO SM2500	HR46	LAV-HR46	40	40	5	4	5	2	2	16	26	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLDD HEAVY DUTY	SAE50	LAV-SAE50-MS	400	100	4	4	4	2	2	20	15	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLDE HEAVY DUTY	SAE50	LAV-SAE50-MS	400	100	4	4	4	2	2	20	15	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLTD HEAVY DUTY	SAE50	LAV-SAE50-MS	400	100	4	4	4	2	2	20	15	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLTE HEAVY DUTY	SAE50	LAV-SAE50-MS	400	100	4	4	4	2	2	20	15	25	25
CARREGAMENTO LAVRA	SM2500 WIRTGEN	ROLO FRESADOR SM2500	SAE90	LAV-SAE90	100	15	5	2	10	10	2	5	15	20	15

Fonte: Próprio Autor.

Dentre os elementos monitorados, a elevada concentração de ferro, alumínio, cobre, estanho e cromo é diretamente associada ao desgaste severo dos componentes internos do sistema mecânico. Desta maneira, a partir do acompanhamento contínuo desses resultados, torna-se possível intervir antes que o equipamento atinja uma falha funcional, mitigando paradas imprevistas (FERNANDES, 2012). O conjunto de dados desses elementos específicos serve como base de dados para o treinamento de algoritmos de aprendizagem de máquina, conforme estudado neste trabalho, automatizando a detecção de padrões de falha e refinando as tomadas de decisão na gestão de ativos.

2.1.3 Análise de Viscosidade

A viscosidade, segundo Lago e Gonçalves (2006), é definida como a resistência ao escoamento de um fluido quando submetido a forças cisalhantes. De acordo com a teoria clássica da lubrificação hidrodinâmica proposta por Reynolds (1886), para uma mesma velocidade de deslizamento, uma viscosidade mais elevada resulta na formação de uma película lubrificante mais espessa entre as superfícies móveis, enquanto uma menor viscosidade origina uma película mais fina. Nesse contexto, trata-se de uma propriedade fundamental, pois indica

a capacidade do lubrificante de manter as superfícies em movimento relativo separadas por uma película de óleo, mesmo quando submetidas a cargas elevadas (BORDIGNON, 2018).

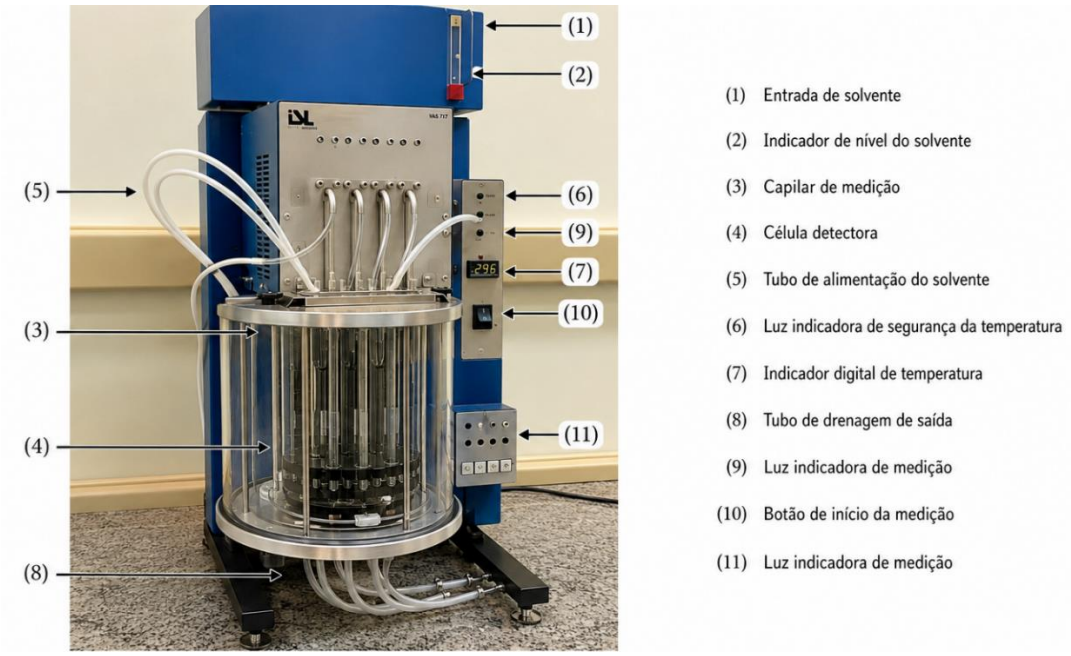
Quanto à origem, os lubrificantes podem ser vegetais, animais, minerais (derivados do petróleo) ou sintéticos. Independentemente da origem, a viscosidade não se mantém constante, visto que varia inversamente com a temperatura, ou seja, à medida que a temperatura aumenta, a viscosidade do líquido diminui (JÚNIOR; ALVES, 2018).

Alterações significativas nessa propriedade podem decorrer de mudanças na base química do óleo — caracterizadas por modificações na sua estrutura molecular — ou pela entrada de contaminantes. Tais variações demandam ensaios adicionais para confirmação do diagnóstico, tais como: o Número de Acidez (AN) e a Espectroscopia no Infravermelho por Transformada de Fourier (FTIR), para identificar oxidação incipiente; testes de contaminação, para detectar a presença de água, fuligem ou glicol; e, de forma menos frequente, a ultracentrifugação ou a cromatografia a gás (GC), com a finalidade de identificar mudanças na base química do produto (MARCASSO, 2017).

Diante disso, o controle dessa propriedade torna-se essencial, pois oscilações na viscosidade podem comprometer a confiabilidade do equipamento. Esse monitoramento pode ser realizado com eficácia e precisão tanto em campo, utilizando instrumentos de análise rápida (*on-site*), quanto por meio do envio frequente de amostras a laboratórios especializados (LAGO; GONÇALVES, 2006).

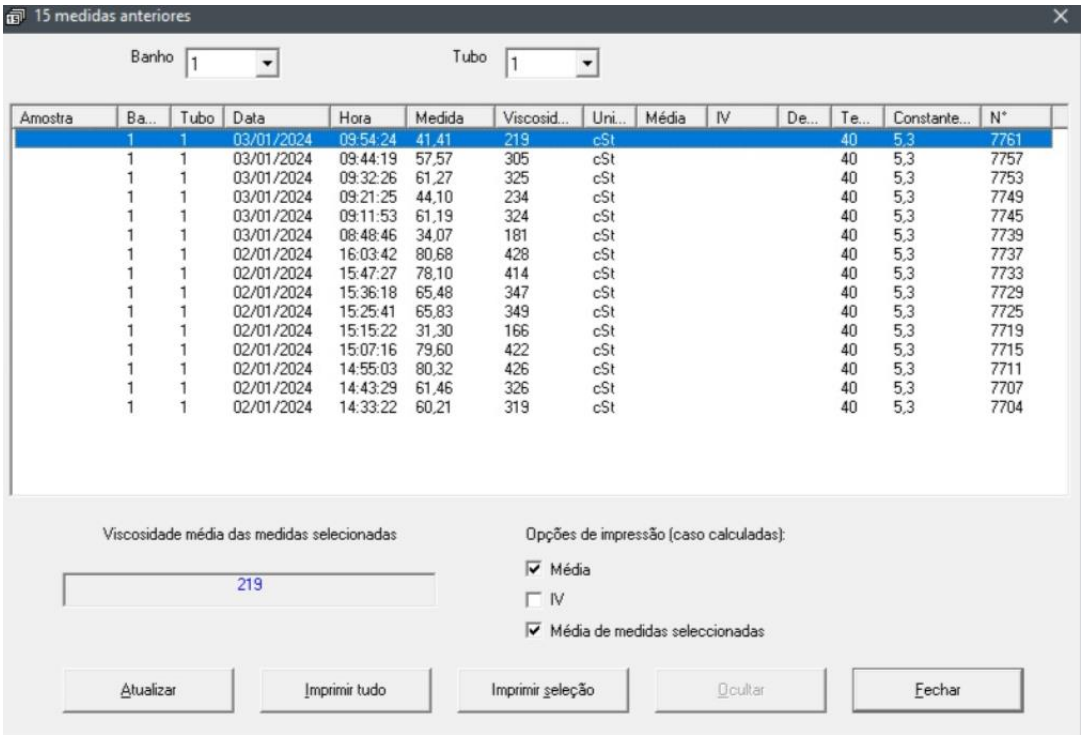
A determinação da viscosidade cinemática neste trabalho foi realizada utilizando o viscosímetro automático Houillon VH1, ilustrado na Figura 10. O equipamento opera segundo o método Houillon, normatizado pela ASTM D7279, que determina a viscosidade por meio da medição automática do tempo de escoamento de um volume conhecido de fluido através de um tubo calibrado mantido em temperatura controlada (ISL, 2013; ASTM INTERNATIONAL, 2020). Durante o ensaio, a amostra é aspirada para o interior do sistema de medição e seu deslocamento é monitorado por sensores ópticos, que registram o tempo necessário para o fluido percorrer uma distância previamente estabelecida. A partir desse tempo e da constante de calibração do capilar, o equipamento calcula automaticamente a viscosidade cinemática da amostra e apresenta o resultado em seu monitor, conforme ilustrado na Figura 11. O método apresenta elevada precisão, reduz a interferência do operador e atende aos requisitos da norma ASTM D7279 para determinação da viscosidade cinemática de produtos derivados de petróleo e lubrificantes (ISL, 2013).

Figura 10 – Viscosímetro automático Houillon VH.



Fonte: Próprio Autor.

Figura 11 – Resultado da viscosidade cinemática.



Fonte: Próprio Autor.

A análise dos dados obtidos permite determinar a viscosidade da substância mediante a comparação com parâmetros preestabelecidos para cada componente, conforme apresentado na Tabela 3. Foram adotados valores de referência correspondentes à viscosidade cinemática do

lubrificante a 40 °C (V40), bem como os respectivos Limites Inferior (L.I.) e Superior (L.S.), que definem a faixa aceitável de variação para cada tipo de fluido. Dessa forma, o resultado do ensaio é enquadrado em status crítico quando o valor medido for inferior ao L.I. — o que pode indicar processos de diluição por combustível ou contaminação — ou superior ao L.S., frequentemente associado à oxidação, envelhecimento do óleo ou presença de partículas de carbono provenientes da combustão. Esse processo de rotulagem e classificação da criticidade, além de fornecer um diagnóstico imediato, gera o conjunto de dados estruturado que será utilizado para o treinamento dos algoritmos de aprendizagem de máquina neste trabalho.

Tabela 3 – Guia de parâmetros de viscosidade.

RESPONSÁVEL	INFORMAÇÕES DO EQUIPAMENTO				PARÂMETRO DE VISCOSIDADE		
SUPERVISÃO	SÉRIE	COMPONENTES	TIPO DE FLUIDO	TAG do Óleo	L.I	L.S	V40
CARREGAMENTO	964 LIEBHERR	MOTOR 964LB	15W40	ESC-15W40-M	91,0	117,0	101
CARREGAMENTO	964 LIEBHERR	HIDRÁULICO 964LB	15W40	ESC-15W40-H	76,0	117,0	101
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO DIR. 964LB	SAE90	ESC-SAE90	157,0	210,0	197
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO ESQ. 964LB	SAE90	ESC-SAE90	157,0	210,0	197
CARREGAMENTO	964 LIEBHERR	REDUTOR DE GIRO 964LB	SAE90	ESC-SAE90	157,0	210,0	197
CARREGAMENTO	964 LIEBHERR	P.T.O 964LB	SAE90	ESC-SAE90	157,0	210,0	197
CARREGAMENTO	R9250 LIEBHERR	MOTOR R9250	15W40	ESC-15W40-M	91,0	117,0	101
CARREGAMENTO	R9250 LIEBHERR	HIDRÁULICO R9250	15W40	ESC-15W40-HR9250	68,0	117,0	101
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO DIR. R9250	VG460	ESC-VG460	382,0	520,0	450
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO ESQ. R9250	VG460	ESC-VG460	382,0	520,0	450
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO DIR. R9250	VG460	ESC-VG460	382,0	520,0	450
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO ESQ. R9250	VG460	ESC-VG460	382,0	520,0	450
CARREGAMENTO	R9250 LIEBHERR	P.T.O R9250	SAE90	ESC-SAE90	157,0	210,0	197
CARREGAMENTO LAVRA	CAT 390FL	MOTOR 390FL	15W40	LAV-15W40-M390	91,0	116,0	101
CARREGAMENTO LAVRA	CAT 390FL	HIDRÁULICO 390FL	15W40	LAV-15W40-H390	59,0	116,0	101
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO DIR. 390FL	SAE 50	LAV-SAE50-390	203,0	248,0	216
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO ESQ. 390FL	SAE 50	LAV-SAE50-390	203,0	248,0	216
CARREGAMENTO LAVRA	CAT 390FL	GIRO DIANT. 390FL	SAE 50	LAV-SAE50-390	203,0	248,0	216
CARREGAMENTO LAVRA	CAT 390FL	GIRO TRAS. 390FL	SAE 50	LAV-SAE50-390	203,0	248,0	216
CARREGAMENTO LAVRA	SM2500 WIRTGEN	MOTOR SM2500	15W40	LAV-15W40-MMS	86,0	116,0	101
CARREGAMENTO LAVRA	SM2500 WIRTGEN	HIDRÁULICO SM2500	HR46	LAV-HR46	36,0	53,0	46
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLDD HEAVY DUTY	SAE50	LAV-SAE50-MS	197,0	248,0	216
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLDE HEAVY DUTY	SAE50	LAV-SAE50-MS	197,0	248,0	216
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLTD HEAVY DUTY	SAE50	LAV-SAE50-MS	197,0	248,0	216
CARREGAMENTO LAVRA	SM2500 WIRTGEN	RLTE HEAVY DUTY	SAE50	LAV-SAE50-MS	197,0	248,0	216

Fonte: Próprio Autor.

2.1.4 Crepitação

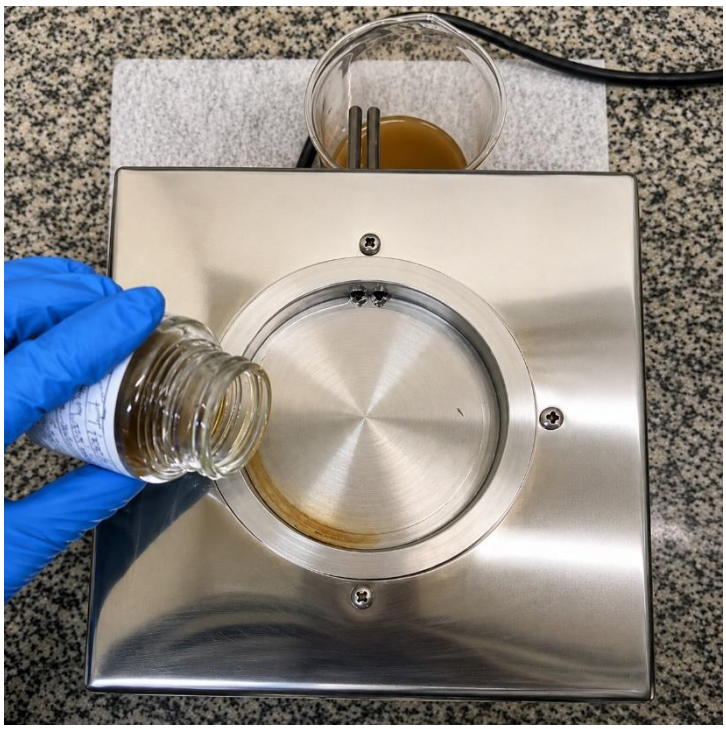
O monitoramento do teor de água no óleo é fundamental, uma vez que a presença desse contaminante no lubrificante é extremamente danosa, provocando a corrosão de componentes metálicos, além de acelerar a degradação dos aditivos e a oxidação do fluido. Um dos métodos

analíticos mais difundidos para a triagem rápida da umidade é o ensaio de crepitação (SPAMER, 2009).

A detecção de água pode indicar falhas operacionais ou de manutenção, tais como vazamentos no sistema de refrigeração, condensação interna devido a temperaturas operacionais abaixo do recomendado, fluxo de ar insuficiente ou operação do equipamento por períodos muito curtos. Adicionalmente, devem-se considerar fatores externos, como o armazenamento inadequado do lubrificante, a exposição do maquinário às intempéries ou falhas nos procedimentos de lavagem do compartimento.

Regulamentado pela norma ABNT NBR 16358 (2015), o ensaio consiste em verter a amostra sobre uma chapa metálica aquecida entre 150 °C e 200 °C, como ilustrado na Figura 12.

Figura 12 – Chapa aquecedora utilizada no ensaio de crepitação.



Fonte: Próprio Autor.

A identificação do teor de água ocorre por meio da inspeção visual e auditiva do comportamento do fluido, classificando-o em três níveis de contaminação:

- **Ausência (N):** Inexistência de bolhas de vapor ou estalos após alguns segundos, indicando amostra isenta de água;
- **Traços (T):** Formação de bolhas muito pequenas que desaparecem rapidamente;

- **Presença (P):** Surgimento de bolhas maiores e persistentes, evidenciando uma contaminação moderada;
- **Excesso (E):** Formação vigorosa de bolhas acompanhada de ruídos característicos de crepitação (aspecto de fritura), indicando contaminação severa.

Diante disso, os resultados qualitativos obtidos são enquadrados segundo os parâmetros de criticidade apresentados na Tabela 4 e, posteriormente, convertidos em dados estruturados na planilha eletrônica. Essa categorização (N, T, P, E) atua como uma variável categórica essencial que servirá como dados de entrada para o treinamento dos algoritmos de aprendizagem de máquina, permitindo correlacionar a presença de umidade com os padrões de falha do ativo.

Tabela 4 – Guia de parâmetros de crepitação.

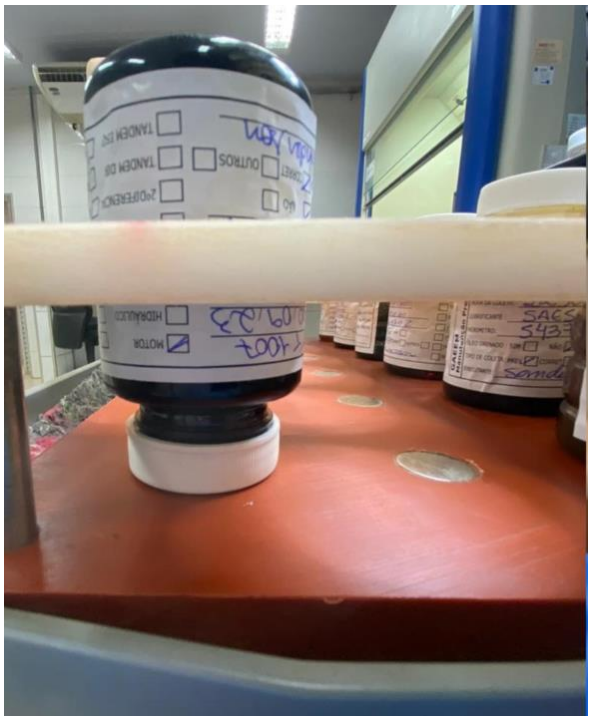
RESPONSÁVEL	INFORMAÇÕES DO EQUIPAMENTO				PARÂMETRO DE CONTAMINANTES - ÁGUA E IMPUREZAS - (ANÁLISE VISUAL)		
SUPERVISÃO	SÉRIE	COMPONENTES	TIPO DE FLUIDO	TAG do Óleo	TRAÇO	PRESENÇA	EXCESSO
CARREGAMENTO	964 LIEBHERR	MOTOR 964LB	15W40	ESC-15W40-M	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	964 LIEBHERR	HIDRÁULICO 964LB	15W40	ESC-15W40-H	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO DIR. 964LB	SAE90	ESC-SAE90	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO	964 LIEBHERR	TRANSLAÇÃO ESQ. 964LB	SAE90	ESC-SAE90	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO	964 LIEBHERR	REDUTOR DE GIRO 964LB	SAE90	ESC-SAE90	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	964 LIEBHERR	P.T.O 964LB	SAE90	ESC-SAE90	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	MOTOR R9250	15W40	ESC-15W40-M	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	HIDRÁULICO R9250	15W40	ESC-15W40-HR9250	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO DIR. R9250	VG460	ESC-VG460	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	TRANSLAÇÃO ESQ. R9250	VG460	ESC-VG460	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO DIR. R9250	VG460	ESC-VG460	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	GIRO LADO ESQ. R9250	VG460	ESC-VG460	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO	R9250 LIEBHERR	P.T.O R9250	SAE90	ESC-SAE90	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	MOTOR 390FL	15W40	LAV-15W40-M390	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	HIDRÁULICO 390FL	15W40	LAV-15W40-H390	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO DIR. 390FL	SAE 50	LAV-SAE50-390	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	TRANSLAÇÃO ESQ. 390FL	SAE 50	LAV-SAE50-390	NORMAL	NORMAL	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	GIRO DIANT. 390FL	SAE 50	LAV-SAE50-390	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	CAT 390FL	GIRO TRAS. 390FL	SAE 50	LAV-SAE50-390	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	SM2500 WIRTGEN	MOTOR SM2500	15W40	LAV-15W40-MMS	NORMAL	MONITORAR	CRÍTICO
CARREGAMENTO LAVRA	SM2500 WIRTGEN	HIDRÁULICO SM2500	HR46	LAV-HR46	NORMAL	MONITORAR	CRÍTICO

Fonte: Próprio Autor.

2.1.5 Análise Visual

A análise visual do óleo lubrificante permite avaliar alterações em sua coloração, bem como identificar a presença de contaminantes externos e partículas ferrosas mediante o processo de decantação. Para a condução desse ensaio, a amostra é posicionada sobre uma base magnética por um período de 6 horas, viabilizando a segregação física dos resíduos por gravidade e atração magnética, conforme ilustrado na Figura 13.

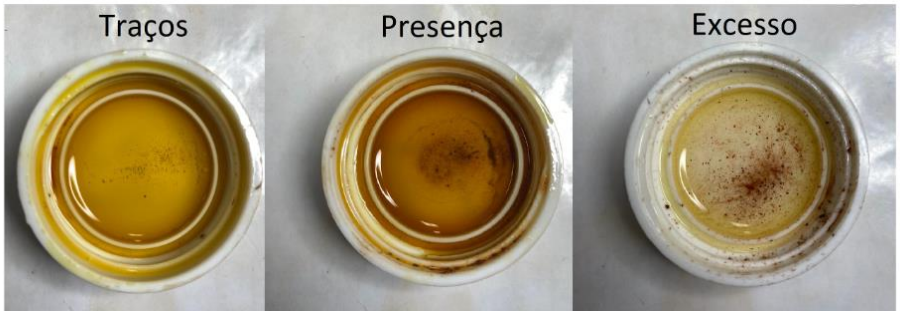
Figura 13 – Disposição das amostras sobre a base magnética para o ensaio de decantação.



Fonte: Próprio Autor.

Após esse período, os sedimentos são avaliados e divididos em duas categorias principais: as impurezas e as limalhas. As impurezas consistem em sedimentos característicos de contaminação externa, sendo classificadas qualitativamente em traços (T), presença (P) e excesso (E), conforme os padrões visuais ilustrados na Figura 14.

Figura 14 – Classificação de impurezas.



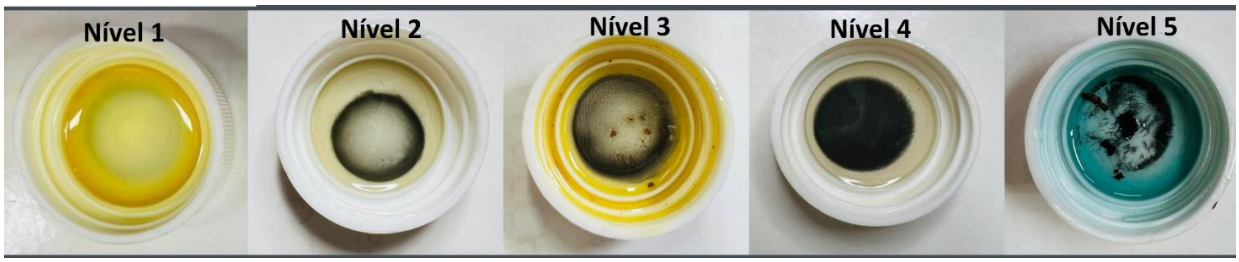
Fonte: Próprio Autor.

As limalhas representam partículas de desgaste metálico atraídas pela base imantada. Esse tipo de sedimento é avaliado com o auxílio de um ímã (Figura 15) e classificado em 5 níveis de severidade, conforme os padrões demonstrados na Figura 16.

Figura 15 – Avaliação do nível de limalhas na amostra com auxílio de ímã.

Fonte: Próprio Autor.

Figura 16 – Classificação das limalhas.



Fonte: Próprio Autor.

Além da análise de sedimentos, a alteração cromática do fluido é avaliada, visto que o desvio da cor original pode ser influenciado por eventos térmicos, reações químicas e contaminações por agentes externos. O escurecimento do óleo é classificado de forma qualitativa em leve (L) ou forte (F), conforme ilustrado na Figura 17. Para todas as análises visuais descritas (impurezas, limalhas e escurecimento), caso não sejam identificadas anomalias, adota-se o status normal (N) no campo do item avaliado.

Figura 17 – Padrão para classificação do escurecimento do óleo lubrificante.

Fonte: Próprio Autor.

O nível de criticidade final dessas variáveis baseia-se em parâmetros que levam em consideração o tipo de componente e o lubrificante utilizado, servindo-se do guia de parâmetros apresentado na Tabela 5. Assim como nos ensaios anteriores, essas classificações qualitativas são computadas na planilha de registro de dados.

Tabela 5 – Guia de parâmetros visual.

INFORMAÇÕES DO EQUIPAMENTO		PARÂMETRO DE LIMALHAS (ANÁLISE VISUAL)					PARÂMETRO DE CONTAMINANTES - ÁGUA E IMPUREZAS - (ANÁLISE VISUAL)			PARÂMETRO DE ESCURECIMENTO - (ANÁLISE VISUAL)	
COMPONENTES	TIPO DE FLUIDO	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5	TRAÇO	PRESENÇA	EXCESSO	LEVE	FORTE
MOTOR 964LB	15W40	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL
HIDRÁULICO 964LB	15W40	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO
TRANSLAÇÃO DIR. 964LB	SAE90	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	MONITORAR	CRÍTICO
TRANSLAÇÃO ESQ. 964LB	SAE90	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	MONITORAR	CRÍTICO
REDUTOR DE GIRO 964LB	SAE90	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	CRÍTICO
P.T.O 964LB	SAE90	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	MONITORAR	CRÍTICO
MOTOR R9250	15W40	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL
HIDRÁULICO R9250	15W40	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO
TRANSLAÇÃO DIR. R9250	VG460	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	MONITORAR	CRÍTICO
TRANSLAÇÃO ESQ. R9250	VG460	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	MONITORAR	CRÍTICO
GIRO LADO DIR. R9250	VG460	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	CRÍTICO
GIRO LADO ESQ. R9250	VG460	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	CRÍTICO
P.T.O R9250	SAE90	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	MONITORAR	CRÍTICO
MOTOR 390FL	15W40	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
HIDRÁULICO 390FL	15W40	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
TRANSLAÇÃO DIR. 390FL	SAE 50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR
TRANSLAÇÃO ESQ. 390FL	SAE 50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR
GIRO DIANT. 390FL	SAE 50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
GIRO TRAS. 390FL	SAE 50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
MOTOR SM2500	15W40	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
HIDRÁULICO SM2500	HR46	NORMAL	NORMAL	MONITORAR	CRÍTICO	CRÍTICO	NORMAL	MONITORAR	CRÍTICO	NORMAL	MONITORAR
RLDD HEAVY DUTY	SAE50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR
RLDE HEAVY DUTY	SAE50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR
RLTD HEAVY DUTY	SAE50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR
RLTE HEAVY DUTY	SAE50	NORMAL	NORMAL	NORMAL	MONITORAR	CRÍTICO	NORMAL	NORMAL	CRÍTICO	NORMAL	MONITORAR

Fonte: Próprio Autor.

Considerando a complexidade da análise de fluidos e o expressivo volume de dados gerados diariamente nas operações de mineração, torna-se indispensável a aplicação de métodos que otimizem esse processo. Nesse cenário, o aprendizado de máquina surge como uma alternativa promissora, capaz de automatizar o diagnóstico, identificar padrões complexos ocultos e realizar previsões com base no histórico das amostras. A seguir, são apresentados os fundamentos teóricos de aprendizado de máquina que embasam o modelo preditivo desenvolvido neste trabalho.

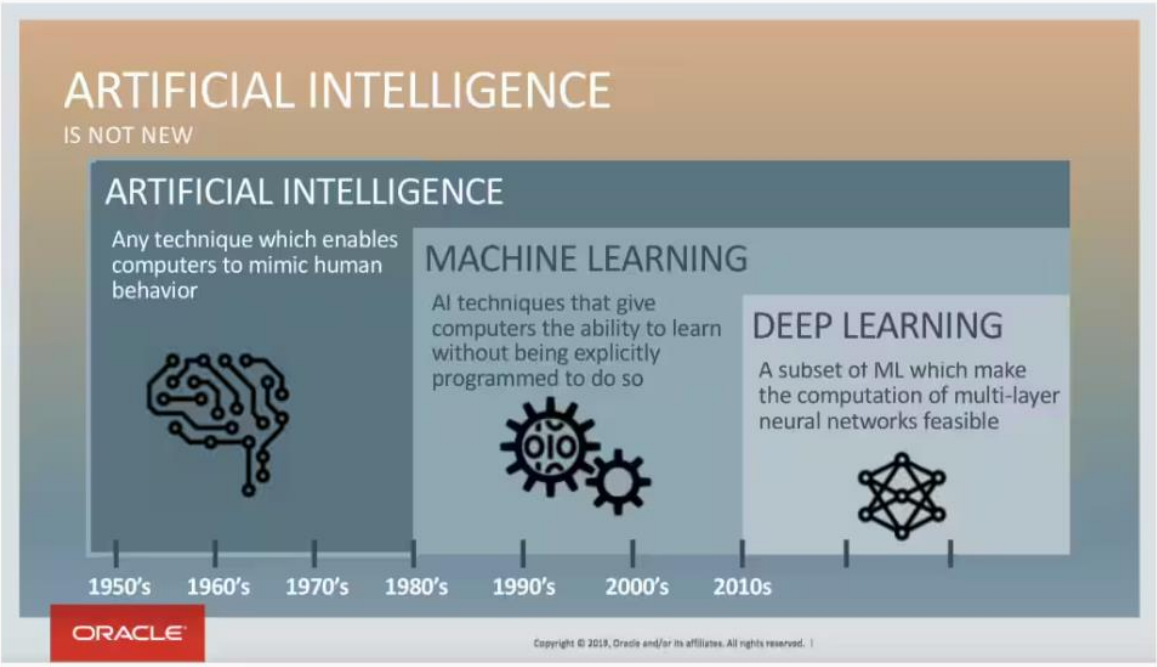
2.2 Aprendizado de Máquina

A inteligência artificial (IA) é um campo recente da ciência da computação que começou a se desenvolver após a Segunda Guerra Mundial. Atualmente, abrange uma ampla gama de subdisciplinas, desde áreas de aplicação geral, como aprendizado e percepção, até tarefas altamente específicas, como jogar xadrez, provar teoremas matemáticos, compor poesia e realizar diagnósticos médicos. O objetivo da IA é organizar e automatizar tarefas intelectuais, o que a torna potencialmente aplicável a qualquer atividade que envolva o raciocínio humano. Por isso, é vista como uma área de conhecimento de caráter universal (RUSSELL; NORVIG, 2004).

A busca por soluções que simulam o raciocínio humano remonta aos estudos filosóficos antigos, que visavam compreender as leis que regem a mente racional. Contudo, esse objetivo começou a se concretizar — ainda que com algumas limitações — a partir de 1950. Assim, o desenvolvimento da IA, impulsionado pelo avanço tecnológico, tornou-se um elemento transformador para a sociedade contemporânea, influenciando diversos âmbitos, como o social, cultural, econômico, político e científico (MARTINS, 2022).

Nesse contexto marcado por rápidas transformações tecnológicas, diversas aplicações emergiram, destacando-se o *Machine Learning* (aprendizado de máquina) e o *Deep Learning* (aprendizado profundo). A linha do tempo da evolução da IA, apresentada pela Oracle (2018), ilustra esse desenvolvimento acelerado. A Figura 18 resume essa evolução cronológica e os seus principais subcampos, destacando como esses conceitos se consolidaram ao longo das décadas.

Figura 18 – Linha do tempo da inteligência artificial.



Fonte: Oracle (2018).

Na primeira seção da imagem, que se refere à década de 1950 em diante, apresenta-se a definição de IA como qualquer técnica que permite aos computadores imitar o comportamento humano. Esse é o campo mais amplo, que engloba uma variedade de abordagens e tecnologias focadas em reproduzir ou simular aspectos do raciocínio e da inteligência humana (ORACLE, 2018).

A partir da década de 1980, o aprendizado de máquina surgiu como uma área específica dentro da IA. Essa vertente é dedicada ao desenvolvimento de algoritmos e sistemas que possibilitam às máquinas aprender e tomar decisões baseadas em dados. Com essa inovação, os sistemas começaram a evoluir por meio de modelos capazes de identificar padrões e realizar predições a partir das informações fornecidas (ORACLE, 2018).

Na década de 2010, o aprendizado profundo passou a ganhar destaque. Este é um subconjunto do aprendizado de máquina que utiliza redes neurais com múltiplas camadas, o que permitiu um salto significativo na capacidade de processamento de dados complexos. O aprendizado profundo é especialmente eficaz para análises e reconhecimento de padrões complexos, aproximando-se ainda mais da forma como o cérebro humano processa informações (ORACLE, 2018).

Para compreender o impacto dessas subáreas, cabe aprofundar o entendimento sobre o Aprendizado de Máquina (AM). De modo geral, o AM visa o desenvolvimento de técnicas de aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma

automática. Um sistema de aprendizado é, essencialmente, um programa de computador que toma decisões apoiado em experiências acumuladas através da solução bem-sucedida de problemas anteriores. Os diversos sistemas de AM possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado (MONARD; BARANAUSKAS, 2003).

Para alcançar esse objetivo, é necessário fornecer ao computador uma grande quantidade de exemplos, que correspondem a hipóteses geradas a partir dos dados. As técnicas de aprendizado de máquina são orientadas por bancos de dados, ou seja, aprendem a partir de grandes volumes de informações, permitindo que os algoritmos formulem hipóteses estatísticas sólidas (LUDERMIR, 2021).

A inferência indutiva é um dos métodos principais no aprendizado de máquina para derivar novos conhecimentos e prever eventos futuros. No entanto, a generalização realizada por meio desse tipo de inferência nem sempre é exata. As chances de gerar generalizações corretas aumentam proporcionalmente à qualidade dos dados: dados mais precisos resultam em modelos e predições mais confiáveis (LUDERMIR, 2021).

Conceitualmente, existem diversos tipos de aprendizado de máquina, sendo os principais o aprendizado supervisionado, o não supervisionado e o por reforço. No aprendizado supervisionado, o algoritmo é treinado com um conjunto de dados rotulado (pares de entrada e saída esperada), permitindo que o modelo faça previsões ou classificações em novos dados. No aprendizado não supervisionado, o algoritmo analisa apenas os dados de entrada, e o objetivo é encontrar padrões ou estruturas ocultas sem a necessidade de rótulos prévios. Já no aprendizado por reforço, o sistema aprende por meio de interações com um ambiente dinâmico, recebendo recompensas ou punições conforme as ações que executa (SILVA, 2024).

Neste trabalho, o interesse concentra-se em algoritmos de AM do tipo supervisionado, dada a natureza do problema sob estudo. Nesta classe de algoritmos, destacam-se duas categorias principais de problemas e abordagens: a classificação e a regressão. A classificação é empregada quando o objetivo é prever uma variável de saída que seja discreta ou categórica. Em outras palavras, o algoritmo analisa as características dos dados de entrada e os atribui a uma classe ou rótulo predefinido (por exemplo, identificar se um e-mail é "spam" ou "não spam", ou diagnosticar se um paciente está "doente" ou "saudável"). Por outro lado, a regressão é utilizada quando o objetivo é prever um valor numérico contínuo. Nesse caso, o modelo busca mapear a relação matemática entre as variáveis de entrada para estimar uma quantidade real (como prever o preço de um imóvel com base no tamanho e localização, ou estimar a

temperatura de amanhã). Enquanto a classificação lida com categorias e decisões qualitativas, a regressão foca em previsões quantitativas e numéricas (LUDERMIR, 2021; SILVA, 2024).

Em suma, o aprendizado de máquina tem transformado diversos setores ao permitir a automação de tarefas complexas, a personalização de serviços, a otimização de processos e a extração de inteligência a partir de grandes volumes de dados. É importante ressaltar que o êxito dessa tecnologia não se baseia apenas na sofisticação dos algoritmos, mas também na disponibilidade de dados confiáveis, na capacidade de interpretar e explicar os resultados gerados e na estrita consideração de questões éticas e sociais relacionadas ao seu uso.

2.3 Aprendizagem Supervisionada

O aprendizado de máquina, conforme definido por Monard e Baranauskas (2003), envolve a construção de algoritmos que permitem que os sistemas aprendam a partir de dados, a fim de fazer previsões ou tomar decisões sem serem programados explicitamentepara essas tarefas. No caso específico da aprendizagem supervisionada, o sistema aprende a partir de conjuntos de dados rotulados, ou seja, exemplos previamente fornecidos, representados por $E = \{E_1, E_2, \ldots, E_N\}$, em que cada exemplo E_i possui um rótulo associado que identifica a classe à qual pertence (VANKOV; GADHOUMI, 2024). De forma mais formal, cada exemplo pode ser descrito como uma tupla $E_i = (x_i, y_i)$, onde x_i é o vetor de características ou atributos, e y_i é o valor da classe correspondente. O objetivo do aprendizado supervisionado é construir uma função $y = f(x)$, capaz de generalizar o relacionamento entre os vetores de entrada x e os valores de saída y, permitindo prever resultados para novos dados.

Os algoritmos de aprendizado supervisionado são projetados para identificar padrões em dados rotulados, destacando-se no escopo deste trabalho as técnicas de Árvore de Decisão, *Random Forest*, KNN, SVM e a rede neural MLP (BATISTA, 2003; FACELI et al., 2011). Cada uma dessas técnicas possui características estruturais próprias que as tornam mais ou menos adequadas conforme o contexto, a dimensionalidade e o tipo de dado analisado (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Segundo Medeiros (2022), as Árvores de Decisão são métodos intuitivos e de fácil interpretação, que representam o processo de decisão por meio de uma estrutura hierárquica de nós e ramificações. A partir desse conceito fundamental, o algoritmo *Random Forest* atua como um método de aprendizado por conjunto (*ensemble*), combinando múltiplas árvores de decisão independentes criadas por meio de subamostragens dos dados. Essa estratégia gera uma

predição final combinada mais robusta e mitiga significativamente o risco de sobreajuste (*overfitting*) (LUDERMIR, 2021).

Para cenários baseados em similaridade e proximidade geométrica, o KNN se destaca por sua simplicidade conceitual, classificando novos dados com base na distância (frequentemente a Euclidiana) em relação aos k vizinhos mais próximos já conhecidos no espaço de atributos (FACELI et al., 2011). Em contrapartida, SVM são amplamente reconhecidas pela alta eficiência em espaços de grande dimensão, operando através da busca por um hiperplano ótimo que separe as classes com a máxima margem de distância possível entre os pontos mais próximos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Por fim, no domínio das redes neurais artificiais, o MLP surge como uma arquitetura de múltiplas camadas com fluxo de informações alimentado adiante (*feedforward*). Utilizando o algoritmo de retropropagação (*backpropagation*) para o ajuste de seus pesos sinápticos, a MLP é capaz de mapear e aproximar relações matemáticas altamente complexas e de natureza não linear entre os dados de entrada e saída (HAYKIN, 2007).

As aplicações do aprendizado supervisionado são vastas e abrangem diversas áreas. Na Ciência da Informação, Martins e Montereis (2022) destacam seu uso em classificação de documentos, recuperação de informações e recomendação de conteúdo. Sistemas supervisionados tornam a categorização de informações mais precisa e melhoram a experiência do usuário. Um exemplo prático é o uso de sistemas de recomendação, como os utilizados por plataformas de *streaming*, que sugerem conteúdos com base em preferências anteriores do usuário.

Apesar de suas inúmeras vantagens, o aprendizado supervisionado enfrenta desafios significativos. Um dos principais é a necessidade de grandes conjuntos de dados rotulados, cuja obtenção pode ser cara e demorada, exigindo conhecimento especializado. Haykin (2007) ressalta que a escassez de dados rotulados de alta qualidade pode comprometer o desempenho dos modelos. Outro desafio importante é o sobreajuste, situação em que o modelo se ajusta excessivamente aos dados de treinamento e falha em generalizar para novos exemplos, conforme apontam Shalev-Shwartz e Ben-David, (2014). Técnicas como validação cruzada e regularização são amplamente utilizadas para mitigar esse problema, avaliando a capacidade de generalização do modelo por meio da divisão do conjunto de dados em partes de treino e teste.

Além disso, a interpretação dos modelos pode ser um obstáculo, principalmente nas redes neurais artificiais, que funcionam como “caixas-pretas”, dificultando a compreensão das decisões tomadas. Essa limitação é especialmente crítica em áreas sensíveis, como saúde e transportes, onde a transparência das decisões é essencial. Nesse sentido, Sagar e

Nanjundeswaraswamy (2019) destacam que o desenvolvimento de modelos mais explicáveis é uma tendência crescente, particularmente em sistemas autônomos, como veículos inteligentes.

Por fim, o aprendizado supervisionado tem evoluído para incluir novas abordagens, como o aprendizado por transferência (*transfer learning*) e o aprendizado ativo (*active learning*), que reduzem a necessidade de grandes volumes de dados rotulados. Segundo Dos Reis (2016), o aprendizado por transferência permite reutilizar o conhecimento adquirido em uma tarefa prévia para outra semelhante, economizando tempo e recursos computacionais, enquanto o aprendizado ativo possibilita que o modelo selecione quais exemplos devem ser rotulados, otimizando o processo de treinamento.

2.4 Classificação Supervisionada

De acordo com Campos (2020) o aprendizado de máquina supervisionado é uma das abordagens mais utilizadas no campo da inteligência artificial. Nessa abordagem, os modelos são treinados com dados rotulados, ou seja, com exemplos cujas respostas são conhecidas. Considera-se uma amostra $(X_1, Y_1), \dots, (X_n, Y_n) \sim (X, Y)$, assumidos como independentes com o objetivo de construir uma função de predição $f(x)$ capaz de classificar corretamente novas observações $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$ (IZBICKI; SANTOS, 2020). Assim, deseja-se que $f(x_{n+1}) \approx y_{n+1}, \dots, f(x_{n+m}) \approx y_{n+m}$, garantindo boa capacidade de generalização para dados novos.

No contexto prático, diferentes estruturas de problemas exigem tipos distintos de classificação. A classificação binária é a forma mais comum e envolve apenas duas classes mutuamente exclusivas, como as decisões entre “positivo” e “negativo” ou “verdadeiro” e “falso” (RODRIGUES et al., 2021). Já a classificação multiclasse expande esse conceito para cenários que lidam com três ou mais categorias, como no reconhecimento de dígitos manuscritos ou na rotulação de imagens (GARZILLO, 2022). Existe ainda a classificação multirrótulo, na qual uma única instância pode receber simultaneamente múltiplas etiquetas não excludentes, a exemplo de um filme categorizado nos gêneros de “ação” e “ficção científica” ao mesmo tempo.

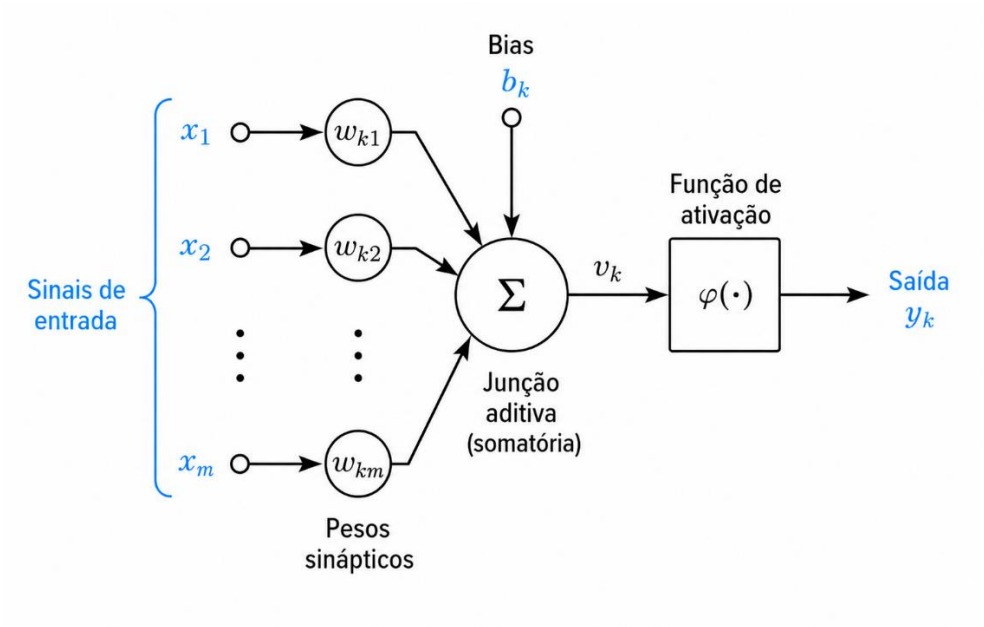
Independentemente do tipo adotado, a avaliação do desempenho dos classificadores se configura como uma etapa essencial do pipeline de desenvolvimento. A métrica de acurácia, embora meça a proporção global de acertos, pode ser severamente inadequada na presença de dados desbalanceados (BATISTA, 2017). Para contornar essa limitação e obter uma visão mais

equilibrada do modelo, recomenda-se a análise conjunta de medidas como precisão, *recall* e *F1-score* (PECCATIELLO, 2023). Além desses índices estatísticos, a matriz de confusão atua como uma ferramenta visual indispensável, permitindo mapear exatamente onde ocorrem os erros de falso positivo e falso negativo para guiar os ajustes e melhorias nos modelos (RODRIGUES, 2015). A seguir, são apresentados os principais algoritmos utilizados e, posteriormente, as medidas de desempenho aplicáveis a esse tipo de problema.

2.4.1 Perceptron de Múltiplas Camadas (MLP)

O perceptron é um dos modelos mais simples de redes neurais artificiais, sendo composto por um único neurônio e uma única camada. Desenvolvido por Rosenblatt (1958), esse modelo introduziu o conceito de treinamento supervisionado em redes neurais, no qual os pesos das conexões são ajustados com base no erro entre a saída desejada e a saída produzida pela rede (FACELI et al., 2011). Apesar de sua simplicidade e de apresentar bons resultados em problemas de classificação elementares, o perceptron possui a limitação geométrica de estar restrito à resolução de problemas linearmente separáveis. A estrutura fundamental desse modelo, evidenciando seus pesos e o neurônio único, é apresentada na Figura 19. Esse arranjo estrutural baseia-se em três elementos básicos descritos por Haykin (2009).

Figura 19 – Modelo não linear de um neurônio.



Fonte: Adaptada de Haykin (2009).

Em termos matemáticos, o funcionamento desse neurônio k é formalizado por meio de um par de equações. Primeiro, cada sinal de entrada x_j é multiplicado por seu respectivo peso sináptico w_{kj}, onde o primeiro índice (k) indica o neurônio em questão e o segundo índice (j) identifica a extremidade de entrada da sinapse. O combinador linear efetua a soma ponderada desses sinais, gerando a saída u_k, conforme apresentado na Equação 2.1.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{2.1}$$

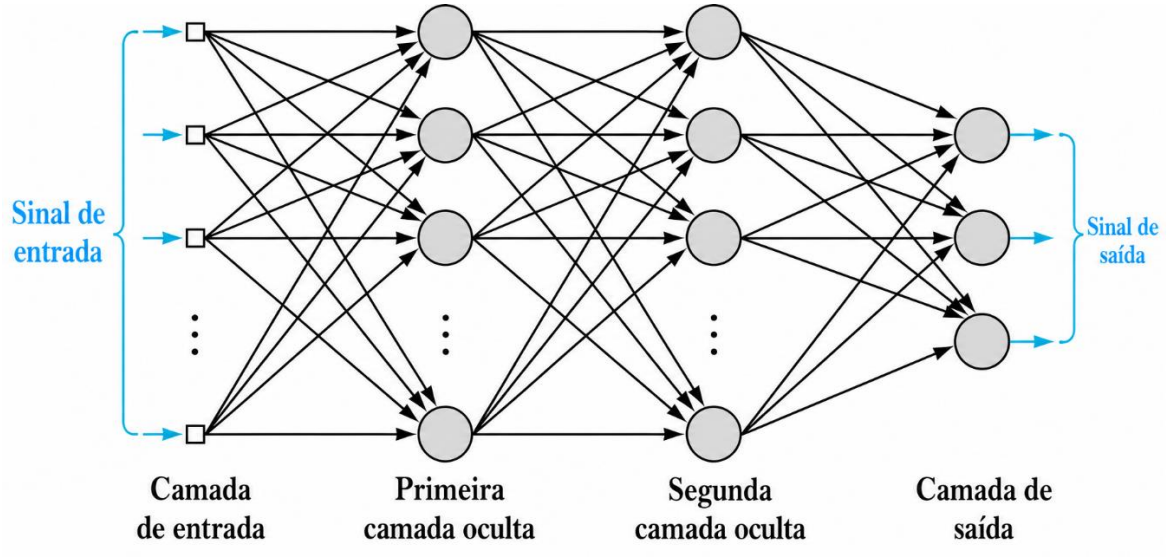
Na sequência, aplica-se um termo de polarização externo (*bias*), denotado por b_k, que gera o potencial de ativação ou campo local induzido $v_k = u_k + b_k$. Esse potencial é submetido a uma função de ativação $\varphi(\cdot)$, que limita a amplitude do sinal e produz a saída final y_k, conforme formalizado na Equação 2.2.

$$y_k = \varphi(v_k) \tag{2.2}$$

Com o objetivo de superar essa limitação, surgiu o MLP, uma rede neural que, por meio de suas camadas intermediárias e funções de ativação não lineares, torna-se capaz de mapear problemas não linearmente separáveis. Sua arquitetura é, em geral, totalmente conectada, o que significa que cada neurônio de uma camada está conectado a todos os neurônios da camada subsequente (MARSLAND, 2009).

O MLP é formado por três tipos fundamentais de camadas: camada de entrada, camadas ocultas (ou intermediárias) e camada de saída. A Figura 20 apresenta a estrutura de um perceptron multicamadas com duas camadas ocultas. Nesse arranjo, o processamento dos sinais ocorre de forma sequencial e direta, avançando da esquerda para a direita, camada por camada, até alcançar a saída, caracterizando o fluxo conhecido como *feedforward* (HAYKIN, 2009).

Figura 20 – Estrutura de um perceptron multicamadas com duas camadas ocultas.



Fonte: Adaptada de Haykin (2009).

O aprendizado da rede neural depende de um mecanismo iterativo denominado retropropagação do erro (*backpropagation*), proposto por Rumelhart, Hinton e Williams (1986). Esse método baseia-se em um processo composto por duas etapas principais: a propagação direta (*forward*), na qual os sinais de entrada são processados sequencialmente pelas camadas da rede até a obtenção das saídas, e a propagação reversa (*backward*), responsável pelo ajuste dos pesos sinápticos a partir do erro observado (MARSLAND, 2009)

Após a fase de propagação direta, as saídas geradas pela rede são comparadas com os valores reais desejados, possibilitando o cálculo de uma função de custo. Para evitar o cancelamento mútuo entre erros positivos e negativos, utiliza-se a função de erro baseada na soma dos quadrados das diferenças entre a saída obtida e o valor alvo, conforme expresso na Equação 2.3.

$$E(t,y) = \frac{1}{2}\sum_{k=1}^{n}(t_k - y_k)^2 \tag{2.3}$$

onde t_k representa o valor alvo (desejado) e y_k denota a saída efetivamente produzida pela rede para o k-ésimo neurônio de saída. A minimização desse erro é realizada por meio do método de descida do gradiente, que busca reduzir gradualmente o valor da função de erro ao ajustar os pesos na direção oposta ao gradiente da função (MARSLAND, 2009; FACELI et al., 2011).

Durante a propagação reversa, o erro calculado na camada de saída é distribuído retroativamente para as camadas anteriores por meio da aplicação de derivadas parciais e da

regra da cadeia, permitindo identificar a contribuição matemática de cada peso para o erro total do sistema. Como as entradas e as funções de ativação permanecem constantes durante aquela iteração de treinamento, apenas os pesos sinápticos e os vieses (*biases*) são ajustados (MARSLAND, 2009). Dessa forma, a retropropagação do erro possibilita o treinamento eficiente de redes neurais com múltiplas camadas, tornando viável a modelagem de problemas complexos de classificação e regressão.

2.4.2 Máquina de Vetores de Suporte (SVM)

As Máquinas de Vetores de Suporte (*Support Vector Machines* – SVM) constituem um dos métodos mais consolidados no aprendizado supervisionado, tendo como objetivo construir um hiperplano que separe duas classes maximizando a margem entre os exemplos mais próximos dessa fronteira, o que confere maior robustez ao processo de classificação (GONÇALVES, 2010). Além desse fundamento geométrico, as SVMs se baseiam na Teoria do Aprendizado Estatístico, segundo a qual a capacidade de generalização do modelo é alcançada por meio da minimização do risco estrutural, reduzindo o *overfitting* e assegurando bom desempenho tanto em dados de treinamento quanto de teste (CORTES; VAPNIK, 1995).

Do ponto de vista matemático, as SVMs estabelecem condições formais para selecionar, entre diversos classificadores possíveis, aquele que apresente melhor desempenho esperado (DOSCIATTI; FERREIRA; PARAISO, 2013). Originalmente formulada para tarefas de classificação binária, a técnica busca o hiperplano ótimo que maximize a distância entre as classes, ou seja, a margem (GONÇALVES, 2010; LORENA et al., 2007). O hiperplano de separação é definido pela função de decisão, conforme apresentado na Equação 2.4:

$$f(x) = wx + b \tag{2.4}$$

onde w representa o vetor de pesos e x o vetor de características da amostra (OLIVEIRA FILHO, 2024). A margem, definida como a menor distância entre os exemplos e o hiperplano, determina diretamente a capacidade de generalização do classificador.

Em problemas linearmente separáveis, utiliza-se a formulação de margem rígida. Entretanto, em cenários com ruído, sobreposição entre classes ou ausência de separabilidade perfeita, emprega-se a margem suave, que introduz variáveis de relaxamento ζ_i, permitindo a

violação controlada das restrições (FACELI et al., 2011). Essas variáveis são definidas para cada amostra sujeitas às restrições apresentadas a seguir.

Para $y_i = +1$:

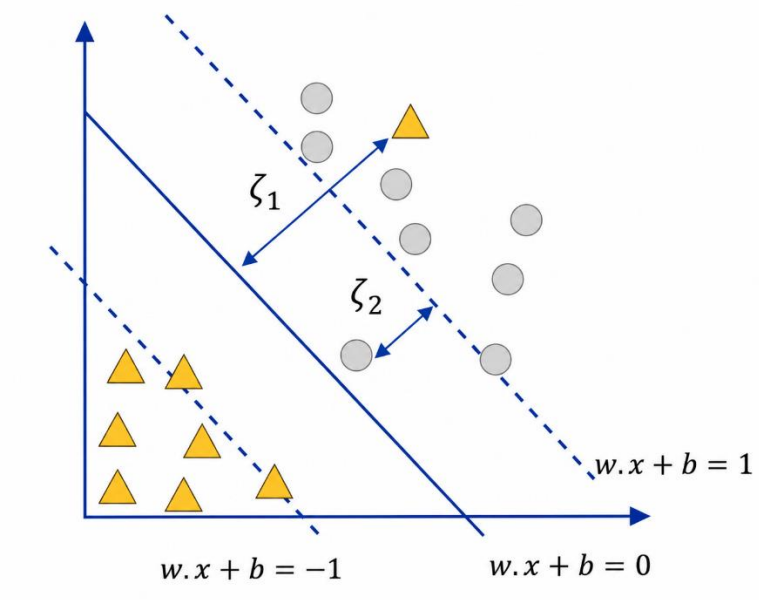
$$\zeta_i(w,b) = \begin{cases} 0 & se\ wx_i + b \geq 1 \\ 1 - wx_i + b & se\ wx_i + b < 1 \end{cases} \quad (2.5)$$

Para $y_i = -1$:

$$\zeta_i(w,b) = \begin{cases} 0 & se\ wx_i + b \geq -1 \\ 1 + wx_i + b & se\ wx_i + b < -1 \end{cases} \quad (2.6)$$

O valor de ζ_i indica o grau de violação da margem: valores iguais a zero representam classificações corretas; valores entre 0 e 1 indicam amostras dentro da margem; valores superiores a 1 correspondem a erros de classificação (OLIVEIRA FILHO, 2024). Esse comportamento e a penalização geométrica das amostras são ilustrados na Figura 21.

Figura 21 – Variáveis de relaxamento.



Fonte: Oliveira Filho (2024).

Assim, o problema de otimização da SVM com margem suave é expresso pela função objetivo apresentada na Equação 2.7:

$$\varepsilon(w,x) = \| w \|^2 + C \sum_{i=1}^{n} \zeta_i(w,b) \quad (2.7)$$

na qual o parâmetro C atua como um hiperparâmetro de regularização, controlando o equilíbrio entre a penalização dos erros de classificação e a complexidade do modelo (FACELI et al., 2011; VILAS BOAS, 2021)

Contudo, muitos problemas reais não apresentam separabilidade linear no espaço original das características. Para lidar com tais casos, as SVMs utilizam funções *Kernel*, capazes de mapear os dados para um espaço de maior dimensionalidade, no qual uma separação linear passa a ser possível. O *Kernel* é definido conforme apresentado na Equação 2.8:

$$K(x_i, x_j) = \Phi(x_i)\Phi(x_j) \quad (2.8)$$

em que Φ é uma transformação não linear para o espaço de características (LORENA et al., 2007). A grande vantagem desse procedimento é que o mapeamento Φ não precisa ser explicitamente conhecido, permitindo operar diretamente em alta dimensionalidade com baixo custo computacional. Entre os *Kernels* mais utilizados destacam-se o polinomial, o Gaussiano de base radial (RBF) e o sigmoide (GONÇALVES, 2010). Tanto o *Kernel* quanto o parâmetro de regularização C são hiperparâmetros essenciais que influenciam significativamente o desempenho do modelo (VILAS BOAS, 2021)

Os vetores de suporte exercem papel central na maximização da margem e na robustez da classificação. As SVMs apresentam vantagens importantes, como boa capacidade de generalização mesmo em alta dimensionalidade, otimização convexa com solução global garantida e flexibilidade para lidar com problemas lineares e não lineares (LORENA et al., 2007). Essas características tornam o método amplamente utilizado em visão computacional, análise de sinais e, em especial, em sistemas de Interface Cérebro-Computador (ICC), onde demonstra desempenho competitivo na classificação de sinais de EEG (NICOLÁS-ALONSO; GOMEZ-GIL,2012).

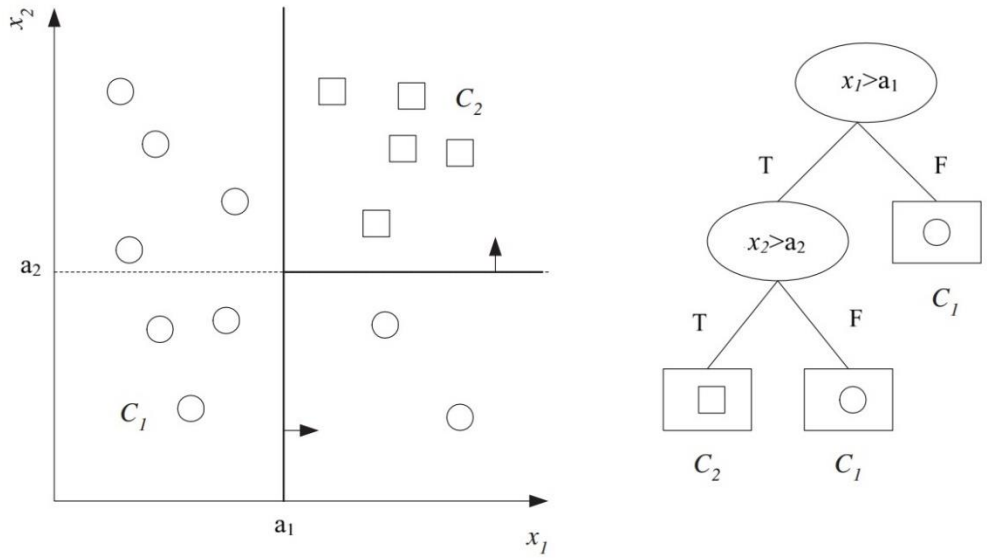
2.4.3 Árvores de Decisão

Uma árvore de decisão é uma estrutura de dados hierárquica que implementa a estratégia de dividir para conquistar. É um método não paramétrico eficiente, que pode ser usado tanto

para classificação quanto para regressão. Trata-se de um algoritmo de aprendizado que constrói a árvore a partir de uma amostra de treinamento rotulada, assim como a forma de converter a árvore em um conjunto de regras simples, fáceis de entender (ALPAYDİN, 2010).

Sua estrutura pode ser representada como um grafo direcionado sem ciclos, composto por nós de decisão, nos quais são aplicados testes condicionais geralmente univariados, e por nós folha responsáveis pela definição da saída do modelo (FACELI et al., 2011). Os nós de decisão m implementam uma função de teste $f_m(x)$ cujos resultados discretos rotulam os ramos da árvore (ALPAYDİN, 2010). O processo tem início no nó raiz e segue até um nó terminal, no qual os exemplos são associados à variável alvo, sendo utilizada, em problemas de classificação, a classe mais frequente, e, em problemas de regressão, valores estatísticos como a média ou a mediana, conforme a função de custo adotada. Essa organização torna as árvores de decisão modelos interpretáveis, permitindo compreender de forma clara como os atributos influenciam o resultado (ALPAYDİN, 2010; FACELI et al., 2011).

Figura 22 – Conjunto de dados da árvore de decisão.



Fonte: Adaptada ALPAYDİN (2010).

A Figura 22 representa o conjunto de dados da árvore de decisão, os símbolos ovais representam os nós de decisão e os retângulos representam os nós folha, sendo que cada nó corresponde a uma região do espaço definido pelos atributos. Os nós de decisão realizam divisões baseadas em um único atributo ao longo de um único eixo, e as divisões sucessivas são ortogonais entre si. Após uma divisão que gera uma região pura, como no conjunto $(x \mid x_1 < w_{10})$, não há necessidade de novas subdivisões (ALPAYDİN, 2010). As regiões

associadas às folhas são mutuamente distintas, não apresentam interseção entre si e, em conjunto, cobrem todo o espaço de atributos (FACELI et al., 2011).

2.4.4 Random Forest

O algoritmo *Random Forest* (Florestas Aleatórias) gera várias árvores de decisão nas quais suas previsões são combinadas por votação uniforme (FACELI, et al., 2011). O classificador utiliza uma técnica de *boostraping*, que é retirar dados aleatoriamente do conjunto de dados original, com reposição, para fazer diversos conjuntos de dados menores e, a partir disto, ele gera as múltiplas árvores de decisão, que, quando colocadas para teste, fazem suas predições uma a uma e o resultado é o voto majoritário (RUIVO, 2023). Dessa forma, cada árvore toma decisões de maneira independente e, ao final, suas predições individuais são agregadas por meio de votação majoritária (para classificação) ou média (para regressão), produzindo uma decisão única e mais confiável (DATA SCIENCE TEAM, 2020).

Como observa Paiva (2025), a aleatoriedade na escolha das características desempenha papel fundamental para a eficiência desse método. Ao limitar o conjunto de atributos disponíveis em cada nó, evita-se que todas as árvores tomem decisões semelhantes, aumentando a variabilidade entre os modelos e reduzindo a correlação entre eles. Esse mecanismo impede que uma única característica domine o processo de divisão, criando árvores estruturalmente distintas e com elevado valor de variância, o que melhora a capacidade de generalização da floresta como um todo.

Apesar de as árvores serem construídas a partir de amostras *bootstrap*, é comum que elas apresentem correlação entre si, o que reduz a eficácia do grupo. Para minimizar esse problema, o método incorpora a seleção aleatória de apenas m atributos dentre o total d disponíveis em cada divisão, sendo $m < d$. A escolha adequada desse parâmetro é importante para promover diversidade entre as árvores. Em geral, valores como $m \approx d/3$ tendem a produzir bons resultados práticos. Quando $m = d$, o método se comporta como bagging, perdendo parte da redução de correlação que caracteriza o *Random Forest* (IZBICKI; SANTOS, 2020).

Outra característica relevante desse algoritmo é sua robustez ao aumento do número de árvores. Diferentemente de outros métodos, a *Random Forest* não apresenta *overfitting* com a elevação desse parâmetro, tornando desnecessário o uso de validação cruzada para definir a quantidade ideal de estimadores, visto que o erro se estabiliza. Implementações modernas,

como o pacote *ranger*, permitem treinar florestas com grande quantidade de observações mantendo um desempenho computacional satisfatório (IZBICKI; SANTOS, 2020).

No geral, as Florestas Aleatórias destacam-se por sua capacidade de produzir predições mais suaves, estáveis e representativas da estrutura real dos dados quando comparadas a uma única árvore. A combinação de técnicas de amostragem, aleatoriedade na seleção de atributos e agregação estatística torna esse algoritmo um dos mais eficientes e amplamente utilizados no aprendizado de máquina contemporâneo.

2.4.3 K-ésimos Vizinhos mais Próximos (KNN)

O algoritmo dos K-ésimos vizinhos mais próximos é uma técnica de aprendizado supervisionado utilizada em problemas de classificação e regressão, destacando-se pela sua simplicidade conceitual e facilidade de interpretação. Seu funcionamento está fundamentado em noções intuitivas, como a medida de distância entre dados (pontos), o que contribui para sua boa acessibilidade mesmo para usuários com menor familiaridade técnica. A ideia central do método é que instâncias semelhantes, localizadas próximas umas das outras no espaço de atributos (*features*), tendem a pertencer à mesma classe. Dessa forma, a classificação de um novo dado ocorre a partir da análise das classes associadas aos seus vizinhos mais próximos (PAIVA, 2025).

Ao contrário de modelos paramétricos, o KNN não constrói um modelo explícito durante a fase de treinamento. Em vez disso, armazena todo o conjunto de dados e realiza previsões diretamente a partir da comparação entre o novo ponto e as instâncias já observadas (PAIVA, 2025). Para isso, calcula-se a distância entre a nova amostra e cada elemento do conjunto de treinamento. A distância Euclidiana é a métrica mais empregada, mas há alternativas, como as distâncias Manhattan, Hamming e Minkowski, que podem ser mais adequadas em cenários envolvendo atributos categóricos ou variáveis com escalas distintas (PAIVA, 2025; OLIVEIRA FILHO, 2024).

A métrica mais utilizada é a distância euclidiana, na qual x_i e y_j são dois objetos representados por vetores no espaço $\mathbb{R}^n$ que correspondem aos valores assumidos por cada coordenada (FACELI et al., 2011). Essa distância é definida conforme apresentado na Equação 2.9:

$$d(x_i, y_j) = \sqrt{\sum_{l=1}^{n} (x_i^l - y_j^l)^2} \tag{2.9}$$

A escolha do parâmetro K exerce influência direta no desempenho do algoritmo KNN. Valores muito baixos de K tendem a tornar o modelo excessivamente sensível a ruídos, uma vez que a classificação passa a depender de um número reduzido de vizinhos, o que pode resultar em *overfitting*. Por outro lado, valores elevados de K podem comprometer a capacidade de previsão do algoritmo, pois incluem vizinhos mais distantes e potencialmente pertencentes a classes menos semelhantes, dificultando a identificação de padrões locais e levando ao underfitting (PAIVA, 2025). A definição de um valor adequado para K depende, portanto, de um equilíbrio entre esses dois extremos, sendo fortemente influenciada pela dispersão e pela sobreposição dos dados no conjunto analisado (VILAS BOAS, 2021; OLIVEIRA FILHO, 2024).

Dentre as suas propriedades operacionais, o KNN é classificado como um algoritmo preguiçoso (*lazy*), armazenando os dados e processando-os apenas no momento da classificação. Ele também é considerado incremental, pois permite incluir novos exemplos sem a necessidade de reprocessar o modelo do zero, construindo aproximações locais que o tornam eficiente em problemas complexos (FACELI et al., 2011). Em contrapartida, entre as principais desvantagens do método está seu alto custo computacional, já que ele precisa calcular a distância entre o ponto a ser classificado e todos os pontos do conjunto de treinamento, repetindo esse processo para cada nova instância (PAIVA, 2025). Adicionalmente, por não gerar uma representação compacta dos dados, o método mostra-se sensível a atributos redundantes ou irrelevantes e sofre com a maldição da dimensionalidade, na qual a eficácia geométrica diminui à medida que o número de atributos cresce, tornando as distâncias entre os vizinhos menos representativas (FACELI et al., 2011).

2.5 Métricas de Desempenho

No contexto do aprendizado de máquina, especialmente em modelos de classificação, é fundamental dispor de métodos que permitam avaliar o desempenho obtido após o treinamento do algoritmo. Embora a acurácia seja uma métrica amplamente utilizada, sua aplicação isolada

pode não ser suficiente para representar adequadamente o comportamento do modelo, principalmente em diferentes tipos de bases de dados. Nesse sentido, as métricas de avaliação consistem em parâmetros quantificáveis que possibilitam analisar e comparar o desempenho do modelo sob distintas perspectivas, incluindo medidas como precisão, recall e F1-score. A utilização de múltiplas métricas permite uma compreensão mais abrangente dos resultados, fornecendo suporte para a tomada de decisão conforme os objetivos da aplicação. A escolha da métrica mais adequada deve considerar os critérios prioritários do problema, possibilitando verificar se o modelo apresenta desempenho satisfatório e se está apto para sua implementação em cenários reais.

2.5.1 Matriz de Confusão

A matriz de confusão consiste em uma estrutura quadrada de dimensão $N \times N$, na qual N corresponde ao total de classes do problema. Nessa matriz, as linhas indicam as classes previstas pelo modelo, enquanto as colunas representam as classes reais dos dados. Essa organização possibilita uma avaliação detalhada do desempenho do classificador, evidenciando os acertos e erros para cada classe individualmente. Esse tipo de análise pode ser aplicado tanto em problemas de classificação binária quanto em cenários com múltiplas classes (SATHYANARAYANAN; TANTRI, 2024). Além disso, permite identificar quantas vezes instâncias de uma determinada classe foram corretamente classificadas ou confundidas com outras, contribuindo para uma análise mais precisa do modelo (GÉRON, 2019). A Tabela 6 mostra a estrutura geral da matriz de confusão.

Tabela 6 - Matriz de Confusão.

		Classes Previstas	
		Positivo	**Negativo**
Classe Real	**Positivo**	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Próprio Autor.

Há dois tipos de previsões: as corretas e as incorretas.

- Verdadeiro Positivo (VP): Tanto os valores reais quanto a previsão são positivos;

- Falso Positivo (FP): Embora a previsão seja positiva, o valor real é negativo;
- Verdadeiro Negativo (VN): O valor real é negativo e a previsão também é negativa;
- Falso Negativo (FN): Embora tenha sido previsto como negativo, a amostra é positiva.

2.5.2 Acurácia de Classificação

A acurácia representa a proporção de previsões corretas em relação ao total de previsões realizadas, indicando com que frequência o modelo acerta. Segundo Sathyanarayanan e Tantri (2024) ela é uma métrica útil quando as classes de um conjunto de dados estão equilibradas. Porém, em conjuntos desbalanceados, a acurácia pode ser enganosa, pois o modelo pode atingir valores elevados apenas prevendo a classe majoritária, sem identificar corretamente as classes minoritárias. Esse fenômeno é conhecido como paradoxo da acurácia, evidenciando que a acurácia, sozinha, nem sempre é suficiente para avaliar adequadamente o desempenho de modelos preditivos. Ela pode ser expressa pela Equação 2.10.

$$\text{Acurácia} = \frac{(\text{Número de instâncias classificadas corretamente})}{(\text{Número total de instâncias})} = \frac{(VP + VN)}{(VP + FP + VN + FN)} \quad (2.10)$$

2.5.3 Precisão

A precisão (precision) é uma métrica utilizada para avaliar a proporção de previsões positivas corretas realizadas por um modelo. Ela é especialmente útil em problemas com múltiplas classes ou em situações nas quais os falsos positivos possuem maior impacto. Sua fórmula é dada pela Equação 2.11:

$$Precisão = \frac{(VP)}{(VP + FP)} \quad (2.11)$$

Um valor alto de precisão indica que o modelo realiza previsões positivas com maior assertividade e apresenta baixa taxa de falsos positivos. Entretanto, essa métrica não considera

os falsos negativos, podendo não representar adequadamente o desempenho do modelo em cenários nos quais identificar corretamente os casos positivos é essencial.

2.5.4 Recall

O *Recall* (também denominado revocação, sensibilidade ou taxa de verdadeiros positivos) é a métrica que mensura a capacidade de um modelo de classificação em identificar corretamente todas as instâncias pertencentes à classe positiva. Em outras palavras, essa métrica avalia a proporção de exemplos positivos reais que foram efetivamente detectados pelo algoritmo. Sua formulação matemática é expressa pela Equação 2.12:

$$Recall = \frac{(VP)}{(VP + FN)} \tag{2.12}$$

Um valor elevado de *recall* indica que o modelo possui uma baixa taxa de falsos negativos (FN), sendo altamente eficiente em não deixar passar casos positivos sem detecção. Conforme apontado por Géron (2019), o *recall* assume papel crítico em cenários nos quais a omissão de um diagnóstico ou de um evento positivo acarreta consequências graves ou custos elevados.

No entanto, há um compromisso natural entre precisão e *recall* (*precision/recall*): a busca por um *recall* próximo a 100% frequentemente induz o modelo a classificar mais instâncias como positivas de forma generalizada, o que tende a aumentar o número de FP e, consequentemente, reduzir a precisão do classificador (SATHYANARAYANAN; TANTRI, 2024). Portanto, a análise isolada do *recall* pode ocultar deficiências no controle de falsos alarmes, exigindo a sua avaliação conjunta com outras métricas.

2.5.5 F1-score

F1-score é uma métrica consolidada que combina a precisão e o *recall* em um único indicador numérico, sendo definida formalmente como a média harmônica entre essas duas

medidas. Ao contrário da média aritmética simples, a média harmônica penaliza severamente desequilíbrios extremos entre os componentes avaliados. Desse modo, para que o modelo obtenha um *F1-score* elevado, é mandatório que ele apresente simultaneamente bons índices de precisão e de *recall*. Matematicamente a métrica é deduzida conforme a Equação 2.13:

$$F1 = 2 \times \frac{Recall \times Precisão}{Recall + Precisão} \tag{2.13}$$

Esta métrica torna-se um parâmetro de avaliação indispensável em problemas que envolvem conjuntos de dados severamente desbalanceados, nos quais a acurácia se mostra uma métrica enganosa devido ao paradoxo da acurácia (SATHYANARAYANAN; TANTRI, 2024). O valor do *F1-score* varia em um intervalo contínuo entre 0 e 1, onde o limite superior indica uma classificação perfeita (ausência completa de falsos positivos e falsos negativos) e o limite inferior representa o colapso preditivo do modelo. Assim, o *F1-score* fornece um panorama robusto e equilibrado sobre o desempenho global do classificador, servindo como uma métrica de consenso para a tomada de decisão em cenários reais (GÉRON, 2019).

3 METODOLOGIA

Este trabalho apresenta o desenvolvimento de um sistema de classificação baseado em algoritmos de aprendizado de máquina, com foco na avaliação do estado de contaminação do óleo lubrificante em equipamentos móveis. Para isso, foram utilizados dados oriundos de análises físico-químicas e inspeções visuais, permitindo a construção de modelos preditivos capazes de categorizar automaticamente as amostras em três classes distintas: normal, monitorar e crítico.

O objetivo da metodologia é descrever os procedimentos de seleção, preparação, modelagem e avaliação dos algoritmos, a fim de identificar o classificador com melhor desempenho. Os critérios de avaliação e os métodos de validação adotados são descritos nas subseções a seguir.

3.1 Conjunto de Dados

A base de dados utilizada neste estudo é composta por 11.853 amostras, coletadas entre os anos de 2021 e 2022 em um laboratório especializado em análise de óleo de uma operação de mineração. Esse volume e a diversidade dos registros fornecem a representatividade e a robustez necessárias para a aplicação das técnicas de aprendizado supervisionado. O dataset é constituído por 19 variáveis (features) de entrada, conforme detalhado na Tabela 7. Para o desenvolvimento e implementação dos modelos preditivos, utilizou-se o ambiente *Google Colaboratory* com a linguagem de programação *Python*, empregando-se a biblioteca *Scikit-learn*, amplamente reconhecida por sua eficácia em tarefas de classificação (SCIKIT-LEARN, 2025).

Tabela 7 – Conjunto de dados.

	FROTA	COMPARTIMENTO	Fe	Cu	Cr	Pb	Sn	Ni	Ag	Al	Si	Na	K	VIS 40°C	ÁGUA	OLEO ESCURO	IMPUREZA	LIMALHAS	SÍLICA	RESULTADO
0	SM2500 WIRTGEN	HIDRÁULICO SM2500	4.66	1.38	0.47	0.65	0.0	0.87	0.35	4.79	4.55	0.32	0.16	45	N	N	T	5	N	CRITICO
1	724K	2º DIFERENCIAL 724K	79.13	5.75	0.48	0.39	0.0	0.32	0.00	1.43	5.46	6.24	0.85	57	P	N	N	5	N	CRITICO
2	JOHN DEERE 850J	COMANDO FINAL ESQ. EXT. 850J	920.31	2.85	4.09	0.00	0.0	4.44	0.00	51.11	55.14	4.03	4.78	74	E	F	N	5	N	CRITICO
3	4844K	CX. DE MARCHA 4844K	135.69	3.14	0.92	0.55	0.0	0.13	0.00	5.12	7.27	1.09	0.00	77	N	N	T	5	N	CRITICO
4	4844K	CX. DE MARCHA 4844K	42.61	4.84	0.41	0.00	0.0	0.07	0.00	12.74	7.05	1.04	0.19	80	N	L	N	5	N	CRITICO

Fonte: Próprio Autor.

A distribuição dessas amostras revela um cenário de desbalanceamento entre as classes da análise, com uma predominância de registros que demandam acompanhamento. A quantificação exata das instâncias correspondente a cada classe está detalhada na Tabela 8.

Tabela 8 – Distribuição das classes no conjunto de dados.

CLASSE	QUANTIDADE
Monitorar	5669
Normal	3319
Crítico	2865

Fonte: Próprio Autor.

O conjunto de dados passou pelas seguintes etapas de pré-processamento:

a) **Análise e tratamento de valores nulos:** realizou-se uma verificação no conjunto de dados por meio da função dropna() da biblioteca *Pandas* para identificar e eliminar eventuais registros com valores ausentes. A eliminação de dados faltantes visa garantir a integridade e a consistência do *dataset* antes da aplicação dos modelos preditivos. Segundo McKinney (2018), essa é uma das estratégias mais diretas e eficazes para garantir que algoritmos supervisionados operem corretamente. Após a varredura, constatou-se que o *dataset* não apresentava valores nulos em suas linhas, mantendo a totalidade de 11.853 amostras preenchidas.

b) **Codificação de variáveis categóricas:** aplicou-se a técnica de *One-Hot Encoding* às variáveis categóricas "FROTA", "COMPARTIMENTO", "ÁGUA", "ÓLEO ESCURO", "IMPUREZA" e "SÍLICA". Essa transformação, realizada por meio da

classe *OneHotEncoder* em conjunto com o *ColumnTransformer*, converte os atributos qualitativos textuais em representações binárias. O procedimento impede que os algoritmos interpretem erroneamente categorias distintas como uma escala ordenada ou atribuam pesos arbitrários entre elas (GÉRON, 2019).

c) **Padronização de variáveis numéricas:** os atributos quantitativos, constituídos pela concentração de metais de desgaste e contaminantes (Fe, Cu, Cr, Pb, Sn, Ni, Ag, Al, Si, Na, K, LIMALHAS) e pela viscosidade cinemática (VIS 40°C), foram normalizados utilizando a classe *StandardScaler*. Essa técnica redimensiona as variáveis numéricas para apresentarem média igual a zero e desvio padrão igual a um. A padronização é indispensável para algoritmos sensíveis à escala dos atributos, como SVM, KNN e Redes Neurais (GÉRON, 2019);

d) **Divisão estratificada dos dados:** o conjuto de dados foi dividido em subconjuntos de treino (80%) e validação (20%) por meio da função *train_test_split* da biblioteca *Scikit-learn*. Utilizou-se o parâmetro stratify referenciado à variável-alvo ("RESULTADO") para mitigar os impactos do desbalanceamento natural das classes (predominância da classe "MONITORAR"). A utilização da divisão estratificada garante a manutenção da proporção original das classes em ambas as partições., esse método é indispensável para mitigar vieses de avaliação e garantir métricas de desempenho confiáveis em cenários com dados desbalanceados.

3.2 Aplicação dos Métodos

Os modelos de classificação supervisionada foram implementados por meio de *pipelines* do *Scikit-learn*, que integram as etapas de pré-processamento e classificação. Essa abordagem assegura a consistência entre treinamento e inferência, além de favorecer a reprodutibilidade dos experimentos (PEDREGOSA et al., 2011). Para evitar vieses decorrentes do uso de configurações padrões (*default*) das bibliotecas, aplicou-se uma variação experimental controlada de hiperparâmetros por meio da técnica de *Grid Search* (busca em grade), combinada com uma validação cruzada de 3 dobras (*3-fold cross-validation*) (GÉRON, 2019). Esse processo testa diferentes combinações de parâmetros para cada algoritmo, selecionando a estrutura ótima baseada no maior desempenho alcançado na métrica *F1-score* Macro, considerada a métrica mais estável para dados desbalanceados.

Além do *F1-score*, a avaliação do desempenho dos modelos otimizados utilizou um conjunto multifacetado das métricas. Foram extraídas a Acurácia Geral, para mensurar a taxa global de acertos, a Precisão, para avaliar a assertividade das predições positivas, e a *Recall*, essencial para medir a capacidade do modelo em capturar a totalidade dos casos reais de cada classe. Complementarmente, foi gerada a Matriz de Confusão para cada classificador, mapeando visualmente os erros e acertos cruzados entre as classes reais e as previstas. A análise conjunta dessas ferramentas é indispensável para diagnosticar o comportamento dos algoritmos diante do desbalanceamento do banco de dados, evidenciando padrões de Falsos Positivos e Falsos Negativos (FACELI et al., 2011).

3.2.1 Random Forest

O modelo baseado em *ensemble* combinou múltiplas árvores de decisão independentes para reduzir a variância global do sistema (BREIMAN, 2001). A busca em grade avaliou o tamanho da floresta alterando a quantidade de estimadores (n_estimators em 50, 100 e 200) combinada à restrição da profundidade máxima das árvores (max_depth em 10, 20 ou None), visando encontrar o ponto ótimo de estabilidade preditiva sem gerar custos computacionais excessivos. A escolha do número de estimadores baseia-se no fato de que a capacidade de generalização do algoritmo tende a se estabilizar a partir de 100 árvores, o que torna conjuntos maiores computacionalmente ineficientes. Paralelamente, os limites de profundidade foram estabelecidos para comparar o comportamento de regras simplificadas (profundidade 10), modelos de maior complexidade (profundidade 20) e o crescimento livre das ramificações (*None*), permitindo identificar na prática o limite do *overfitting* para as características do óleo.

3.2.2 K-Nearest Neighbors (KNN)

O classificador baseado em vizinhança teve o número de vizinhos próximos (n_neighbors) variado entre 3, 5, 7 e 11, avaliando-se também a função de ponderação dos votos (weights) entre uniforme e o inverso da distância. O ajuste busca equilibrar a sensibilidade do modelo a ruídos e a perda de capacidade discriminativa nas fronteiras das classes (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A adoção de valores ímpares para os vizinhos visa evitar

empates na votação das classes, enquanto a distância Euclidiana foi selecionada devido à sua adequação matemática a atributos numéricos que foram previamente padronizados.

3.2.3 Support Vector Machine (SVM)

Na fase de treinamento, o algoritmo baseado em vetores de suporte teve seu hiperplano de separação otimizado a partir da variação do parâmetro de regularização (C em $0.1,1.0,10.0$) e da alternância entre as funções de *kernel* do tipo Linear e RBF. O ajuste do parâmetro C foi delineado para controlar matematicamente o equilíbrio entre o tamanho da margem de separação e a tolerância aos erros de classificação no conjunto de treino. A inclusão do *kernel* RBF justifica-se pela necessidade de mapear correlações não lineares complexas entre os atributos físico-químicos do lubrificante, permitindo que o algoritmo estabeleça fronteiras de decisão curvas e flexíveis para delimitar as condições de desgaste do ativo (GÉRON, 2019).

3.2.4 Árvore de Decisão

O classificador hierárquico teve suas divisões internas avaliadas sob os critérios de impureza de Gini e Entropia. A fim de controlar a tendência natural do algoritmo ao *overfitting*, a profundidade máxima da árvore (max_depth) foi limitada experimentalmente nos níveis 5, 10, 15 ou mantida livre (None). Essa variação atua como uma estratégia de pré-poda (*pre-pruning*), forçando o modelo a gerar regras de decisão mais generalistas diante das variáveis do lubrificante (QUINLAN, 1986).

3.2.5 Multi-Layer Perceptron (MLP)

A rede neural artificial *feedforward* foi configurada com um limite de 1.000 iterações (max_iter=1000) para assegurar margem computacional suficiente para a estabilização e convergência dos pesos sinápticos via algoritmo de *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986). Seguindo as diretrizes de investigação empírica de arquiteturas de rede, a capacidade de mapeamento não linear foi testada sob quatro topologias distintas de camadas ocultas (hidden_layer_sizes): uma configuração rasa de 64 neurônios; uma expansão horizontal de 128 neurônios; duas camadas dispostas em formato de funil decrescente (64, 32),

visando a compressão de características; e uma estrutura profunda de três camadas (128, 64, 32) para capturar correlações ocultas de alta complexidade nos atributos do lubrificante.

3.3 Critérios de Avaliação e Seleção do Modelo

A homologação e a comparação do desempenho dos classificadores otimizados foram realizadas a partir da extração de métricas analíticas no conjunto de validação independente. O diagnóstico do comportamento dos algoritmos baseou-se na análise conjunta da Acurácia Geral, da Precisão, da *Recall* e do *F1-score*, além do mapeamento visual de erros e acertos por meio da Matriz de Confusão.

Contudo, a definição do classificador definitivo para o monitoramento preditivo das frotas não utilizou a acurácia global como critério soberano de escolha. Diante do desbalanceamento natural do banco de dados, onde a classe "MONITORAR" detém a maior concentração de amostras, a taxa isolada de acertos totais pode mascarar falhas graves na identificação das categorias minoritárias. Para contornar essa limitação estatística, adotou-se o índice *F1-score* com a estratégia de agregação *Macro Average* (Média Macro) como o fator decisório e de desempate no processo de *Grid Search*. A média macro garante que as três classes (Normal, Monitorar e Crítico) possuam o mesmo peso e relevância matemática na avaliação final, penalizando severamente os modelos que negligenciem as classes com menor volume de registros.

A escolha do *F1-score* Macro como critério de seleção alinha-se diretamente ao viés operacional e econômico da engenharia de manutenção em ambiente de mineração. Nesse cenário, o impacto financeiro de um Falso Negativo na classe "CRÍTICO" — isto é, classificar incorretamente um componente em iminência de colapso como "NORMAL" — acarreta quebras catastróficas imprevistas, paradas prolongadas de produção e custos elevados de reparo corretivo. Portanto, buscou-se selecionar o modelo que maximizasse a capacidade de capturar os alvos severos (alta revocação) sem comprometer a confiabilidade das predições (alta precisão). O algoritmo que apresentou o maior equilíbrio harmônico quantificado por essa métrica foi definido como a solução ideal para o problema proposto.

4 RESULTADOS E DISCUSSÕES

Após a execução do ambiente experimental, os cinco classificadores supervisionados — *Random Forest*, SVM, MLP, Árvore de Decisão e KNN — foram submetidos ao conjunto de validação independente, composto por 2.371 instâncias estrategicamente estratificadas.

A análise a seguir expõe inicialmente o comportamento dinâmico dos erros e acertos interclasses por meio das Matrizes de Confusão e de seus respectivos relatórios de sensibilidade por categoria, permitindo um diagnóstico operacional antes da consolidação das métricas globais de desempenho.

4.1 Análise de Desempenho por Classe

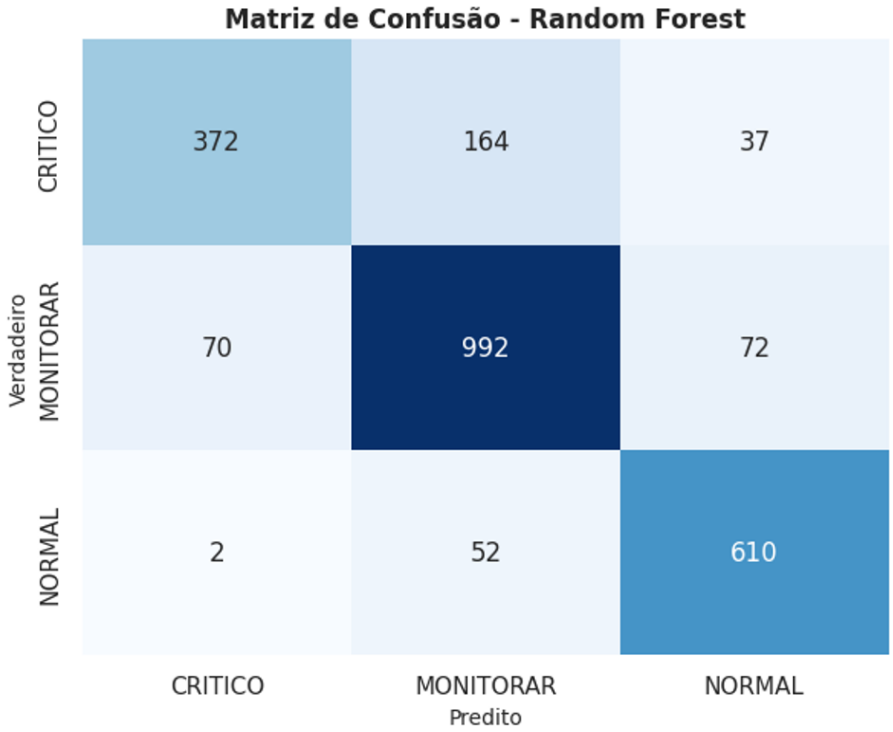
A avaliação isolada por índices globais pode omitir a distribuição real dos erros cometidos pelos algoritmos. Para mitigar esse viés e salvaguardar a confiabilidade exigida pela engenharia de manutenção em ambiente de mineração, aplicou-se o diagnóstico detalhado por classe.

Sob a perspectiva industrial, o foco analítico reside na minimização dos Falsos Negativos associados à classe “CRÍTICO” (avaliada pelo *Recall*), dado que a não identificação de um ativo em iminência de colapso deflagra quebras catastróficas, paradas operacionais imprevistas e custos elevados de manutenção corretiva.

4.1.1 Random Forest

O modelo *Random Forest*, operando com sua configuração ótima de 200 estimadores e crescimento livre das ramificações (max_depth: None), ratificou sua robustez operacional ao registrar altos índices de assertividade nas extremidades do espectro de desgaste. Conforme os dados de suporte da Figura 23, o modelo alcançou sua maior eficiência na classe “NORMAL”, com um *Recall* de 92% e *Precision* de 85%. Para a classe de maior severidade operacional (CRÍTICO), o modelo demonstrou alta precisão 84%, indicando que os alarmes gerados são altamente confiáveis, embora seu *Recall* tenha se fixado em 65%. O erro predominante concentrou-se estritamente na transição para a classe “MONITORAR”, comportamento tolerável e seguro dentro de um plano preditivo.

Figura 23 – Matriz de confusão *Random Forest*.

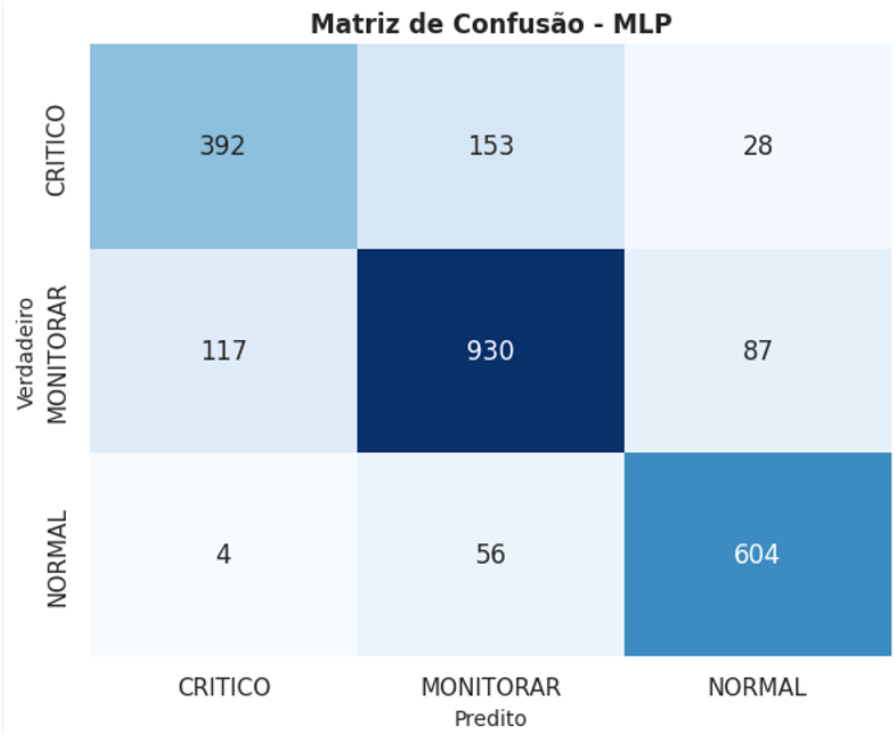


Fonte: Próprio Autor.

4.1.2 Multi-Layer Perceptron (MLP)

A rede neural MLP, cuja topologia ótima selecionada pelo *Grid Search* consistiu em uma única camada horizontal expandida de 128 neurônios ocultos (hidden_layer_sizes: (128)), exibiu um comportamento promissor ao igualar o maior nível de *Recall* na classe "CRÍTICO" (*Recall* de 68%). No entanto, o modelo manifestou uma sutil zona de sombreamento (*overlap*) na transição dessa classe, apresentando uma precisão ligeiramente inferior (76%) em relação aos líderes do experimento. Sob o ponto de vista da engenharia de manutenção, esse erro assume um caráter conservador e aceitável, pois induz a inspeções preventivas por excesso de zelo, distanciando o ativo do risco de quebra inesperada.

Figura 24 – Matriz de confusão MLP.

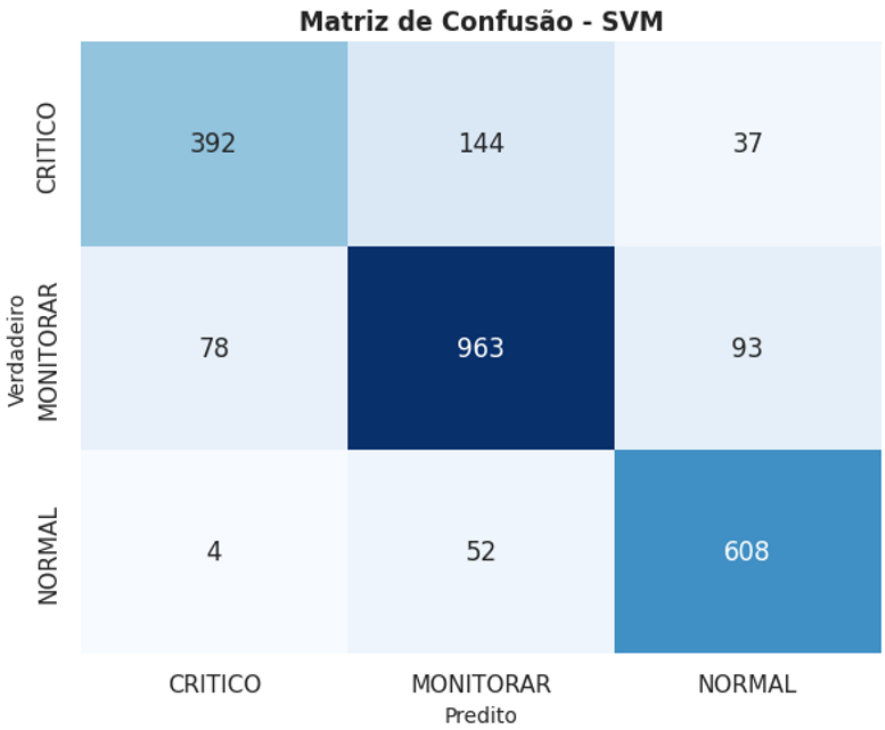


Fonte: Próprio Autor.

4.1.3 Support Vector Machine (SVM)

O classificador SVM, calibrado com o parâmetro de regularização $C = 10$ e função de *kernel* RBF (gamma: 'scale'), apresentou o comportamento mais equilibrado para a detecção de anomalias severas, conforme detalhado na matriz de confusão da Figura 25. Na classe "CRÍTICO", o algoritmo obteve um *Recall* de 68% (o maior índice de captura empírico empatado com o MLP) associado a uma precisão de 83%. Esse resultado demonstra que o mapeamento de fronteiras curvas via RBF foi eficaz para separar o espaço amostral não linear dos metais de desgaste. Além disso, o modelo obteve excelente discriminação na classe "NORMAL" (*Recall* de 92%), consolidando-se como um forte candidato para a operação industrial.

Figura 25 – Matriz de confusão SVM.

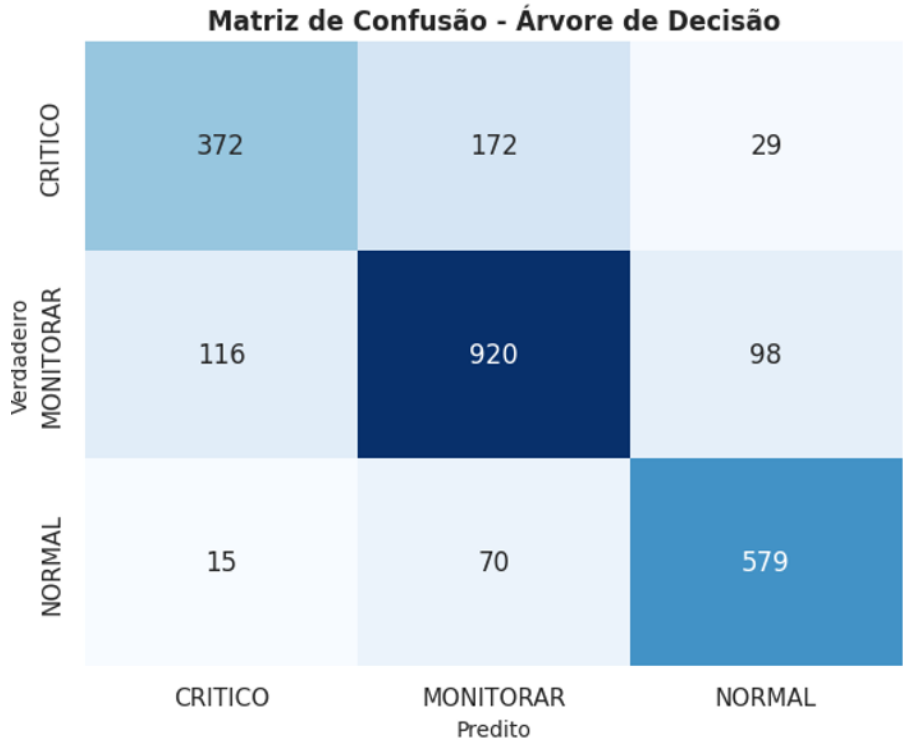


Fonte: Próprio Autor.

4.1.4 Árvore de Decisão

A distribuição de erros na estrutura da Árvore de Decisão isolada — otimizada sob o critério de *Gini* e profundidade máxima limitada em 15 níveis (max_depth: 15) — expõe os riscos práticos de sua aplicação sem uma arquitetura de comitê. O modelo apresentou a menor precisão para a classe “CRÍTICO” (74%) combinada a um *Recall* de 65%, conforme mapeado na Figura 26. Essa vulnerabilidade diagnóstica decorre da rigidez geométrica de suas regras de decisão ortogonais, que tendem a sofrer *overfitting* local diante de variações sutis na concentração de contaminantes do óleo, comprometendo a confiabilidade e elevando o índice de alarmes falsos nas classes intermediárias.

Figura 26 – Matriz de confusão Árvore de Decisão.

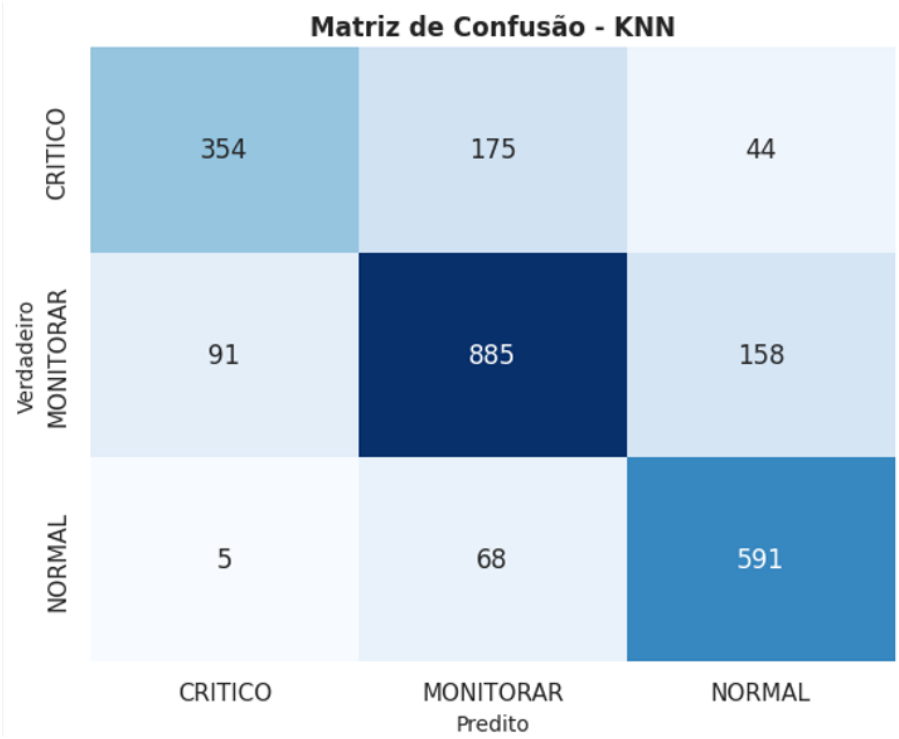


Fonte: Próprio Autor.

4.1.5 K-Nearest Neighbors (KNN)

Consistente com seus baixos índices, o KNN configurado com 5 vizinhos próximos e ponderação ponderada pelo inverso da distância (weights: 'distance') revelou limitações estruturais para segregar as amostras. Conforme ilustrado na Figura 27, o modelo foi severamente penalizado na classe “CRÍTICO”, onde registrou o menor índice de captura do experimento (*Recall* de 62%). Devido ao desbalanceamento original do banco de dados, os hiperplanos de vizinhança foram distorcidos pelo volume massivo da classe majoritária “MONITORAR”, fazendo com que amostras críticas fossem sistematicamente assimiladas por vizinhanças inadequadas, gerando um volume perigoso de classificações equivocadas para a manutenção.

Figura 27 – Matriz de confusão KNN.



Fonte: Próprio Autor.

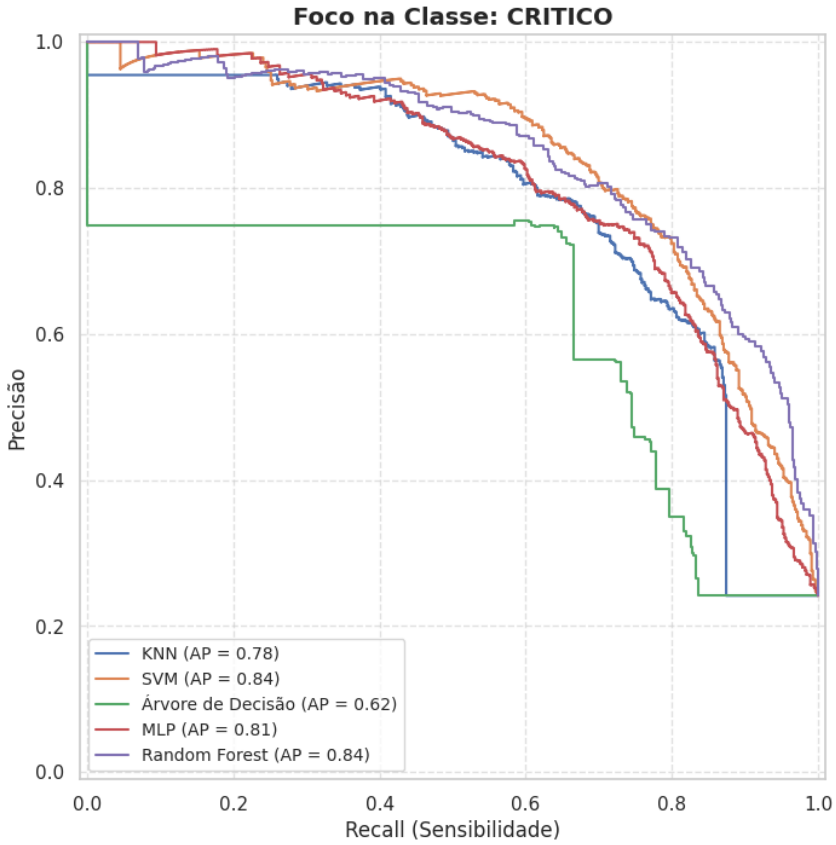
4.2 Avaliação Global e Análise Precisão-Recall

Embora as matrizes de confusão permitam identificar a distribuição dos acertos e erros dos classificadores em um limiar específico de decisão, elas não evidenciam o comportamento do modelo ao longo de diferentes limiares. Dessa forma, com o objetivo de complementar a avaliação experimental e quantificar formalmente os dinâmicas interclasses, os modelos foram submetidos a uma análise conjunta por meio das curvas *Precisão-Recall* (PR) e de um painel consolidado de métricas estatísticas globais.

A utilização das curvas PR mostra-se particularmente adequada em problemas com desbalanceamento entre classes, evidenciando o compromisso entre Precisão e Recall. O rendimento de cada curva é sintetizado pela métrica *Average Precision* (AP), cujos valores mais próximos de 1 indicam maior capacidade discriminativa. Paralelamente, adotou-se o F1-score Macro como principal métrica para a seleção final do modelo, por considerar igualmente as predições em todas as classes, independentemente da distribuição das amostras.

A Figura 28 ilustra as curvas PR para a identificação da classe CRÍTICO (categoria de maior relevância operacional neste estudo), enquanto a Tabela 9 e a Figura 29 sintetizam os resultados das métricas globais no conjunto de validação independente.

Figura 28 – Curvas Precisão-Recall para a classe CRÍTICO.



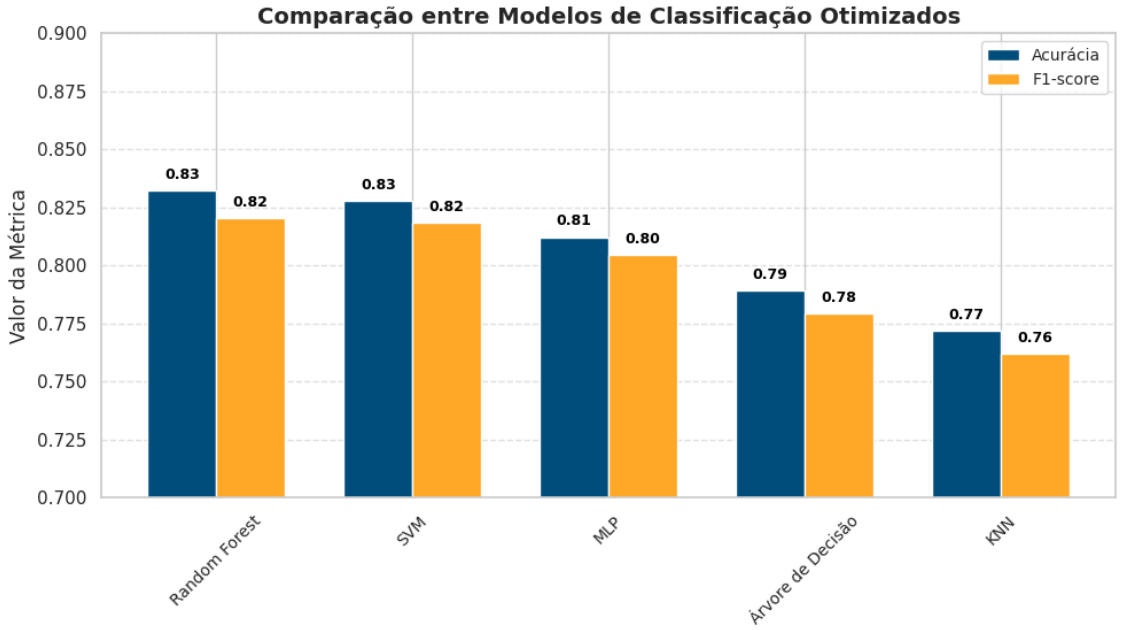
Fonte: Próprio Autor.

Tabela 9 - Comparação de desempenho entre os modelos aplicados.

Modelo	Acurácia Geral (%)	Precisão (%)	Recall (%)	F1-score (%)
Random Forest	83,25	83,58	81,42	82,02
SVM	82,79	82,72	81,63	81,86
MLP	81,23	80,68	80,46	80,45
Árvore de Decisão	78,91	78,38	77,74	77,93
KNN	77,18	77,21	76,27	76,19

Fonte: Próprio Autor.

Figura 29 – Comparação entre os classificadores.



Fonte: Próprio Autor.

Os índices experimentais indicam que o algoritmo *Random Forest* estabeleceu o limite superior de desempenho preditivo global, alcançando um *F1-score Macro* de 82,02% e a maior acurácia global de 83,25%, além de exibir excelente estabilidade ao longo dos limiares de decisão, com AP igual a 0,84 na classe CRÍTICO. Essa superioridade decorre de sua arquitetura baseada em *ensemble*, que reduz a variância do modelo por meio da combinação de múltiplas árvores de decisão construídas sobre subconjuntos aleatórios dos dados (BREIMAN, 2001).

É mandatório destacar que o classificador SVM estabeleceu uma disputa acirrada, fixando seu F1-score Macro em 81,63% (uma diferença infinitesimal de apenas 0,0015 em relação ao líder), além de igualar o *Random Forest* no AP da classe CRÍTICO (AP = 0,84). O SVM superou o *Random Forest* na métrica de Recall Macro (81,63% contra 81,46%), provando que o hiperplano de separação gerado com regularização possui excelente generalização estatística para capturar as instâncias com menor representatividade volumétrica, mitigando os efeitos do desbalanceamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

O modelo MLP consolidou-se como a terceira força do estudo, atingindo um F1-score Macro de 80,45% e AP de 0,81. Embora tenha mantido índices competitivos acima de 80%, a convergência da rede demandou um custo computacional significativamente superior, validando a premissa de que arquiteturas de redes neurais exigem treinamento iterativo rigoroso para ajustar seus pesos sinápticos via retropropagação do erro (RUMELHART; HINTON; WILLIAMS, 1986).

Na base do painel experimental, a Árvore de Decisão isolada (F1-score de 77% e AP de 0,62) e o KNN (F1-score de 76% e AP de 0,78) confirmaram suas limitações teóricas. O KNN apresentou a menor capacidade discriminativa da rodada experimental, evidenciando que a classificação baseada puramente em distâncias espaciais sofre forte degradação quando submetida a espaços vetoriais expandidos pelo *One-Hot Encoding* — um reflexo prático da "maldição da dimensionalidade" que torna o modelo altamente sensível a ruídos operacionais (GÉRON, 2019).

5 CONCLUSÃO

O presente trabalho cumpriu o objetivo de desenvolver e avaliar um sistema automatizado de classificação baseado em algoritmos de aprendizado de máquina para o monitoramento do estado de contaminação e degradação do óleo lubrificante em frotas de equipamentos móveis. A modelagem preditiva, estruturada sob a ótica de proteger os ativos contra quebras catastróficas, provou ser uma alternativa metodológica viável, científica e de alta confiabilidade. Frente ao método convencional, no qual a equipe de laboratório demandava um tempo médio de 1 hora para a classificação manual das amostras nas classes Normal, Monitorar e Crítico, a automatização via *Machine Learning* elimina o gargalo do tempo de análise, mitigando o risco de falhas catastróficas durante o intervalo de espera do diagnóstico.

A abordagem experimental baseada em *Grid Search* e validação cruzada permitiu mapear os limites operacionais de cinco arquiteturas distintas frente a um banco de dados real e desbalanceado composto por 11.853 registros. Os resultados evidenciaram que os modelos baseados em comitês de decisão e margens estatísticas estabeleceram os melhores limiares discriminativos para o problema proposto.

O algoritmo *Random Forest* consolidou-se como a solução ideal em termos globais, alcançando o limite superior de desempenho com uma Acurácia Geral de 83,25% e um *F1-score Macro* de 82,02% e notável estabilidade na curva PR (AP de 0,84 para a classe CRÍTICO). Sua arquitetura de múltiplas árvores independentes demonstrou robustez singular para processar as 19 variáveis heterogêneas do sistema sem incorrer em sobreajuste.

Paralelamente, o modelo SVM destacou-se como um forte concorrente operacional, fixando seu *F1-score Macro* em 81,86% e superando o líder na métrica de *Recall Macro* (81,63% contra 81,42%) e igualando seu poder discriminativo nas falhas mais severas (AP de 0,84). Sob a perspectiva da engenharia de manutenção, a capacidade do SVM em capturar uma fração superior de amostras da classe “CRÍTICO” (sensibilidade de 0,68) mitiga significativamente a ocorrência de Falsos Negativos, blindando a operação contra falhas inesperadas de alto impacto financeiro. Os modelos baseados em Redes Neurais apresentaram desempenho competitivo, porém acompanhados de maior custo computacional, enquanto abordagens isoladas ou puramente espaciais (Árvore de Decisão e KNN) mostraram-se vulneráveis à alta dimensionalidade e ao desbalanceamento dos dados.

Como principais contribuições deste estudo, destaca-se a estruturação de um *pipeline* reprodutível de pré-processamento que converteu dados qualitativos textuais e quantitativos de

metais de desgaste em um espaço vetorial padronizado, permitindo tomadas de decisão rápidas e mitigando a subjetividade humana. O modelo proposto atua diretamente na transição da manutenção corretiva para a preditiva analítica, otimizando os intervalos de troca de lubrificantes e elevando a disponibilidade física dos equipamentos da mina.

Para a continuidade desta pesquisa e o aprimoramento dos modelos desenvolvidos, recomendam-se as seguintes vertentes de trabalhos futuros:

- A aplicação de técnicas de sobreamostragem sintética (como o SMOTE) para balancear artificialmente o conjunto de treinamento, buscando expandir a sensibilidade (*Recall*) sobre a classe "CRÍTICO";

- Investigar a integração das variáveis físico-químicas do óleo lubrificante com dados de telemetria em tempo real dos ativos (como temperatura de operação, pressão do sistema hidráulico, rotação do motor e níveis de vibração), construindo um modelo preditivo multimodal de maior fidelidade;

- A integração do *pipeline* vencedor em um sistema de *dashboard* em tempo real para a equipe de engenharia de confiabilidade da operação de mineração.

6 REFERÊNCIAS

ABNT NBR 16358:2015. Óleos lubrificantes — Determinação de água por crepitação. Rio de Janeiro: ABNT, 2015.

ALPAYDİN, Ethem. *Introduction to machine learning*. 2. ed. Cambridge; London: The MIT Press, 2010. ISBN 978-0-262-01243-0.

ARAÚJO, Gustavo Santos Rodrigues de; OLIVEIRA, Lucas Batista de; FONSECA JÚNIOR, Valter Ângelo da; MACHADO, Jefferson. *Análise dos óleos lubrificantes em equipamentos móveis de grande porte na mineração*. In: SIMPÓSIO DE TCC DAS FACULDADES FINOM E TECSOMA, 1., 2019. Anais [...]. 2019. p. 1154–1167.

ASTM INTERNATIONAL. ASTM D7279-20: Standard test method for kinematic viscosity of transparent and opaque liquids by automated Houillon viscometer. West Conshohocken, PA: ASTM International, 2020. Disponível em: ASTM D7279-20. Acesso em: 05 jun. 2026.

BATISTA, Gustavo Enrique de Almeida Prado Alves. *Pré-processamento de dados em aprendizado de máquina supervisionado*. 2003. 206 f. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003. Disponível em: http://www.teses.usp.br/teses/disponiveis/55/55134/tde-06102003-160219/. Acesso em: 28 set. 2025. DOI: https://doi.org/10.11606/T.55.2003.tde-06102003-160219.

BATISTA, Rodrigo de Abreu. *Classificação automática de códigos NCM utilizando o algoritmo Naïve Bayes*. 2017. Trabalho de Conclusão de Curso (Pós-Graduação em Engenharia de Software, ênfase em Soluções de Governo) – Universidade de Santa Cruz do Sul, Porto Alegre, 2017. Disponível em: https://repositorio.unisc.br/jspui/handle/11624/1994. Acesso em: 8 nov. 2025.

BORDIGNON, R. *Desempenho tribológico de grafeno funcionalizado como aditivo em óleo lubrificante de baixa viscosidade*. 2018. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis, 2018.

BREIMAN, Leo. *Random forests*. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. Disponível em: https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf. Acesso em: 15 jan. 2026.

CAMPOS, Rafael Saraiva. Desmistificando a inteligência artificial: uma breve introdução conceitual ao aprendizado de máquina. *Aoristo – International Journal of Phenomenology, Hermeneutics and Metaphysics*, v. 3, n. 1, p. 106–123, 2020. DOI: https://doi.org/10.48075/aoristo.v3i1.24880. Disponível em: https://e-revista.unioeste.br/index.php/aoristo/article/view/24880. Acesso em: 8 out. 2025.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995. DOI: https://doi.org/10.1007/BF00994018. Disponível em: https://link.springer.com/content/pdf/10.1007/BF00994018.pdf. Acesso em: 8 nov. 2025.

CUNHA, R. C. *Redutor de velocidade através da técnica de partículas de desgaste no óleo lubrificante auxiliada pela análise de vibrações*. 2005. Tese (Doutorado) – Universidade Estadual Paulista, São Paulo, 2005.

DATA SCIENCE TEAM. *Random Forest explained: a visual guide with code examples*. *Towards Data Science*, 2020. Disponível em: https://towardsdatascience.com/random-forest-explained-a-visual-guide-with-code-examples-9f736a6e1b3c/. Acesso em: 2 dez. 2025.

DOSCIATTI, Mariza Miola; FERREIRA, Lohann Paterno Coutinho; PARAISO, Emerson Cabrera. Identificando emoções em textos em português do Brasil usando máquina de vetores de suporte em solução multiclasse. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC), 10., 2013, Fortaleza. *Anais...* Fortaleza: Sociedade Brasileira de Computação, 2013. p. 1–12. Disponível em: https://www.ppgia.pucpr.br/~paraiso/mineracaodeemocoes/recursos/emocoesENIAC2013.pdf. Acesso em: 16 nov. 2025.

DOS REIS, Denis Moreira. *Classificação de fluxos de dados com mudança de conceito e latência de verificação*. 2016. Tese (Doutorado em Ciências de Computação e Matemática

Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2016.

DUDA, Richard O.; HART, Peter E.; STORK, David G. *Pattern classification*. 2. ed. Hoboken, NJ: Wiley, 2001.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. *Inteligência artificial: uma abordagem de aprendizagem de máquina*. Rio de Janeiro: LTC, 2011.

FAÇANHA, Steffany de Oliveira. *Prospecção do uso de machine learning nas corretoras brasileiras*. 2019. Trabalho de Conclusão de Curso (Graduação em Finanças) – Faculdade de Economia, Administração, Atuária e Contabilidade, Universidade Federal do Ceará, Fortaleza, 2019. Disponível em: https://repositorio.ufc.br/bitstream/riufc/60322/1/2019_tcc_sofa%C3%A7anha.pdf. Acesso em: 8 out. 2025.

FERNANDES, F. R. *Ensaios com amostras de óleos lubrificantes, como ferramenta auxiliar no desenvolvimento de novos motores de combustão interna: revisão sistemática da literatura*. 2012. 61 f. Monografia (Especialização) – Curso de Engenharia Automotiva, Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, 2012.

GARZILLO, Monique Joaquim Witt. *Classificação de tumores cerebrais com algoritmos de machine learning*. 2022. Dissertação (Mestrado em Tecnologias de Física Médica) – Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Lisboa, 2022. Disponível em: http://hdl.handle.net/10400.21/15042. Acesso em: 12 out. 2025.

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.

GONÇALVES, André Ricardo. *Máquina de vetores suporte*. São Paulo: autor, 2010. Disponível em: https://andreric.github.io/files/pdfs/svm.pdf. Acesso em: 15 nov. 2025.

GONÇALVES, Aparecido Carlos; SILVA, João Batista Campos. Predictive maintenance of a reducer with contaminated oil under an excentrical load through vibration and oil analysis. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, v. 33, n. 1, p. 1–7, 2011. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/4ee7a1ec-8f73-4a6a-97ed-d22b0ac52ac8/content. Acesso em: 29 maio 2026.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. New York, NY: Springer, 2009.

HAYKIN, Simon. Fundamental issues in cognitive radio. In: HOSSAIN, Ekram; BHARGAVA, Vijay K. (ed.). *Cognitive wireless communication networks*. New York: Springer, 2007. p. 1–43.

HAYKIN, Simon. *Neural networks and learning machines*. 3. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

HYDRO now owns 100% of Mineração Paragominas S.A. MPSA's shares. **EuropaWire**, 16 dez. 2016. Disponível em: https://news.europawire.eu/hydro-now-owns-100-of-mineracao-paragominas-s-a-mpsas-shares-7643213456789/eu-press-release/2016/12/16/16/04/13/49356/. Acesso em: 28 mai. 2026.

ISL – INSTRUMENTS DE SERVICES LABORATOIRE. *Houillon Viscometer Bath VH1/VH2: Installation, Operation and Maintenance Manual*. Houilles, France: ISL, 2013. Disponível em: Manual VH1/VH2. Acesso em: 05 jun. 2026.

IZBICKI, Rafael; SANTOS, Tiago Mendonça dos. *Aprendizado de máquina: uma abordagem estatística*. São Carlos, SP: Rafael Izbicki, 2020. Livro eletrônico (PDF). ISBN 978-65-00-02410-4. Disponível em: https://tiagoms.com/publications/ame/AME.pdf. Acesso em: 8 out. 2025.

JÚNIOR, G. F. P. M.; ALVES, I. H. S. *Análise de viscosidade do óleo SAE 20W50 pelo método de Stokes*. 2018. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural do Semiárido, Mossoró, RN, 2018.

KHAYAL, Osama Mohammed Elmardi Suleiman. Study of hydraulic system failures in gold mining machinery: a case study of local mining operations. Excellence Journal for Engineering Sciences, v. 2, n. 5, p. 60–66, ago. 2025. ISSN 1858-9448.

LAGO, D. F.; GONÇALVES, A. C. *Manutenção preditiva de um redutor usando análise de vibrações e de partículas de desgaste*. In: POSMEC – SIMPÓSIO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA, 16., 2006, Uberlândia. Anais [...]. Uberlândia: Universidade Federal de Uberlândia, 2006.

LORENA, Ana Carolina; CARVALHO, André Carlos P. L. F. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, Porto Alegre, v. 14, n. 2, p. 43–67, dez. 2007. Disponível em: https://seer.ufrgs.br/index.php/rita/article/view/rita_v14_n2_p43-67. Acesso em: 8 out. 2025.

LUDERMIR, Teresa Bernarda. *Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências*. Estudos Avançados, São Paulo, v. 35, n. 101, p. 85–94, 2021. DOI: 10.1590/s0103-4014.2021.35101.007.

MARCASSO, G. *Estudo de caso da falha de um motor diesel 9.0 L John Deere de uma colhedora de cana-de-açúcar John Deere 3522 2L*. 2017. Trabalho de Conclusão de Curso (Graduação) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Guaratinguetá, SP, 2017.

MARSLAND, Stephen. *Machine learning: an algorithmic perspective*. Boca Raton: Chapman & Hall/CRC, 2009. (Chapman & Hall/CRC Machine Learning & Pattern Recognition Series). ISBN 978-1-4200-6718-7.

MARTINS, Dalton Lopes; MONTEREI, Rafaella. *Aprendizagem de máquina e Ciência da Informação: contribuições para uma agenda de pesquisa e ensino na CI brasileira*. Informação & Informação, v. 27, n. 3, p. 751–774, 2022.

MCKINNEY, Wes. *Python for Data Analysis: Data Wrangling with pandas, NumPy, and IPython*. 2. ed. Sebastopol: O'Reilly Media, 2018.

MEDEIROS, Leonardo Santiago de. *Modelo de predição da condição de estabilidade de taludes via algoritmo de árvore de decisão*. 2022. 44 f. Monografia (Graduação em Engenharia de Minas) – Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, MG, 2022. Disponível em: http://www.monografias.ufop.br/handle/35400000/6945. Acesso em: 28 set. 2025.

MONARD, M. C.; BARANAUSKAS, J. A. *Conceitos sobre aprendizado de máquina*. In: REZENDE, S. O. (Org.). Sistemas inteligentes: fundamentos e aplicações. Barueri, SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168.

MOREIRA, H. C. *Um estudo preliminar da relação entre ferrografia e análise de desgaste em sistemas lubrificados de metais com variados níveis de oxidação*. 2022. Trabalho de Conclusão de Curso (Graduação) – Universidade Tecnológica Federal do Paraná, Curitiba, 2022.

NICOLÁS-ALONSO, L. F.; GÓMEZ-GIL, J. Brain-computer interfaces: a review. *Sensors*, v. 12, n. 2, p. 1211–1279, 2012. ISSN 1424-8220. Disponível em: https://www.mdpi.com/1424-8220/12/2/1211. Acesso em: 20 nov. 2025.

OLIVEIRA FILHO, Otacílio Rodrigues de. *Diagnóstico inteligente de faltas em transformadores baseado na análise de gás dissolvido em óleo*. 2024. 105 f. Dissertação (Mestrado em Computação Aplicada) – Núcleo de Desenvolvimento Amazônico em Engenharia, Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pará, Tucuruí, 2024. Disponível em: https://repositorio.ufpa.br/handle/2011/18015. Acesso em: 8 mar. 2025.

ORACLE. *What's the difference between AI, Machine Learning, and Deep Learning?* [S. l.]: Oracle, 11 jul. 2018. Disponível em: https://blogs.oracle.com/bigdata/post/whatx27s-the-difference-between-aimachine-learning-and-deep-learning. Acesso em: 03 set. 2025.

OSMARI, Junior Garlet; SANTOS, Cesar Gabriel dos; HERZOG, Daniela. Análise do desgaste de motores ciclo diesel submetidos a diferentes regimes através da análise do lubrificante. In: SIMPÓSIO GAÚCHO DE ENGENHARIA AEROESPACIAL E MECÂNICA, 1., 2022, Santa Maria, RS. Anais [...]. Santa Maria: Universidade Federal de Santa Maria, 2022.

PAIVA, Victor Hugo Silva de. *Aprendizado de máquina aplicado à elipsometria espectroscópica para o diagnóstico sorológico de malária*. 2025. Dissertação (Mestrado em Física) – Instituto de Ciências Exatas, Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, 2025. Disponível em: https://repositorio.ufmg.br/handle/1843/80831. Acesso em: 2 dez. 2025.

PECCATIELLO, Rafael Bruno. *Detecção de comportamentos maliciosos de usuários internos em redes utilizando análise de fluxos de dados*. 2023. Dissertação (Mestrado Profissional em Computação Aplicada) – Instituto de Ciências Exatas, Departamento de Ciência da Computação, Universidade de Brasília, Brasília, 2023. Disponível em: http://repositorio.unb.br/handle/10482/47823. Acesso em: 8 nov. 2025.

PEDREGOSA, F. et al. *Scikit-learn: machine learning in Python*. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Disponível em: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf. Acesso em: 11 jun. 2025.

QUINLAN, J. R. *Induction of decision trees*. *Machine Learning*, v. 1, n. 1, p. 81–106, 1986. Disponível em: https://link.springer.com/article/10.1007/BF00116251. Acesso em: 4 jan. 2026.

REYNOLDS, Osborne. On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil. *Philosophical Transactions of the Royal Society of London*, Londres, v. 177, p. 157–234, 1886

RODRIGUES, Carolini Nascimento Martins; GONÇALVES, Ariadne Barbosa; SILVA, Gercina Gonçalves da; PISTORI, Hemerson. Evaluation of Machine Learning and Bag of Visual Words Techniques for Pollen Grains Classification. *IEEE Latin America Transactions*, v. 13, n. 10, p. 3498–3504, 2015. Disponível em: https://ieeexplore.ieee.org/document/7379927. Acesso em: 8 nov. 2025.

RODRIGUES, Fabiano; RODRIGUES, Francisco Aparecido; RODRIGUES, Thelma Valéria Rocha. Modelos de machine learning para predição do sucesso de startups. *Revista de Gestão de Projetos*, São Paulo, v. 12, n. 2, p. 28–55, 2021. DOI:

https://doi.org/10.5585/gep.v12i2.18942. Disponível em: https://periodicos.uninove.br/gep/article/view/18942. Acesso em: 11 out. 2026.

RODRIGUES, Pedro Nevton Sangalo; OLIVEIRA NETO, José Leal de; CERQUEIRA, Claudivânia Santos de; JESUS, Flávio dos Santos de; SANTOS, Paulo Vitor Figueiredo. *A importância da análise de óleo*. In: CONGRESSO NACIONAL DE ESTUDANTES DE ENGENHARIA MECÂNICA, 26., 2019, Ilhéus, BA. Anais […]. 2019. Disponível em: https://www.sistema.abcm.org.br/articleFiles/download/23473. Acesso em: 03 fev. 2025.

ROSENBLATT, Frank. The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386–408, 1958.

RUIVO, Eduardo Acrani. *Ciência de dados aplicada para manutenção preditiva*. 2023. Trabalho de Conclusão de Curso (Graduação em Engenharia de Controle e Automação) – Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, 2023.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L. (org.). *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press, 1986. v. 1.

RUSSELL, Stuart J.; NORVIG, Peter. *Inteligência artificial: tradução da segunda edição*. Rio de Janeiro: Elsevier Campus, 2004. 1021 p. ISBN 978-85-352-1177-1.

SAGAR, Vinyas D.; NANJUNDESWARASWAMY, T. S. Artificial intelligence in autonomous vehicles: a literature review. *i-manager's Journal on Future Engineering and Technology*, v. 14, n. 3, p. 56–62, 2019. DOI: https://doi.org/10.26634/jfet.14.3.15149.

SANTOS, Victor Alves dos; FERREIRA, Ronaldo Lourenço. *Análise de óleo lubrificante por espectros de raios X*. 2016. Trabalho de Conclusão de Curso (Graduação em Engenharia Mecânica) – Universidade de Rio Verde, Rio Verde, GO, 2016. Disponível em: https://www.unirv.edu.br/conteudos/fckfiles/files/Victor%20Alves%20dos%20Santos.pdf. Acesso em: 03 fev. 2025.

SATHYANARAYANAN, S.; TANTRI, B. R. Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, v. 27, n. 4S, p. 4023–4031, nov. 2024. DOI: https://doi.org/10.53555/AJBR.v27i4S.4345.

SCIKIT-LEARN. *Site oficial do pacote Scikit-Learn*. Disponível em: http://scikit-learn.org/stable/. Acesso em: 27 jun. 2025.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. *Understanding machine learning: from theory to algorithms*. New York: Cambridge University Press, 2014.

SILVA, Filipe Ramalho da. *Criptografia homomórfica no aprendizado de máquina*. 2024. 14 f. Artigo (Bacharelado em Ciência da Computação) – Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Campina Grande, PB, 2024. Disponível em: https://dspace.sti.ufcg.edu.br/handle/riufcg/38116. Acesso em: 22 set. 2025.

SILVA, Paula Fernanda da. Desgaste e fadiga térmica de ligas “aço matriz + NbC”. 2006. Dissertação (Mestrado em Engenharia Metalúrgica e de Materiais) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2006. Disponível em: https://teses.usp.br/teses/disponiveis/3/3133/tde-13032007-170403/publico/DissertacaoPaulaFernandadaSilva.pdf. Acesso em: 1 jun. 2026.

SOUZA, Eggleston Patricio de Oliveira. *Análise comparativa de modelos de machine learning e convolutional neural networks na detecção de falhas em máquinas rotativas*. 2024. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Pernambuco, Caruaru, 2024. Disponível em: https://repositorio.ufpe.br/handle/123456789/55791. Acesso em: 10 out. 2026.

SPAMER, Fernanda Rosa. *Técnicas preditivas de manutenção de máquinas rotativas*. 2009. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) – Escola Politécnica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 2009. Disponível em: http://hdl.handle.net/11422/7208. Acesso em: 20 mar. 2025.

SPECTRO SCIENTIFIC. *SpectrOil 100 Series – RDE-OES Elemental Analyzer*. Spectro Scientific. Disponível em: https://www.spectrosci.com/Product/SpectrOil-100-Series---RDE-OES-Elemental-Analyzer. Acesso em: 02 jun. 2026.

SPECTRO SCIENTIFIC. *Overview of Rotating Disc Electrode (RDE) Optical Emission Spectroscopy for In-Service Oil Analysis*. Chelmsford, MA: Spectro Scientific, 2014. Disponível em: https://www.spectrosci.com/www.spectrosci.com/-/media/project/ameteksxa/spectroscientific/ametekspectro/spectro-products/documents/spectroil-q100/overview-of-rotating-disc-electrode-rde-optical-emission-spectroscopy-for-inservice-oil-analysis.pdf. Acesso em: 2 jun. 2026.

STACHOWIAK, Gwidon W.; BATCHELOR, Andrew W. *Engineering tribology*. 4. ed. Oxford: Butterworth-Heinemann, 2013.

VANKOV, E. R.; GADHOUMI, K. *Supervised learning*. In: CHIANG, Sharon; RAO, Vikram; VANNUCCI, Marina (eds.). Statistical Methods in Epilepsy. Boca Raton: Chapman & Hall/CRC, 2024. p. 30. ISBN 978-1-00382-931-7.

VILAS BOAS, Vitor Mendes. *AutoBCI: interface cérebro-máquina com configuração hiperparamétrica automatizada*. 2021. 249 f. Dissertação (Mestrado em Computação Aplicada) – Núcleo de Desenvolvimento Amazônico em Engenharia, Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pará, Tucuruí, 2021. Disponível em: https://repositorio.ufpa.br/jspui/handle/2011/13211. Acesso em: 8 mar. 2025.

ZHANG, C. et al. Machine learning-enabled predictive maintenance: framework, challenges, and perspectives. *IEEE Access*, v. 7, p. 177634–177646, 2019. Disponível em: https://ieeexplore.ieee.org/document/8833058. Acesso em: 8 jan. 2026.