

Desenvolvimento de um modelo de aprendizagem de máquina para análise de qualidade de água de caldeira

Anne Karoline Cardoso Macedo¹, Márcio Andrade Rocha²

¹Instituto Federal de Educação, Ciência e Tecnologia da Bahia - Campus Jequié, Brasil. Email: karolcardoso7@gmail.com

²Instituto Federal de Educação, Ciência e Tecnologia da Bahia - Campus Jequié, Brasil. Email: marcioandrade@ifba.edu.br

Resumo

A água é um recurso vital para processos industriais que dependem de caldeiras para produzir vapor, como geração de energia e aquecimento. A qualidade da água utilizada é fundamental para garantir a segurança operacional, bem como para evitar a corrosão e o acúmulo de incrustações nos equipamentos. A proposta deste estudo é avaliar a qualidade da água destinada para alimentação de caldeiras, auxiliando na identificação de problemas de qualidade, como a presença de impurezas ou contaminantes. Desta forma, a análise permitirá que as indústrias tomem medidas para evitar danos e interrupções na operação e possíveis sanções regulatórias, provenientes da não conformidade com as normas. Foi desenvolvido um modelo de aprendizado de máquina para classificar a qualidade da água com base em parâmetros físico-químicos, como ferro, cobre, sílica, alcalinidade, dureza total e condutividade, fornecendo uma categorização sobre a adequação da água, verificando sua conformidade para uma operação eficiente.

Palavras-chave: máquinas térmicas; análise de dados; python; dados sintéticos.

Abstract

Water is a vital resource for industrial processes that rely on boilers to produce steam for purposes such as energy generation and heating. The quality of the water used is crucial for ensuring operational safety and preventing corrosion and scaling in equipment. This study aims to assess the quality of water used in boiler feed, helping to identify quality issues such as impurities or contaminants. The analysis allows industries to take measures to prevent damage and operational interruptions, as well as potential regulatory sanctions due to non-compliance with standards. A machine learning model was developed to classify water quality based on physicochemical parameters, such as iron, copper, silica, alkalinity, total hardness, and conductivity, providing a categorization of water suitability and verifying its compliance for efficient operation.

Keywords: thermal machines; data analysis; python; synthetic data.

1. Introdução

Em 2021, a escassez de chuvas em certas regiões do Brasil levou à diminuição dos níveis dos reservatórios das principais usinas hidrelétricas, resultando em uma queda na produção de hidreletricidade [1]. Para contornar essa redução, aumentou-se a oferta de outras fontes energéticas, como carvão, vapor, gás natural, energia eólica e solar fotovoltaica.

Assim, a participação de renováveis na matriz elétrica brasileira caiu de 83,8% em 2020 para 78,1% em

2021 [2]. Esse cenário impulsionou a produção de termoeletricidade, principalmente a gás natural, que cresceu 24% em comparação ao ano anterior, elevando a participação da geração termelétrica de 26,2% para 31,1% do total da geração de energia elétrica no Brasil [2].

A geração termelétrica é baseada na conversão de energia térmica em energia mecânica, e da conversão desta em energia elétrica. E com o crescimento desse setor no Brasil, faz-se necessário a busca por eficiência operacional na produção de energia, a fim

de ter um maior aproveitamento energético dos recursos.

Um dos tipos de geração de termoenergia mais comuns é o ciclo a vapor, onde o calor da queima de combustível aquece água, transformando-a em vapor de alta pressão e temperatura [3]. A qualidade da água usada nesse processo é crucial, pois suas propriedades de solvente e transporte de partículas podem introduzir impurezas que afetam o rendimento e a eficiência [4]. Impurezas na água das caldeiras podem causar depósitos e incrustações, diminuindo a transferência de calor e aumentando perdas de energia, além de problemas graves como superaquecimento e corrosão [5].

Para prevenir esses problemas, é essencial analisar regularmente a purga da caldeira e a água de alimentação, aplicando medidas corretivas quando necessário. Com a automação industrial, esse monitoramento pode ser aprimorado. Por exemplo, um estudo da Universidade Estadual de Feira de Santana (UEFS) propôs o monitoramento automatizado e remoto da qualidade da água, considerando parâmetros como condutividade, pH, temperatura e oxigênio dissolvido, com dados armazenados em um banco de dados para consulta contínua [6]. Outro estudo mostrou que é possível identificar regimes de ebulição em máquinas térmicas usando técnicas de aprendizado de máquina em imagens de baixa resolução e velocidade, de forma não intrusiva e de baixo custo [7].

Este estudo propõe desenvolver um modelo de aprendizado de máquina para monitorar a qualidade da água de caldeiras, ajudando a reduzir os impactos causados pela má gestão dos parâmetros físico-químicos.

2. Metodologia

Para que um algoritmo classifique a qualidade do fluido de operação de caldeiras, ele deve ser treinado para entender quando essa qualidade é ideal ou não. Este experimento baseia-se na análise dos parâmetros estabelecidos pela The American Society of Mechanical Engineers (ASME), sendo eles: ferro, cobre, condutividade, alcalinidade, dureza total e sílica [8].

O algoritmo classificador deve ser capaz de indicar, em classificação binária, se a qualidade da água está dentro dos padrões estabelecidos. Desta forma, o modelo de aprendizado de máquina poderá ser implementado em sistemas de controle existentes, colaborando com o diagnóstico através dos dados do sistema de água da caldeira.

Na primeira etapa do desenvolvimento, identificaram-se os requisitos do projeto. Para o

ambiente de desenvolvimento, utilizou-se o software Visual Studio Code como editor de código-fonte e Python como linguagem de desenvolvimento.

Para a construção e treinamento do modelo de aprendizado de máquina, é necessário dispor de dados de treino consistentes para que o modelo possa identificar os padrões necessários para classificar novos dados de teste.

Para obter os dados, há algumas alternativas:

- Coletar dados de caldeiras em funcionamento;
- Utilizar dados abertos disponíveis em plataformas de dados abertos ou de organizações governamentais;
- Sintetizar dados pseudo-aleatórios com base em regras que delimitam o cenário e o comportamento dos valores gerados.

Para este estudo, optou-se por sintetizar os dados utilizados no treino do modelo devido à necessidade de uma amostra grande e variada de caldeiras, que cobrisse o máximo de cenários possíveis. A coleta de dados reais demandaria mais tempo e custo para análises químicas das amostras. Além disso, não foram encontrados dados consistentes em plataformas de dados abertos que contivessem os parâmetros necessários para a análise desejada.

Apesar de optar por um modelo classificador, testes serão realizados com modelos de regressão linear para avaliar o comportamento e decidir pelo modelo com melhores previsões. Para aumentar o número de possibilidades ao selecionar o modelo a ser utilizado, a biblioteca Python Lazy Predict será aplicada, permitindo construir testes de variados modelos de aprendizado de máquina com pouco código, gerando um relatório com as métricas de avaliação de cada um.

Logo, para os modelos que serão testados, deve-se considerar que:

- Classificar a qualidade da água da caldeira de forma binária - 0 (qualidade negativa) ou 1 (qualidade positiva);
- Ser avaliado com base nas métricas de F1 Score, ROC AUC e acurácia;
- Modelos com métricas abaixo ou próximas a 50% serão descartados, pois para uma classe binária, 50% de precisão representa uma previsão pseudo-aleatória.

2.1. Geração e Pré-processamento dos Dados

Os dados utilizados serão sintetizados a partir dos parâmetros físico-químicos estabelecidos para faixas de pressão pela ASME (Tabela 1). Para cada intervalo de pressão, serão gerados dados pseudo-aleatórios com base nos critérios definidos.

Para cada parâmetro em um determinado intervalo, será considerado seu valor máximo como o limite suportado para que a qualidade da água esteja dentro do ideal.

Para que amostras de água com baixa qualidade também sejam geradas, calcula-se uma variação entre o valor máximo e mínimo do parâmetro com base no desvio-padrão (Equação 1).

$$tolerância = desvio\ padrão = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (1)$$

O valor mínimo dos parâmetros, para cada faixa de pressão, não são descritos na norma representada na Tabela 1. Para o cálculo da tolerância, será considerado o menor valor para o parâmetro, independente do intervalo de pressão.

Tabela 1. Parâmetros de qualidade da água de caldeira

Pressão (psi)	Ferro (ppm Fe)	Cobre (ppm Cu)	Dureza total (ppm CaCO ₃)	Sílica (ppm SiO ₂)	Alcalinidade total (ppm CaCO ₃)	Condutividade específica (micro-ohms/cm) (não neutralizada)
0 a 300	0,1	0,05	0,3	150	700	7000
301 a 450	0,05	0,025	0,3	90	600	6000
451 a 600	0,03	0,02	0,2	40	500	5000
601 a 750	0,025	0,02	0,2	30	400	4000
751 a 900	0,02	0,015	0,1	20	300	3000
901 a 1000	0,02	0,015	0,05	8	200	2000
1001 a 1500	0,01	0,01	0	2	0	150
1501 a 2000	0,01	0,01	0	1	0	100

Fonte: Guidelines for Water Quality in Modern Industrial Water Tube Boilers for Reliable Continuous Operation [8].

Com a tolerância, calculam-se os novos limites máximo e mínimo para o parâmetro:

$$limite\ máxima = valor\ máximo + tolerância \quad (2)$$

$$limite\ mínimo = valor\ mínimo + tolerância \quad (3)$$

Em casos onde o novo limite inferior seja menor que zero, o valor é corrigido para ser igual a zero. Números pseudo-aleatórios são escolhidos entre esses limites e organizados para formar os novos dados sintéticos. O processo é repetido para todas as faixas de pressão, gerando 50 amostras por intervalo e um total de 400 amostras sintéticas.

2.2. Balanceamento dos Dados

Ao sintetizar os dados, entende-se que as proporções entre amostras com qualidade de água ideal e não ideal ficam desbalanceadas, pois cada amostra é um número qualquer em um intervalo possível dentro ou fora do padrão recomendado.

Modelos de aprendizado de máquina podem ser sensíveis a este desbalanceamento, tendendo a ficarem enviesados com a classe com maior proporção de dados.

Para balancear os dados, utilizou-se o *Synthetic Minority Oversampling Technique* (SMOTE) do pacote *Python imbalanced-learn*. O SMOTE é uma técnica de *oversampling* que gera novos dados sintéticos da classe minoritária, combinando linearmente dados existentes e seus vizinhos mais próximos, com os pesos determinados aleatoriamente [9].

Pode-se observar na Figura 1 que, antes da aplicação do método SMOTE há um claro desequilíbrio entre as classes de qualidade de água conforme (*confirming*) e desconforme (*nonconfirming*). Após a aplicação da técnica, os dados sintéticos gerados ganham uma proporção balanceada.

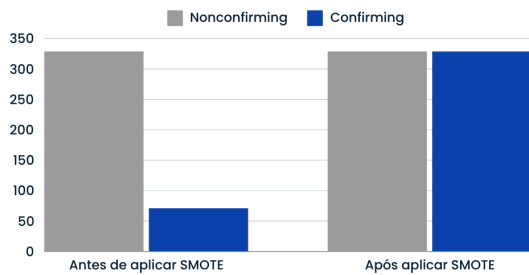


Figura 1. Proporção do volume de dados entre as classes de conformidade. Fonte: elaboração própria.

2.3. Seleção do Modelo de Machine Learning

A biblioteca *Lazy Predict* permitirá a testagem e avaliação de vários modelos ao mesmo tempo, fornecendo o teste de 61 algoritmos, sendo 32 de regressão linear e 29 classificadores. Os seguintes passos serão seguidos para avaliar os algoritmos classificadores:

- Dividir as amostras sintéticas em um conjunto de teste e outro para treino, numa proporção de 30% e 70%, respectivamente.
- Realizar o treinamento supervisionado do algoritmo com os dados de treino.
- Com o algoritmo treinado, aplica-se os dados de testes para predição.
- Comparar a classificação predita com a classificação real, gerando as medidas de desempenho do modelo.

As medidas de desempenho avaliadas serão a acurácia, o ROC AUC e o F1 Score do modelo de aprendizado de máquina e, com base no resultado de cada algoritmo, será escolhido aquele com melhores indicadores.

2.4. Medidas de desempenho

Um classificador binário pode ter quatro resultados de classificação, sendo eles: verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo. Entre elas, as duas primeiras indicam uma classificação correta, enquanto, por sua vez, as duas últimas indicam uma classificação incorreta. Essas classificações incorretas são conhecidas, respectivamente, como erro tipo 1 (Falso positivo) e erro tipo 2 (Falso negativo) [10].

Sabendo a contabilização de cada classificação, pode-se construir uma de matriz de confusão, onde cada linha da matriz representa uma classe real e cada coluna representa uma classe predita. Um classificador perfeito teria valores diferentes de zero

apenas na diagonal principal (do superior esquerdo até o inferior direito) [11].

A acurácia é o percentual de classificações corretas, sendo sua fórmula definida pela Equação 4.

$$acurácia = \frac{tp+tn}{tp+tn+fp+fn} \quad (4)$$

Onde tp é o total de verdadeiros positivos, tn é o total de verdadeiros negativos, o fp é o total de falsos positivos e o fn é o total de falsos negativos. A medida resultante, indica quanto o classificador está acertando a predição de classes [10]. Quanto mais próximo do valor 1, ou 100%, mais acurado é o classificador.

Outro meio de avaliar o desempenho de um modelo classificador é observar a acurácia de predições positivas, ou seja, quantas predições positivas estavam corretas [10]. Esta medida é nomeada de precisão, varia entre os valores de 0 a 1, sendo expressa pela Equação 5.

$$precisão = \frac{tp}{tp+fp} \quad (5)$$

A revocação, ou também chamada de sensibilidade do classificador, mede a taxa de verdadeiros positivos, ou seja, quantas classes positivas foram corretamente detectadas pelo modelo [11]. A revocação é definida pela Equação 6, e também varia entre os valores de 0 a 1.

$$revocação = \frac{tp}{tp+fn} \quad (6)$$

O F1 Score (pontuação F1) é a média harmônica entre a precisão e a revocação, pertinente ao combinar as duas métricas para simplificar a comparação de desempenho entre classificadores. Diferente da média comum, a harmônica dá um peso maior aos valores mais baixos, logo, o modelo só terá uma pontuação F1 alta se a revocação e a precisão forem altas [10]. O F1 Score é definido pela Equação 7.

$$F_1 = \frac{2}{\frac{1}{precisão} + \frac{1}{revocação}} = \frac{tp}{tp + \frac{fn+fp}{2}} \quad (7)$$

A curva de característica de operação (ROC) representa a taxa de verdadeiros positivos (revocação) em relação à taxa de falsos positivos [11]. A taxa de falsos positivos (FPR) é dada pela Equação 8.

$$FPR = 1 - especificidade \quad (8)$$

Sendo a especificidade a taxa de verdadeiros negativos (TNR), dada pela proporção de instâncias negativas corretamente classificadas. Logo, a curva ROC irá representar graficamente a relação da sensibilidade (revocação) e a medida que especificidade invertida (FPR) muda [10].

Uma forma de utilizar a curva ROC para comparação de desempenho é calculando a área sob a curva (AUC). A ROC AUC com valor 1 indica um classificador perfeito, enquanto um ROC AUC de 0,5 indica um comportamento aleatório na predição [11].

2.5. Testes com os dados sintéticos

Para este trabalho, há maior interesse na aplicação de classificadores ao invés de modelos de regressão linear, pois a regressão tem melhor desempenho com dados numéricos em uma mesma escala linear e sem muitos outliers. Como os dados foram gerados de forma pseudo-aleatória, é esperado que haja outliers e cenários incomuns em comparação com a operação real.

Modelos de floresta de árvores aleatórias possuem suporte para dados não numéricos e para relacionamentos não lineares, além de não exigir muitas etapas de preparação e tratamento dos dados [10].

Seu ponto fraco se dá na sensibilidade às variações dos dados de treinamento: caso ocorra uma alteração nos dados, o modelo resultante pode ter um desempenho completamente diferente do anterior à mudança. Isso se dá pelo comportamento estocástico aplicado na maioria dos algoritmos de árvores de decisão [11].

Em outras palavras, estes modelos são mais propensos ao *overfitting* (sobreajuste).

Para comprovar a hipótese, serão testados tanto modelos de regressão linear quanto classificadores, utilizando o Lazy Predict. Dos 61 modelos treinados, serão considerados os melhores desempenhos em teste para cada um dos seguintes cenários:

- dados sem tratamento
- dados com balanceamento
- dados normalizados
- dados normalizados e balanceamento

Sendo estes cenários escolhidos para avaliar a influência dessas etapas de tratamento de dados no desempenho dos modelos.

Com os resultados dos testes, o modelo com melhor ajuste será escolhido para construção do classificador de qualidade de água de caldeira. Será então realizada uma nova etapa de avaliação através de validação cruzada (k-fold), que consiste em repartir os dados de treinamento em um número k de subconjuntos, garantindo que todos sejam utilizados para teste e treino durante a validação.

Após a validação cruzada, o modelo pode passar por ajustes de parâmetros para refino do desempenho e redução de erros.

3. Resultados

O primeiro teste foi realizado com os dados gerados sem nenhum tipo de tratamento ou balanceamento. Sem o balanceamento, espera-se que os modelos tendem a enviesar seus resultados para a classe majoritária, resultando em medidas de desempenho desequilibradas.

Como pode ser visto na Tabela 2, apesar de uma acurácia média de 88% entre os cinco melhores classificadores, a média da ROC AUC foi de apenas 78%, indicando perda de sensibilidade e precisão devido ao desbalanceamento dos dados.

Tabela 2. Medidas de desempenho dos classificadores.

Modelo	Acurácia	ROC AUC	F1 Score
AdaBoostClassifier	88%	83%	89%
LGBMClassifier	91%	80%	91%
LabelPropagation	86%	79%	87%
QuadraticDiscriminantAnalysis	80%	79%	82%
XGBClassifier	88%	78%	89%
Média	88%	79%	89%

Fonte: elaboração própria.

Para os modelos de regressão linear, conforme observado na Tabela 3, os melhores desempenhos ainda mostram predições muito imprecisas. O modelo com a melhor medida de R-Squared conseguiu explicar apenas 53% da variância dos dados.

Dado que o R-Squared é uma medida baseada na correlação dos dados, quanto mais variância for explicada pelo modelo, mais ajustados os dados estarão em relação à função de regressão.

Tabela 3. Medidas de desempenho dos regressores.

Modelo	R-Squared	RMSE
ExtraTreesRegressor	53%	24%
SVR	47%	25%
MLPRegressor	44%	26%
NuSVR	36%	28%
RandomForestRegressor	34%	28%
Média	44%	26%

Fonte: elaboração própria.

No segundo teste, os dados foram balanceados utilizando a técnica de SMOTE, resultando em classes equilibradas. Espera-se que, com esse balanceamento, os modelos façam previsões mais precisas e acuradas, produzindo medidas de desempenho mais equilibradas.

Na Tabela 4, observa-se que os modelos classificadores apresentaram medidas de desempenho mais altas, com uma média de acurácia de 90%. Modelos como *ExtraTreesClassifier* e *RandomForestClassifier* mostraram um bom equilíbrio entre acurácia, sensibilidade e precisão, com ROC AUC de 94% e 92%, respectivamente.

Tabela 4. Medidas de desempenho dos classificadores após balanceamento.

Modelo	Acurácia	ROC AUC	F1 Score
ExtraTreesClassifier	93%	94%	93%
RandomForestClassifier	91%	92%	91%
DecisionTreeClassifier	90%	90%	90%
XGBClassifier	90%	90%	90%
LGBMClassifier	90%	90%	90%
Média	90%	90%	90%

Fonte: elaboração própria.

Na Tabela 5, observa-se que, apesar dos modelos de regressão linear terem alcançado melhores médias após o balanceamento, eles ainda apresentam um comportamento de predição impreciso em comparação aos classificadores.

Tabela 5. Medidas de desempenho dos regressores após balanceamento.

Modelo	R-Squared	RMSE
ExtraTreesRegressor	79%	23%
HistGradientBoostingRegressor	70%	27%
LGBMRegressor	70%	27%
XGBRegressor	70%	27%
RandomForestRegressor	70%	27%
Média	70%	27%

Fonte: elaboração própria.

Nos testes anteriores, a hipótese era de que o balanceamento dos dados influenciaria o desempenho

dos modelos. Contudo, um outro fator que pode impactar as métricas, especialmente nos modelos de regressão linear, é a diferença de escala entre as classes. Parâmetros como condutividade e cobre possuem escalas numéricas desproporcionais, o que pode dar maior peso às classes com valores mais altos no resultado do modelo. Para mitigar esse efeito, aplica-se a técnica de normalização.

No terceiro teste, para manter a validade da hipótese de que os classificadores tendem a ter melhor desempenho com as características dos dados observados, foram avaliados resultados para dados normalizados sem a aplicação da técnica de SMOTE.

Na Tabela 6, observa-se que entre os modelos classificadores, o desempenho geral decaiu em comparação com os dados balanceados, o que era esperado devido ao comportamento enviesado com classes desiguais.

Tabela 6. Medidas de desempenho dos classificadores após normalização.

Modelo	Acurácia	ROC AUC	F1 Score
AdaBoostClassifier	88%	83%	89%
LGBMClassifier	91%	80%	91%
LabelPropagation	86%	79%	87%
QuadraticDiscriminantAnalysis	80%	79%	82%
XGBClassifier	88%	78%	89%
Média	88%	79%	89%

Fonte: elaboração própria.

Para os modelos de regressão (Tabela 7), os resultados foram inesperados, com valores preditos e reais se distanciando conforme indicado pela métrica RMSE. A expectativa era que os regressores tivessem melhores resultados com a normalização, mas a presença de outliers, causada pela randomização na geração dos dados, pode ter impactado negativamente o desempenho.

Tabela 7. Medidas de desempenho dos regressores após normalização.

Modelo	R-Squared	RMSE
ExtraTreesRegressor	53%	24%
SVR	47%	25%
MLPRegressor	44%	26%
NuSVR	36%	28%

RandomForestRegressor	34%	28%
Média	44%	26%

Fonte: elaboração própria.

Dado o resultado do terceiro teste, o quarto teste se propõe a validar se a queda no desempenho ocorreu pelo desbalanceamento dos dados, ou se foi uma influência direta da técnica de normalização aplicada. Para isso, os modelos serão treinados e testados com dados balanceados e, logo após, normalizados.

Tabela 8. Medidas de desempenho dos classificadores após normalização e normalização.

Modelo	Acurácia	ROC AUC	F1 Score
ExtraTreesClassifier	93%	93%	93%
LGBMClassifier	90%	91%	90%
RandomForestClassifier	90%	91%	90%
XGBClassifier	89%	89%	89%
BaggingClassifier	88%	88%	88%
Média	90%	91%	90%

Fonte: elaboração própria.

Tabela 9. Medidas de desempenho dos regressores após normalização e normalização.

Modelo	R-Squared	RMSE
ExtraTreesRegressor	78%	24%
HistGradientBoostingRegressor	71%	27%
LGBMRegressor	70%	27%
BaggingRegressor	69%	28%
RandomForestRegressor	69%	28%
Média	70%	27%

Fonte: elaboração própria.

Como pode ser visto nas Tabelas 8 e 9, tanto os modelos classificadores, quanto os regressores, não obtiveram uma diferença significativa no desempenho em relação ao teste com dados apenas balanceados.

Contudo, diferente do terceiro teste, o RSME nos modelos regressores apresenta valores baixos e mais próximos aos alcançados no teste dois. Ou seja, o tratamento de balanceamento tem influência direta no desempenho dos modelos, diferentemente da normalização, que, para este caso, não teve impacto positivo significativo.

Como observado nos quatro testes realizados, o modelo de melhor desempenho com os dados foi o *ExtraTreesClassifier* e, por isto, este foi escolhido para construção do modelo para categorização da qualidade da água de caldeira.

3.1. Matriz de confusão

Para o modelo escolhido (*ExtraTreesClassifier*) foram construídas duas matrizes: uma com dados desbalanceados e outra com dados balanceados.

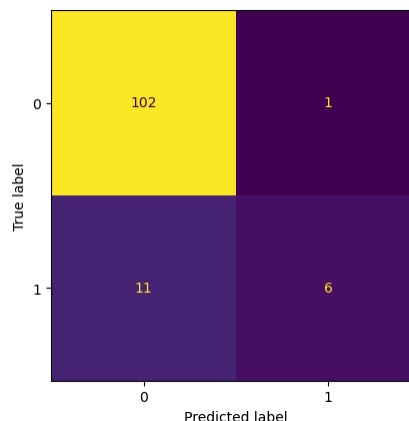


Figura 2. Matriz de confusão do treinamento com dados desbalanceados. Fonte: elaboração própria.

Na Figura 2, com dados desbalanceados, entre 120 amostras de teste, 103 estavam fora do padrão de qualidade e 17 atendiam ao padrão. Observa-se um erro tipo 1 predominante, com baixa sensibilidade do modelo em identificar corretamente amostras que atendem o padrão, resultando em uma sensibilidade de 35% e uma precisão de 86%. Esse desequilíbrio é esperado devido ao volume maior de amostras fora do padrão.

Na Figura 3, com dados balanceados, entre 198 amostras de teste, 101 estavam fora do padrão e 97 atendiam ao padrão. Observa-se um equilíbrio entre as classes e uma significativa redução no erro tipo 1, resultando em uma sensibilidade de 98,9% e uma precisão de 89,7%. Esse cenário é mais favorável em comparação com o anterior, embora uma sensibilidade muito próxima de 100% possa indicar *overfitting*, sugerindo a necessidade de ajustes para melhorar a precisão.

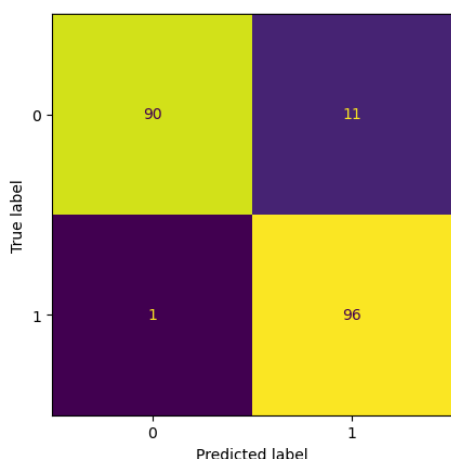


Figura 3. Matriz de confusão do treinamento com dados balanceados. Fonte: elaboração própria.

Para erros do tipo 2, tem-se um baixo volume de erros em relação ao total. Não é trivial controlar estas duas medidas de erro, pois, em geral, elas não são aperfeiçoadas simultaneamente.

Entende-se que, para este estudo, há um maior peso no erro tipo 2, pois em um cenário de manutenção e monitoramento preventivo, o aumento de erros tipo 1 significa mais falsos alertas, porém o tipo 2, significaria que amostras fora do padrão não estariam sendo corretamente detectadas.

3.2. Validação cruzada e *overfitting*

Para validar se há *overfitting* no modelo, foi utilizado o método de *k-fold* em 10 iterações para avaliar o modelo. Vale enfatizar que, cada vez que for executado, a validação cruzada irá retornar métricas diferentes, pois os subconjuntos são repartidos de forma aleatória.

Na Tabela 10, observa-se o resultado obtido em cada iteração. A depender de como os subconjuntos foram separados, vê-se que o comportamento do modelo alcança medidas abaixo de 50% para a revocação e a precisão, enquanto o melhor resultado alcança valores acima de 90%.

Essa discrepância entre os resultados das iterações é um grande indicativo de que o modelo se sobreajustou aos ruídos dos dados e, por isso, não soube lidar com os novos dados de forma satisfatória.

Tabela 10. Resultado da validação cruzada com dados balanceados.

Iteração	Precisão	Sensibilidade/ Revocação	Acurácia
1	41,0%	48,5%	80,0%

2	92,3%	57,1%	85,0%
3	92,3%	57,1%	85,0%
4	41,3%	50,0%	82,5%
5	41,3%	50,0%	82,5%
6	58,6%	54,1%	80,0%
7	41,0%	48,5%	80,0%
8	95,8%	78,6%	92,5%
9	91,3%	91,3%	95,0%
10	91,0%	56,3%	82,5%

Fonte: elaboração própria.

Dado o resultado, foram ajustados os parâmetros do modelo, reduzindo o número de árvores de 100 para 50, aumentando o número mínimo de amostras para dividir um nó interno de 2 para 15, e limitando a profundidade máxima das árvores a 30. Além disso, foi definida a semente como 42 para controlar a aleatoriedade.

Para facilitar a análise de *overfitting* do modelo, foram construídos dois gráficos de curva de aprendizado do modelo, um antes e outro após os ajustes. O gráfico compara a curva de aprendizado do modelo no treinamento e no teste, e, como pode ser visto na Figura 4, para o modelo sem ajustes a curva de treinamento é linear e possui uma pontuação constante no valor 1 (ou 100%).

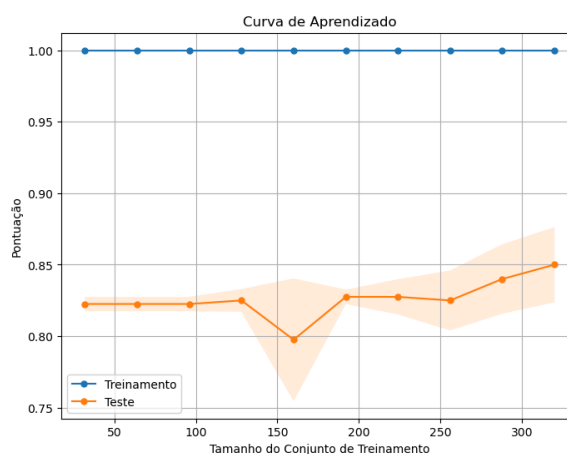


Figura 4. Curvas de aprendizado do modelo treinado com *overfitting*. Fonte: elaboração própria.

Essa aprendizagem constante é um indicativo do comportamento já notado de *overfitting*, dado que o modelo se adaptou completamente aos dados de treinamento. Nesse cenário, o modelo está apenas se especializando nos dados de treinamento, em vez de generalizar a partir deles.

Na Figura 5 já observa-se um comportamento mais generalizado do modelo, as curvas de aprendizado de treinamento e teste variam e tendem a se aproximar à medida que o tamanho do conjunto de treinamento cresce.

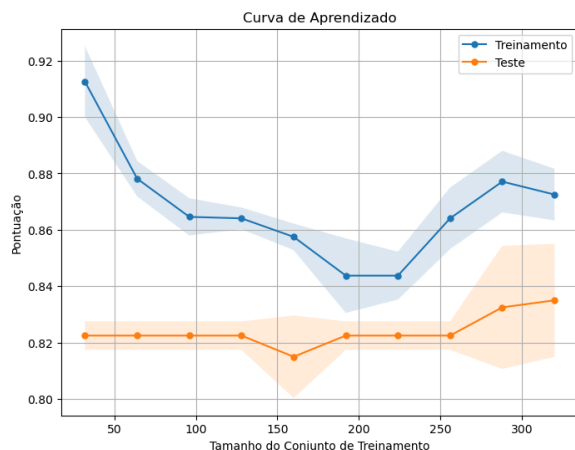


Figura 5. Curvas de aprendizado do modelo treinado com ajustes. Fonte: elaboração própria.

Apesar de obter medidas finais de acurácia e sensibilidade inferiores ao modelo pré-ajustado (Tabela 10), isso se traduz numa melhor generalização do modelo aos dados. Isso pode resultar em um melhor comportamento ao ser implementado em máquinas em operação para testes

Tabela 10. Medidas de desempenho do modelo antes e após ajustes.

Métrica	Antes do ajuste	Após o ajuste
Acurácia	93%	90%
ROC AUC	94%	98%
F1 Score	93%	90%

Fonte: elaboração própria.

Ao repetir a validação cruzada com os ajustes, obtém-se um cenário de valores mais constantes durante as primeiras sete iterações, onde o modelo tem um desempenho baixo (inferior a 50%) em sensibilidade e precisão. Contudo, as iterações oito e nove alcançam os melhores resultados, com métricas, em sua maioria, acima de 80%.

Isso pode indicar que, apesar das iterações iniciais não funcionarem bem, há uma melhoria na resposta do modelo, que levou a iterações significativamente melhores, o que condiz com o comportamento observado na curva de aprendizagem.

Tabela 11. Resultado da validação cruzada após ajustes no modelo

Iteração	Precisão	Sensibilidade/ Revocação	Acurácia
1	41,25%	50,00%	82,50%
2	41,25%	50,00%	82,50%
3	41,25%	50,00%	82,50%
4	41,25%	50,00%	82,50%
5	41,25%	50,00%	82,50%
6	41,25%	50,00%	82,50%
7	41,25%	50,00%	82,50%
8	95,83%	78,57%	92,50%
9	97,14%	85,71%	95,00%
10	91,03%	56,25%	82,50%

Fonte: elaboração própria.

4. Conclusões

O experimento mostrou-se eficiente na automatização do processo de análise de dados e tomada de decisão, para a classificação normativa da qualidade da água, evidenciando o potencial da aplicação de modelagem de aprendizado de máquina no auxílio no controle de qualidade e na regulamentação de processos industriais.

O modelo em sua primeira versão resultou em um cenário de sobreajuste (overfitting), onde a curva de aprendizado estava constante em um cenário perfeito - igual a 100%. Caso este modelo fosse aplicado em cenários reais, sua incapacidade de generalizar os dados resultaria em falsos alarmes do sistema e classificações incorretas, que poderiam prejudicar a operação.

Após os ajustes, o segundo modelo gerado já foi mais capaz de generalizar, e, apesar de ter uma menor acurácia, pelos resultados obtidos, entende-se que isso se torna uma oportunidade de melhoria. Pois o modelo já consegue ser testado em ambientes operacionais e ter um desempenho razoável, podendo gerar insumos para ser aperfeiçoado.

Um cenário ideal seria testar o modelo em diferentes caldeiras em operação, armazenar os dados analisados e os resultados de classificação, para que pudessem ser avaliados e serem base para um novo treinamento do modelo.

Ainda há possibilidade de entender profundamente as análises de validação cruzada, com um número mais adequado de iterações ou aplicando outros tipos de análise de erro, a fim de encontrar as razões que

levaram o modelo ajustado a ter um baixo desempenho em suas primeiras iterações.

Ao identificar esses problemas, podem ser aplicadas penalidades ao modelo, para evitar repetir estes cenários. Além também, de poder apontar para falhas no pré-processamento dos dados.

Um importante ponto é que os dados serem gerados de forma pseudo-aleatório é um fator limitante para o algoritmo, e este pode ser o principal fator para ter ocorrido o overfitting inicial. Logo, obter novos dados de operações industriais é indispensável para a evolução deste estudo.

Uma evolução deste trabalho seria explorar outros algoritmos, com intuito de retornar sugestões de tratativas para melhorar a qualidade da água, para a caldeira operar em um cenário idealmente recomendado. Dessa forma, o responsável técnico pela caldeira teria não apenas a informação de conformidade, como também um auxílio com relação a quais e como os parâmetros deveriam ser regulados.

Vale enfatizar ainda que são necessários testes mais rigorosos, principalmente em cenários reais, para verificar se as hipóteses levantadas são válidas. Para classificação de qualidade de água de caldeira, a automatização pode ser integrada a sistemas de controle, dado que ele exige pouco recurso computacional para ser executado.

Também podem ser pensadas outras formas de disponibilizar o modelo, como em uma Application Programming Interface (API), que facilitaria não só o compartilhamento do recurso entre distintos sistemas, como também facilitaria a captação de dados de diferentes máquinas para evolução do algoritmo.

Outra oportunidade seria utilizar modelos de aprendizado por reforço, onde o algoritmo poderia aprender durante seu funcionamento, a partir da resposta de retorno do usuário ou do sistema sobre suas predições.

5. Referências

- [1] BRASIL. Resolução ANA nº 77, de 1º de junho de 2021. [S.l.: s.n.], 2021. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 25 abr.
- [2] EPE, Empresa de Pesquisa Energética. Balanço Energético Nacional 2022. EPE, Rio de Janeiro, 2022.
- [3] TOLMASQUIM, Mauricio Tiomno. Energia Termelétrica: Gás Natural, Biomassa, Carvão, Nuclear. Rio de Janeiro: Empresa de Pesquisa Energética - EPE, 2016.
- [4] ZARPELON, Willian; AZZOLINI, José Carlos. Caldeiras de alta pressão: caracterização e avaliação da qualidade do tratamento das águas de abastecimento. *Unoesc & Ciência - ACET Joaçaba*, v. 6, n. 2, p. 141–154, 2015.
- [5] IBP, Instituto Brasileiro de Petróleo e Gás. Inspeção de caldeiras. 3. ed. Rio de Janeiro: IBP, 2020.
- [6] LIMA, Ellen Lima de. Módulo de sensores para monitoramento da qualidade da água com transmissão sem fio utilizando plataforma de prototipagem. Dissertação (Mestrado em Computação Aplicada). Universidade Estadual de Feira de Santana, Feira de Santana, 2018.
- [7] HOBOLD, Gustavo M.; DA SILVA, Alexandre K. Machine learning classification of boiling regimes with low speed, direct and indirect visualization. *International Journal of Heat and Mass Transfer*, v. 171, p. 121166, Jan. 2021.
- [8] ASME, American Society of Mechanical Engineers. Guidelines for Water Quality in Modern Industrial Water Tube Boilers for Reliable Continuous Operation. New York: ASME Press, 2011.
- [9] LEMAITRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, v. 18, n. 17, p. 1–5, 2017.
- [10] HARRISON, Matt. Machine learning – guia de referência rápida: trabalhando com dados estruturados em Python. Califórnia, EUA: O'Reilly, 2019. Traduzido por Lúcia A. Kinoshita - São Paulo: Novatec, 2020.
- [11] GÉRON, Aurélien. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow. 2ª ed. Califórnia: O'Reilly, 2019. Traduzido por Cibelle Ravaglia - Rio de Janeiro: Alta Books, 2021.