

# AUTOMAÇÃO DE DIAGNÓSTICO PARA MANUTENÇÃO PREDITIVA BASEADA EM ANÁLISE DE FLUIDOS DE EQUIPAMENTOS COM *MACHINE LEARNING* \*

Rafael Wilson Soares Ladeira<sup>1</sup>

Ivan Rosa Soares Júnior<sup>2</sup>

Álvaro Conde Lemos Neto<sup>3</sup>

Rafael Padovezi Miranda<sup>4</sup>

Eduardo Henrique Suski Torres<sup>5</sup>

## Resumo

Realizar manutenções preditivas de forma cada vez mais assertiva é um dos benefícios da atual evolução dos sistemas produtivos industriais conhecida como Indústria 4.0. A Sotreq - revendedora Caterpillar - oferece um serviço de análise de fluidos a partir da coleta periódica para máquinas operantes em campo. Com os resultados das medições, especialistas atribuem às amostras um nível de criticidade, justificativas e recomendações. Neste trabalho propõe-se uma solução que aumentou a eficiência desse processo através da utilização de modelos de *Machine Learning* treinados com as análises já realizadas que automatizam a classificação e a geração do texto de diagnóstico para novas amostras. As predições do sistema foram consideradas bem sucedidas, com métricas de sensibilidade e especificidade acima de 88%.

**Palavras-chave:** Aprendizado de máquina; Manutenção baseada em condição; Análise de fluidos de equipamentos.

## DIAGNOSTIC AUTOMATION FOR PREDICTIVE MAINTENANCE BASED ON EQUIPMENT FLUID ANALYSIS USING MACHINE LEARNING

### Abstract

Improving predictive maintenance is a top priority for the 4.0 Industry. Sotreq - a Caterpillar dealer - offers a fluid analysis service that relies on periodic fluid sampling for in-service machinery. After processing the samples, a team of fluid interpreter experts provides a criticality level and a diagnosis text for each sample. This work proposes a automated approach for this process using Machine Learning models. The system provided successful results, with sensitivity and specificity greater than 88%.

**Keywords:** Machine Learning; Condition-based maintenance; Equipment fluid analysis.

<sup>1</sup> Cientista de Dados e Mestrando em Inteligência Computacional, Universidade Federal de Minas Gerais, Belo Horizonte/MG, Brasil.

<sup>2</sup> Graduando em Ciência da Computação, Universidade Federal de Minas Gerais, Estagiário em Engenharia de Dados, Radix Engenharia e Software S.A, Belo Horizonte/MG, Brasil.

<sup>3</sup> Graduando em Engenharia Elétrica, Universidade Federal de Minas Gerais, Estagiário em Engenharia de Dados, Radix Engenharia e Software S.A, Belo Horizonte/MG, Brasil.

<sup>4</sup> Eng. de Controle e Automação, Radix Engenharia e Software S.A, Belo Horizonte/MG, Brasil.

<sup>5</sup> Eng. de Controle e Automação, Radix Engenharia e Software S.A, Belo Horizonte/MG, Brasil

## 1 INTRODUÇÃO

Entre as tecnologias presentes na tendência atual de automação industrial conhecida como Indústria 4.0, estão avanços nas técnicas de coleta, armazenamento e, principalmente, extração automatizada de conhecimento a partir de dados, o que tem possibilitado a criação de soluções inovadoras e altamente customizáveis. No atual contexto de Big Data, indústria e profissionais desejam aumento de eficiência, tomada de decisões baseadas em evidências e a redução de desperdício de materiais e serviço. Com os avanços da Inteligência Artificial, tarefas que antes eram exclusivamente executadas por operadores humanos são agora passíveis de serem endereçadas por computadores. No entanto, o processamento de dados não estruturados permanece ainda como um problema desafiador. Nesse contexto, durante um projeto para o Grupo Sotreq, a Radix desenvolveu uma solução híbrida, chamada de Oil X!pert, combinando diferentes tipos de técnicas de Machine Learning, para endereçar um problema de manutenção preditiva relacionado à análise de fluidos coletados de máquinas industriais.

A análise de óleos lubrificantes e líquidos de arrefecimento é um indicador das condições protetoras do fluido e nível de desgaste de componentes como transmissões, motores diesel e seus componentes e sistemas hidráulicos. Por meio da análise de composição dos fluidos é possível gerar um diagnóstico preditivo em relação à máquina da qual a amostra foi coletada.

A Sotreq - revendedora de equipamentos, serviços e sistemas industriais - oferece um serviço de análise de fluidos para seus clientes a partir de amostras coletadas de diferentes reservatórios de máquinas operantes em campo. São realizados testes para obtenção de variáveis como índices de Ferro, Cromo e Titânio contagem de particulados e análise de contaminantes, análises estas que são utilizadas para o preparo do diagnóstico. A partir dos resultados das medições, um especialista treinado na análise de fluido é responsável por categorizar o nível de criticidade da amostra em três possíveis estados:

1. Normal: não é necessária intervenção na máquina e as análises posteriores podem seguir a programação normal;
2. Monitorar: é necessária a realização de uma nova análise em período especificado;
3. Crítico: é necessária intervenção imediata na máquina.

Além disso, os especialistas redigem para cada amostra um breve texto detalhando e explicando o resultado do diagnóstico, antecipando potenciais falhas e sugerindo ações preventivas.

Algumas tentativas de automatização desse processo através de sistemas especialistas não atingiram bons resultados, dada a complexidade da atividade que envolve a contextualização de múltiplas séries temporais a informações específicas do equipamento em questão.

Observando a grande quantidade de dados armazenada pela Sotreq, foi proposto utilizar algoritmos de Machine Learning para automatização do processo de classificação das amostras coletadas e geração do texto de diagnóstico. A solução final desenvolvida pela Radix utiliza dezenas de modelos supervisionados, em conjunto com um modelo hierárquico não supervisionado para processamento de linguagem natural. Essa diversidade de modelos foi aplicada para automaticamente lidar com as diferentes etapas do processo de geração do diagnóstico. Além disso, são utilizadas algumas técnicas para extração de características com o objetivo de

umentar o poder informativo das amostras coletadas. Ao longo deste artigo, são apresentados a metodologia utilizada na construção da solução, os resultados obtidos e as métricas aplicadas na validação dos modelos.

## 2 MATERIAIS E MÉTODOS

### 2.1 Conjunto de dados

Utilizando um banco de dados histórico, com informações sobre centenas de milhares de análises previamente realizadas no laboratório da Sotreq, foram coletados dados para alimentar os modelos matemáticos de Inteligência Artificial. Além das variáveis de resposta (nível de criticidade da peça e diagnóstico textual), foram recuperadas para cada amostra informações sobre o cliente em questão, o equipamento (fabricante, tipo, modelo e horímetro), o compartimento, o fluido (fabricante, tipo, horímetro, massa, adições, data de amostragem e data de entrega), dentre diversas outras medições, detalhadas a seguir, para cada tipo de fluido.

#### 2.1.1 Amostras de óleo

Foram extraídas 379.767 amostras de óleo previamente interpretadas referentes a um período de aproximadamente 15 meses. Entre as informações disponíveis para cada amostra estão:

- Concentrações de elementos de desgaste (em partes por milhão): Cu, Fe, Cr, Al, Pb, Sn, Si, Na, K, B, Mo, Ni, Ag, Ca, Mg, Zn, P e Ba;
- Indicadores de presença (classificados em negativo, positivo e em excesso) de detritos, água, anticongelantes, combustíveis, fuligem, oxidação, nitração e sulfatação; Viscosidade a 100°C; Número de basicidade total;
- Contagem de partículas por granularidade (4µ, 6µ, 10µ, 14µ, 18µ, 21µ, 38µ, 50µ) e um código de classificação ISO 4406 (referente a contaminação por partículas sólidas).

#### 2.1.2 Amostras de líquido de arrefecimento

Quanto às amostras de líquido de arrefecimento, as variáveis consolidadas são:

- Ponto de congelamento, nível de glicol e pH;
- Aspectos sensoriais: aparência (límpido, turvo, opaco), coloração (ex.: laranja, vermelho, roxo), odor (normal, anormal, combustível, amônia, queimado, éter, solvente, acetona), nível de espuma (normal, moderado, em excesso).

Foram coletadas 103.888 amostras desse tipo, relativas a um período de quatro anos e três meses.

## 2.2 Aprendizado supervisionado

Historicamente, as técnicas de Inteligência Artificial que tem causado maior impacto na indústria se enquadram em um paradigma denominado aprendizado supervisionado. No contexto da solução desenvolvida pela Radix, modelos desse tipo desempenham um papel chave na extração automática de conhecimento a partir dos dados.

O aprendizado supervisionado tem como objetivo encontrar uma função de mapeamento entre variáveis de entrada e uma variável de resposta. Neste trabalho,

todos modelos desenvolvidos são para classificação, situação em que a variável de resposta apresenta distribuição discreta. Um exemplo é o modelo de criticidade, cuja variável de resposta apresenta sempre um dos três valores descritos em 1.

Para resolver esse problema de estimação de funções, um algoritmo de aprendizado é aplicado a um conjunto de dados, utilizado para treinar o modelo [1]. Concretamente, o processo de aprendizado reside na aplicação de um otimizador, que busca pela melhor função que ajuste aos dados fornecidos. Nesse processo, é preciso validar o modelo encontrado segundo sua capacidade de generalização, para dados não conhecidos na etapa de treinamento. Dessa maneira, estabelece-se garantias de que o modelo vai operar com boa performance em ambiente produtivo.

O modelo escolhido para a versão produtiva do sistema utiliza uma variação da técnica *Gradient Boosting*, que constrói de maneira sequencial um *Ensemble* de classificadores/regressores individuais. A ideia consiste em gerar um modelo mais poderoso com redução de *bias* e variância, através da combinação de vários modelos mais simples. Usualmente, *Ensembles* utilizam como unidades fundamentais modelos baseados em árvores.

### 2.2.1 Árvores de Classificação e Regressão

Originalmente propostos por Breiman et al. [2], modelos baseados em árvores são obtidos através de partições recursivas no espaço dos dados, que definem regras simples de predição. Como resultado desse processo, constrói-se um modelo que pode ser representado graficamente através de uma árvore de decisão, como ilustrado na Figura 1.

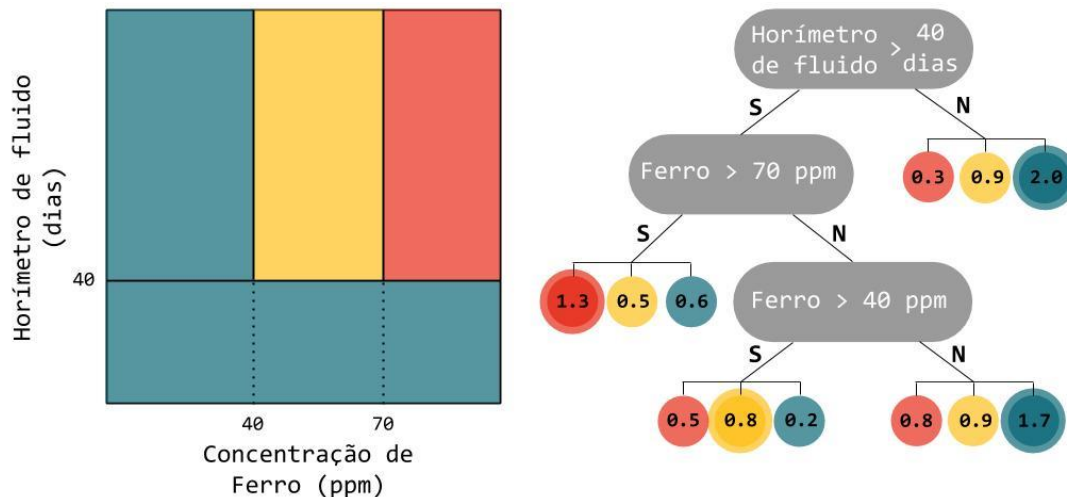


Figura 1. Partição do espaço amostral e sua representação em árvore.

Modelos baseados em árvores têm a desvantagem de apresentar alta variância. Isto é, treinamentos com diferentes subconjuntos de amostras tendem a gerar modelos com estruturas significativamente distintas.

### 2.2.2 Gradient Boosting

O algoritmo *Gradient Boosting* busca um melhor compromisso entre *bias* e variância [3], oferecendo um desempenho mais competitivo que modelos baseados em uma única árvore. Através da construção de um modelo aditivo, combina-se

sequencialmente múltiplos modelos complementares, que de maneira conjunta apresentam maior poder preditivo.

Na solução desenvolvida, foi utilizada uma variação da proposição original do *Gradient Boosting* denominada *XGBoost* [4], que provê bom desempenho para o treinamento com o uso de paralelismo e inclui funcionalidades como tratamento eficiente de dados esparsos e regularizações de diversos tipos. Durante o processo de treinamento de um modelo desse tipo com múltiplas árvores, a seguinte função de custo é minimizada:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(\mathbf{x}_i)) + \Omega(f_t) \quad (1)$$

onde  $l$  é a função que computa o custo (foi utilizada a *cross entropy*) para uma predição,  $(x_i, y_i)$  a  $i$ -ésima amostra do conjunto de treino,  $\hat{y}_i^{t-1}$  a predição da  $i$ -ésima amostra na última árvore ajustado e  $\Omega(f)$  um termo regularizador.

A otimização é feita de forma gulosa, enumerando para cada árvore, em cada nível, todas as partições. A partir da Equação 1, deve ser escolhida a partição que minimiza a expressão:

$$\mathcal{L}_{split} = \frac{1}{2} \left[ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (2)$$

onde  $g_i$  e  $h_i$  são o gradiente e o Hessiano de  $l$  (ambos em relação à  $\hat{y}_i^{(t-1)}$ ), respectivamente,  $\lambda$  um termo regularizador,  $I_L$  e  $I_R$  as partições das amostras em  $I$ , ramificadas para a esquerda e para a direita da partição, respectivamente.

Ainda considerando a Equação 1, a predição ótima para uma partição pode ser obtida por:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (3)$$

### 2.2.3 Modelos supervisionados alternativos para comparação

Para comparação com o modelo principal, são apresentadas métricas relativas a modelos treinados com outras técnicas, a saber: Árvore de classificação tradicional; "Nearest neighbor" (em que a predição é feita procurando a amostra mais semelhante entre os dados de treino); regressão logística multinomial [3]. Tal composição forma um bom *baseline* para comparação com uma técnica não linear, como o Gradient Boosting.

### 2.3 Distância de Levenshtein e Clustering Hierárquico

Devido a existência de diagnósticos padronizados, há frases que aparecem com alta frequência nas análises. Porém, como é possível alterar o texto, podem surgir erros de digitação, que fazem com que variações de poucos caracteres sejam tratadas como frases diferentes, o que dificulta a criação dos modelos de geração de relatórios de diagnóstico. Para mitigar tal problema, foi utilizado o método de Ward [3], agrupando as frases de forma a minimizar a dissimilaridade *intra-cluster* medida com o algoritmo de Levenshtein [5], que pondera o número de inserções, deleções e substituições necessárias para transformar uma cadeia de caracteres em outra.

## 2.4 Protocolo de validação dos resultados

### 2.4.1 Validação cruzada e *k-fold*

A validação cruzada é um método de reamostragem para estimativa do desempenho de generalização de um modelo supervisionado [3]. São geradas  $k$  partições e o algoritmo de aprendizado é aplicado  $k$  vezes, cada uma utilizando uma partição diferente para teste (e as demais para treinamento). Todas as validações realizadas neste trabalho foram feitas com 5 partições. As métricas exibidas são obtidas calculando a média e são disponibilizados também intervalos de confiança.

### 2.4.2 Validação *Out-of-time*

O uso da validação cruzada tradicional para reamostragem ignora potenciais aspectos temporais nos dados. No entanto, estimativas completamente atemporais podem produzir resultados enganosos. Por exemplo, dentro do contexto do sistema desenvolvido, um cenário de mudanças nos diagnósticos ao longo do tempo é possível. Tais mudanças podem estar relacionadas ao surgimento de novos clientes, lançamento de novos equipamentos, ou até mesmo a possíveis correções de diagnósticos não precisos. Dessa maneira, para se ter uma estimativa mais confiável do desempenho produtivo, aplicou-se também a técnica de validação *Out-of-time*, utilizando como conjunto de teste 20% das análises mais recentes disponíveis. Essa configuração simula melhor a potencial degradação, ao longo do tempo, da performance do modelo em ambiente produtivo.

### 2.4.3 Curvas ROC e a métrica AUC

Um gráfico de curva ROC é uma técnica para visualização e seleção de modelos de classificação binária, que exhibe, no eixo Y, a sensibilidade (proporção de casos positivos classificados corretamente) e, no eixo X, o complemento da especificidade (i.e. 1-especificidade, onde a especificidade é proporção de casos negativos classificados corretamente). Um método comum para ter um único valor para comparação é computar a área sob a curva ROC (AUC), que é equivalente a probabilidade do classificador atribuir um valor maior para uma amostra positiva do que para uma negativa, considerando um par aleatório de amostras [6].

$$AUC_{total} = \frac{2}{|C|(|C|-1)} \sum_{\{c_i, c_j \in C\}} AUC(c_i, c_j) \quad (4)$$

Foram utilizadas generalizações das curvas ROC e da AUC para modelos multi-classe considerando as predições para cada par de classes. A AUC é calculada mantendo as propriedades de indiferença ao custo do erro e desbalanceamento de acordo com a Equação 4. Para resumir os resultados na validação cruzada, foi utilizado o algoritmo de média vertical [6].

### 2.4.4 Precisão, revocação e *n-grams*

Para avaliação automática dos textos de diagnósticos produzidos pelos modelos, foram utilizadas as métricas de precisão e revocação com 1, 2 e 3-grams, com a particularidade de cada frase representar um *gram* individual. A Figura 2 mostra esses conjuntos para a referência, o texto predito e o cálculo de precisão e revocação considerando o seguinte exemplo:

**Referência:** Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.

**Predição:** Nível normal de desgaste. Primeira amostra. Óleo em condições de uso. Monitorar nos intervalos regulares.

#### 1-gram

Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.  
Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.  
Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.

#### 2-gram

Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.  
Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.

#### 3-gram

Nível normal de desgaste. Monitorar nos intervalos regulares. Foi trocado o óleo.



Figura 2. Conjunto de 1, 2 e 3-grams (esquerda) e cálculo de precisão e revocação (direita).

## 3 RESULTADOS E DISCUSSÃO

### 3.1 Sistema desenvolvido

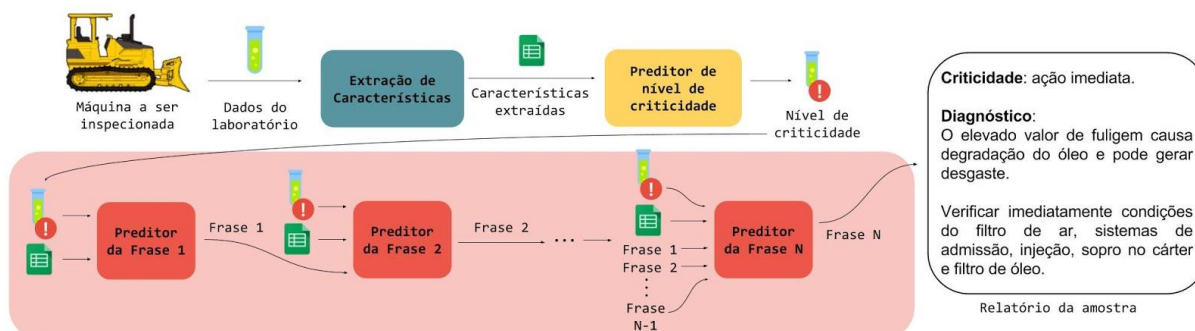
#### 3.1.1 Visão geral

Os dados obtidos das análises laboratoriais são processados para extração de características. Em seguida, são treinados dois modelos de predição de nível de criticidade, um para amostras de óleo e outro para amostras de líquido de arrefecimento. A partir dos resultados de nível de criticidade, modelos de geração de texto são então treinados sequencialmente, para que, ao final do treinamento, relatórios de diagnóstico possam ser gerados.

Os modelos selecionados durante o treinamento são serializados para gerar predições em produção, que podem ser acionadas através de uma interface HTTP disponibilizada pelo sistema.

#### 3.1.2 Extração de características

Durante o pré-processamento dos dados, são feitos tratamentos que visam aumentar o poder preditivo das informações disponíveis. Algumas das etapas implementadas se baseiam em conhecimento específico do domínio, como a criação de escores para “velocidade” de desgaste da peça, segundo vários critérios. Os cálculos realizados utilizam diversos índices de desgaste divididos pela medição do horímetro associado à peça em questão. Além disso, técnicas mais tradicionais também foram aplicadas nessa fase, como *one-hot encoding*, contagem e codificação de valores ausentes, codificação de frequências relativas, bem como quantização e agrupamento de variáveis.



**Figura 3.** Predição em cascata para geração do relatório diagnóstico.

### 3.1.3 Processamento de texto

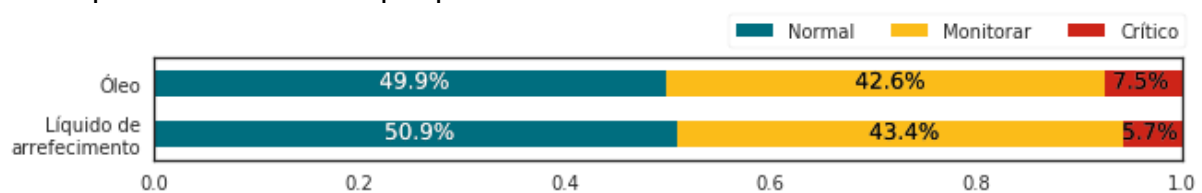
Para lidar com o caráter não estruturado dos dados textuais, há, antes do treinamento dos modelos supervisionados, uma etapa que inclui I) divisão em frases, II) agrupamento: é feito o processo descrito em 2.3 e a frase com maior representatividade *intra-cluster* substitui as demais e III) descarte de frases customizadas com ocorrência muito baixa (evitando inferências com poucas amostras). As frases selecionadas são então usadas para compor a variável de resposta dos modelos supervisionados subsequentes. Dessa maneira, reduz-se o problema para o paradigma de aprendizado de sequências, sendo cada frase uma unidade discreta dentro das sequências de resposta.

Com o objetivo de mapear a nova variável de resposta, são realizados treinamentos encadeados para modelar a probabilidade condicional de ocorrência de uma certa frase dadas as medições resultantes das análises, o nível de criticidade atribuído e as frases nas posições anteriores (Figura 3). Exemplo: se, no conjunto, entre as amostras com o nível de criticidade “monitorar”, o maior número de frases que aparece em um texto é 8, são treinados 8 modelos: o 1º, que prevê a frase inicial a partir das medições, o 2º que prevê a próxima frase com as medições e a 1ª frase e assim sucessivamente. Para permitir o funcionamento com frases de tamanho variável, é criada uma frase fictícia que funciona como um marcador: se ela for predita, é o fim do texto. No sistema, a predição é feita de forma sequencial, com modelos de texto selecionados a partir do nível de criticidade predito.

### 3.2 Análise das variáveis de resposta

Como as predições são realizadas a partir do treinamento de múltiplos modelos de classificação, informações sobre as distribuições marginais das variáveis de resposta são úteis para uma avaliação melhor informada dos resultados.

A Figura 4 mostra que, para predição do nível de criticidade, há um desbalanceamento: amostras rotuladas com o nível “Crítico” são menos frequentes, como esperado. A avaliação dos modelos com curvas ROC e AUC é apropriada nesse contexto, já que consideram a especificidade e permitem visualizar o desempenho dos modelos por par de classes.



**Figura 4.** Frequências dos níveis de criticidade por tipo de amostra.

A Figura 5 e a Figura 6 permitem comparar as distribuições marginais das variáveis de resposta dos classificadores usados para geração dos textos entre as amostras de óleo e as de líquido de arrefecimento. Há, para as primeiras, uma maior diversidade: não só o número de frases distintas é maior, mas as frequências de ocorrência também são mais variadas.



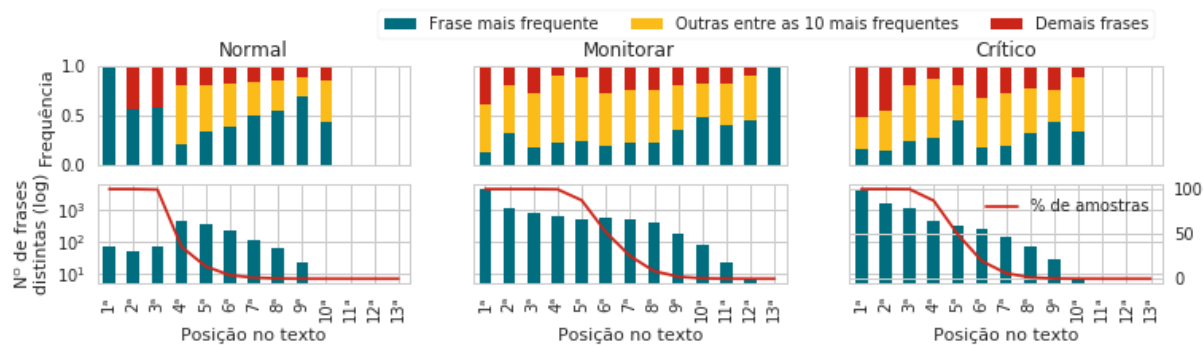


Figura 5. Estatísticas das frases por posição nos relatórios de diagnóstico (Óleo).

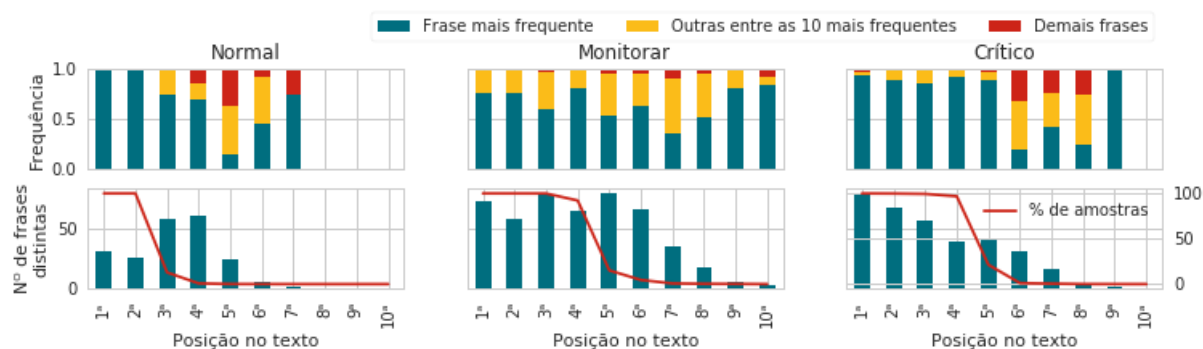


Figura 6. Estatísticas das frases por posição nos relatórios de diagnóstico (Líquido de arrefecimento).

### 3.3 Métricas para predição do nível de criticidade

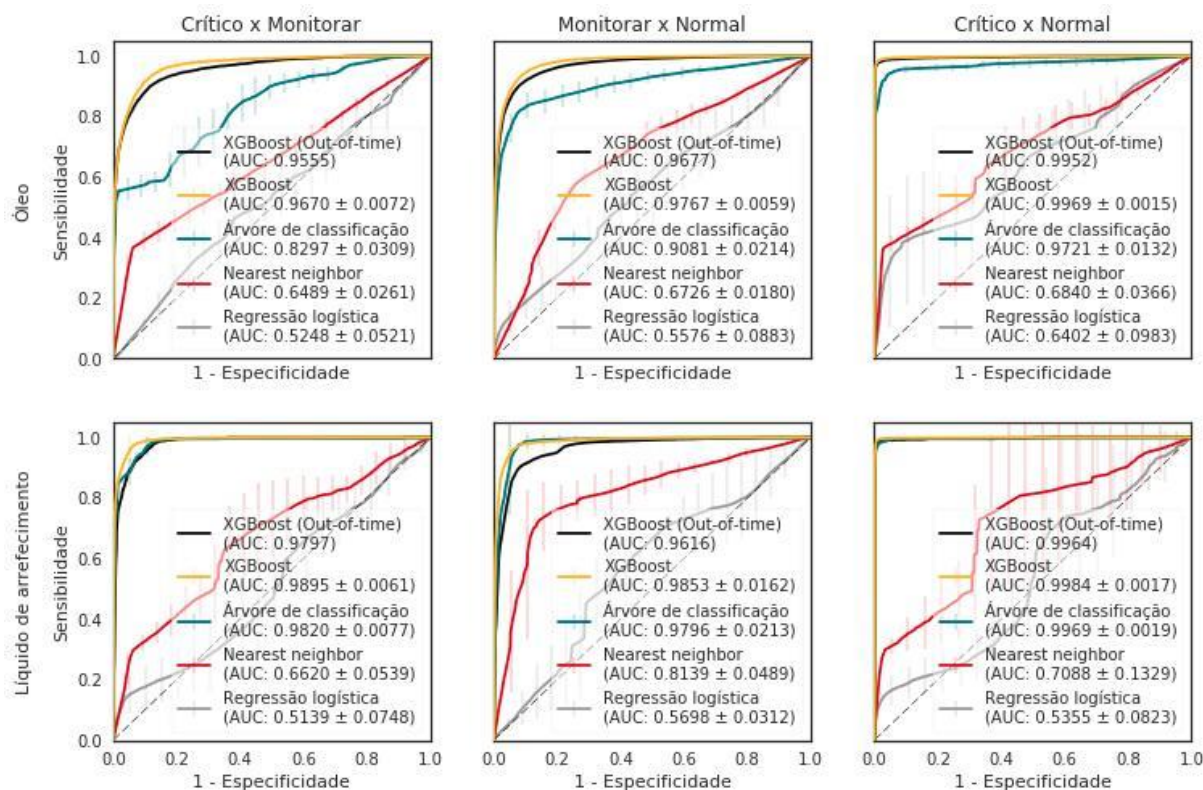


Figura 7. Predição do nível de criticidade. Curvas ROC por par de classes (I.C. 95%, t-Student).

As amostras de líquido de arrefecimento aparecem em menor número no conjunto de dados e sua análise é feita considerando menos características (2.1.2) e adotando critérios mais simples. Como consequência, o desempenho do modelo principal nesse caso é maior do que o para amostras de óleo, como pode ser visto na Figura 7. A menor dificuldade do problema de classificação para essas amostras também pode ser constatada pela comparação dos resultados do modelo com Árvore de classificação: para as de óleo, esse modelo apresenta desempenho claramente menor do que o principal, já para as de líquido de arrefecimento, o desempenho é comparável.

Entre os pares de classe que os modelos precisam distinguir, o caso com o desempenho geral mais baixo é o “Crítico x Monitorar”. As razões para isso incluem além do desbalanceamento (3.2), a necessidade de avaliar um número maior de características para a decisão. Tal fato é visível através da queda de desempenho entre o modelo principal (XGBoost) e o com Árvore de classificação para as amostras de óleo. Ou seja, é esse o caso em que o *boosting* traz o maior ganho.

**Tabela 1.** Predição do nível de criticidade: AUC média por modelo

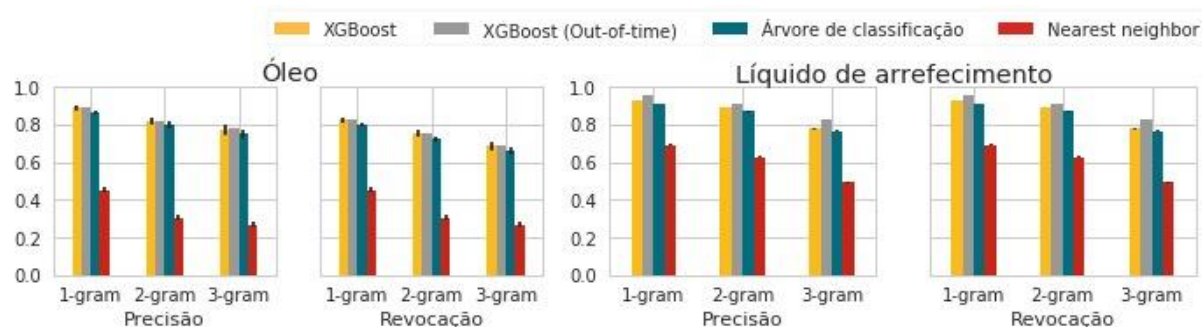
	<b>XGBoost</b>	<b>XGBoost (Out-of-time)</b>	<b>Árvore de classificação</b>	<b>Nearest neighbor</b>	<b>Regressão logística</b>
Óleo	<b>0.980 ± 0.003</b>	0.973	0.903 ± 0.013	0.669 ± 0.016	0.574 ± 0.047
Líquido de arrefecimento	<b>0.991 ± 0.006</b>	0.979	0.986 ± 0.008	0.728 ± 0.051	0.540 ± 0.039

Obs.: I.C. calculado com propagação do erro padrão considerando a variação entre os folds (95%, *t*-Student).

Além disso, a Tabela 1 permite verificar que a degradação dos resultados do modelo principal considerando a validação *Out-of-time* é relativamente pequena para o período analisado e que os modelos alternativos com estratégias de classificação muito simples (“*Nearest neighbor*” e regressão logística multinomial) apresentam desempenho pouco satisfatório com as características utilizadas no sistema.

### 3.4 Métricas para geração de relatórios de diagnóstico

Foram calculadas as métricas de precisão e revocação para os *n*-grams com *n* entre 1 e 3 (inclusive). Valores maiores de *n* não foram considerados, pois o percentual de amostras com mais que 3 frases cai significativamente (Figura 5 e Figura 6). E como os modelos com regressão logística multinomial não apresentaram bons resultados para predição do nível de criticidade, eles não foram considerados nessa etapa.



**Figura 8.** Métricas para os relatórios de diagnóstico (I.C. 95%, *t*-Student).

Como ocorreu na predição do nível de criticidade, as métricas para as amostras de líquido de arrefecimento (Figura 8) são mais altas que as para as amostras de óleo

(3.2). E, apesar do ganho em relação ao modelo com Árvore de classificação ter sido menor nessa etapa, o desempenho do modelo principal é o melhor. Não houve degradação dos resultados considerando a validação *Out-of-time*.

**Tabela 2.** Métricas para os relatórios de diagnóstico

Modelo	1-gram				2-gram				3-gram			
	Óleo		L. de arref.		Óleo		L. de arref.		Óleo		L. de arref.	
	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
<b>XGBoost</b>	<b>0.888</b>	<b>0.824</b>	<b>0.929</b>	<b>0.932</b>	<b>0.82</b>	<b>0.752</b>	<b>0.888</b>	<b>0.896</b>	<b>0.771</b>	<b>0.688</b>	<b>0.776</b>	<b>0.757</b>
<i>Out-of-time</i>	0.889	0.823	0.95	0.969	0.821	0.751	0.909	0.936	0.776	0.689	0.827	0.827
Árvore de classificação	0.866	0.796	0.91	0.915	0.797	0.721	0.868	0.877	0.75	0.657	0.764	0.736
<i>Nearest neighbor</i>	0.451	0.451	0.691	0.691	0.303	<u>0.303</u>	0.628	0.628	0.266	0.266	0.493	0.493

## 4 CONCLUSÃO

As predições do sistema foram consideradas bem sucedidas: na validação *Out-of-time*, usando os limiares ótimos, o modelo principal tem especificidade 88%, a menor sensibilidade é 89.6% (Crítico x Monitorar, para as amostras de óleo) e a maior é 99.3% (Crítico x Normal, para as amostras de líquido de arrefecimento) para a predição do nível de criticidade. Para os relatórios de diagnóstico, considerando 1-gram, a precisão mínima é de 88.8% e a revocação é 82.4%.

Com as predições sendo utilizadas em ambiente de produção na Sotreq, a empresa obteve um aumento representativo na velocidade de realização das análises de óleo. Considerando o processo completo, compreendido pelo recebimento físico da amostra, realização dos ensaios físico-químicos, execução da análise automatizada pelo Oil X!pert e entrega dos resultados, houve uma diminuição do tempo de análise de 72 horas para 48 horas.

Por fim, o sistema permitiu uma redução de 30% na equipe do laboratório, além de possibilitar a padronização dos relatórios e reduzir o prazo de entrega dos resultados das análises para os clientes.

## REFERÊNCIAS

1. DOMINGOS, Pedro. A few useful things to know about machine learning. Communications of the ACM, v. 55, n. 10, p. 78-87, 2012.
2. BREIMAN, Leo et al. Classification and regression trees. CRC press, 1984.
3. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. 11 Edição. Berlin: Springer; 2001.
4. Chen, Tianqi, and Guestrin, Carlos. "Xgboost: A scalable tree boosting system." Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
5. Van der Loo MP. The stringdist package for approximate string matching. The R. 2014 Jun 1.
6. Fawcett, Tom. "An introduction to ROC analysis." Pattern recognition letters 27.8. 2006: 861-874.