

UNIVERSIDADE FEDERAL DE ITAJUBÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
PRODUÇÃO

**Predição do teor de óleos e graxas em água  
produzida em uma Plataforma de Petróleo,  
utilizando redes neurais artificiais**

**Deyvid Martins Santos**

Itajubá

Agosto 2024

Deyvid Martins Santos

# **Predição do teor de óleos e graxas em água produzida em uma Plataforma de Petróleo, utilizando redes neurais artificiais**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para a obtenção do título de Mestre em Ciências em Engenharia de Produção.

**Área de concentração:** Engenharia de Produção

**Orientador:** Prof. Pedro Paulo Balestrassi, Dr.

**Coorientador:** Aloisio Euclides Orlando Jr., Dr.

Itajubá

Agosto 2024

## AGRADECIMENTOS

Cresci em um pequeno bairro da cidade de Campos dos Goytacazes, em frente a uma vasta plantação de cana-de-açúcar. Ao fim dessa plantação, alguns quilômetros adentro, havia um córrego, sobre o qual repousava um tronco que servia de ponte, permitindo o acesso ao que se escondia além daquele canavial. Durante minhas caminhadas de infância, jamais ultrapassava aquele ponto; além do tronco, tudo era mistério e desconhecido. Essa foi a primeira fronteira de mundo, para mim.

Os anos passaram, e outras fronteiras se estabeleceram em minha vida. Em muitas delas, fui vitorioso; em outras, nem tanto. Assim como aquele tronco me dava acesso ao desconhecido, outras figuras surgiram em minha trajetória, desempenhando papéis semelhantes. A educação pública e de qualidade, que me proporcionou uma profissão; o concurso público, que me deu um emprego; e meu trabalho, que me permite explorar o mundo e, a cada dia, aprender um pouco mais.

Foi em uma iniciativa no meu trabalho que conheci o Professor Pedro Balestrassi e o Aloísio Euclides. Eles me convidaram a tentar o processo seletivo de mestrado da UNIFEI. Ali, eu me deparei com uma nova fronteira, dessa vez no campo do conhecimento. Formado como Técnico e Bacharel em Química, decidi enfrentar o desafio de ingressar em um programa na área de Engenharia de Produção. Fiz o processo, fui aprovado e ultrapassei mais essa fronteira. Hoje, posso afirmar que ainda há muito a descobrir, mas cada dia menos do que no anterior. Portanto, meu agradecimento inicial vai para eles, que me permitiram estar aqui hoje, escrevendo este trabalho.

Agradeço aos professores da UNIFEI, que tanto me ensinaram ao longo dessa jornada, especialmente àqueles com quem tive a honra de ser aluno. Agradeço imensamente a Simone e ao Fabrício pelo auxílio neste trabalho e por toda dedicação ao projeto que realizamos juntos. Também sou grato à Petrobras, a empresa da qual tenho o orgulho de fazer parte, por ter possibilitado a execução deste trabalho.

Não menos importante, agradeço à minha mãe, Maria Auxiliadora, e à minha esposa, Bruna. Minha mãe, por me proporcionar o meu maior bem: a educação. Minha esposa, pela parceria diária e pelo carinho constante.

Por fim, agradeço a tudo que é etéreo, intangível e inexplicável. Forças estas que habitam uma fronteira muito mais difícil de compreender.

## RESUMO

A utilização de análises físico-químicas em processos industriais é uma realidade dissociável para a indústria contemporânea. Determinações dessa natureza são eficientes e necessárias para verificação da qualidade de produção, na criação de padrões consistentes e na promoção da segurança. Além disso, as empresas comprometidas com a sustentabilidade procuram minimizar ao máximo seu impacto no meio ambiente, no que diz respeito a emissão de agentes poluentes e alterações que possam alterar a dinâmica do local na qual estão inseridas. É responsabilidade das autoridades governamentais regulamentar as ações relacionadas à preservação do meio ambiente, com o intuito de harmonizar práticas e garantir a minimização do impacto ambiental mencionado. Em 8 de agosto de 2007, foi regulamentada a resolução CONAMA n° 393. Que dispõe sobre o descarte contínuo de água de processo ou de produção em plataformas marítimas de petróleo e gás natural. O artigo 5° resolve: O descarte de água produzida deverá obedecer à concentração média aritmética simples mensal de óleos e graxas (TOG) de até 29 mg/L, com valor máximo diário de 42 mg/L. (CONAMA, 2007). Este trabalho teve como objetivo central desenvolver um método preditivo para estimar os valores de TOG, utilizando variáveis de processo de uma planta de tratamento de petróleo e medições online de TOG. O estudo foi dividido em etapas que contemplaram a obtenção e seleção de variáveis, análise e validação da base de dados, definição de limites para classificação de TOG, balanceamento dos dados e modelagem preditiva. Inicialmente, foi realizada a seleção das variáveis mais relevantes com o apoio de especialistas, garantindo que o modelo considerasse parâmetros pertinentes ao processo. A base de dados passou por uma análise exploratória detalhada para tratamento de inconsistências e validação do comportamento das variáveis em relação ao TOG gravimétrico. Na sequência, limites foram definidos para criação de classes binárias (enquadrado/desenquadrado), e técnicas de balanceamento foram aplicadas para assegurar um conjunto de dados consistente. A modelagem preditiva utilizou redes neurais artificiais, resultando em um modelo robusto, com precisão de 97,3%, validado por métricas estatísticas e testes de generalização. Como principais contribuições deste presente trabalho, destaca-se a identificação das variáveis de maior impacto no processo e a proposta de uma estratégia alternativa baseada em ciência de dados para monitoramento do TOG. Além disso, o trabalho reforça a importância da transformação digital na otimização de processos produtivos e na promoção de práticas sustentáveis.

Palavras-chave; Teor de óleos e graxas; redes neurais; variáveis de processo

## ABSTRACT

*The use of physicochemical analyses in industrial processes is an inseparable reality for the contemporary industry. Such determinations are efficient and necessary for verifying production quality, establishing consistent standards, and promoting safety. Furthermore, companies committed to sustainability strive to minimize their environmental impact, particularly concerning the emission of pollutants and changes that could alter the dynamics of the ecosystems where they operate. It is the responsibility of government authorities to regulate actions related to environmental preservation, aiming to harmonize practices and ensure the minimization of the mentioned environmental impacts. On August 8, 2007, CONAMA Resolution No. 393 was enacted, addressing the continuous discharge of process or production water in offshore oil and natural gas platforms. Article 5 stipulates: "The discharge of produced water must comply with a simple monthly arithmetic mean concentration of oils and greases (TOG) of up to 29 mg/L, with a maximum daily value of 42 mg/L." (CONAMA, 2007). The central objective of this study was to develop a predictive method for estimating TOG values, using process variables from an oil treatment plant and online TOG measurements. The study was divided into stages encompassing data acquisition and variable selection, database analysis and validation, definition of TOG classification limits, data balancing, and predictive modeling. Initially, the selection of the most relevant variables was carried out with the support of specialists, ensuring that the model accounted for parameters pertinent to the process. The database underwent a detailed exploratory analysis to address inconsistencies and validate the behavior of the variables concerning gravimetric TOG. Subsequently, limits were defined to create binary classes (compliant/non-compliant), and balancing techniques were applied to ensure a consistent dataset. The predictive modeling employed artificial neural networks, resulting in a robust model with 97.3% accuracy, validated through statistical metrics and generalization tests. The main contributions of this study include identifying the variables with the greatest impact on the process and proposing an alternative strategy based on data science for TOG monitoring. Additionally, the study emphasizes the importance of digital transformation in optimizing production processes and promoting sustainable practices.*

*Keywords: Oil and grease content; neural networks; process variables.*

## Lista de figuras

|   |    |
|---|----|
| Figura 1: Na esquerda exemplo de um separador bifásico e na direita um trifásico (THOMAS, 2004).....  | 16 |
| Figura 2: Esquema simplificado do tratamento primário do petróleo (NEVES, 2008) .....   | 17 |
| Figura 3: Planta de descarte (NEVES, 2008) .....  | 18 |
| Figura 4: Desenho esquemático de um hidrociclone, mostrando o movimento espiralado do líquido em seu interior (adaptado de MURTHY & BHASKAR, 2012)..... | 19 |
| Figura 5: Parte 1 da análise de TOG gravimétrico (Autoria própria).....   | 20 |
| Figura 6: Parte 2 da análise de TOG gravimétrico (Autoria própria).....   | 20 |
| Figura 7: Parte 3 da análise de TOG gravimétrico (Autoria própria).....   | 21 |
| Figura 8: Exemplo de uma ANN com uma camada escondida (STREITENBERGER, 2022)  | 22 |
| Figura 9: Rede Multylayer Perceptron (feedforward) .....  | 23 |
| Figura 10: Processo de treinamento de uma RNA para problemas não lineares usando DOE.   | 24 |
| Figura 11: Representação de uma gota de água de uma emulsão do tipo água/óleo. (BOEIRA; MOYSÉS; AMORIM, 2015).....                                      | 25 |
| Figura 12: Etapas para a elaboração do modelo (Autoria própria).....  | 29 |
| Figura 13: Comportamento histórico da variável vazão da água produzida .....  | 34 |
| Figura 14: Comportamento histórico da variável Vazão de óleo produzido .....  | 35 |
| Figura 15: Comportamento histórico da variável Temperatura do SG.....   | 36 |
| Figura 16: Comportamento histórico da variável vazão do SG.....   | 36 |
| Figura 17: Comportamento histórico da variável Temperatura do TO.....   | 37 |
| Figura 18: Comportamento histórico dos LVs.....   | 38 |
| Figura 19: Comportamento histórico da variável Vazão no header de descarga .....  | 39 |
| Figura 20: Comportamento histórico da variável Pressão do flotor B .....  | 39 |
| Figura 21: Contribuição % dos poços na produção diária.....   | 40 |
| Figura 22: Comportamento histórico da variável Desemulsificante no header.....  | 41 |
| Figura 23: Comportamento histórico da variável Polieletrólito .....   | 42 |
| Figura 24: Variação do TOG online .....   | 42 |
| Figura 25: Exemplo de planilha e dados enviados para elaboração do modelo .....   | 43 |
| Figura 26: Exemplos de pontos inconsistentes na base de origem.....   | 44 |
| Figura 27: Passas para gerar a análise de regressão .....   | 45 |
| Figura 28: Passos para gerar a análise de gráficos fatoriais .....  | 46 |
| Figura 29: Relação de influência das variáveis independente no TOGG .....   | 46 |

|  |    |
|--|----|
| Figura 30: Exemplo de gráfico de efeitos principais das variáveis envolvidas na regressão gerada no Minitab®.....              | 48 |
| Figura 31: Análise de variância .....  | 49 |
| Figura 32: Gráfico de Pareto dos efeitos padronizados.....   | 50 |
| Figura 33: Coletas realizadas x registros da base .....  | 51 |
| Figura 34: Exemplo de extração dos dados de data, hora e minuto no Minitab®.....   | 52 |
| Figura 35: Seleção de registros associados aos horários fiscais no MS Excel®.....  | 52 |
| Figura 36: Exemplo de validação do TOGG com os dados do laboratório realizados no MS Excel®.....                               | 53 |
| Figura 37: Criação da classe de enquadramento (0) ou desenquadramento (1) com base no limiar estabelecido.....                 | 54 |
| Figura 38: Cálculo das quantidades e percentuais das classes na base de dados .....  | 54 |
| Figura 39: Desbalanceamento da base de dados .....   | 55 |
| Figura 40: Exemplo de código para geração do arquivo balanceado através da técnica de oversampling (SMOTE) .....               | 56 |
| Figura 41: Base de dados balanceada.....   | 56 |
| Figura 42: Planilha no formato compatível com o programa Statistica .....  | 57 |
| Figura 43 - Módulo Automated Neural Networks do Statistica.....  | 57 |
| Figura 44 - Seleção das variáveis preditoras definidas junto aos especialistas para previsão do desenquadramento do TOGG ..... | 58 |
| Figura 45: Aba Quick .....   | 58 |
| Figura 46: MLP activation functions.....   | 60 |
| Figura 47: Weight decay.....   | 60 |
| Figura 48: Initialization .....  | 61 |
| Figura 49 - Exemplo de rede gerada no módulo de treinamento do Statistica.....   | 61 |
| Figura 50 - Seleção da opção de métricas geradas pelo Statistica.....  | 63 |
| Figura 51: Resumo da classificação (matriz de confusão) .....  | 64 |
| Figura 52: Tabela comparativa da classificação .....   | 65 |
| Figura 53: Curva ROC.....  | 65 |
| Figura 54: Save networks .....   | 66 |
| Figura 55; Parte do código Python .....  | 66 |
| Figura 56 - Extração dos pesos para transposição para o código em Python.....  | 67 |

## LISTA DE ABREVIATURAS E SIGLAS

|        |  |
|--------|--|
| CONAMA | Conselho Nacional do Meio Ambiente                 |
| mg/L   | Miligramas por litro                               |
| TOG    | Teor de óleos e graxas                             |
| SMOTE  | Synthetic Minority Over-sampling Technique         |
| PI     | Processo Information                               |
| SG     | Separador de Gás                                   |
| TO     | Tratador de óleo                                   |
| TOGG   | Teor de óleos e graxas determinado por gravimetria |
| SM     | Standard Methods                                   |
| UTGN   | Unidade de Tratamento de Gás Natural               |
| LIMS   | Laboratory Information Management System           |
| ANOVA  | Analysis of variance                               |
| MLP    | Multilayer Perceptron                              |
| RBF    | Radial Basis Function                              |
| ROC    | Receiver Operating Characteristic                  |

## Sumário

|  |    |
|--|----|
| 1. Introdução.....   | 10 |
| 1.1. Justificativa e relevância.....   | 10 |
| 1.2. Objetivos.....  | 12 |
| 1.2.1. Objetivo Geral .....  | 12 |
| 1.2.2. Objetivos específicos .....   | 12 |
| 1.3. Método de pesquisa .....  | 13 |
| 1.4. Delimitações do estudo.....   | 13 |
| 2. Fundamentação teórica.....  | 15 |
| 2.1. Processamento primário do Petróleo .....  | 15 |
| 2.2. Análise TOG pelo Standard Methods 5520-B.....                                       | 19 |
| 2.3. Redes neurais artificiais.....  | 21 |
| 2.4. Parâmetros de processo relevantes para o modelo .....                               | 24 |
| 2.4.1. Vazão de água produzida .....   | 24 |
| 2.4.2. Polieletrólitos e desemulsificantes .....   | 24 |
| 2.4.3. TOG online .....  | 26 |
| 3. Metodologia .....   | 27 |
| 3.1. Metodologia.....  | 27 |
| 3.2. Etapas de execução .....  | 28 |
| 3.2.1. Etapa 1: Obtenção dos dados e seleção das variáveis .....                         | 29 |
| 3.2.2. Etapa 2: Análise da base de dados e validação do sentido dos efeitos .....        | 29 |
| 3.2.3. Etapa 3: Definição dos limites e registros fiscais .....                          | 30 |
| 3.2.4. Etapa 4: Criação das classes e balanceamento da base de dados .....               | 30 |
| 3.2.5. Etapa 4: Modelagem.....   | 31 |
| 3.2.6. Etapa 5: Criação de um modelo preditivo através de Redes Neurais Artificiais..... | 31 |
| 3.3. Complexidades do trabalho .....   | 32 |

|          |  |    |
|----------|--|----|
| 4.       | Resultados .....   | 33 |
| 4.1.     | Etapa 1: Obtenção dos dados e seleção das variáveis .....                          | 33 |
| 4.2.     | Etapa 2: Análise da base de dados e validação do sentido dos efeitos .....         | 43 |
| 4.2.1.   | Tratamento da base de dados .....  | 43 |
| 4.2.2.   | Validação do sentido dos efeitos principais .....                                  | 44 |
| 4.2.3.   | Etapa 3: Definição dos limites e registros fiscais .....                           | 51 |
| 4.2.4.   | Etapa 4: Criação das classes e balanceamento da base de dados .....                | 53 |
| 4.2.5.   | Etapa 5: Criação de um modelo preditivo através de Redes Neurais Artificiais ..... | 56 |
| 4.2.5.1. | Parametrização das Redes Neurais Artificiais .....                                 | 56 |
| 4.2.5.2. | Treinamento e execução das Redes Neurais Artificiais .....                         | 61 |
| 4.2.5.3. | Avaliar o desempenho das Redes Neurais Artificiais .....                           | 62 |
| 4.2.5.4. | Selecionar a Rede Neural com melhor desempenho. ....                               | 63 |
| 4.2.5.5. | Gerar modelo robusto de previsão por Rede Neural. ....                             | 63 |
| 4.2.5.6. | Converter previsor C/C++ para Python .....   | 66 |
| 5.       | Considerações.....   | 68 |
| 5.1.1.   | Conclusões gerais .....  | 68 |
| 5.1.2.   | Contribuições do trabalho.....   | 69 |
| 5.1.3.   | Sugestões para estudos futuros .....   | 69 |

# 1. Introdução

Para as indústrias modernas, manter processos limpos e alinhados com normas ambientais não é apenas uma questão de conformidade, mas uma estratégia crucial para garantir rentabilidade e reputação. A implementação de práticas sustentáveis e a redução da poluição são essenciais para otimizar operações, reduzir custos e evitar multas. Além disso, adotar medidas responsáveis para minimizar o impacto ambiental fortalece a imagem da empresa no mercado, promovendo uma percepção positiva entre consumidores e investidores e traduzindo-se em vantagens financeiras e competitivas significativas.

Além disso, essa necessidade vai além de uma escolha estratégica e se torna uma exigência legal para a continuidade das operações. A Resolução CONAMA 393, por exemplo, estabelece diretrizes rigorosas para o gerenciamento, tratamento e descarte de resíduos, obrigando as empresas a seguir práticas responsáveis. O não cumprimento dessas normas pode resultar em sanções legais e multas, além de ameaçar sua licença de operação. Portanto, cumprir essa regulamentação não é apenas uma questão de boa prática, mas uma condição indispensável para a viabilidade dos negócios.

Este trabalho visa explorar como as Redes Neurais Artificiais, especialmente redes neurais artificiais, podem ser utilizadas para auxiliar as indústrias na adaptação a essas exigências legais e ambientais. A análise e a aplicação dessas tecnologias podem proporcionar soluções eficientes para o gerenciamento de resíduos e otimização de processos, ajudando as empresas a se alinharem com as regulamentações e a alcançarem uma operação mais sustentável, mais conforme, e até mesmo que antevêm possíveis necessidades.

## 1.1. Justificativa e relevância

Em média para cada barril/dia de petróleo produzido são gerados três a quatro barris/dia de água. Há campos que essa relação se eleva sete vezes ou mais (THOMAS, 2004). Segundo o boletim da produção de petróleo e Gás natural de 2022, da ANP, a produção de petróleo no Brasil foi de 3,021 milhões de barris/dia (ANP, Boletim Mensal da Produção de Petróleo e Gás). Relacionando-se essas duas informações pode-se dimensionar o quão grande é a produção de água gerada pela cadeia produtiva da indústria de petróleo, aqui denominada água produzida. Ela desempenha um papel significativo no agravamento da poluição por petróleo na indústria de petróleo e gás (indústria *offshore*). A ameaça ao ecossistema oceânico surge em decorrência da complexa toxicidade bioquímica do petróleo bruto, seja em concentrações elevadas ou

reduzidas ( PITRE, 1984 ; NEFF ET AL., 2011 ).

A água produzida apresenta teor de salinidade pelo menos quatro vezes superior ao da água do mar e, comumente, contém uma variedade de toxinas, metais pesados e elementos radioativos, cuja presença depende das características geológicas da formação de onde foi extraída (FARAG; HARPER, 2014).

A exposição à água produzida pode acarretar riscos à saúde humana, abrangendo desde problemas dermatológicos até riscos graves de intoxicação em caso de ingestão acidental. Além disso, essa água pode ter impactos negativos sobre a fauna, chegando até mesmo a causar a morte de animais (VIEIRA, 2011).

Para além dos efeitos ambientais imediatos, as considerações econômicas também desempenham um papel crucial neste debate. Sanções e penalidades são impostas quando os descartes não estão em conformidade com os limites e condições estabelecidos por regulamentações. Um caso exemplar ocorreu em 2017, quando o IBAMA aplicou uma multa à Petrobras, no valor de R\$ 14,28 milhões devido a divergências na metodologia de determinação do teor de óleo e graxa na água produzida. (EXAME, 2018)

Esse episódio culminou na assinatura do chamado termo de compromisso, onde entre às questões estabelecidas encontra-se a determinação da metodologia a ser empregada nas análises do Teor de Óleos e Graxas, TOG. Estabelece-se a Standard Methods (SM) 5520-B.

É essencial ressaltar que as análises físico-químicas para determinar o valor de TOG, utilizando o método Standard Methods 5520-B, são conduzidas em terra, não nas Unidades Marítimas. Isso se deve ao fato de que este é um método gravimétrico que requer o uso de balanças de alta precisão. Para garantir medidas confiáveis, é fundamental proporcionar um ambiente de estabilidade, o qual não é naturalmente alcançado em plataformas marítimas. Essa restrição desencadeia um atraso de 20 dias entre a coleta da amostra e disponibilização dos resultados.

Estes pontos evidenciam a relevância de monitorar cuidadosamente o descarte da água produzida e reduzir a concentração de óleos e graxas em sua composição. Além disso, podemos avançar, destacando a importância de antecipar os níveis de TOG na água produzida e agir prontamente ajustando parâmetros de processo e até mesmo suspendendo o descarte.

Esse conjunto de necessidades abre portas para o desenvolvimento de ferramentas avançadas, que, combinadas com a inteligência de dados, podem auxiliar a indústria petrolífera

na predição dos valores de TOG, com precisão e tempo hábil suficiente para tomada de decisões no que diz respeito ao processo.

## 1.2. Objetivos

A partir da elaboração anterior do problema de pesquisa, torna-se factível delinear os objetivos, tanto o objetivo geral quanto os objetivos específicos, que norteiam este trabalho.

### 1.2.1. Objetivo Geral

O objetivo central deste estudo é criar um método de previsão destinado a estimar os valores do TOG, a luz do Standard Methods 5520-B, com base nas variáveis do processo da planta de tratamento de petróleo e nas medições online do TOG. Pretende-se identificar os parâmetros mais pertinentes para esse propósito dentre os disponíveis e, a partir disso, elaborar um modelo de previsão confiável capaz de acertar de forma significativa os valores do TOG. Essa precisão é fundamental para que o modelo seja relevante na orientação das decisões operacionais.

### 1.2.2. Objetivos específicos

- Realizar a mineração de dados, coletando e compilando dados históricos de medições de TOG, bem como as variáveis de processo relevantes.
- Aplicar técnicas de análise multivariada de dados para identificar as variáveis de processo mais fortemente associadas aos valores do TOG, a fim de selecionar as mais relevantes para o modelo de previsão.
- Desenvolver um modelo preditivo utilizando técnicas estatísticas avançadas, como regressão multivariada ou aprendizado de máquina, incorporando as variáveis de processo identificadas como relevantes.
- Avaliar a eficácia do modelo de previsão por meio de testes de validação cruzada e análises de erro, garantindo que ele seja capaz de prever os resultados de TOG com precisão significativa.
- Disponibilizar o modelo em código que permita a implementação em ferramenta interna da empresa proprietária dos dados.

### 1.3.Método de pesquisa

Para alcançar os objetivos estabelecidos neste estudo, propõe-se a aplicação de diversas abordagens, incluindo técnicas de estatística multivariada, bem como utilização de técnicas sofisticadas, como as redes neurais artificiais.

Esta pesquisa caracteriza-se pela sua natureza aplicada devido ao seu interesse prático, isto é, os resultados serão aplicados ou utilizados imediatamente na solução de problemas que ocorrem na realidade. Quanto aos seus objetivos, a pesquisa pode ser classificada como normativa. Neste caso, primariamente, tem-se interesse no desenvolvimento de políticas, estratégias e ações para aperfeiçoar os resultados disponíveis na literatura existente, para encontrar uma solução ótima para novas definições de problemas ou para comparar várias estratégias relativas a um problema específico (BERTRAND e FRANSSO, 2002). Esta pesquisa possui uma Abordagem Quantitativa, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las, através de técnicas estatísticas (percentagem, média, moda, mediana, desvio padrão, coeficiente de correlação, análise de regressão etc.).

A técnica para a coleta de dados deve ser principalmente a observação estruturada ou sistemática. Esta técnica realiza-se em condições controladas, para responder a propósitos preestabelecidos. Para se gerar uma informação sobre a correlação entre os valores de TOGs *offshore* e *onshore*, por exemplo, uma organização das informações para que o estudo seja conduzido corretamente é requerida. Alguns questionamentos típicos são:

- Qual a técnica de coleta de dados, considerando variáveis controláveis e fatores aleatórios?
- Quantas medidas são necessárias no estudo?
- Como fazer um balanço entre replicações e repetições?
- Como estabelecer a confiabilidade da correlação?
- Como fazer a mudança da correlação em função de parâmetros dinâmicos?

### 1.4.Delimitações do estudo

O estudo em questão foi realizado em uma plataforma de petróleo pertencente à empresa Petróleo Brasileiro SA. Os dados utilizados são representativos da realidade do processo e

pertencentes a três origens distintas na empresa. A primeira são as variáveis de operação da planta, esses dados foram integrados ao sistema PI e por lá consumidos. Este sistema é uma solução que permite o monitoramento e automação da planta, possibilitando a aquisição de informações em tempo real ou histórico, dispensando a necessidade de presença física na planta.

As outras duas fontes de dados são de natureza analítica. A primeira é proveniente do monitoramento online do teor de óleo e graxa (TOG), realizado por meio de amostradores e analisadores automáticos. A segunda fonte é a análise gravimétrica, conduzida de acordo com o método padronizado Standard Methods 5520-B. Estas fontes adicionais de dados garantem a abrangência e precisão necessárias para a pesquisa.

Este trabalho propõe um método que proporcione uma resposta rápida para a possibilidade de desenquadramento do descarte de água produzida. Dessa forma, ao adotar medidas corretivas oportunamente, é possível garantir que os resultados oficiais do método gravimétrico, que normalmente demoram cerca de 20 dias para serem disponibilizados, estejam em total conformidade com as exigências legais estipuladas.

O enfoque esteve em determinar as variáveis de processo determinantes no resultado e TOG, em seguida construindo um modelo matemático e, por fim, implementá-lo em uma aplicação de fácil acesso ao público de interesse.

## 2. Fundamentação teórica

No processo de extração de petróleo, o fluido resultante é composto por petróleo, água e gás. Dentro desses três componentes, apenas a porcentagem de água não possui valor econômico. Em campos de petróleo maduros, a quantidade de água gerada representa mais de 80% do líquido residual produzido, chegando a atingir até 95%. Normalmente, a proporção entre o volume de óleo e água é de 1:3. A água associada a esse processo é conhecida como "água produzida" ou "água de produção" (FAKHRU'L, 2009).

No estágio inicial do processamento do petróleo, as três fases (óleo, água e gás) são separadas. A fração correspondente à água produzida passa por um processo de tratamento que inclui a adição de produtos químicos para remover quaisquer hidrocarbonetos e óleos ainda presentes. Após atender aos requisitos necessários, essa água pode ser descartada no oceano ou reinjetada nos poços ou em outras formações geológicas apropriadas (ABASS, 2020).

O impacto do descarte da água de produção no ambiente marítimo tem sido objeto de extensas pesquisas, devido à presença de uma considerável quantidade de compostos orgânicos dissolvidos, metais pesados e outros produtos químicos. Até o momento, o impacto a longo prazo dessas substâncias nos organismos marinhos ainda não é totalmente compreendido (EBENEZER, 2014).

No Brasil, o descarte contínuo de água produzida em plataformas marítimas de petróleo e gás natural é regulamentado pela Resolução CONAMA 393, datada de 8 de agosto de 2007. Esta resolução estabelece diretrizes para a prevenção, controle e fiscalização da poluição decorrente do lançamento de óleo e outras substâncias nocivas em águas sob jurisdição nacional. De acordo com o Artigo 5º, o descarte de água produzida deve atender a uma concentração média aritmética simples mensal de TOG de até 29 mg/L, com um valor máximo diário de 42 mg/L. O Artigo 15º estabelece que o não cumprimento dessas disposições sujeita os infratores às sanções previstas na legislação em vigor (CONAMA, 2007).

### 2.1. Processamento primário do Petróleo

Na produção de petróleo, onde o foco principal recai sobre a obtenção de hidrocarbonetos, é economicamente viável que a própria instalação disponha de mecanismos destinados a tratar e separar os fluidos provenientes dos poços. Este procedimento visa a eliminação de impurezas, bem como a segregação de diferentes fases, tais como petróleo, gás

e água, preparando, desse modo, o petróleo para etapas posteriores de processamento. Esse estágio é efetuado sob rigoroso controle e é denominado de processamento primário, e seu propósito consiste em separar o gás, em condições controladas, ao mesmo tempo em que se retira a água, as impurezas e os contaminantes, assegurando a estabilidade do óleo para posterior transferência à unidade de refino (THOMAS, 2004).

Inicialmente, os fluidos produzidos atravessam vasos separadores, que podem ser bifásicos ou trifásicos, podendo operar em sequência ou simultaneamente. No separador bifásico, a separação ocorre entre os componentes gasosos e líquidos, ao passo que no separador trifásico, efetua-se a segregação entre água, óleo e gás. Esses separadores podem ser observados na Figura 1. (THOMAS, 2004).

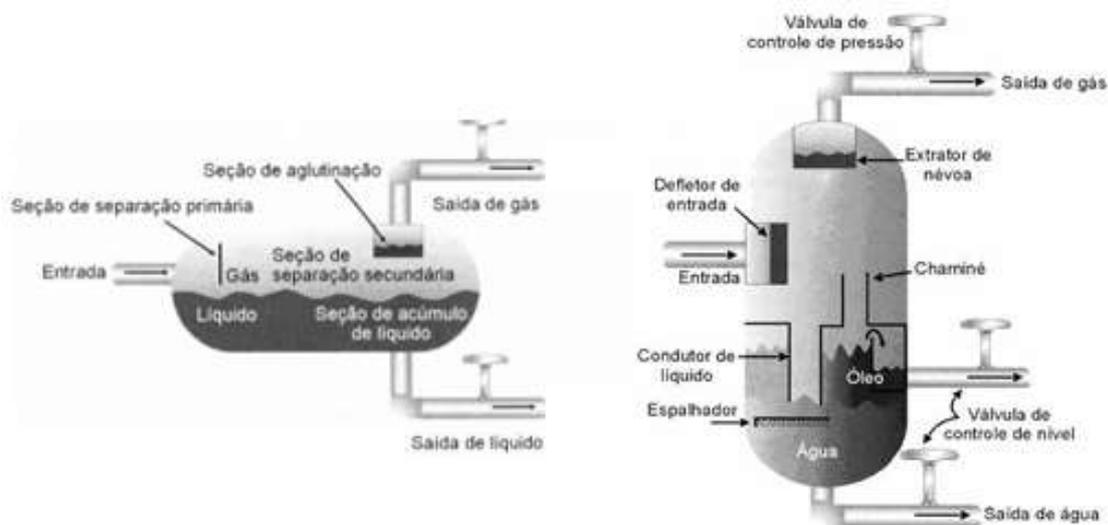


Figura 1: Na esquerda exemplo de um separador bifásico e na direita um trifásico (THOMAS, 2004)

Cada uma das fases é encaminhada para tratamento nas instalações da planta. O gás natural passa por um processo de desidratação e dessulfurização antes de ser comprimido e encaminhado às Unidades de Tratamento de Gás Natural (UTGNs) para a realização de tratamentos adicionais e subsequente venda. Por outro lado, o óleo é submetido a um tratamento focalizado na remoção de gotículas de água emulsionadas, que inclui a aplicação de campos elétricos de alta voltagem (entre 15.000V e 50.000V) nos dispositivos conhecidos como

tratadores eletroestáticos. (PEREIRA, 2004)

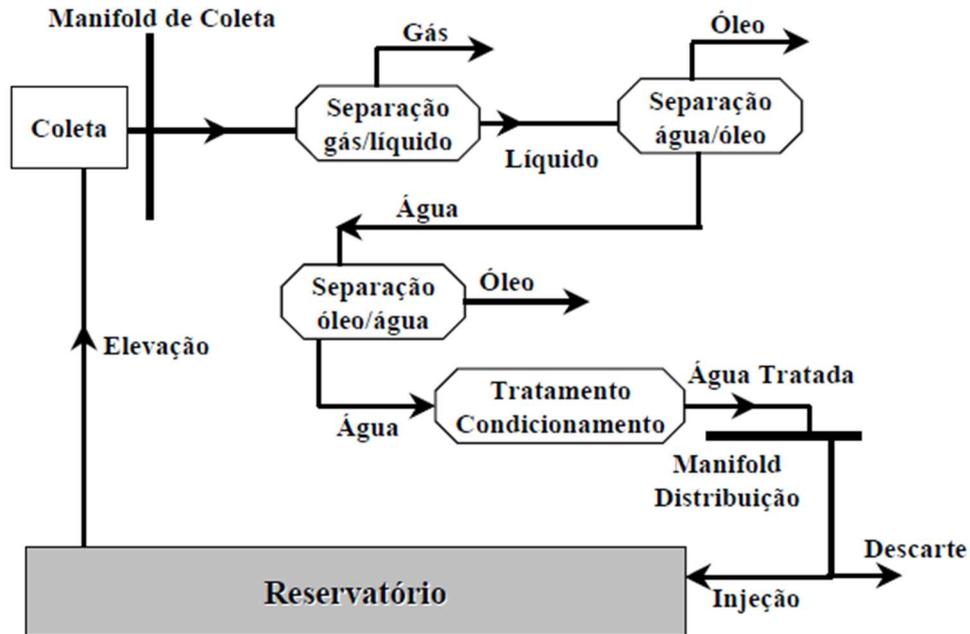


Figura 2: Esquema simplificado do tratamento primário do petróleo (NEVES, 2008)

Neste contexto, o foco central reside no tratamento da água produzida. Como indicado na Figura 3, todo o óleo recuperado ao longo das várias etapas é coletado em um tanque recuperador para ser reintegrado ao processo. O esquema delineia o percurso da água originada nos separadores e nos tratadores de óleo, direcionando-a inicialmente para um vaso desgaseificador, prosseguindo então para um separador água/óleo e, por fim, para um canal de descarte. É importante salientar que a maioria das unidades da petrolífera adota a abordagem de tratamento da água produzida com vistas à reinjeção (para recuperação secundária de petróleo), o que representa a alternativa mais favorável em termos ambientais, especialmente quando não afeta adversamente o reservatório, resolvendo, assim, a questão da destinação da

água produzida conjuntamente com o óleo (THOMAS, 2004).

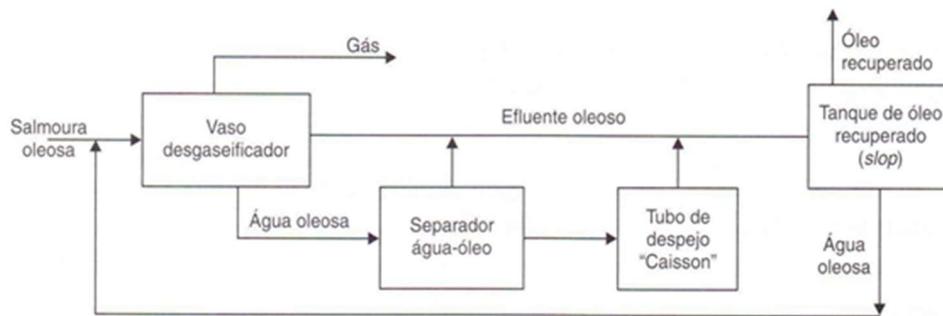


Figura 3: Planta de descarte (NEVES, 2008)

A escolha do método de tratamento da água produzida depende dos componentes a serem removidos, os quais, por sua vez, são determinados pelo destino planejado para a água tratada, que, como mencionado anteriormente, pode abranger opções de descarte, injeção ou reutilização. O processo adotado é, portanto, adaptado às exigências específicas de purificação da água em consonância com o propósito final de sua utilização.

Hidrociclones e a flotação são os processos de separação óleo/água atualmente mais utilizados pela indústria do petróleo. A flotação atua através da separação gravitacional, enquanto os hidrociclones procuram acelerar esse processo. A água oleosa entra no hidrociclone no trecho de maior diâmetro, sendo direcionada em espiral para o trecho de menor diâmetro. (VIEIRA, 2006)

No fim do percurso encontra-se o tubo de despejo, que apresenta câmaras e anteparos de retenção para promover tempo extra de resistência para separar óleo remanescente oriundo dos hidrociclones, a água oleosa recuperada é enviada para um tanque recuperador, enquanto o restante é descartado no meio ambiente. A Figura 4 demonstra um modelo de hidrociclone. (THOMAS, 2004)

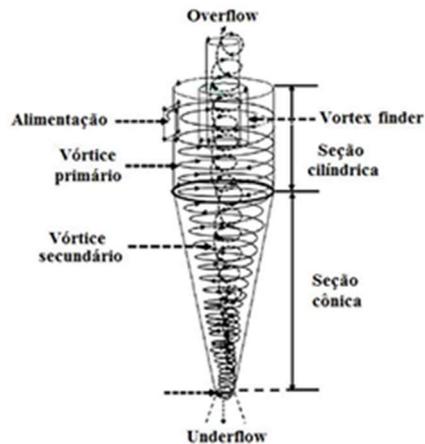


Figura 4: Desenho esquemático de um hidrociclone, mostrando o movimento espiralado do líquido em seu interior (adaptado de MURTHY & BHASKAR, 2012)

## 2.2. Análise TOG pelo Standard Methods 5520-B

O processo analítico do TOG inicia-se com a coleta de 4 alíquotas de 250 mL da água produzida, no ponto de descarte homologado pelo IBAMA. Essas coletas são realizadas em horários definidos, no geral 2h, 8h, 14h e 20h. Em caso de não haver descarte esses horários podem ser alterados ou até mesmo repetidos.

Esses frascos são enviados semanalmente para terra, acidificados com pH menor que 2 e sempre refrigerados de 0 – 6°C e possuem validade de 28 dias. (Standard Methods 5520-B).

A rota analítica para determinação do TOG, resumidamente, inicia-se com a adição do conteúdo dos quatro frascos de amostra a um funil de decantação. Nesse funil são adicionados 120 mL de n-hexano. Em seguida, para promover a homogeneização essa mistura é agitada por 2 minutos. (Standard Methods 5520-B)



Figura 5: Parte 1 da análise de TOG gravimétrico (Autoria própria)

O n-hexano é empregado na metodologia por ser uma substância química que é amplamente utilizada na indústria devido às suas propriedades como solvente. É apolar e, portanto, é especialmente útil na extração de óleos e gorduras. Quando utilizado como solvente, ele interage principalmente com as moléculas de óleo, permitindo que elas se dissolvam nele. Em seguida, o solvente é separado do óleo, restando assim o óleo puro. (O'NEIL, 2006)

Após a agitação, é necessário que o funil esteja em repouso. Desse modo a fase orgânica ficará na parte superior do funil e a água na parte inferior. A fase orgânica é levada para o aparelho evaporador rotativo. (Standard Methods 5520-B). Esse aparelho é uma ferramenta de laboratório amplamente utilizada para a destilação a vácuo e evaporação de solventes em escala reduzida. (COOK, 2004)



Figura 6: Parte 2 da análise de TOG gravimétrico (Autoria própria)

Após a evaporação o balão secará na estufa, a 70 °C, por 30 minutos. Por fim, este balão é armazenado em um ambiente sem umidade até atingir a temperatura ambiente, para ser, então, pesado. (Standard Methods 5520-B)



Figura 7: Parte 3 da análise de TOG gravimétrico (Autoria própria)

A determinação do TOG é dada pela diferença entre a massa inicial e final do balão, dividido pelo volume inicial de amostra. A Equação 1, demonstra essa relação matemática. (Standard Methods 5520-B)

$$\text{TOG} = \frac{(m_2 - m_1) \times 1.000.000}{V}$$

Onde:

TOG: Teor de óleo e graxa, em mg/L

m<sub>2</sub>: Massa final do balão após destilação, em g

m<sub>1</sub>: Massa inicial do balão após destilação, em g

V: Volume de amostra, em mL

1.000.000: Fator de conversão de unidades.

### 2.3. Redes neurais artificiais

As redes neurais artificiais (ANN), são uma abordagem amplamente adotada em diversos domínios, incluindo classificação, regressão, análise de séries temporais e agrupamento. Essas ANNs são compostas por neurônios organizados em camadas,

compreendendo uma camada de entrada, uma ou mais camadas intermediárias (escondidas) e uma camada de saída (BALESTRASSI et al., 2009).

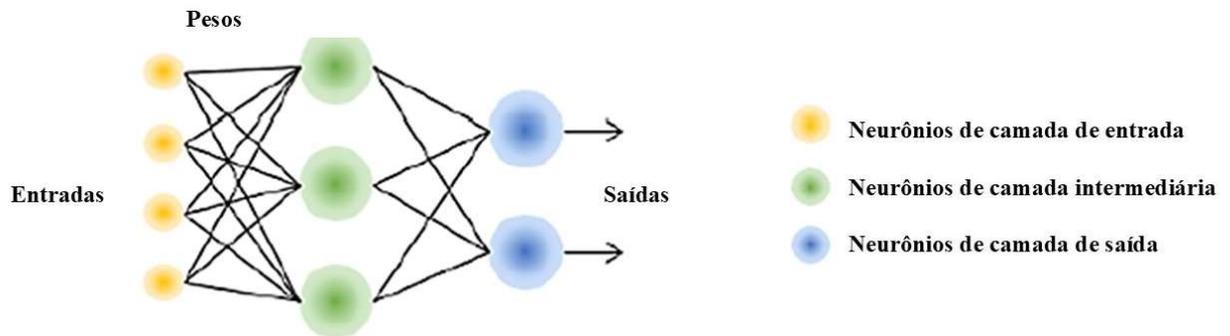


Figura 8: Exemplo de uma ANN com uma camada escondida (STREITENBERGER, 2022)

À rede neural de múltiplas camadas composta por neurônios com funções sigmoidais nas camadas intermediárias dá-se o nome de Perceptron de Múltiplas Camadas (MLPs – MultiLayer Perceptron) (BRAGA, 2007). Os MLPs demonstraram ser eficazes na resolução de uma ampla variedade de problemas desafiadores por meio do treinamento supervisionado com o algoritmo de retropropagação de erro (HAYKIN, 2001).

Algumas observações a respeito da aplicação de redes neurais, para este estudo, devem ser ressaltadas:

- É demonstrado que as redes neurais do tipo *feedforward* são aproximadores de função universais (teoremas de Kolmogorov, Sprecher e Hecht-Nielsen). Esse resultado teórico é acompanhado por investigações numéricas que tem mostrado experimentalmente a grande capacidade de redes neurais relativamente simples em aproximar mapas não-lineares. Esses achados estimulam a aplicação de redes neurais no campo de regressões lineares e não lineares, tópico principal dessa pesquisa.
- No projeto de uma rede neural os resultados de desempenho, e especialmente a sua capacidade de generalização da solução, dependem da configuração inicial dos parâmetros da rede neural. Isso é, geralmente, baseado em heurística e muitas vezes pode tornar o método extremamente complexo. O uso de Delineamento de Experimentos (DOE) pode otimizar a parametrização de ANN.
- Uma rede neural pode construir um modelo linear a partir de uma função de

transferência não-linear. A equação de uma rede neural pode ser dada pela equação 2:

$$x_t = w_0 + \sum_{l=1}^H w_l \tanh\left(\sum_{k=1}^I a_{lk} x_{t-k} + a_{l0}\right)$$

Onde:

H: número de camadas ocultas

I: número de entradas da rede

- A rede neural da equação anterior, pode ser usada como um aproximador universal de funções. Ela pode aproximar qualquer função com um número finito de descontinuidades.

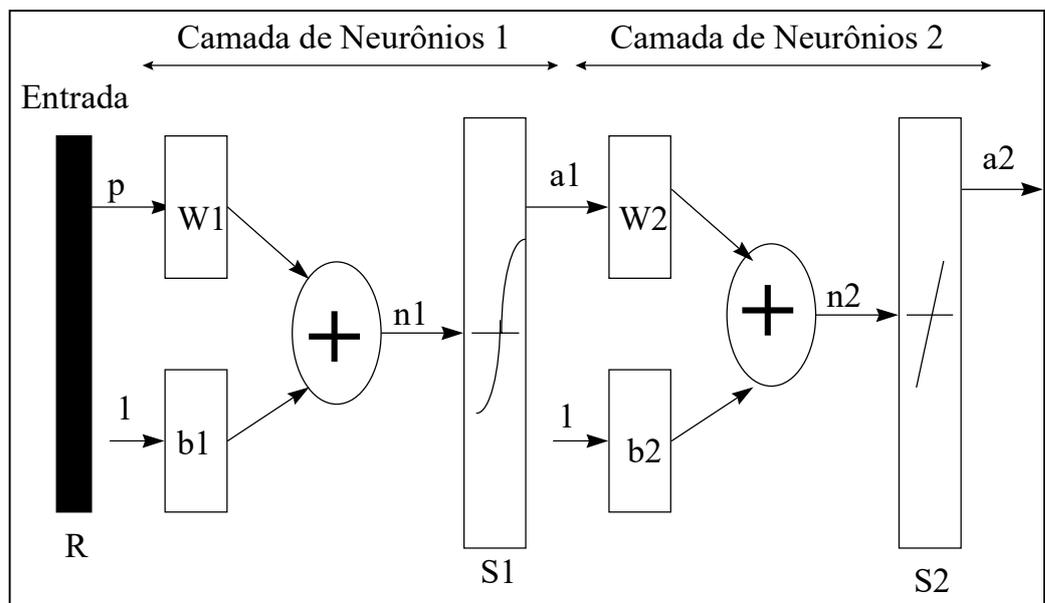


Figura 9: Rede *Multilayer Perceptron (feedforward)*

Onde:

P: Pesos iniciais

R: Número de Entradas

Si: Número de Neurônios da Camada i

B: Bias

W: Pesos

$A_i$ : Saídas após função de transferência

$N_i$ : Saídas após somatório com pesos e bias

O processo de aprendizagem da rede pode ser implementado utilizando um programa como, por exemplo, o *Statistica*. Em problemas não lineares, a serem abordados nessa pesquisa, a Figura 10 apresenta o problema prático em se parametrizar as redes neurais.

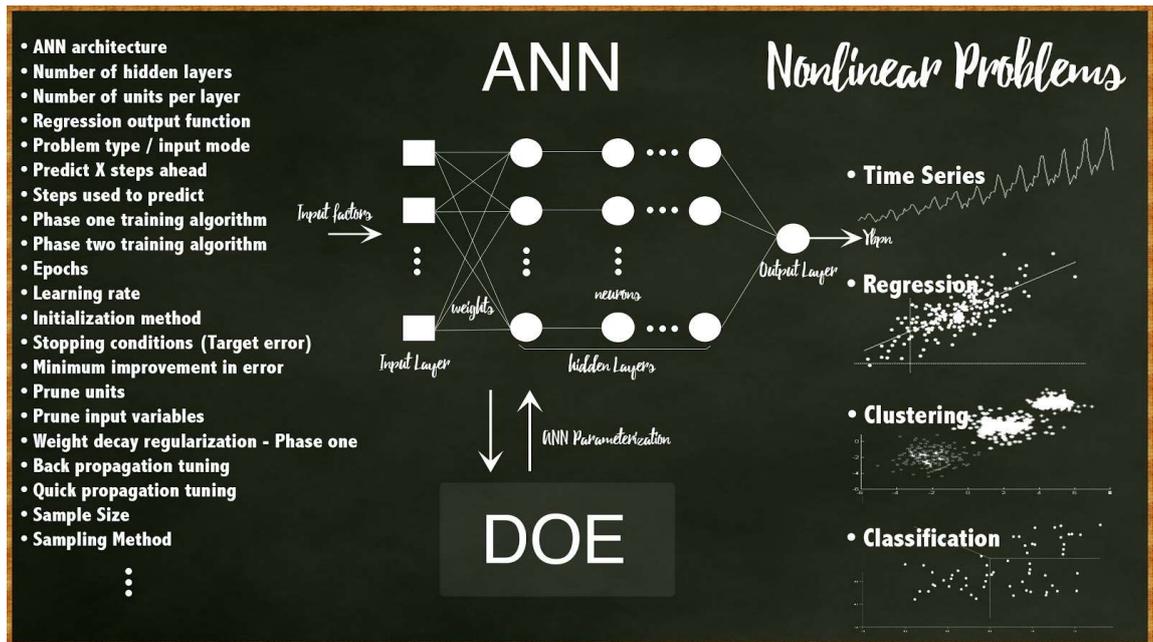


Figura 10: Processo de treinamento de uma RNA para problemas não lineares usando DOE.

## 2.4. Parâmetros de processo relevantes para o modelo

### 2.4.1. Vazão de água produzida

A água produzida refere-se à água que é trazida à superfície junto com o petróleo durante a produção de reservatórios de petróleo, sua vazão de descarte em uma plataforma de petróleo pode influenciar o valor de TOG (ADEWUNMI, 2018). A água produzida precisa ser separada do óleo, para isso aplicam-se os separadores e em seguida equipamentos mais específicos, como os hidrociclones. Quanto maior a vazão de água produzida, mais recursos são necessários para realizar essa separação de maneira eficiente, mais exige-se dos hidrociclones e separadores. Se a capacidade de tratamento não acompanhar a vazão de água, pode ocorrer mais emulsificação (mistura de óleo e água) e, conseqüentemente, mais óleo pode ser descartado associado a essa água. (SPEIGHT, 2019).

### 2.4.2. Polieletrólitos e desemulsificantes

As emulsões desempenham um papel significativo em diversos estágios da indústria petrolífera, abrangendo desde a perfuração de poços, a subdivisão do petróleo bruto até a extração subsequente. Uma emulsão se caracteriza como um sistema líquido heterogêneo composto por dois líquidos imiscíveis, nos quais um deles é disperso na forma de minúsculas gotículas no outro. A formação de emulsões requer a ação de agentes emulsificantes e a aplicação de energia, seja de natureza mecânica ou outro método, para possibilitar a dispersão dos componentes (SCHRAMM, 1992).

Em especial, na indústria de petróleo, as emulsões detêm um papel de considerável relevância, pois podem elevar o conteúdo de TOG na água descartada. Para tratar essa questão, faz-se uso de desemulsificantes, que se configuram como substâncias químicas cuidadosamente desenvolvidas para desmembrar emulsões de óleo na água. Os desemulsificantes atuam rompendo a estabilidade das emulsões, o que possibilita que as gotas de óleo coalesçam e se separem da água. Este processo simplifica a subsequente separação eficaz do óleo da água, contribuindo para a redução das concentrações de óleo na água produzida (BOEIRA; MOYSÉS; AMORIM, 2015).

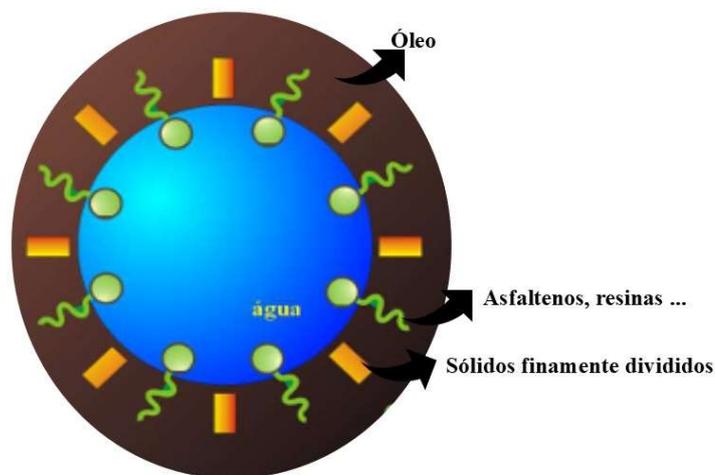


Figura 11: Representação de uma gota de água de uma emulsão do tipo água/óleo. (BOEIRA; MOYSÉS; AMORIM, 2015)

Concomitantemente, polieletrólitos, que se caracterizam como polímeros dotados de carga elétrica, desempenham um papel essencial no tratamento da água produzida. Essas substâncias podem ser classificadas em catiônicas (com carga positiva), aniônicas (com carga negativa) ou não-iônicas (sem carga) e são comumente empregadas devido às suas propriedades benéficas no processo de tratamento. Os polieletrólitos podem ser aplicados na coagulação e floculação, aglutinando partículas suspensas na água, incluindo sólidos e pequenas gotas de

óleo, formando flocos maiores que podem ser facilmente removidos por sedimentação ou flotação. Além disso, a adição de polieletrólitos a processos de separação, como hidrociclones e separadores, aumenta a eficiência da separação entre a água e o óleo, minimizando as perdas de óleo na água produzida e, conseqüentemente, influenciando no valor de TOG. Por fim, os polieletrólitos contribuem para a melhoria da qualidade da água, tornando-a adequada para descarte ou reinjeção no reservatório, em conformidade com as regulamentações ambientais (KUNERT *et al.*, 2007) (MOTTA, 2013).

### 2.4.3. TOG online

Analisadores online de TOG na água produzida são instrumentos projetados para monitorar e medir a concentração de óleos e graxas dissolvidos ou dispersos na água produzida em tempo real. Normalmente, operam com base em princípios de análise, como espectroscopia, fluorescência e infravermelho (SPEIGHT, 2019).

No entanto esses métodos não são robustos e confiáveis o suficiente para serem aceitos como método de referência para medir a concentração de TOG. Geralmente exigem calibração adequada e um ambiente controlado para serem confiáveis. Entretanto, em instalações em campo, grandes variações de misturas, pressões, produtos químicos e temperaturas criam um ambiente hostil e não ideal para os monitores. Sendo assim o método online é mais empregado como uma avaliação da tendência do TOG (SEVERIN *et al.*, 2020).

## 3. Metodologia

### 3.1. Metodologia

O propósito primordial deste estudo foi desenvolver um método preditivo para estimar o teor de TOG em plataformas de exploração marítima. Isso assume uma relevância notável, pois as medições gravimétricas oficiais, geralmente conduzidas em instalações terrestres, enfrentam consideráveis atrasos na disponibilização dos resultados, devido a questões logísticas e à complexidade inerente das análises.

Neste contexto, este trabalho adotou-se inicialmente a abordagem do sistema "Seis Sigma", uma metodologia enraizada em medidas estatísticas de variação que descrevem a qualidade de um processo. Este estudo se propôs a explorar a aplicação, os impactos e os desafios associados à implementação, empregando as melhores práticas inerentes.

A metodologia Seis Sigma é ancorada em um ciclo de melhoria contínua conhecido como DMAIC, que, neste contexto, corresponde às seguintes etapas:

- Definir (Define): Identificação do problema (impacto do descarte inadequado de água) e estabelecimento de metas para criar um modelo preditivo preciso.
- Medir (Measure): Coleta de dados, incluindo TOG gravimétrico e outras variáveis que influenciam o resultado.
- Analisar (Analyze): Análise de dados para identificar causas raiz e áreas que precisam de melhorias.
- Melhorar (Improve): Implementação de soluções, considerando a criação de um modelo de redes neurais.
- Controlar (Control): Implementação de sistemas de controle e escolha de *software* para monitorar o processo, como o *Spotfire* da TIBCO ou o *Smart Monitor* interno.

É fundamental observar que o Seis Sigma emprega uma gama de ferramentas estatísticas, como gráficos de controle, análise de regressão e testes de hipóteses, para tomar decisões embasadas em dados e aprimorar a qualidade do processo. É importante ressaltar que o Seis Sigma é uma abordagem altamente orientada por dados e metas, destinada a otimizar os processos e satisfazer as expectativas dos clientes.

Segundo Marconi e Lakatos (2006), “um objeto que mereça ser investigado

cientificamente e tenha condições de ser formulado e delimitado em função da pesquisa” será aqui explorado em três formas: dados da literatura, dados simulados e dados reais obtidos na empresa. A correlação entre medidas de TOG obtidas em plataforma (*offshore*) e oficializadas em solo (*onshore*), como objeto de pesquisa, foram exploradas a partir de dados de literatura, através de uma extensa pesquisa bibliográfica. Simulações computadorizadas foram desenvolvidas para obter o modelo de treinamento das redes neurais artificiais considerando as possíveis variáveis que impactam nessa correlação. Diversas estruturas de dados foram testadas pelo algoritmo considerando os valores de TOG *offshore* e *onshore*.

### 3.2. Etapas de execução

A complexidade inerente à criação de modelos de previsão demandou uma abordagem metódica e interdisciplinar, que compreendeu desde a aquisição dos dados brutos até a validação de um modelo final. Antes de apresentar os detalhes específicos das etapas da metodologia utilizada, é crucial compreender a estrutura geral que norteia todo o processo de mineração e preparação dos dados e, posteriormente, a criação do modelo de previsão baseado em redes neurais artificiais. O fluxograma delineado na Figura 12 apresenta uma visão das etapas desde o acesso inicial aos dados até a validação final do modelo.

Este fluxograma destaca as interconexões entre cada etapa do processo e evidencia a lógica subjacente à construção do modelo, a partir de uma visão ampla e generalizada das etapas.

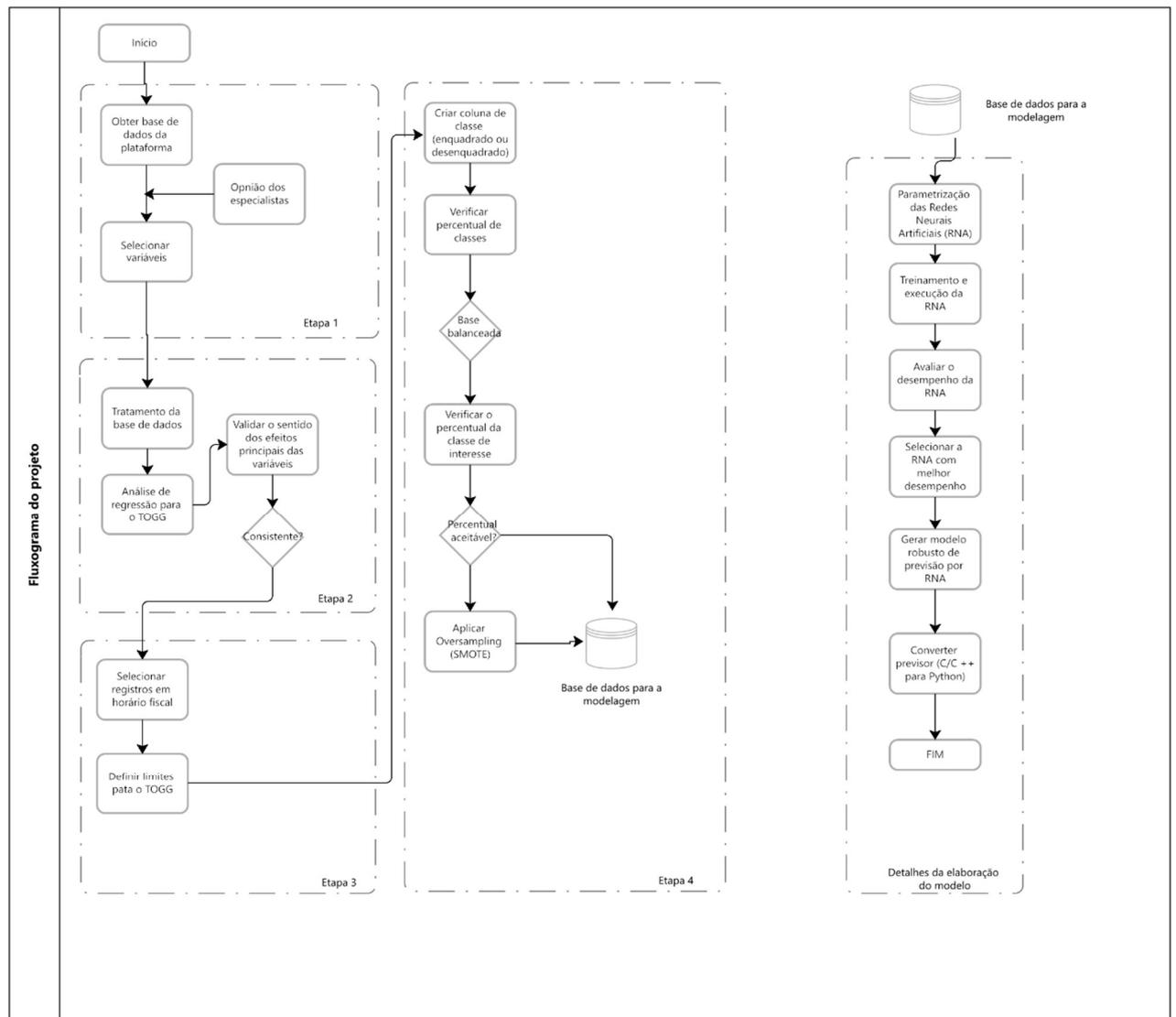


Figura 12: Etapas para a elaboração do modelo (Autoria própria)

### 3.2.1. Etapa 1: Obtenção dos dados e seleção das variáveis

Nessa fase o objetivo é realizar o mapeamento dos dados relevantes para a elaboração do modelo. A empresa proprietária dos dados possui uma variedade de informações, diretamente proporcionais a complexidade da indústria petroquímica. São dezenas de parâmetros, que devem ser submetidos ao crivo dos especialistas da área, de modo a responder com base na experiência e expertise quais desses são preliminarmente relevantes para o resultado de TOG, e podem contribuir de modo eficaz para o modelo proposto.

### 3.2.2. Etapa 2: Análise da base de dados e validação do sentido dos efeitos

A etapa seguinte é o tratamento dos dados levantados e selecionados na fase inicial do projeto. Ela compreende a avaliação e correção de dados faltantes e inconsistências, garantindo a qualidade e a integridade para o desenvolvimento do modelo. Essa avaliação consiste na realização de uma análise exploratória. É relevante avaliar a natureza dessas inconsistências, se são de forma aleatória ou se há um padrão subjacente, pois essa compreensão ajudará na decisão da melhor abordagem.

Nessa fase propõe-se a validação do comportamento das variáveis pré-selecionadas, criando um modelo de regressão (linear ou não-linear) em relação ao TOG gravimétrico para discussão do sentido de influência das variáveis do processo na resposta de interesse. Possíveis ajustes na base de dados podem ser necessários para evitar vieses na regressão. Esses ajustes podem incluir a revisão e limpeza dos dados, a remoção de *outliers* (valores discrepantes) ou a normalização das variáveis.

Ao final do processo, a opinião dos especialistas será considerada, pois, para se obter um modelo satisfatório, é essencial considerar as modificações às quais o processo está poderá estar exposto. Isso ajudará na garantia de elaborar um modelo o mais abrangente possível, e não que represente apenas um recorte histórico.

### 3.2.3. Etapa 3: Definição dos limites e registros fiscais

Nessa etapa do trabalho o objetivo é definir o limite do TOG gravimétrico para que posteriormente se possam criar classes binárias de enquadramento e desenquadramento.

Além disso, dada a característica de amostragem do TOG gravimétrico, faz-se necessário selecionar os registros da base de dados que reflitam os momentos mais próximos dos horários de coleta das amostras de TOG. Esta etapa propõe a realizar esta seleção, obtendo, na sequência, as medições oficiais diárias de TOGG do laboratório, da empresa em questão, para validação da base e alocação adequada aos respectivos horários fiscais definidos, caso necessário.

### 3.2.4. Etapa 4: Criação das classes e balanceamento da base de dados

Nessa fase o objetivo consiste em criar classes binárias de enquadramento e desenquadramento do TOGG, baseado nas definições do limiar delineado na etapa 3.2.3. Posteriormente, faz-se necessário verificar o percentual de cada tipo de classe, para tomada de

decisão sobre a necessidade de balanceamento.

Uma vez havendo essa necessidade, sugere-se avaliar a técnica de balanceamento mais adequada que expanda a base de dados ou suprima registros selecionados como, por exemplo, as técnicas SMOTE (*Synthetic Minority Over-sampling Technique*) e *NearMiss*. O produto desejado será uma base de dados balanceada em termos dos percentuais de enquadramento e desenquadramento, com finalidade de obter um conjunto de dados consistente e equilibrado para extração do comportamento do TOG gravimétrico através da modelagem futura, por redes neurais artificiais.

### 3.2.5. Etapa 4: Modelagem

Na fase de modelagem, os dados são divididos em conjuntos de treino, teste e validação para garantir que o modelo desenvolvido seja robusto e generalize bem para novos dados. O conjunto de treino é usado para ajustar o modelo, permitindo que ele aprenda os padrões e relações presentes nos dados. Durante essa etapa, diferentes algoritmos de modelagem são aplicados, ajustando os parâmetros para minimizar os erros de previsão.

Após o treinamento, o modelo é aplicado ao conjunto de teste. Este conjunto de dados, desconhecido pelo modelo, ajuda a verificar se o modelo está superajustado (*overfitting*) ou subajustado (*underfitting*). Métricas como o coeficiente de determinação ( $R^2$ ), erro quadrático médio (MSE) e análise de resíduos são utilizadas para avaliar o desempenho do modelo no conjunto de teste.

Finalmente, o conjunto de avaliação, que também é separado dos dados de treino e teste, é utilizado para uma avaliação final e imparcial do desempenho do modelo. Esta fase final garante que o modelo não apenas performa bem nos dados selecionados anteriormente, mas também mantenha a precisão em dados completamente novos. Isso é crucial para assegurar a confiabilidade e assertividade do modelo par futuros dados.

### 3.2.6. Etapa 5: Criação de um modelo preditivo através de Redes Neurais Artificiais

Essa fase consiste em efetivamente desenvolver o modelo proposto, ou seja, o modelo preditivo para o TOG gravimétrico baseado em redes neurais. Os resultados a serem preditos deverão ser classificados como enquadrado/desenquadrado e um código, com usabilidade

adequada fornecido para a empresa proprietária do processo e dos dados.

### 3.3.Complexidades do trabalho

Abordando o problema de pesquisa, entende-se que as seguintes complexidades procuram justificar o trabalho:

- Não existe um mesmo padrão de correlação entre as variáveis e processo para todas as plataformas
- Não existe um método de avaliação segura do TOG em tempo real
- Existe variabilidade no sistema de medição do TOG
- As variáveis que influenciam o TOG são complexas
- Os métodos de controle do TOG são complexos e devem ser feitos preventivamente
- Valores de TOG altos resultam em fortes sanções para a empresa e em potenciais impactos ambientais

## 4. Resultados

### 4.1. Etapa 1: Obtenção dos dados e seleção das variáveis

O primeiro passo desse trabalho contou com a participação dos especialistas da empresa detentora dos dados. Essa participação foi crucial, pois uma planta de processamento primário de petróleo conta com diversas variáveis, que por sua vez descrevem os diversos vasos e etapas do processamento e tratamento da água produzida, que é a matriz onde determina-se o TOG.

A primeira definição necessária foi acerca dos *softwares* de onde seriam extraídos os dados. Dois foram indicados e utilizados: o primeiro deles foi o PI *ProcessBook*, que é uma ferramenta amplamente utilizada na indústria para realizar o monitoramento do processo e de plantas. Desenvolvido pela *OSIsoft*, que atualmente integra AVEVA, ele permite que os usuários visualizem dados históricos e em tempo real coletados pelo sistema PI. Esta coleta de dados históricos que permitiram a utilização dessa aplicação na execução do trabalho. O segundo sistema utilizado foi *SampleManager* que é um sistema de gerenciamento de informações laboratoriais LIMS, desenvolvido pela *Thermo Fisher Scientific*. Ele é utilizado para gerenciar e otimizar as operações de laboratórios em diversos setores, como farmacêutico, biotecnologia e como em nosso estudo, petroquímica. Nesse sistema estão registradas todas as informações relativas às análises químicas realizadas pelo Laboratório da indústria em questão. Vale lembrar que o objeto de estudo desta pesquisa, o TOG, é uma análise físico-química.

Após um mês de reuniões semanais com os especialistas da empresa, foram selecionados 28 parâmetros para compor a base de dados que gerou o modelo. De acordo com o conhecimento do corpo técnico e a bibliografia interna, cada um desses parâmetros pode se relacionar e contribuir para o TOG, conforme resumido a seguir. Além destas informações, pôde-se observar o comportamento histórico dos dados, esboçados em *boxplots* e/ou séries temporais, realizados através do Spotfire.

**Vazão da água produzida** – A análise de TOG é realizada no fluido água produzida, o que, por si só, já sugere a importância desse parâmetro para o valor final da análise. Grande parte do objetivo do tratamento primário do Petróleo é tratar essa água, promovendo a máxima separação entre fase água e fase óleo. Esse tratamento levado ao aspecto ótimo permite maior aproveitamento do óleo produzido e menos prejuízos ao meio ambiente. Uma vazão alta demais, ou em descontrole, sugere que o tratamento não está sendo eficaz em realizar essa separação, isso por consequência pode gerar altos valores de TOG no descarte. Esse dado foi

obtido no sistema PI. A Figura 13 indica o perfil desses dados.

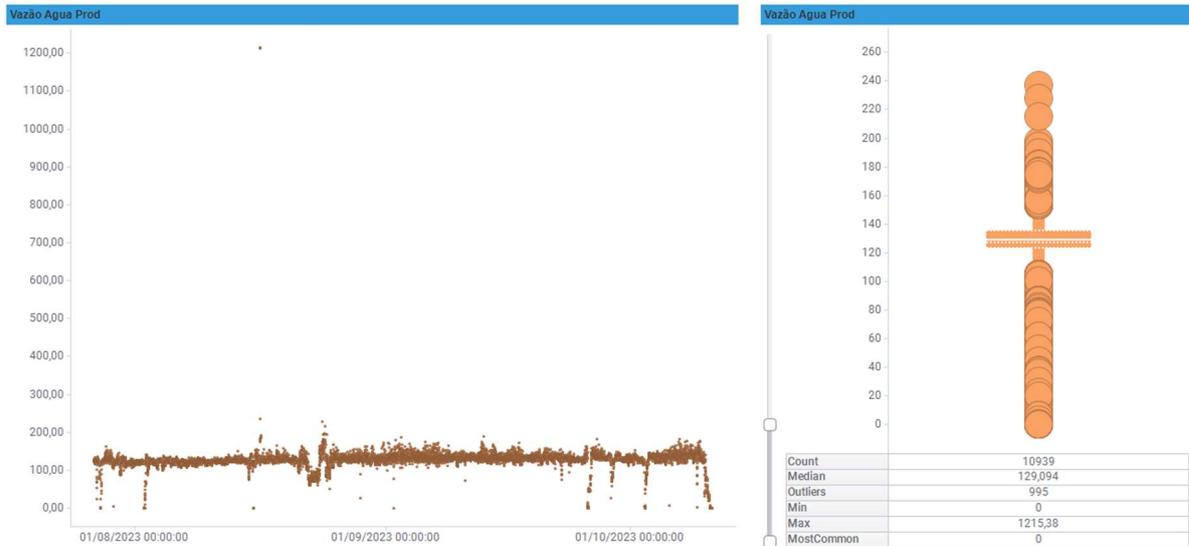


Figura 13: Comportamento histórico da variável vazão da água produzida

**Vazão de óleo produzido** – A fonte dos óleos e graxas presentes na água produzida, onde se determina o TOG, é o óleo que estava presente na formação juntamente com essa água. Tão importante quanto a verificação da vazão de água produzida é a vazão do óleo de onde se separou esse fluido. Sua vazão é um indicador crucial da eficiência do processo de extração e separação. Uma vazão muito alta sugere que o processo de extração, pode não estar recuperando o máximo de hidrocarbonetos, enquanto uma vazão de óleo baixa pode sugerir problemas na eficiência de recuperação ou mesmo questões relacionadas à formação geológica do

reservatório. A Figura 14 mostra o perfil dos dados, disponibilizados pelos especialistas.

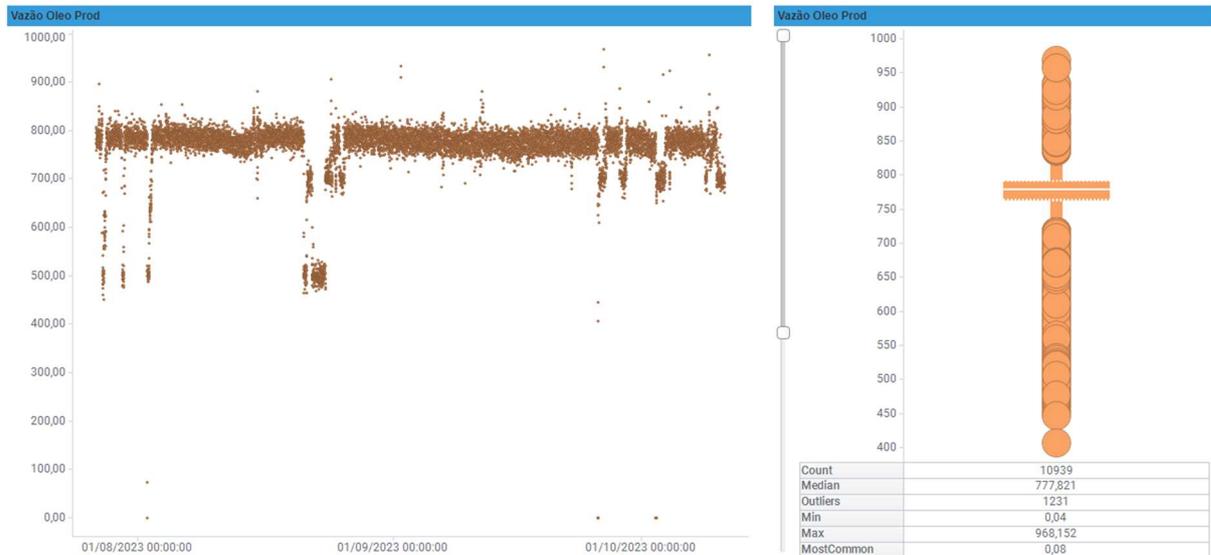


Figura 14: Comportamento histórico da variável Vazão de óleo produzido

**Temperatura do SG** – Um separador de gás, presente no tratamento primário, é um equipamento utilizado na indústria de petróleo e gás para separar o gás natural dos outros componentes presentes na produção, como líquidos e sólidos. Ele é responsável por realizar a separação física desses componentes, permitindo que o gás seja direcionado para processos de tratamento e transporte, enquanto os líquidos e sólidos são encaminhados para outros sistemas de processamento. De acordo com o estabelecido nas reuniões, o SG é uma possível fonte de influência no resultado do TOG. A temperatura afeta propriedades físico-químicas do petróleo, como sua densidade, viscosidade e estabilidade das emulsões. Portanto esse parâmetro merece atenção ao longo do processamento, uma vez que sua variação pode alterar a eficiência da separação, que é influenciado pelas propriedades anteriormente citadas. No entanto, é importante ressaltar que a temperatura ideal pode variar dependendo das especificidades do processo e das características do fluxo de produção e do fluido. No perfil dos dados, demonstrado na Figura 15, a análise preliminar já indica uma grande variabilidade e um perfil

não tão bem ajustado como os dos dados anteriores.

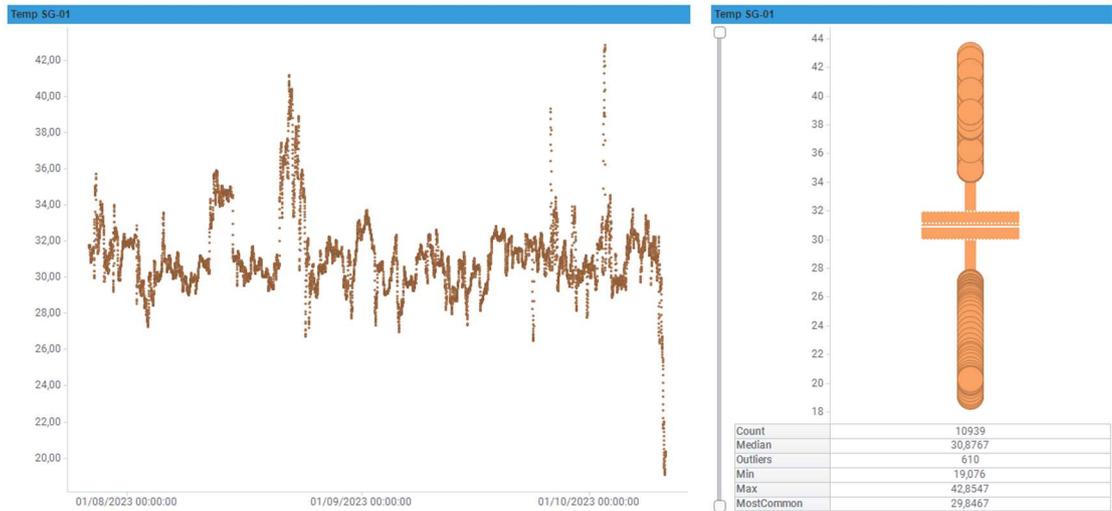


Figura 15: Comportamento histórico da variável Temperatura do SG

**Vazão do SG** – A vazão do SG pode influenciar no TOG devido à sua relação com a eficiência da separação dos componentes presentes na produção de petróleo. Se a vazão do SG estiver inadequada, pode ocorrer uma menor eficiência na separação, o que pode resultar em uma maior presença de óleos e graxas na corrente de saída. Portanto, é importante monitorar e controlar sua vazão, para garantir uma separação eficiente e reduzir o TOG na corrente de produção. Essa influência tornou este parâmetro relevante para o modelo proposto. Em relação ao perfil dos dados, pode-se observar na Figura 16 uma presença de *outliers* abaixo do valor de mediana. Essa questão foi tratada nas fases seguintes do projeto.

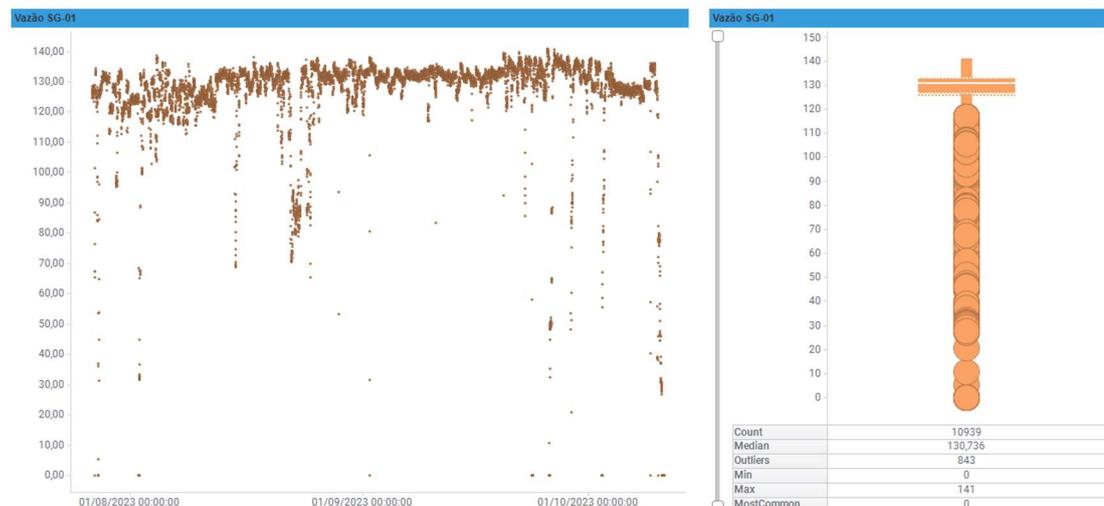


Figura 16: Comportamento histórico da variável vazão do SG

**Temperatura do TO** – O tratador eletrostático é aplicado ao processamento primário do petróleo e tem como objetivo separar a água e o gás do óleo. O petróleo passa por uma etapa de separação primária no SG, em seguida é direcionado para o TO, onde ocorre o tratamento eletrostático para separar o máximo de água do óleo. Essa etapa do tratamento é uma das mais relevantes para o valor final de TOG, é a primeira etapa do processamento primário do petróleo que visa promover a separação água/óleo.

Assim como a temperatura influencia o processo de separação no SG, no TO não é diferente, ela pode vir a afetar a viscosidade e a densidade dos líquidos, incluindo o óleo e até mesmo da água, influenciando diretamente a eficiência da separação. Em temperaturas mais altas, a viscosidade do óleo diminui, o que pode facilitar a separação do óleo da água. No entanto, temperaturas excessivamente altas podem causar problemas como a volatilização de compostos leves, que podem resultar em maior conteúdo de óleo na fase de água produzida. Por outro lado, temperaturas muito baixas podem aumentar a viscosidade do óleo, dificultando a separação e potencialmente resultando em maiores concentrações de óleo na água produzida. Em relação ao perfil dos dados, pôde observar na Figura 17 uma presença de *outliers* abaixo do valor de mediana, assim como ocorreu em relação a vazão do SG.

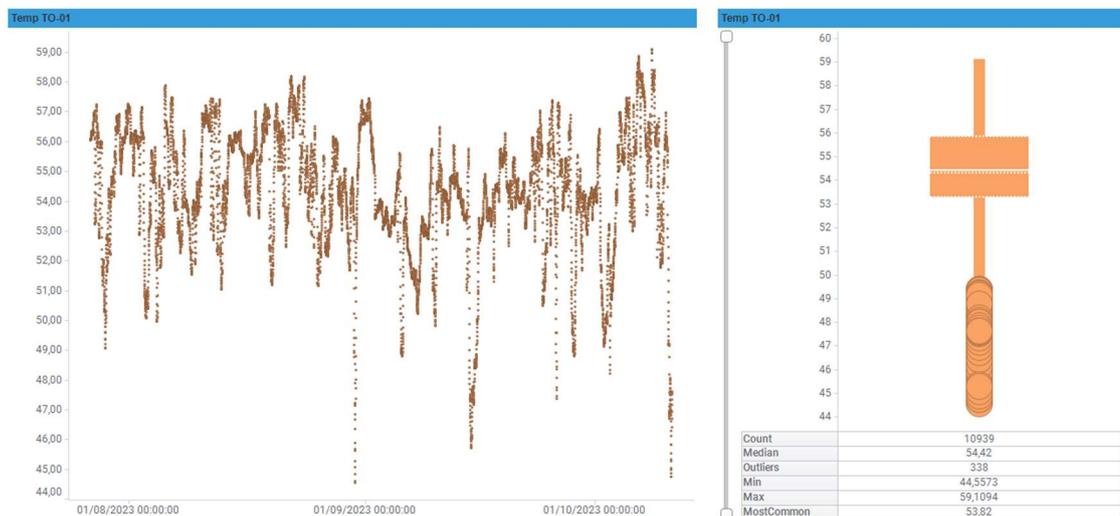


Figura 17: Comportamento histórico da variável Temperatura do TO

**LVs** - A válvula de nível do Pré TO é chamada de LV, pois deriva do termo em inglês que traduzido é "Válvula de Nível", dentro do contexto do sistema de tratamento de petróleo. Essa válvula é responsável por controlar o nível de água no pré-tratador eletrostático, garantindo que ele esteja dentro dos parâmetros definidos. A correta operação dessa válvula é fundamental para

o bom funcionamento do sistema de tratamento, pois um nível de água inadequado pode afetar negativamente a eficiência do processo, resultando em uma menor remoção de água e impurezas do óleo. Para a execução do modelo foi levado em consideração duas LVs, que podem influenciar de forma semelhante ao TOG como apontado nos parâmetros anteriores. Em laranja, na Figura 18, o perfil de dados do LV do TO-2; em azul o perfil do parâmetro LV do TO-1.

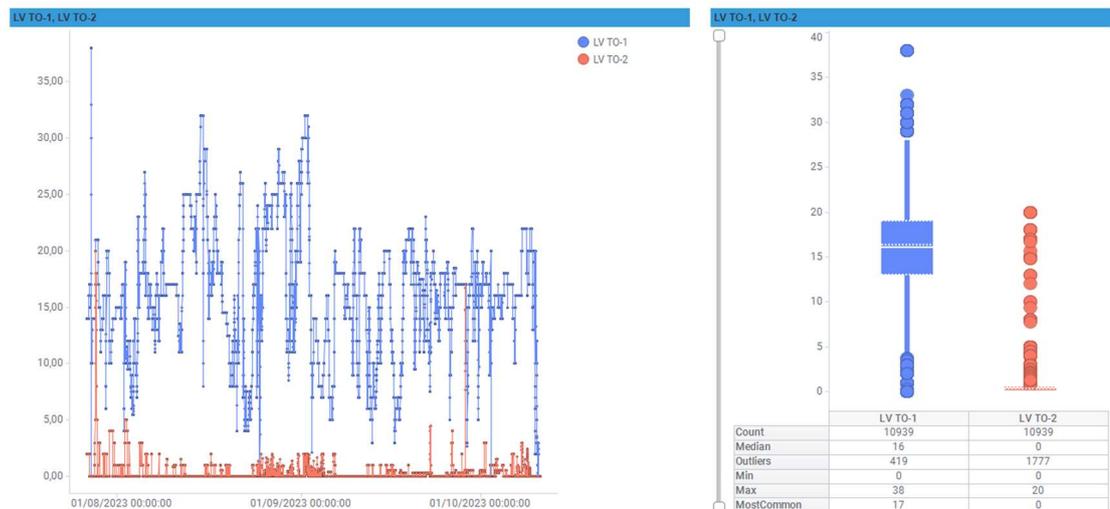


Figura 18: Comportamento histórico dos LVs

**Vazão no header de descarga** - A vazão no header de descarga das bombas, na planta da unidade em questão, é a quantidade de fluido bombeado e que é direcionada para o sistema de destino. A vazão é medida em volume por unidade de tempo, metros cúbicos por hora ( $m^3/h$ ) e pode variar dependendo das condições operacionais e das necessidades do sistema em que estão inseridas.

Esse é um parâmetro importante a ser monitorado e controlado, pois afeta diretamente o desempenho e a eficiência do sistema. Uma vazão inadequada pode resultar em problemas como baixa pressão, falta de alimentação adequada para os equipamentos subsequentes ou até mesmo falhas no processo. É importante garantir que a vazão no header de descarga esteja dentro dos limites operacionais estabelecidos, de acordo com as necessidades do sistema e as especificações do processo. Isso pode ser alcançado por meio do ajuste das bombas, controle de válvulas e monitoramento contínuo da vazão. Na Figura 19 é possível observar o perfil dos

dados, que apresenta alguns *outliers* acima do último quartil.

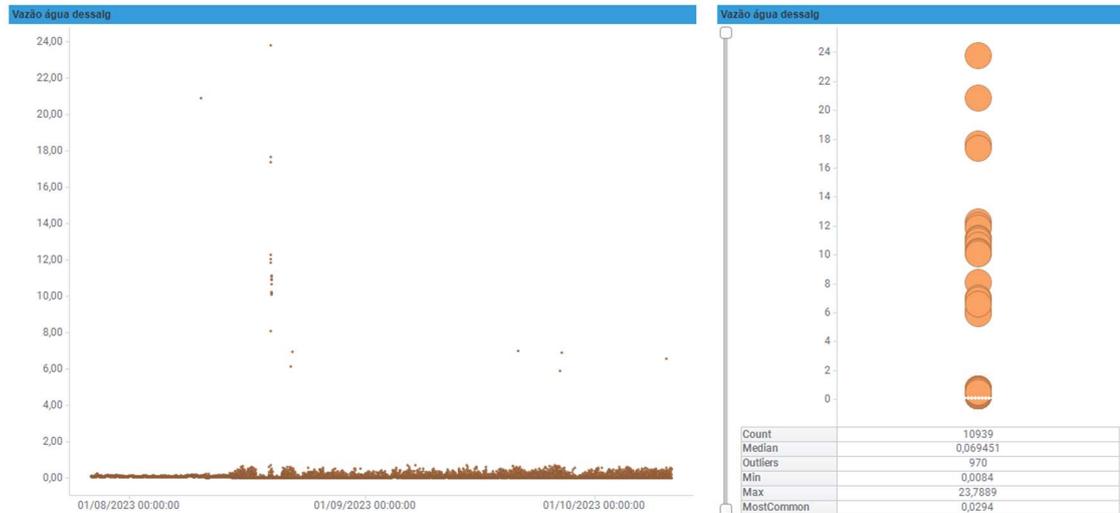


Figura 19: Comportamento histórico da variável Vazão no header de descarga

**Pressão do flotor B** - Os flotores são equipamentos utilizados para separar óleos e graxas presentes na água produzida, através do princípio da flotação. Sua pressão é controlada para garantir a eficiência do processo de separação. É importante ressaltar que a pressão adequada nos flotores é essencial para garantir a correta separação dos óleos e graxas da água produzida. Uma pressão inadequada pode comprometer a eficiência do processo e resultar em uma maior concentração de óleos e graxas na água tratada. Na Figura 20, o perfil dos dados extraídos do PI.

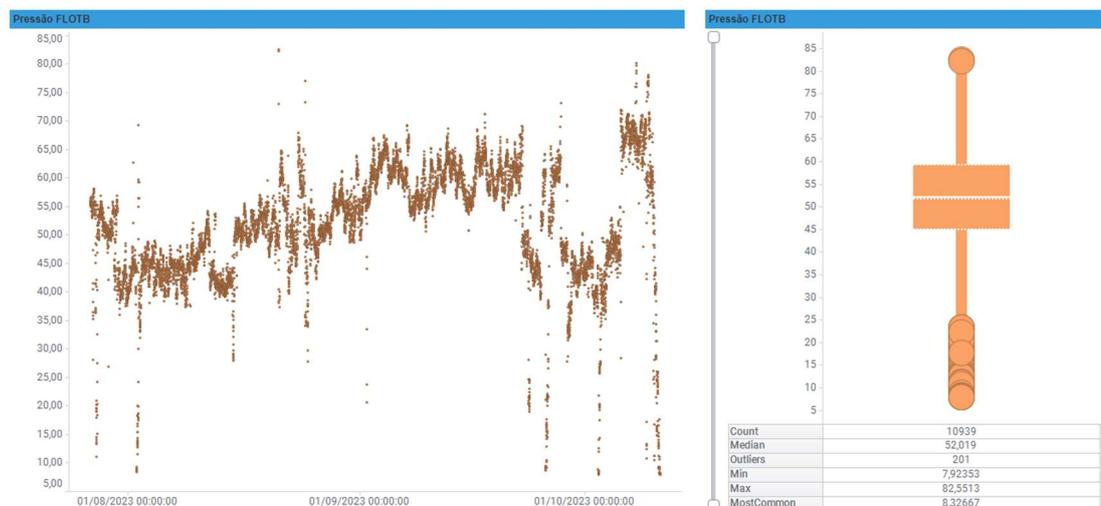


Figura 20: Comportamento histórico da variável Pressão do flotor B

**Alinhamento dos poços** – Oito poços produtores foram considerados no banco de dados do modelo. A composição dos poços próximos pode variar, dependendo das características geológicas e das formações presentes em cada poço. Embora possam existir semelhanças na composição entre poços próximos, é importante ressaltar que cada poço é único e pode apresentar diferenças em termos de reservatórios, fluidos produzidos e propriedades do petróleo. Dito isto, o alinhamento dos poços cria blends de petróleo que podem performar de modos diferentes ao longo do processo, resultando em diferentes possibilidades relacionadas ao valor final de TOG. Por exemplo, se houver poços que gerem altos teores de TOG alinhados para o descarte, isso pode aumentar a concentração total de TOG no sistema. Por outro lado, se esses poços forem alinhados para a reinjeção, a concentração total de TOG no sistema pode ser reduzida.

Na Figura 21 podemos observar um gráfico de barras que mostra a contribuição % de cada poço produtor, cada um representado por uma coloração distinta.

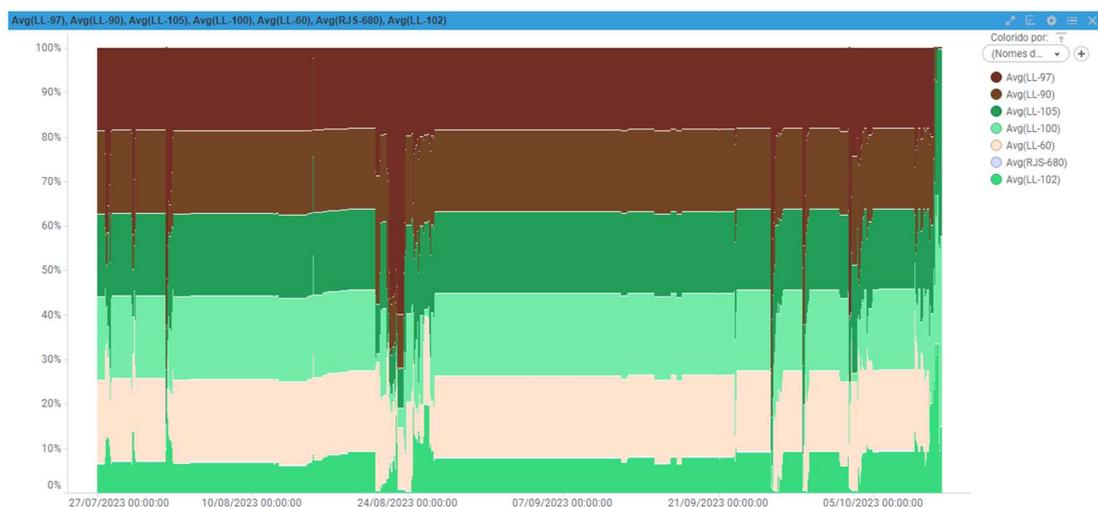


Figura 21: Contribuição % dos poços na produção diária

**Desemulsificante no header de produção e de testes** - A vazão do desemulsificante para o header tem influência direta na eficiência do processo de separação das emulsões presentes no petróleo, uma vez que esse produto é utilizado para quebrá-las, facilitando a separação em dessas fases. Ao ajustar a vazão do desemulsificante para o header, é possível controlar a quantidade de produto químico adicionado ao sistema. Um fluxo adequado ajuda a reduzir a formação de emulsões estáveis, permitindo uma separação mais eficiente do óleo e da água. É importante ressaltar que a vazão do desemulsificante deve ser ajustada de acordo com as características do blend e as condições operacionais específicas. Um fluxo excessivo de

desemulsificante pode levar a problemas como a formação de espuma e a redução da eficiência da separação. Por outro lado, uma vazão insuficiente pode resultar na persistência de emulsões e na baixa qualidade do óleo produzido. Portanto, esse parâmetro torna-se relevante para o modelo uma vez que esse parâmetro influencia diretamente no descarte de água produzida. Na Figura 22 o perfil dos dados.

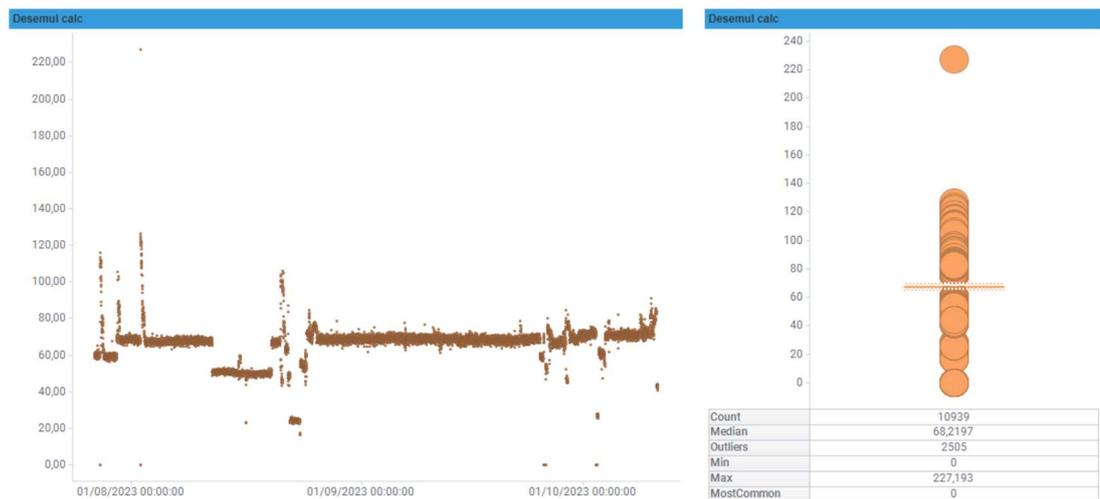


Figura 22: Comportamento histórico da variável Desemulsificante no header

**Polieletrólito** - O polieletrólito atua como um agente coagulante e floculante, ajudando a aglomerar as partículas de óleo e sólidos em maiores tamanhos, facilitando sua remoção. A influência do polieletrólito no TOG está relacionada à sua eficiência na remoção do óleo presente na água produzida. Quando utilizado corretamente, pode contribuir para a redução do TOG, promovendo uma separação mais eficiente do óleo da água. No entanto, é importante ressaltar que a dosagem e o tipo de polieletrólito utilizado devem ser adequados às

características da água produzida e às condições operacionais do sistema de tratamento.

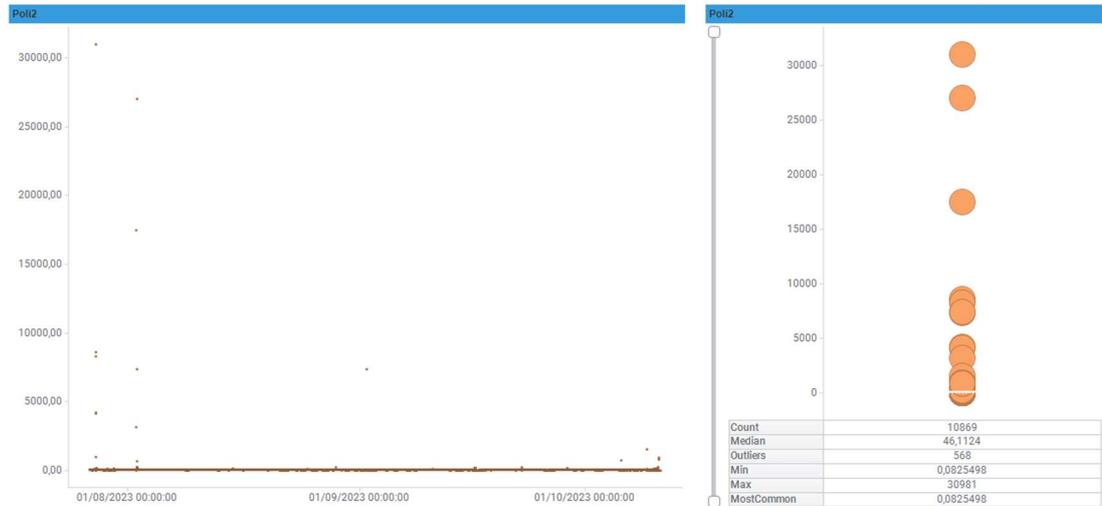


Figura 23: Comportamento histórico da variável Polieletrólito

**TOG online** – Esse sistema consiste em um analisador de TOG instalado em linha, que realiza medições contínuas e em tempo real do teor de óleo e graxas na água. Sua presença não influencia em nada o resultado do TOG do descarte, mas pode oferecer ao modelo informações relevantes sobre a tendência do processo e consequente desenquadramento do TOG. A Figura 24, por meio de um gráfico de dispersão mostra a variação do TOG online a cada 5 minutos.

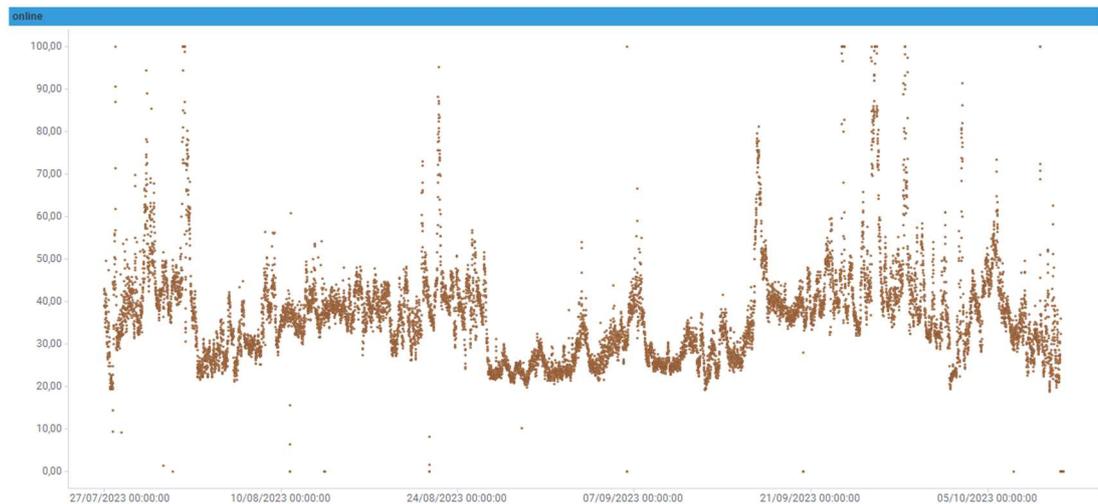


Figura 24: Variação do TOG online

**TOG gravimétrico** – Todas as informações mencionadas anteriormente foram extraídas do PI, sendo apenas o resultado analítico de TOG gravimétrico oriundo do SampleManager. Esse dado é de grande importância para o modelo, uma vez que é o objeto a

ser determinado por ele. Esses dados foram essenciais para avaliar a assertividade do modelo, treiná-lo e testá-lo.

Dada essas definições, iniciou-se a coleta dos dados históricos, a partir de um período pré-selecionado. Durante o desenvolvimento do modelo de previsão para a Plataforma em estudo. Utilizou-se o software MS Excel® para conhecimento inicial dos dados que foram disponibilizados em formato “*comma-separated values*” (csv) ou ainda planilha xlsx, como mostra a Figura 25.

|    | A                   | B               | C               | D          | E          | F           | G       | H       | I                 | J             | K     | L          | M      | N     | O     | P     | Q                |
|----|---------------------|-----------------|-----------------|------------|------------|-------------|---------|---------|-------------------|---------------|-------|------------|--------|-------|-------|-------|------------------|
| 1  |                     | Vazão Água Prod | Vazão Oleo Prod | Temp SG-01 | Temp TO-01 | Vazão SG-01 | LV TO-1 | LV TO-2 | Vazão água dessal | Pressão FLOTB | LL-97 | LL-105     | LL-100 | LL-60 | LL-90 | LL-69 | RJS-             |
| 2  | 2023-07-27 00:00:00 | 123,2001114     | 759,8756104     | 31,7626534 | 56,0666656 | 126,085953  | 14      | 2       | 0,102970004       | 55,17632675   | 100   | 99,2600021 | 100    | 100   | 100   | 100   | 99,7300034 0,469 |
| 3  | 2023-07-27 00:10:00 | 128,9294891     | 782,0737305     | 31,7512741 | 56,0466652 | 124,854134  | 14      | 2       | 0,096781448       | 55,37240982   | 100   | 99,2603836 | 100    | 100   | 100   | 100   | 99,7366638 0,469 |
| 4  | 2023-07-27 00:20:00 | 120,6190414     | 807,0828247     | 31,7626801 | 56,0466652 | 128,65416   | 14      | 2       | 0,100825004       | 55,94096756   | 100   | 99,268013  | 100    | 100   | 100   | 100   | 99,7376556 0,46  |
| 5  | 2023-07-27 00:30:00 | 123,4540329     | 772,618103      | 31,7580051 | 56,0666656 | 127,030563  | 14      | 2       | 0,111230277       | 55,57687759   | 100   | 99,2600021 | 100    | 100   | 100   | 100   | 99,7346725 0,458 |
| 6  | 2023-07-27 00:40:00 | 128,7561646     | 796,196106      | 31,6973438 | 56,087986  | 124,606308  | 14      | 2       | 0,11963889        | 55,96412506   | 100   | 99,2623444 | 100    | 100   | 100   | 100   | 99,7300034 0,462 |
| 7  | 2023-07-27 00:50:00 | 123,7649536     | 796,1733398     | 31,6686611 | 56,0973206 | 125,868927  | 14      | 2       | 0,090379111       | 55,73890886   | 100   | 99,260994  | 100    | 100   | 100   | 100   | 99,7376556 0,46  |
| 8  | 2023-07-27 01:00:00 | 123,0078888     | 774,9851685     | 31,6539364 | 56,0733337 | 127,839333  | 14      | 2       | 0,081900001       | 55,72442245   | 100   | 99,2680588 | 100    | 100   | 100   | 100   | 99,7366638 0,473 |
| 9  | 2023-07-27 01:10:00 | 127,529892      | 802,6901245     | 31,668005  | 56,0733337 | 126,532021  | 14      | 2       | 0,113423206       | 56,47370148   | 100   | 99,2629852 | 100    | 100   | 100   | 100   | 99,7346802 0,461 |
| 10 | 2023-07-27 01:20:00 | 124,8199005     | 774,0932617     | 31,669342  | 56,0773239 | 127,391289  | 14      | 2       | 0,102892646       | 55,90971375   | 100   | 99,2590103 | 100    | 100   | 100   | 100   | 99,7346725 0,464 |
| 11 | 2023-07-27 01:30:00 | 123,5426788     | 785,0994873     | 31,6913414 | 56,0800018 | 126,606926  | 14      | 2       | 0,098767899       | 55,95072937   | 100   | 99,2623444 | 100    | 100   | 100   | 100   | 99,7306366 0,468 |
| 12 | 2023-07-27 01:40:00 | 124,212532      | 763,972168      | 31,6920223 | 56,117321  | 126,664009  | 14      | 2       | 0,109142318       | 55,80589294   | 100   | 99,2699966 | 100    | 100   | 100   | 100   | 99,7300034 0,476 |
| 13 | 2023-07-27 01:50:00 | 119,1305542     | 793,6997681     | 31,7033329 | 56,1399994 | 126,991882  | 14      | 2       | 0,077847004       | 56,25004578   | 100   | 99,2629852 | 100    | 100   | 100   | 100   | 99,7346802 0,465 |
| 14 | 2023-07-27 02:00:00 | 120,7353973     | 765,9280396     | 31,7306862 | 56,1773186 | 126,903351  | 14      | 2       | 0,148879483       | 55,26647568   | 100   | 99,2536163 | 100    | 100   | 100   | 100   | 99,7346802 0,461 |
| 15 | 2023-07-27 02:10:00 | 125,7941666     | 784,1462402     | 31,7306557 | 56,1733322 | 125,067757  | 14      | 2       | 0,092448518       | 54,97177505   | 100   | 99,260994  | 100    | 100   | 100   | 100   | 99,7300034 0,45  |
| 16 | 2023-07-27 02:20:00 | 122,0547409     | 786,2878418     | 31,7373199 | 56,1800003 | 125,293884  | 14      | 2       | 0,081995554       | 54,8801498    | 100   | 99,2633362 | 100    | 100   | 100   | 100   | 99,7329788 0,463 |
| 17 | 2023-07-27 02:30:00 | 118,9692764     | 767,0158691     | 31,7306824 | 56,1800003 | 126,511742  | 14      | 2       | 0,069377005       | 54,85899353   | 100   | 99,2576599 | 100    | 100   | 100   | 100   | 99,7346725 0,464 |
| 18 | 2023-07-27 02:40:00 | 127,4192429     | 788,5195923     | 31,7099991 | 56,2000008 | 127,94487   | 14      | 2       | 0,119601004       | 55,51527786   | 100   | 99,2543335 | 100    | 100   | 100   | 100   | 99,7366638 0,456 |
| 19 | 2023-07-27 02:50:00 | 119,1196976     | 769,3526611     | 31,7186623 | 56,2133331 | 126,523354  | 14      | 2       | 0,113391459       | 55,88775635   | 100   | 99,2600021 | 100    | 100   | 100   | 100   | 99,7366638 0,466 |
| 20 | 2023-07-27 03:00:00 | 125,3268083     | 782,3293457     | 31,7493114 | 56,2400017 | 125,907745  | 14      | 2       | 0,098646238       | 55,98973465   | 100   | 99,2580185 | 100    | 100   | 100   | 100   | 99,7370224 0,45  |
| 21 | 2023-07-27 03:10:00 | 125,0993195     | 799,5401001     | 31,75      | 56,2400017 | 126,980721  | 14      | 2       | 0,114742723       | 56,38197708   | 100   | 99,2613449 | 100    | 100   | 100   | 100   | 99,7336807 0,453 |
| 22 | 2023-07-27 03:20:00 | 126,9223557     | 777,6022339     | 31,7679768 | 56,246666  | 126,70681   | 14      | 2       | 0,092436753       | 56,35745239   | 100   | 99,2633362 | 100    | 100   | 100   | 100   | 99,731987 0,456  |
| 23 | 2023-07-27 03:30:00 | 117,2657013     | 791,1613159     | 31,7507038 | 56,2666664 | 127,944382  | 14      | 2       | 0,081861503       | 56,65958023   | 100   | 99,2600021 | 100    | 100   | 100   | 100   | 99,7319946 0,462 |
| 24 | 2023-07-27 03:40:00 | 122,6020584     | 791,850647      | 31,5566959 | 56,3026772 | 124,956764  | 14      | 2       | 0,092494376       | 56,2235527    | 100   | 99,2633198 | 100    | 100   | 100   | 100   | 99,731987 0,464  |
| 25 | 2023-07-27 03:50:00 | 132,5132446     | 771,8696979     | 31,3753529 | 56,3400002 | 125,55323   | 14      | 2       | 0,113414839       | 55,24575424   | 100   | 99,2600021 | 100    | 100   | 100   | 100   | 99,7366638 0,46  |
| 26 | 2023-07-27 04:00:00 | 125,0051041     | 793,065918      | 31,2680073 | 56,3547096 | 124,718697  | 14      | 2       | 0,111101858       | 55,2976265    | 100   | 99,268013  | 100    | 100   | 100   | 100   | 99,739006 0,468  |
| 27 | 2023-07-27 04:10:00 | 134,2218018     | 784,947937      | 31,1940155 | 56,3733333 | 123,512604  | 14      | 2       | 0,10918425        | 55,44429016   | 100   | 99,2586594 | 100    | 100   | 100   | 100   | 99,7366638 0,467 |
| 28 | 2023-07-27 04:20:00 | 121,5214386     | 766,1843872     | 31,1120262 | 56,3693466 | 126,179054  | 14      | 2       | 0,092431501       | 55,28007889   | 100   | 99,2623444 | 100    | 100   | 100   | 100   | 99,731987 0,46   |
| 29 | 2023-07-27 04:30:00 | 124,2973099     | 801,6782227     | 31,0580063 | 56,3400002 | 128,410309  | 14      | 2       | 0,108957529       | 54,84888458   | 100   | 99,2633362 | 100    | 100   | 100   | 100   | 99,7329865 0,464 |
| 30 | 2023-07-27 04:40:00 | 136,2766666     | 771,7310336     | 31,0010061 | 56,3333331 | 136,947307  | 14      | 2       | 0,100731446       | 54,80883484   | 100   | 99,2633362 | 100    | 100   | 100   | 100   | 99,7329865 0,465 |

Figura 25: Exemplo de planilha e dados enviados para elaboração do modelo

A consistência dos dados é fundamental para o sucesso do modelo de previsão proposto. A atenção dedicada à seleção e preparação adequada proporciona um sólido alicerce para as fases posteriores do método. Com essa fase inicial superada, pode-se caminhar para a fase seguinte, que será apresentada no tópico a seguir.

## 4.2. Etapa 2: Análise da base de dados e validação do sentido dos efeitos

### 4.2.1. Tratamento da base de dados

O objetivo dessa etapa foi realizar os ajustes necessários na base de dados disponibilizada, garantindo a qualidade das informações, a partir do tratamento de valores faltantes, outliers e inconsistências que possam prejudicar as análises estatísticas. Considerou-se como pontos relevantes para consistência da base as diretrizes abaixo.

- Formato dos dados: garantiu-se que as colunas contivessem apenas informações do tipo correspondente ao campo a que elas se referem. E informações qualitativas foram convertidas por escala numérica, por exemplo Sim = 1 e Não

= 0.

- Campos em branco: De acordo com o mais pertinente a cada variável, adotou-se um critério para preenchimento de valores faltantes como, média, mediana, valor mais próximo, valor mais frequente ou ainda eliminou-se o registro do banco de dados, caso conveniente.
- Intervalo de variação: garantiu-se que o intervalo de variação histórica da variável fosse consistente, questionando os especialistas sobre o tratamento mais adequado quando identificada inconsistência. Alguns desses tratamentos estão representados na Figura 26, que é uma captura de tela da aplicação MS Excel®

|      | P           | Q           | R           | S                          | T                           | U      | V             | W     | X      | Y    | Z        | AA                  | AB           | AC          |            |
|------|-------------|-------------|-------------|----------------------------|-----------------------------|--------|---------------|-------|--------|------|----------|---------------------|--------------|-------------|------------|
| 1    | LL-69       | RIS-680     | LL-102      | Desemulficante header prod | Desemulficante header teste | online | Limite Online | TOG-F | TOG-IV | TOGG | Descarte | Polieletrólito - PI | Desemul calc | Poliz       |            |
| 9113 | 37,57801437 | 0,463333338 | 100         | 29,45914268                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,167482376  | 84,3844579  | inf        |
| 9114 | 37,58099365 | 0,46233961  | 100         | 28,99334908                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 5,797465324  | 83,28633585 | inf        |
| 9115 | 32,86000061 | 0,464041322 | 100         | 29,32935333                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 5,914168835  | 84,41324857 | inf        |
| 9116 | 32,86000061 | 0,453333348 | 100         | 29,20287895                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 5,9903965    | 83,41437387 | inf        |
| 9117 | 32,86234283 | 0,460000008 | 100         | 29,31799507                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,059687138  | 83,93181037 | inf        |
| 9118 | 32,86333466 | 0,45801422  | 100         | 29,41775703                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,138762951  | 83,99000919 | inf        |
| 9119 | 32,86333466 | 0,456311673 | 100         | 29,08451271                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,18401432   | 83,21728642 | inf        |
| 9120 | 9,515694618 | 0,460994422 | 100         | 29,03199959                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,358736038  | 83,04338139 | inf        |
| 9121 | 0,514326096 | 0,462340564 | 59,97773361 | 28,91669464                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,501935005  | 82,7780535  | inf        |
| 9122 | 0,519006252 | 0,464327097 | 48,65241241 | 14,92234802                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 6,177360535  | 42,95875928 | inf        |
| 9123 | 0,516666651 | 0,463333338 | 25,04999924 | 14,98274899                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014335945  | 43,1167187  | 4,55407688 |
| 9124 | 0,513333321 | 0,462351799 | 0,931370258 | 15,12325096                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,015        | 42,7129655  | inf        |
| 9125 | 0,515673995 | 0,457304001 | 0,927659333 | 14,99887466                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,013333334  | 42,79538844 | inf        |
| 9126 | 0,514325976 | 0,462340683 | 0,930992663 | 15,009161                  |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,015        | 43,8705616  | inf        |
| 9127 | 0,514326215 | 0,463333338 | 0,930000007 | 15,20770168                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,01433689   | 43,18448753 | inf        |
| 9128 | 0,514328122 | 0,46099478  | 6,887661457 | 15,06887531                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,015        | 42,54908939 | inf        |
| 9129 | 0,513333321 | 0,462355763 | 8,032259941 | 14,99206352                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014335967  | 43,22541357 | inf        |
| 9130 | 0,515669823 | 0,460000008 | 20,90333366 | 15,08539295                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,013997222  | 43,35695172 | inf        |
| 9131 | 0,515673995 | 0,460992664 | 22,0143261  | 15,14676094                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014343296  | 43,70339176 | inf        |
| 9132 | 0,513333321 | 0,457015991 | 0,873333335 | 15,12755203                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,013997333  | 43,71150219 | inf        |
| 9133 | 0,516666651 | 0,455318898 | 5,013471603 | 14,92477417                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,013333334  | 41,02419778 | inf        |
| 9134 | 0,513333321 | 0,458014667 | 20,16666603 | 14,98395634                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,015        | 43,2690416  | inf        |
| 9135 | 0,514327407 | 0,44865489  | 24,90432739 | 14,99433613                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014336279  | 43,39893533 | inf        |
| 9136 | 0,519009113 | 0,455315113 | 33,02567673 | 15,12339687                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,015        | 41,69831241 | inf        |
| 9137 | 0,516666651 | 0,460355341 | 34,86099243 | 15,06316376                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014337     | 43,96212344 | 18,212008  |
| 9138 | 0,516666651 | 0,452340662 | 34,86000061 | 14,92886162                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,013333334  | 41,86499035 | inf        |
| 9139 | 0,514326036 | 0,453333348 | 34,86333466 | 15,01011467                |                             | 0      | 0             | 57    | 12     | 30   | 16       | Fechada             | 0,014336967  | 42,40849938 | inf        |

Figura 26: Exemplos de pontos inconsistentes na base de origem

Para a construção da base de dados, referente à plataforma em questão, foram identificados e eliminados os registros que continham ao menos um campo vazio, convertidas colunas de dados nominais em escala numérica (localizar e substituir no MS Excel®), identificados outliers (boxplot no Minitab®) e tratados de acordo com orientação dos especialistas.

#### 4.2.2. Validação do sentido dos efeitos principais

A partir da seleção de variáveis da etapa anterior, foi possível modelar a relação entre elas (aqui denominadas preditores) e a resposta de interesse TOGG, utilizando a técnica de regressão. Para aplicação dessa técnica utilizou-se o software Minitab®.

Foram utilizados dois recursos nesse processo, sendo o primeiro deles o “Ajuste de Modelo de Regressão”. Essa ferramenta permite gerar um modelo matemático para indicar o

comportamento do TOGG para os valores históricos das variáveis, além de permitir incluir termos polinomiais e de interação entre os fatores ou mesmo transformar a resposta, se necessário, e oferece parâmetros valiosos ( $R^2$ ,  $R^2_{adj}$ , valor-p entre outros) para avaliação do modelo. Em termos práticos após acessar o *Minitab*® e importar a base de dados, basta acessar a sequência, *Estat* > *Regressão* > *Regressão* > *Ajuste de Modelo de Regressão*, como indicado na Figura 27.

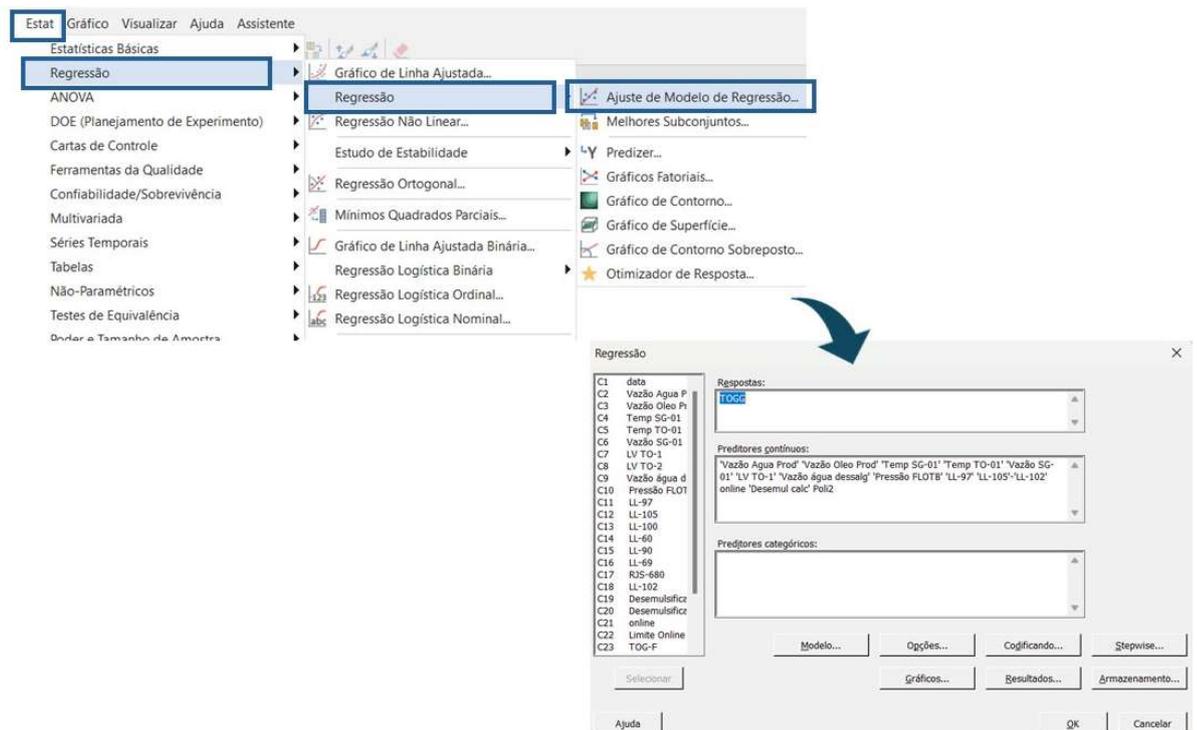


Figura 27: Passas para gerar a análise de regressão

O segundo recurso utilizado foi o de “*Gráficos Fatoriais*”. Esse recurso plota as médias ajustadas ou as médias dos dados, para examinar como os fatores influenciam a resposta e se as variáveis interagem. No *Minitab*® o caminho para gerá-lo é semelhante ao anteriormente

citado, *Estat > Regressão > Regressão > Gráficos fatoriais*, como indicado na Figura 28.

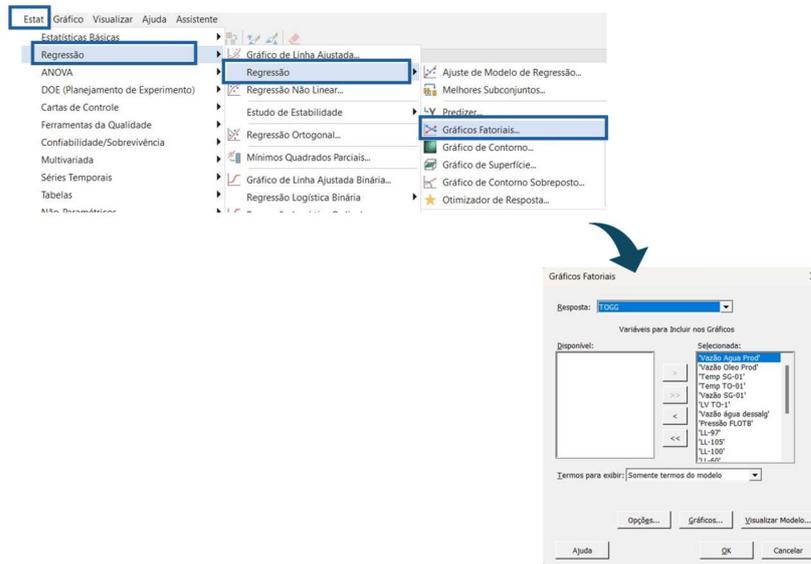


Figura 28: Passos para gerar a análise de gráficos fatoriais

Gerando-se os relatórios explicitados acima, informações relevantes para o modelo foram obtidas, destacando-se para o trabalho proposto as descritas a seguir.

Em relação à equação da regressão e seus coeficientes a Figura 29 forneceu informações cruciais sobre a influência de cada variável independente na variável dependente (TOGG).

#### Equação de Regressão

TOGG = -1,94 - 0,00361 Vazão Água Prod - 0,01362 Vazão Oleo Prod + 0,1094 Temp SG-01 - 0,1762 Temp TO-01 - 0,00672 Vazão SG-01 + 0,00822 LV TO-1 + 0,3097 Vazão água dessalg + 0,04828 Pressão FLOTB - 0,00045 LL-97 - 0,00658 LL-105 - 0,00872 LL-100 - 0,00501 LL-60 + 0,01748 LL-90 + 0,02662 LL-69 + 69,45 RJS-680 - 0,01842 LL-102 + 0,07272 online - 0,04179 Desemul calc + 0,000161 Poli2

#### Coeficientes

| Termo              | Coef     | EP de Coef | Valor-T | Valor-P | VIF  |
|--------------------|----------|------------|---------|---------|------|
| Constante          | -1,94    | 4,42       | -0,44   | 0,660   |      |
| Vazão Água Prod    | -0,00361 | 0,00202    | -1,79   | 0,074   | 1,81 |
| Vazão Oleo Prod    | -0,01362 | 0,00107    | -12,75  | 0,000   | 3,52 |
| Temp SG-01         | 0,1094   | 0,0298     | 3,67    | 0,000   | 2,79 |
| Temp TO-01         | -0,1762  | 0,0191     | -9,22   | 0,000   | 1,27 |
| Vazão SG-01        | -0,00672 | 0,00399    | -1,68   | 0,092   | 3,46 |
| LV TO-1            | 0,00822  | 0,00793    | 1,04    | 0,300   | 1,67 |
| Vazão água dessalg | 0,3097   | 0,0624     | 4,96    | 0,000   | 1,01 |
| Pressão FLOTB      | 0,04828  | 0,00492    | 9,81    | 0,000   | 1,77 |
| LL-97              | -0,00045 | 0,00613    | -0,07   | 0,941   | 1,73 |
| LL-105             | -0,00658 | 0,00620    | -1,06   | 0,289   | 8,86 |
| LL-100             | -0,00872 | 0,00270    | -3,23   | 0,001   | 3,30 |
| LL-60              | -0,00501 | 0,00327    | -1,53   | 0,126   | 3,41 |
| LL-90              | 0,01748  | 0,00455    | 3,84    | 0,000   | 5,22 |
| LL-69              | 0,02662  | 0,00421    | 6,32    | 0,000   | 4,97 |
| RJS-680            | 69,45    | 9,33       | 7,44    | 0,000   | 1,03 |
| LL-102             | -0,01842 | 0,00360    | -5,12   | 0,000   | 1,97 |
| online             | 0,07272  | 0,00339    | 21,46   | 0,000   | 1,26 |
| Desemul calc       | -0,04179 | 0,00423    | -9,88   | 0,000   | 1,64 |
| Poli2              | 0,000161 | 0,000080   | 2,02    | 0,044   | 1,09 |

Figura 29: Relação de influência das variáveis independente no TOGG

- **Vazão Água Prod:** Coeficiente -0,00361 com valor-p de 0,074, mostrando uma influência negativa não muito significativa.
- **Vazão Óleo Prod:** Coeficiente -0,01362 com valor-p de 0,000, mostrando uma influência negativa e significativa.
- **Temp SG-01:** Coeficiente 0,1094 com valor-p de 0,000, indicando uma influência positiva e significativa.
- **Temp TO-01:** Coeficiente -0,1762 com valor-p de 0,000, mostrando uma forte influência negativa e significativa.
- **Vazão SG-01:** Coeficiente -0,00672 com valor-p de 0,092, mostrando uma influência negativa não muito significativa.
- **LV TO-1:** Coeficiente 0,00822 com valor-p de 0,300, não significativo.
- **Vazão água dessalg:** Coeficiente 0,3097 com valor-p de 0,000, mostrando uma forte influência positiva e significativa.
- **Pressão FLOTB:** Coeficiente 0,04828 com valor-p de 0,000, indicando uma forte influência positiva e significativa.
- **LL-97:** Coeficiente -0,00045 com valor-p de 0,941, não significativo.
- **LL-105:** Coeficiente -0,00658 com valor-p de 0,289, não significativo.
- **LL-100:** Coeficiente -0,00872 com valor-p de 0,001, indicando não significativa.
- **LL-60:** Coeficiente -0,00501 com valor-p de 0,126, não significativo.
- **LL-90:** Coeficiente 0,01748 com valor-p de 0,000, indicando uma influência positiva significativa.
- **LL-69:** Coeficiente 0,02662 com valor-p de 0,000, mostrando uma forte influência positiva e significativa.
- **RJS-680:** Coeficiente 69,45 com valor-p de 0,000, indicando uma influência positiva significativa.
- **LL-102:** Coeficiente -0,01842 com valor-p de 0,000, mostrando uma influência negativa significativa.
- **online:** Coeficiente 0,07272 com valor-p de 0,000, indicando uma forte influência positiva e significativa.
- **Desemul calc:** Coeficiente -0,04179 com valor-p de 0,000, mostrando uma forte influência negativa e significativa.
- **Poli2:** Coeficiente 0,000161 com valor-p de 0,044, indicando uma influência

positiva significativa.

É importante mencionar que para algumas variáveis, mesmo a regressão indicando que não havia significância (considerando o recorte histórico), foi solicitado pelos especialistas que estas determinadas variáveis fossem mantidas.

No *Minitab*®, gerou-se o gráfico de efeitos principais, apresentado na Figura 30, cujo objetivo é identificar como a mudança nos níveis dos parâmetros de processo afetam a resposta média do TOGG, permitindo uma interpretação visual fácil dos efeitos individuais de cada fator para o conjunto de dados.

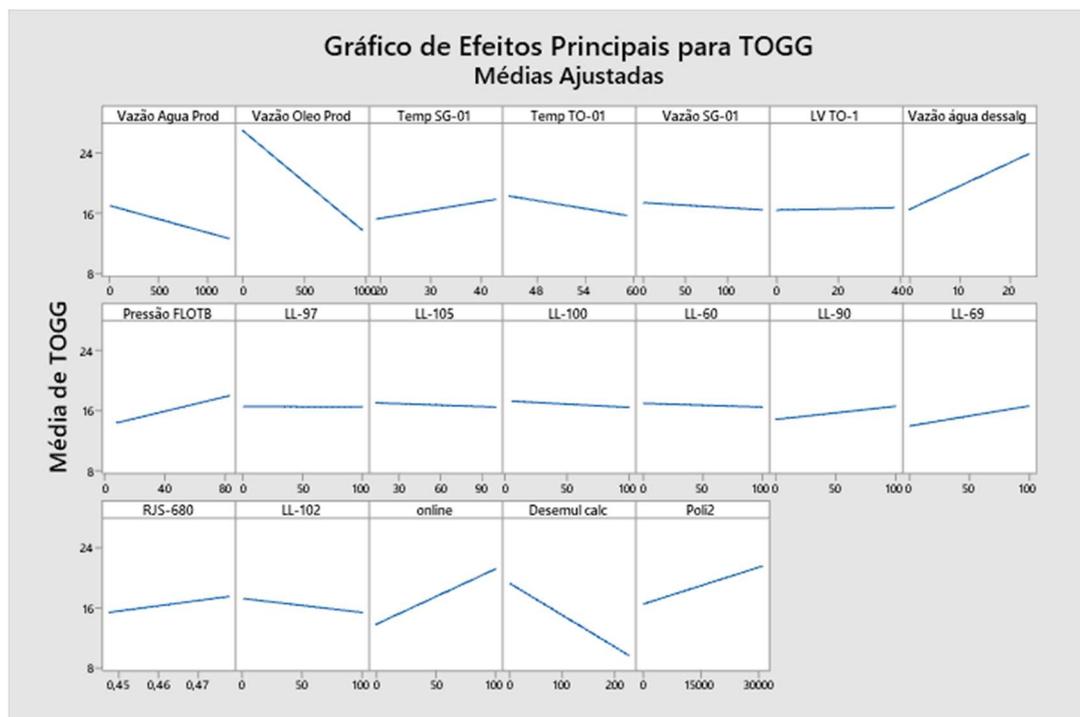


Figura 30: Exemplo de gráfico de efeitos principais das variáveis envolvidas na regressão gerada no *Minitab*®

O gráfico corrobora as informações obtidas pela regressão, visto que as linhas horizontais significam que o parâmetro em questão não tem um efeito significativo sobre o TOGG. Diferentemente das linhas inclinadas que indicam o efeito da variável de processo sobre o TOGG, onde a inclinação representa a magnitude desse efeito. Por exemplo: quando a variável “Desemulsificante Calculado” diminui, o valor de “TOGG” aumenta, indicando o sentido dessa variável.

A Análise de Variância (ANOVA) é uma técnica estatística fundamental para avaliar a

significância global de modelos de regressão e o impacto individual de suas variáveis independentes. Ao analisar a tabela de resumo provenientes desta análise, conforme mostrado na Figura 31, observamos um Valor F de regressão significativo de 92,39, com um valor-p associado de 0,000. Esses resultados indicam que o modelo de regressão é altamente significativo. Um valor-p tão baixo quanto 0,000 sugere que a probabilidade de os resultados serem devidos ao acaso é extremamente baixa, validando assim a robustez estatística do modelo proposto.

Além disso, ao analisar os valores-p individuais das variáveis independentes no modelo, observamos que várias delas são estatisticamente significativas. Em particular, “Vazão Óleo Prod”, “Temp SG-01”, “Temp TO-01”, “Vazão água dessalg”, “Pressão FLOTB”, “LL-100”, “LL-90”, “LL-69”, “RJS-680”, “LL-102”, “online”, “Desemul calc” e “Poli2” têm valores-p muito baixos. Isso indica que essas variáveis são significativamente diferentes de zero e têm um impacto estatisticamente significativo sobre a variável dependente (TOGG).

Portanto, com base na ANOVA, foi possível concluir que o modelo de regressão é globalmente significativo, e muitos dos coeficientes individuais das variáveis explicativas são estatisticamente robustos, contribuindo para explicar a variabilidade na variável de interesse na planta de indústria petroquímica estudada.

#### Análise de Variância

| Fonte              | GL    | SQ (Aj.) | QM (Aj.) | Valor F | Valor-P |
|--------------------|-------|----------|----------|---------|---------|
| Regressão          | 19    | 23418    | 1232,52  | 92,39   | 0,000   |
| Vazão Agua Prod    | 1     | 43       | 42,60    | 3,19    | 0,074   |
| Vazão Oleo Prod    | 1     | 2167     | 2167,05  | 162,45  | 0,000   |
| Temp SG-01         | 1     | 179      | 179,34   | 13,44   | 0,000   |
| Temp TO-01         | 1     | 1134     | 1133,78  | 84,99   | 0,000   |
| Vazão SG-01        | 1     | 38       | 37,81    | 2,83    | 0,092   |
| LV TO-1            | 1     | 14       | 14,34    | 1,07    | 0,300   |
| Vazão água dessalg | 1     | 329      | 328,69   | 24,64   | 0,000   |
| Pressão FLOTB      | 1     | 1285     | 1284,94  | 96,32   | 0,000   |
| LL-97              | 1     | 0        | 0,07     | 0,01    | 0,941   |
| LL-105             | 1     | 15       | 15,02    | 1,13    | 0,289   |
| LL-100             | 1     | 139      | 138,93   | 10,41   | 0,001   |
| LL-60              | 1     | 31       | 31,29    | 2,35    | 0,126   |
| LL-90              | 1     | 197      | 197,13   | 14,78   | 0,000   |
| LL-69              | 1     | 532      | 532,41   | 39,91   | 0,000   |
| RJS-680            | 1     | 739      | 739,01   | 55,40   | 0,000   |
| LL-102             | 1     | 350      | 350,15   | 26,25   | 0,000   |
| online             | 1     | 6145     | 6145,01  | 460,65  | 0,000   |
| Desemul calc       | 1     | 1301     | 1300,88  | 97,52   | 0,000   |
| Poli2              | 1     | 54       | 54,20    | 4,06    | 0,044   |
| Erro               | 10849 | 144725   | 13,34    |         |         |
| Total              | 10868 | 168143   |          |         |         |

Figura 31: Análise de variância

Dando continuidade à análise, obteve-se também o gráfico de Pareto, Figura 32, que

mostra os valores absolutos dos efeitos padronizados desde o maior efeito até o menor efeito. Os efeitos padronizados são estatísticas t que testam a hipótese nula de que o efeito é 0. O gráfico também traça uma linha de referência para indicar quais efeitos são estatisticamente significativos, no caso em questão essa linha está em 1,96. Ou seja, todas as barras superiores a esse valor são fatores significativos.

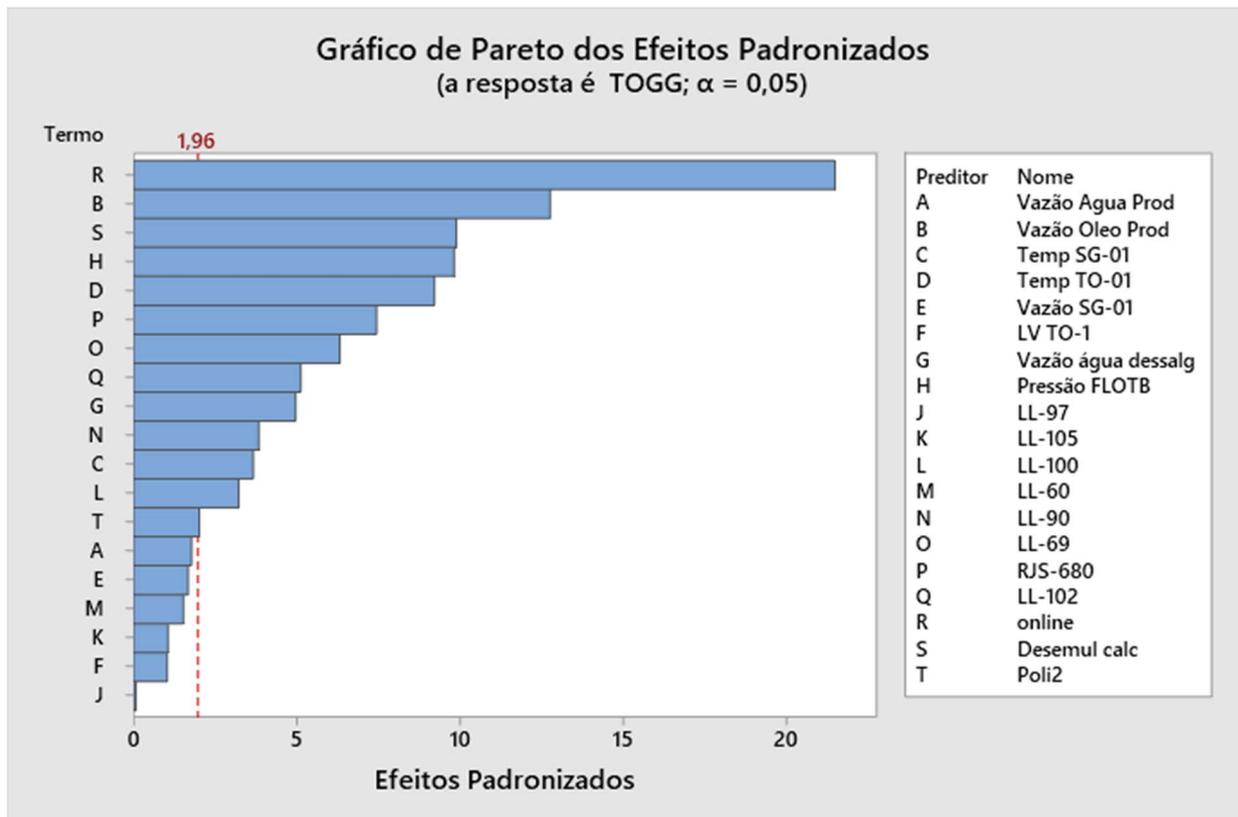


Figura 32: Gráfico de Pareto dos efeitos padronizados

Em posse dessas informações, a validação foi realizada a partir da expertise dos especialistas envolvidos no processo. Como o sentido das variáveis apresentaram-se coerentes, seguiu-se para a próxima etapa. Se uma ou mais variáveis apresentassem sentido oposto ao esperado, a metodologia adotada propõe voltar à etapa anterior e reavaliar o conjunto de preditores.

Como já mencionado, para a execução dessa etapa, utilizaram-se os pacotes estatísticos voltados à Regressão do software *Minitab*®, mas alternativas poderiam ser adotadas como por exemplo a linguagem de programação *Python*®, através do pacote *Sklearn*, capaz de produzir os mesmos resultados e gráficos necessários para o diagnóstico e a tomada de decisão desta

etapa.

#### 4.2.3. Etapa 3: Definição dos limites e registros fiscais

Os registros de horário fiscal foram selecionados de acordo com a indicação de coleta fornecida pelo laboratório que realiza as análises de TOGG, visto que os horários de coleta e a quantidade de material são meticulosamente controlados e registrados. Desta forma, é possível acessar possíveis variações em torno dos horários fiscais determinados para a plataforma.

Por padrão, para uma análise diária de TOG gravimétrico são analisados 1 L de água produzida, coletados em quatro parcelas de 250 mL, em 4 horários definidos e espaçados igualmente ao longo do dia. A Figura 33 representa um parâmetro da base de dados e sua quantidade de registros diária. É fácil perceber que apenas quatro desses pontos são correspondentes a horários em que se ocorre a amostragem do TOG gravimétrico. Nessa fase a base de dados foi saneada de modo que só fossem mantidos os 4 registros diários, sendo estes relacionados aos horários fiscais.

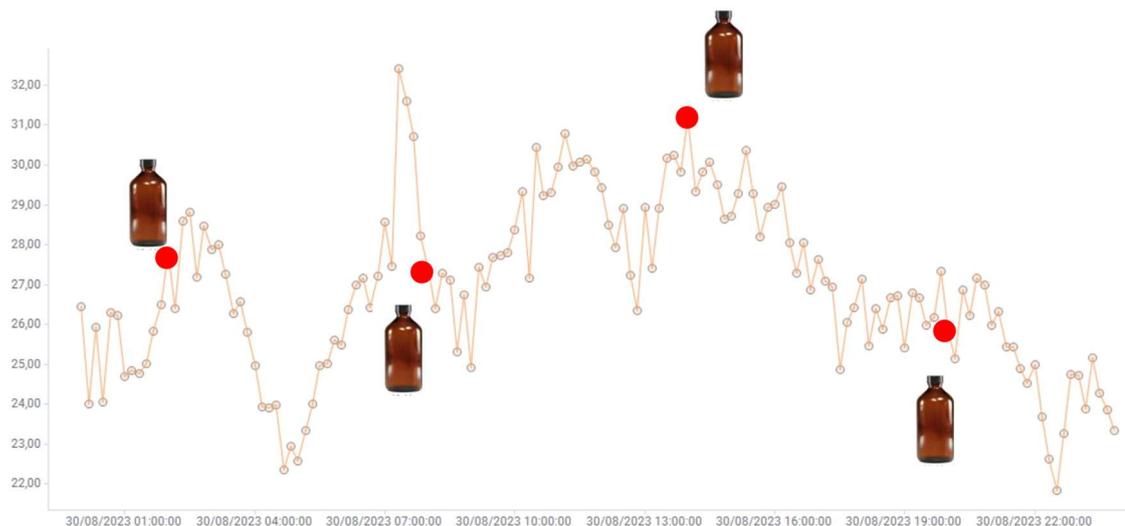


Figura 33: Coletas realizadas x registros da base

A partir dos dados de TOGG selecionados, adotou-se um critério de classificação para enquadramento/desenquadramento a partir de um valor limite estabelecido, sua escolha foi relativamente simples, uma vez que o valor de 29 mg/L é estabelecido pelo CONAMA 393.

A seleção dos horários fiscais foi realizada no MS Excel® através das funções de filtro aplicadas às informações de data, hora e minuto extraídas por função do Minitab®, conforme

indicado nas Figuras 34 e 35.

|    | C1-D                | C2       | C3 | C4 | C5          | C6     | C7      | C8      | C9      | C10     | C11     | C12 |
|----|---------------------|----------|----|----|-------------|--------|---------|---------|---------|---------|---------|-----|
| 1  | 2023-07-27 00:00:00 | 20230727 | 0  | 0  | 2.02307E+11 | 123.20 | 759,876 | 31,7627 | 56,0667 | 126,086 | 14,0000 | 2,0 |
| 2  | 2023-07-27 00:10:00 | 20230727 | 0  | 10 | 2.02307E+11 | 128,93 | 782,074 | 31,7513 | 56,0467 | 124,854 | 14,0000 | 2,0 |
| 3  | 2023-07-27 00:20:00 | 20230727 | 0  | 20 | 2.02307E+11 | 120,62 | 807,083 | 31,7627 | 56,0467 | 128,654 | 14,0000 | 2,0 |
| 4  | 2023-07-27 00:30:00 | 20230727 | 0  | 30 | 2.02307E+11 | 123,45 | 772,618 | 31,7580 | 56,0667 | 127,031 | 14,0000 | 2,0 |
| 5  | 2023-07-27 00:40:00 | 20230727 | 0  | 40 | 2.02307E+11 | 128,76 | 796,196 | 31,6973 | 56,0880 | 124,606 | 14,0000 | 2,0 |
| 6  | 2023-07-27 00:50:00 | 20230727 | 0  | 50 | 2.02307E+11 | 123,76 | 796,173 | 31,6687 | 56,0973 | 125,869 | 14,0000 | 2,0 |
| 7  | 2023-07-27 01:00:00 | 20230727 | 1  | 0  | 2.02307E+11 | 123,01 | 774,985 | 31,6539 | 56,0733 | 127,839 | 14,0000 | 2,0 |
| 8  | 2023-07-27 01:10:00 | 20230727 | 1  | 10 | 2.02307E+11 | 127,53 | 802,690 | 31,6680 | 56,0733 | 126,532 | 14,0000 | 2,0 |
| 9  | 2023-07-27 01:20:00 | 20230727 | 1  | 20 | 2.02307E+11 | 124,82 | 774,093 | 31,6693 | 56,0773 | 127,391 | 14,0000 | 2,0 |
| 10 | 2023-07-27 01:30:00 | 20230727 | 1  | 30 | 2.02307E+11 | 123,54 | 785,099 | 31,6913 | 56,0800 | 126,607 | 14,0000 | 2,0 |

Figura 34: Exemplo de extração dos dados de data, hora e minuto no Minitab®

|     | A                   | B        | C    | D   | E             | F             | G           | H           | I     |
|-----|---------------------|----------|------|-----|---------------|---------------|-------------|-------------|-------|
| 1   |                     | data     | hora | min | Vazão Agua Pr | Vazão Oleo Pr | Temp SG-C   | Temp TO-C   | Vazão |
| 134 | 2023-07-27 22:00:00 | 20230727 | 22   | 0   | 120,2542953   | 717,9547119   | 34,28285599 | 56,10936737 | 126,  |
| 140 | 2023-07-27 23:00:00 | 20230727 | 23   | 0   | 121,1081009   | 555,4227905   | 31,84261131 | 55,92666626 | 124,6 |
| 146 | 2023-07-28 00:00:00 | 20230728 | 0    | 0   | 125,0952377   | 613,7435913   | 32,88866425 | 56,31333316 | 127,2 |
| 152 | 2023-07-28 01:00:00 | 20230728 | 1    | 0   | 126,1006165   | 572,3905029   | 33,23262405 | 55,42269516 | 126,3 |
| 158 | 2023-07-28 02:00:00 | 20230728 | 2    | 0   | 124,8169632   | 583,3826904   | 33,43068314 | 53,25469971 | 126,7 |

Figura 35: Seleção de registros associados aos horários fiscais no MS Excel®

A alocação dos valores de TOGG de acordo com informação fornecida pelo laboratório de aferição foi realizada a partir da extração de datas no Minitab® e subsequente aplicação da

função “procv” no MS Excel® (Figura 36).

|    | A                   | AC       | AD             | AE | AF       | AG               | AH | AI                                    | AJ            | AK              | AL |
|----|---------------------|----------|----------------|----|----------|------------------|----|---------------------------------------|---------------|-----------------|----|
| 1  |                     | data     | TOGG Petrobras |    | data     | Gravimétrico Lab |    |                                       | TOGG Validado | Inconsistência? |    |
| 2  | 2023-07-27 00:00:00 | 20230727 | 12             |    | 20230701 | 40               |    | 2023=PROCV(AC2;SAF51:SAG5103;2;FALSO) |               |                 |    |
| 3  | 2023-07-27 00:10:00 | 20230727 | 12             |    | 20230702 | 19               |    | 2023-07-27 00:10:00                   | 24            | SIM             |    |
| 4  | 2023-07-27 00:20:00 | 20230727 | 12             |    | 20230703 | 20               |    | 2023-07-27 00:20:00                   | 24            | SIM             |    |
| 5  | 2023-07-27 00:30:00 | 20230727 | 12             |    | 20230704 | 18               |    | 2023-07-27 00:30:00                   | 24            | SIM             |    |
| 6  | 2023-07-27 00:40:00 | 20230727 | 12             |    | 20230705 | 15               |    | 2023-07-27 00:40:00                   | 24            | SIM             |    |
| 7  | 2023-07-27 00:50:00 | 20230727 | 12             |    | 20230706 | 16               |    | 2023-07-27 00:50:00                   | 24            | SIM             |    |
| 8  | 2023-07-27 01:00:00 | 20230727 | 12             |    | 20230707 | 14               |    | 2023-07-27 01:00:00                   | 24            | SIM             |    |
| 9  | 2023-07-27 01:10:00 | 20230727 | 12             |    | 20230708 | 15               |    | 2023-07-27 01:10:00                   | 24            | SIM             |    |
| 10 | 2023-07-27 01:20:00 | 20230727 | 12             |    | 20230709 | 16               |    | 2023-07-27 01:20:00                   | 24            | SIM             |    |
| 11 | 2023-07-27 01:30:00 | 20230727 | 24             |    | 20230710 | 22               |    | 2023-07-27 01:30:00                   | 24            |                 |    |

|    | A                   | AC       | AD             | AE | AF       | AG               | AH | AI                  | AJ                    | AK              | AL |
|----|---------------------|----------|----------------|----|----------|------------------|----|---------------------|-----------------------|-----------------|----|
| 1  |                     | data     | TOGG Petrobras |    | data     | Gravimétrico Lab |    |                     | TOGG Validado         | Inconsistência? |    |
| 2  | 2023-07-27 00:00:00 | 20230727 | 12             |    | 20230701 | 40               |    | 2023-07-27 00:00:00 | 24                    | SIM             |    |
| 3  | 2023-07-27 00:10:00 | 20230727 | 12             |    | 20230702 | 19               |    | 2023-07-27 00:10:00 | 24                    | SIM             |    |
| 4  | 2023-07-27 00:20:00 | 20230727 | 12             |    | 20230703 | 20               |    | 2023-07-27 00:20:00 | 24                    | SIM             |    |
| 5  | 2023-07-27 00:30:00 | 20230727 | 12             |    | 20230704 | 18               |    | 2023-07-27 00:30:00 | 24                    | SIM             |    |
| 6  | 2023-07-27 00:40:00 | 20230727 | 12             |    | 20230705 | 15               |    | 2023-07-27 00:40:00 | 24                    | SIM             |    |
| 7  | 2023-07-27 00:50:00 | 20230727 | 12             |    | 20230706 | 16               |    | 2023-07-27 00:50:00 | 24                    | SIM             |    |
| 8  | 2023-07-27 01:00:00 | 20230727 | 12             |    | 20230707 | 14               |    | 2023-07-27 01:00:00 | 24                    | SIM             |    |
| 9  | 2023-07-27 01:10:00 | 20230727 | 12             |    | 20230708 | 15               |    | 2023-07-27 01:10:00 | 24                    | SIM             |    |
| 10 | 2023-07-27 01:20:00 | 20230727 | 12             |    | 20230709 | 16               |    | 2023-07-27 01:20:00 | =SE(AJ10=AD10;"SIM";) |                 |    |
| 11 | 2023-07-27 01:30:00 | 20230727 | 24             |    | 20230710 | 22               |    | 2023-07-27 01:30:00 | 24                    |                 |    |

Figura 36: Exemplo de validação do TOGG com os dados do laboratório realizados no MS Excel®

A definição dos horários fiscais é crucial para o planejamento da base de dados, visto que a medição diária do TOGG é composta por frações coletadas ao longo do dia. Assim, torna-se possível aproximar o comportamento do processo aos momentos de coleta.

#### 4.2.4. Etapa 4: Criação das classes e balanceamento da base de dados

Com a base de dados e valores de TOGG ajustados, criaram-se classes binárias considerando o limiar definido de 29 mg/L. Em termos práticos, para um registro de TOG gravimétrico igual a 31 mg/L, por exemplo, considerou-se como sendo da classe 1 (um), ou seja, desenquadrado; já um registro associado a um valor de TOGG de 15mg/L, por sua vez, seria de classe 0 (zero), ou seja, enquadrado.

Posteriormente, verificou-se, pelo total de registros, qual o percentual de ocorrências de enquadramento e de desenquadramento. Em um cenário ideal para modelos de previsão, os dados devem estar balanceados, o que significa ter 50% de ocorrências para enquadramentos e 50% para desenquadramento. Outras composições razoáveis, seriam de 60%–40% ou 40%–60%. Caso o comportamento das classes apresente esses percentuais, tem-se uma base de dados aceitável para a modelagem de previsão e pode-se seguir para a etapa final, ou seja, iniciar a criação do modelo por Redes Neurais Artificiais.

Os valores das classes do TOGG foram inseridos na base a partir da classificação com valor 0 para enquadramento e 1 para desenquadramento a partir do limiar estabelecido, através

da função “se” do MS Excel® (Figura 37). Da mesma forma, é possível criar a classificação através do operador lógico “if” disponível nas linguagens de programação, como o Python®.

|    | W                             | X           | Y             | Z     | AA     | AB       | AC                 | AD           | AE          | AF       | AG             | AH                  | AI | AJ      | AK | AL |
|----|-------------------------------|-------------|---------------|-------|--------|----------|--------------------|--------------|-------------|----------|----------------|---------------------|----|---------|----|----|
| 1  | Desemulsificante header teste | online      | Limite Online | TOG-F | TOG-IV | Descarte | Poleietrolito - PI | Desemul calc | Poliz2      | data     | TOGG Petrobras | Classe              |    | Limiar: | 29 |    |
| 2  | 4,124047279                   | 42,93493652 | 57            | 7     | 42     | Aberta   | 6,411001682        | 61,25135048  | 52,03730427 | 20230727 |                | =SE(AG2<=SAKS1;0;1) |    |         |    |    |
| 3  | 4,114683628                   | 33,04070282 | 57            | 13    | 33     | Aberta   | 6,507054806        | 58,56746751  | 54,8673874  | 20230727 | 24             | 0                   |    |         |    |    |
| 4  | 4,286746979                   | 19,86325073 | 57            | 1     | 20     | Aberta   | 7,621666431        | 61,31342162  | 61,23218919 | 20230727 | 24             | 0                   |    |         |    |    |
| 5  | 4,155200005                   | 34,28854752 | 57            | 1     | 32     | Fechada  | 0,017005334        | 108,2517922  | 0,702463749 | 20230727 | 24             | 0                   |    |         |    |    |
| 6  | 4,043897152                   | 30,51357651 | 57            | 5     | 31     | Fechada  | 6,623670578        | 74,86188481  | 52,94902267 | 20230728 | 24             | 0                   |    |         |    |    |
| 7  | 4,230927467                   | 31,39205933 | 57            | 10    | 35     | Aberta   | 6,242344856        | 58,63148779  | 48,93639899 | 20230728 | 23             | 0                   |    |         |    |    |
| 8  | 3,850931644                   | 41,26508331 | 57            | 10    | 35     | Aberta   | 6,463673592        | 58,87468928  | 45,25771186 | 20230728 | 23             | 0                   |    |         |    |    |
| 9  | 3,984223843                   | 33,56373215 | 57            | 10    | 65     | Aberta   | 6,073691368        | 59,78556329  | 48,76786225 | 20230728 | 23             | 0                   |    |         |    |    |
| 10 | 3,808456659                   | 44,69966507 | 57            | 5     | 26     | Aberta   | 1,773516655        | 58,64373446  | 13,65269013 | 20230729 | 23             | 0                   |    |         |    |    |
| 11 | 3,421874523                   | 40,31859207 | 57            | 6     | 41     | Aberta   | 6,387814522        | 58,79022087  | 48,66593993 | 20230729 | 19             | 0                   |    |         |    |    |
| 12 | 3,558600903                   | 45,79532242 | 57            | 16    | 38     | Fechada  | 3,631176949        | 59,59176491  | 31,64223443 | 20230729 | 19             | 0                   |    |         |    |    |
| 13 | 3,527447462                   | 35,51942825 | 57            | 12    | 44     | Aberta   | 3,51148057         | 60,25534503  | 28,63104807 | 20230729 | 19             | 0                   |    |         |    |    |
| 14 | 0                             | 38,54107285 | 57            | 7     | 30     | Aberta   | 6,571243763        | 69,33396251  | 49,90559331 | 20230730 | 19             | 0                   |    |         |    |    |
| 15 | 0                             | 61,26160812 | 57            | 11    | 52     | Fechada  | 3,431875706        | 89,02630541  | 37,29751288 | 20230730 | 31             | 1                   |    |         |    |    |
| 16 | 0                             | 47,49819565 | 57            | 10    | 43     | Fechada  | 5,914574146        | 68,18578501  | 45,5041376  | 20230730 | 31             | 1                   |    |         |    |    |
| 17 | 0                             | 51,37936401 | 57            | 25    | 38     | Fechada  | 7,671773434        | 67,85200512  | 59,70779907 | 20230730 | 31             | 1                   |    |         |    |    |
| 18 | 0                             | 47,98287582 | 57            | 17    | 60     | Fechada  | 8,533581734        | 67,91550696  | 77,22070866 | 20230731 | 31             | 1                   |    |         |    |    |
| 19 | 0                             | 40,16287613 | 57            | 6     | 14     | Aberta   | 9,256997108        | 70,67915256  | 75,37467016 | 20230731 | 14             | 0                   |    |         |    |    |
| 20 | 0                             | 41,26873398 | 57            | 4     | 28     | Aberta   | 8,182412148        | 68,83510726  | 68,91829833 | 20230731 | 14             | 0                   |    |         |    |    |
| 21 | 0                             | 43,89496613 | 57            | 5     | 29     | Aberta   | 8,511476517        | 68,07560433  | 66,24849939 | 20230731 | 14             | 0                   |    |         |    |    |

Figura 37: Criação da classe de enquadramento (0) ou desenquadramento (1) com base no limiar estabelecido

Também no MS Excel® foram contabilizadas as quantidades de ambas as classes, de maneira a avaliar o percentual de composição da base, conforme exemplo da Figura 38.

|    | W                             | X           | Y             | Z     | AA     | AB       | AC                 | AD           | AE          | AF       | AG             | AH     | AI | AJ      | AK | AL | AM | AN | AO | AP |  |
|----|-------------------------------|-------------|---------------|-------|--------|----------|--------------------|--------------|-------------|----------|----------------|--------|----|---------|----|----|----|----|----|----|--|
| 1  | Desemulsificante header teste | online      | Limite Online | TOG-F | TOG-IV | Descarte | Poleietrolito - PI | Desemul calc | Poliz2      | data     | TOGG Petrobras | Classe |    | Limiar: | 29 |    |    |    |    |    |  |
| 2  | 4,124047279                   | 42,93493652 | 57            | 7     | 42     | Aberta   | 6,411001682        | 61,25135048  | 52,03730427 | 20230727 |                |        |    |         |    |    |    |    |    |    |  |
| 3  | 4,114683628                   | 33,04070282 | 57            | 13    | 33     | Aberta   | 6,507054806        | 58,56746751  | 54,8673874  | 20230727 | 24             | 0      |    |         |    |    |    |    |    |    |  |
| 4  | 4,286746979                   | 19,86325073 | 57            | 1     | 20     | Aberta   | 7,621666431        | 61,31342162  | 61,23218919 | 20230727 | 24             | 0      |    |         |    |    |    |    |    |    |  |
| 5  | 4,155200005                   | 34,28854752 | 57            | 1     | 32     | Fechada  | 0,017005334        | 108,2517922  | 0,702463749 | 20230727 | 24             | 0      |    |         |    |    |    |    |    |    |  |
| 6  | 4,043897152                   | 30,51357651 | 57            | 5     | 31     | Fechada  | 6,623670578        | 74,86188481  | 52,94902267 | 20230728 | 24             | 0      |    |         |    |    |    |    |    |    |  |
| 7  | 4,230927467                   | 31,39205933 | 57            | 10    | 35     | Aberta   | 6,242344856        | 58,63148779  | 48,93639899 | 20230728 | 23             | 0      |    |         |    |    |    |    |    |    |  |
| 8  | 3,850931644                   | 41,26508331 | 57            | 10    | 35     | Aberta   | 6,463673592        | 58,87468928  | 45,25771186 | 20230728 | 23             | 0      |    |         |    |    |    |    |    |    |  |
| 9  | 3,984223843                   | 33,56373215 | 57            | 10    | 65     | Aberta   | 6,073691368        | 59,78556329  | 48,76786225 | 20230728 | 23             | 0      |    |         |    |    |    |    |    |    |  |
| 10 | 3,808456659                   | 44,69966507 | 57            | 5     | 26     | Aberta   | 1,773516655        | 58,64373446  | 13,65269013 | 20230729 | 23             | 0      |    |         |    |    |    |    |    |    |  |
| 11 | 3,421874523                   | 40,31859207 | 57            | 6     | 41     | Aberta   | 6,387814522        | 58,79022087  | 48,66593993 | 20230729 | 19             | 0      |    |         |    |    |    |    |    |    |  |
| 12 | 3,558600903                   | 45,79532242 | 57            | 16    | 38     | Fechada  | 3,631176949        | 59,59176491  | 31,64223443 | 20230729 | 19             | 0      |    |         |    |    |    |    |    |    |  |
| 13 | 3,527447462                   | 35,51942825 | 57            | 12    | 44     | Aberta   | 3,51148057         | 60,25534503  | 28,63104807 | 20230729 | 19             | 0      |    |         |    |    |    |    |    |    |  |
| 14 | 0                             | 38,54107285 | 57            | 7     | 30     | Aberta   | 6,571243763        | 69,33396251  | 49,90559331 | 20230730 | 19             | 0      |    |         |    |    |    |    |    |    |  |
| 15 | 0                             | 61,26160812 | 57            | 11    | 52     | Fechada  | 3,431875706        | 89,02630541  | 37,29751288 | 20230730 | 31             | 1      |    |         |    |    |    |    |    |    |  |
| 16 | 0                             | 47,49819565 | 57            | 10    | 43     | Fechada  | 5,914574146        | 68,18578501  | 45,5041376  | 20230730 | 31             | 1      |    |         |    |    |    |    |    |    |  |

| AJ      | AK | AL | AM       | AN                    | AO | AP |
|---------|----|----|----------|-----------------------|----|----|
| Limiar: | 29 |    |          |                       |    |    |
|         |    |    |          | Qtde                  | %  |    |
|         |    |    |          | =CONT.SE(AH1:AH306;0) |    |    |
|         |    |    | Classe 1 | 4                     | 1% |    |

| AJ      | AK | AL | AM       | AN                       | AO  | AP |
|---------|----|----|----------|--------------------------|-----|----|
| Limiar: | 29 |    |          |                          |     |    |
|         |    |    |          | Qtde                     | %   |    |
|         |    |    | Classe 0 | 301                      | 99% |    |
|         |    |    | Classe 1 | =AN4/CONT.NUM(AH2:AH306) |     |    |

Figura 38: Cálculo das quantidades e percentuais das classes na base de dados

Para a base de dados em questão, constatou-se um grande desbalançamento, relação “99%-1%”. Esse alto nível de discrepância prejudica a qualidade da modelagem, sendo

necessário a aplicação de técnicas de balanceamento, delineadas nos passos seguintes.

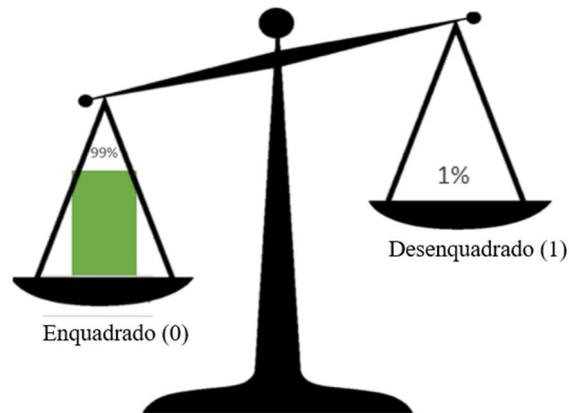


Figura 39: Desbalanceamento da base de dados

Para esse cenário, optou-se pela estratégia de "*Oversampling*", que visa aumentar a base de dados original criando registros sintéticos que imitam o comportamento dos registros reais da classe com menor frequência. Essa abordagem é especialmente promissora quando a classe de interesse possui menos ocorrências, pois mantém todos os registros reais e cria novos registros sintéticos que tendem a se aproximar do comportamento dos registros reais através de interpolação linear, mantendo-os o mais distante possível da fronteira entre as classes. A alternativa de "*Undersampling*" propõe remover registros da classe majoritária até que ambas as classes estejam equilibradas, mas essa técnica reduziria a base de dados para aproximadamente 10 registros, o que não seria ideal para a robustez da análise.

Para prosseguir nessa etapa, utilizou-se, a linguagem *Python*<sup>®</sup> que apresenta bibliotecas e pacotes específicos para essa aplicação. Para a técnica de "*Undersampling*", pode-se aplicar o algoritmo *NearMiss* disponível através da biblioteca *imbalanced-learn*. Para o "*Oversampling*", técnica adotada, utilizou-se a técnica SMOTE da mesma biblioteca, aplicada conforme demonstrado na Figura 40. Como resultado, após aplicação do algoritmo obteve-se uma base perfeitamente balanceada, cada classe passou a contar com 50% de registros, simbolizado na Figura 41. Desse modo, a etapa 4 produziu uma base de dados plenamente

preparada para o desenvolvimento do modelo via redes neurais.

```
import pandas as pd
from imblearn.over_sampling import SMOTE
from imblearn.under_sampling import NearMiss
import random

# Carregue o arquivo XLSX
df = pd.read_excel('dados2023_paraBalanceamento.xlsx')

# Separe as linhas com classe 0 e classe 1
df_class_0 = df[df.iloc[:, -1] == 0]
df_class_1 = df[df.iloc[:, -1] == 1]

# Combine as classes novamente
df_filled = pd.concat([df_class_0, df_class_1])

# Aplicar SMOTE para oversampling da classe 1
smote = SMOTE(sampling_strategy='auto')
X_resampled, y_resampled = smote.fit_resample(df_filled.iloc[:, :-1], df_filled.iloc[:, -1])

# Ajuste manual do random_state para garantir a reprodutibilidade
random.seed(42)

df_balanced_oversampled = pd.concat([pd.DataFrame(X_resampled), pd.DataFrame(y_resampled)], axis=1)
```

Figura 40: Exemplo de código para geração do arquivo balanceado através da técnica de *oversampling* (SMOTE)

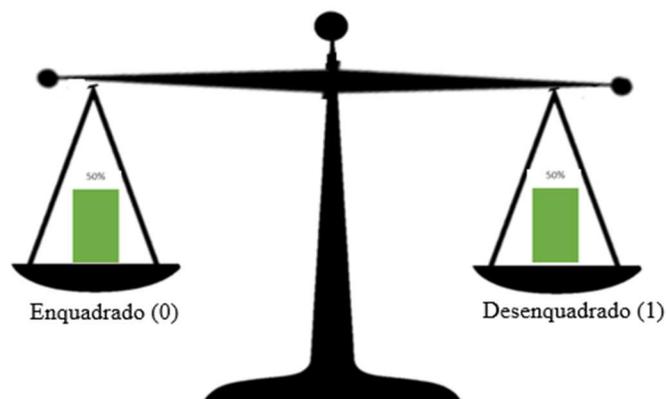


Figura 41: Base de dados balanceada

#### 4.2.5. Etapa 5: Criação de um modelo preditivo através de Redes Neurais Artificiais

Após a definição da base de dados para modelagem, com o procedimento de *oversampling*, o programa *Statistica* foi utilizado para a criação da Rede Neural e seleção do modelo mais adequado aos objetivos do projeto. Os principais passos na definição das redes são apresentados a seguir:

##### 4.2.5.1. Parametrização das Redes Neurais Artificiais

As RNAs requerem um conjunto específico de parâmetros, que precisam ser definidos para sua implementação eficiente em tempo real. Esses parâmetros incluem pesos, biases, a estrutura da rede (número de camadas e neurônios por camada), funções de ativação, entre outros. O treinamento dessas redes pôde ser realizado offline, utilizando-se, para o fim aqui proposto, o conjunto de dados tratados nas etapas anteriores. Esse processo de treinamento pode ser intensivo em termos de tempo e recursos computacionais. No entanto, uma vez que os parâmetros são ajustados e definidos, a rede neural está preparada para uso imediato. Isso significa que ela poderá processar e fornecer resultados rapidamente quando novos dados são inseridos.

Realizaram-se então os seguintes passos nessa parametrização:

- a) Uma planilha em formato *.sta* foi criada, conforme demonstrado na Figura 42. Tal planilha pôde ser facilmente importada de um arquivo MS Excel®, possuindo todas as colunas das variáveis independentes e a variável de resposta de interesse que, neste caso, trata-se da classe do TOGG.

| Vazão Agua Prod | Temp SG-01 | Temp TO-01 | Desemul calc | Poli2     | TOGG | Classe |
|-----------------|------------|------------|--------------|-----------|------|--------|
| 116,244583      | 32,465359  | 58,586666  | 20,917356    | 47,062038 | 18   | 0      |
| 84,4757         | 31,26664   | 58,25732   | 20,568353    | 67,553941 | 18   | 0      |
| 120,231926      | 32,86005   | 58,374729  | 21,040882    | 47,560237 | 18   | 0      |
| 110,022926      | 29,897255  | 55,572941  | 20,885038    | 51,716494 | 18   | 0      |
| 129,543884      | 32,396645  | 57,277325  | 20,6887      | 49,875012 | 8    | 0      |
| 115,912239      | 28,479372  | 57,031967  | 21,29959     | 53,580425 | 8    | 0      |
| 92,550385       | 32,42268   | 57,674637  | 21,74502     | 69,407108 | 8    | 0      |
| 108,605507      | 30,330683  | 57,24268   | 20,662481    | 59,816158 | 8    | 0      |
| 91,665024       | 30,459507  | 56,06284   | 21,710907    | 72,117319 | 7    | 0      |
| 122,69294       | 32,650013  | 55,32267   | 21,401568    | 45,085655 | 7    | 0      |

Figura 42: Planilha no formato compatível com o programa *Statistica*

- b) O módulo de *Automated Neural Networks* (Figura 43) foi aberto para o problema de classificação:

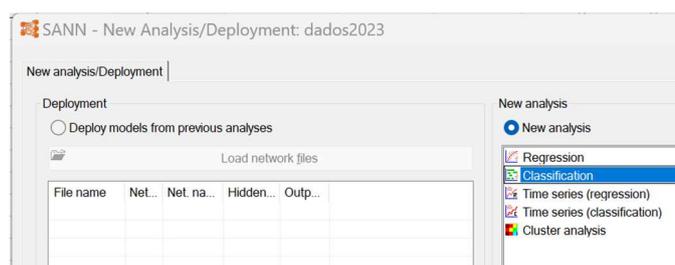


Figura 43 - Módulo *Automated Neural Networks* do *Statistica*

- c) As variáveis do problema de classificação foram selecionadas, como apresentado na Figura 44:

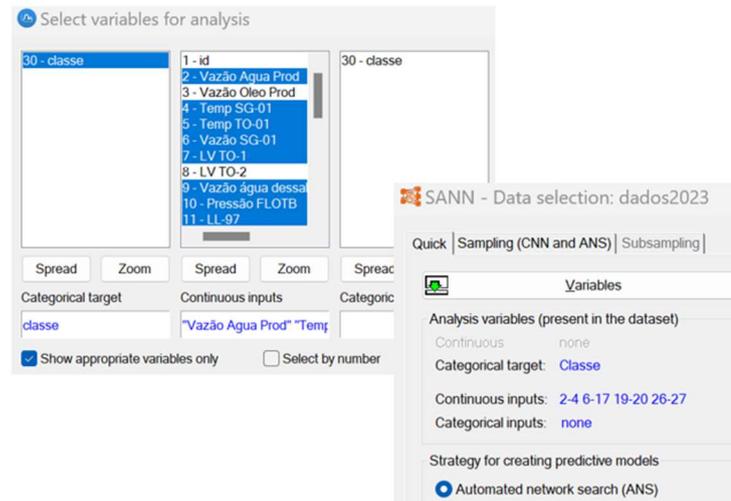


Figura 44 - Seleção das variáveis predictoras definidas junto aos especialistas para previsão do desenquadramento do TOGG

- d) Os parâmetros das redes podem ser agora definidos, na Figura 45 pode-se observar as opções selecionadas e os parâmetros informados para obtenção do modelo.

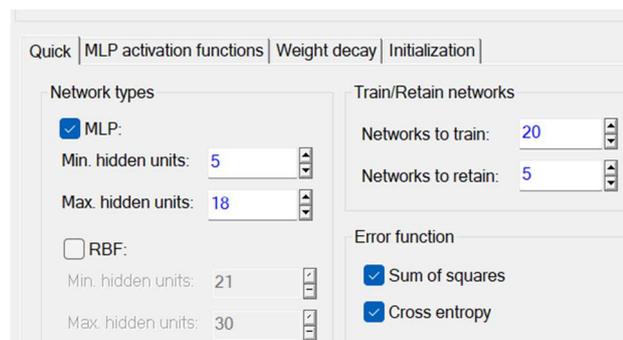


Figura 45: Aba *Quick*

Entre as opções “*MLP*” (*Multilayer Perceptron*) e “*RBF*” (*Radial Basis Function*), elegeu-se a opção “*MLP*”. O número mínimo de neurônios em cada camada oculta da rede foram 5, e o máximo 18, para permitir uma exploração equilibrada entre complexidade e simplicidade, garantindo ao mesmo tempo que as redes neurais tivessem capacidade suficiente para capturar padrões significativos nos dados, sem se tornarem excessivamente complexas e suscetíveis a

*overfitting*. A opção “*RBF*” não é ideal para o conjunto de dados em questão, uma vez que estes não são esfericamente simétricos ao redor dos centros das funções de base radial.

No campo “*Train/retain network*” na célula “*Networks to train*” escolheram-se 20 redes para que o software treinasse, e dessas, 5 serão retidas após o treinamento. Essa última informação está indicada no campo “*Networks to retain*”. Essa proporção 20:5 é uma estratégia que equilibra a exploração de diferentes configurações de modelo com a necessidade de eficiência computacional e robustez do desempenho. Essa abordagem aumenta as chances de encontrar uma arquitetura de rede que ofereça o melhor desempenho possível para o problema do TOGG, ao mesmo tempo em que permite a criação de modelos que são mais generalizáveis e resistentes a variações nos dados.

A função de erro “*Sum of Squares*” foi utilizada para minimizar a soma dos quadrados das diferenças entre as saídas previstas pela rede e os valores reais. Esta função de erro é particularmente útil para problemas de regressão onde o objetivo é prever valores contínuos, pois penaliza severamente grandes erros de previsão, incentivando a rede a fazer ajustes mais precisos. Por outro lado, a função de erro *Cross Entropy*, que também está selecionada, é frequentemente utilizada em problemas de classificação, especialmente quando as saídas são binárias ou categóricas. Utilizar ambas as funções de erro no modelo TOGG é benéfico, pois permite que a rede neural seja treinada tanto para minimizar grandes discrepâncias em valores contínuos quanto para melhorar a precisão das previsões em tarefas de classificação (enquadrado/desenquadrado), proporcionando um modelo mais robusto e adaptável.

Na janela MLP “*activation functions*”, Figura 46, selecionaram-se as funções de ativação de interesse. Em “*Hidden Neurons*” (neurônios ocultos) estão localizados nas camadas intermediárias da rede, que no modelo serão responsáveis por processar e transformar os sinais recebidos e passá-los para a próxima camada. Esse passo é importante, pois a quantidade e configuração dos neurônios ocultos podem afetar significativamente a capacidade da rede de aprender e generalizar. Já os “*Output Neurons*” (neurônios de saída), estão localizados na camada final da rede e produzem os resultados, utilizados para fazer previsões ou classificações. Para o trabalho proposto selecionou-se “*Identity*”, “*logistic*”, “*Tanh*” e “*Exponential*”. A

“*sine*” é mais aplicável em problemas que envolvem padrões periódicos ou cíclicos, característico da função seno.

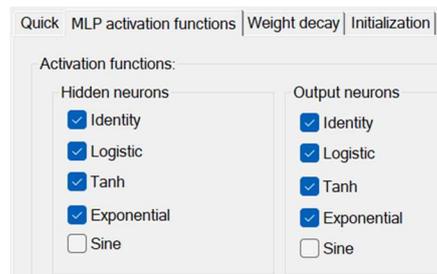


Figura 46: MLP activation functions

Na aba “*Weight decay*”, Figura 47, selecionou-se “*Use Weight Decay*” para ambas as camadas (oculta e de saída) e definiu-se os coeficientes de decaimento entre 0,0001 e 0,001, ou seja, aplicou-se uma regularização que impõe restrições leves a moderadas sobre os pesos da rede neural. O objetivo foi auxiliar o modelo TOG gravimétrico a prevenir o *overfitting* e incentivar a rede a aprender padrões mais simples e generalizáveis nos dados. Definir esses valores foram essenciais para atingir um equilíbrio entre flexibilidade no ajuste dos pesos (com um valor mínimo de 0,0001) e restrições (com um valor máximo de 0,001) para evitar ajustes excessivos.

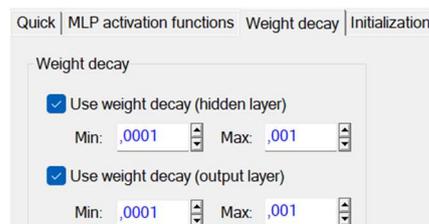


Figura 47: Weight decay

Por fim, na aba “*inicialization*”, Figura 48, seria possível inserir um valor inicial de sementes no algoritmo. Para a modelagem em questão, optou-se por não informar esse número.

Preferiu-se aproveitar a aleatoriedade natural, para garantir que o modelo fosse exposto a diferentes condições.

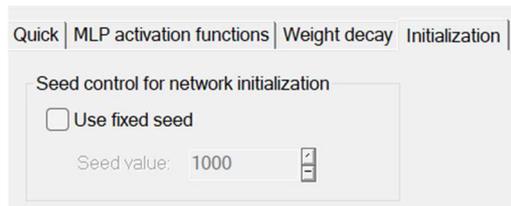


Figura 48: Initialization

#### 4.2.5.2. Treinamento e execução das Redes Neurais Artificiais

O módulo de treinamento da rede (*Train*) permitiu a geração de inúmeras configurações de rede, a partir das quais foram selecionadas as melhores. A Figura 49 exemplifica um quadro resumo da rede selecionada, e logo na sequência da figura há uma breve explicação dos parâmetros apontados pelo *software*. Vale lembrar que o programa *Statistica* gera uma série de redes, e que, neste passo, foi escolhida a rede MLP 18-14-2 como a mais favorável para o problema do TOGG.

| Net... | Net. na... | Trainin... | Test ... | Validatio... | Algori... | Error ... | Hidden... | Outp... |
|--------|------------|------------|----------|--------------|-----------|-----------|-----------|---------|
| 1      | MLP 18...  | 99,712...  | 97,30... | 95,9641...   | BFG...    | CE        | Expon...  | Soft... |

Figura 49 - Exemplo de rede gerada no módulo de treinamento do *Statistica*

- *Net. Name* (Nome da Rede): Retornou o nome da rede neural criada, MLP 18-14-2
- *Training Perf.* (Desempenho no Treinamento): Exibiu a performance da rede no conjunto de dados de treinamento. O desempenho de 99,71% sugere que a rede está acertando 99,71% das previsões nos dados usados para treiná-la. Isso sugere que a rede aprendeu bem os padrões presentes nos dados de treinamento do TOGG.
- *Test Perf.* (Desempenho no Teste): Exibiu a performance no conjunto de dados de teste. Um desempenho de 97,30% mostra que a rede está acertando 97,30%

das previsões nos dados de teste. Isso é um bom indicador de que a rede está generalizando bem, mas ligeiramente menor que a performance no treinamento.

- *Validation Perf.* (Desempenho na Validação): Este valor apresentado pelo software indica a performance da rede neural no conjunto de dados de validação, usado para ajustar hiperparâmetros e evitar *overfitting*. Um desempenho de 95,96% significa que a rede está acertando 95,96% das previsões nos dados de validação.
- *Training Algorithm* (Algoritmo de Treinamento): Indicou a escolha do algoritmo de otimização. Na RN proposta para o TOGG foi o algoritmo de treinamento BFGS 97, que se refere ao uso do método *Broyden-Fletcher-Goldfarb-Shanno* (BFGS), uma técnica de otimização numérica utilizada para ajustar os pesos da rede neural durante o treinamento.
- *Error Function* (Função de Erro): Mede a discrepância entre previsões e valores reais, para essa RN a função indicada pelo software foi a "*Entropy*".
- *Hidden Activation* (Ativação Oculta): Esse parâmetro indica a função de ativação usada nos neurônios das camadas ocultas. Para essa RN o software indicou a "*Exponential*".
- *Output Activation* (Ativação de Saída): Esse parâmetro indica a função de ativação usada na camada de saída. Para essa RN o software indicou a Softmax (classificação *multiclasse*)

#### 4.2.5.3. Avaliar o desempenho das Redes Neurais Artificiais

Diversas métricas de desempenho estão disponíveis para a validação da rede selecionada. As estatísticas de Test são extremamente úteis, pois verificam o poder de

generalização da rede, para dados não vistos durante a fase de treinamento. A Figura 50 apresenta os parâmetros que podem ser selecionados.

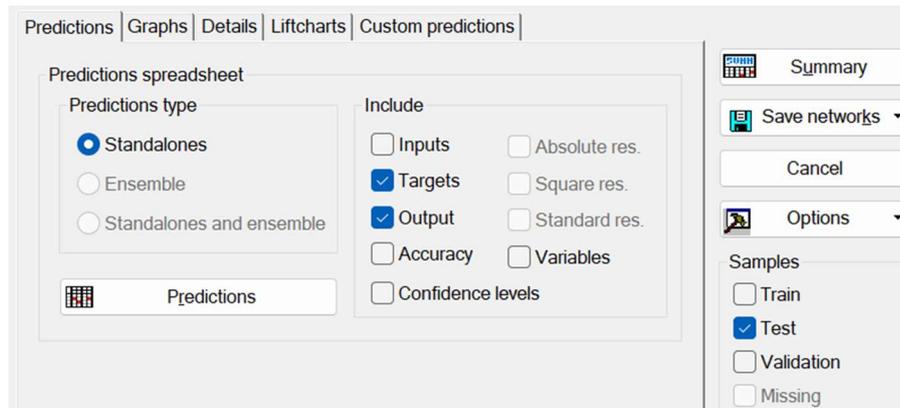


Figura 50 - Seleção da opção de métricas geradas pelo *Statistica*

Em *prediction types* a opção *Standalones* foi selecionada. Esses modelos são chamados de "*Standalones*" porque operam de forma autônoma, sem interação ou colaboração com outros modelos. *Ensemble*, por outro lado, refere-se a uma técnica em que vários modelos individuais, *Standalones*, são combinados para formar um conjunto ou grupo. Optou-se por não adotar essa abordagem.

Os valores de *predictions*, o gráfico da curva ROC e a matriz de confusão, atestam o desempenho da rede escolhida. Nessas análises, os valores de *prediction* devem ser similares aos valores de classe. A curva ROC deve ser aproximadamente no formato abaixo e as métricas corretas da matriz de confusão devem ser altas (quanto mais próximas de 100% tanto melhor). Estas informações são apresentadas conforme modelo da Figura 51.

#### 4.2.5.4. Selecionar a Rede Neural com melhor desempenho.

Os processos de parametrização, treinamento e avaliação de desempenho da Rede Neural podem ser feitos recursivamente até a obtenção da rede final. Tal processo deve ser feito sempre que dados novos estejam disponíveis ou quando uma mudança de processo nos dados ocorre.

#### 4.2.5.5. Gerar modelo robusto de previsão por Rede Neural.

Um modelo de previsão robusto baseado em Redes Neurais requer a definição de parâmetros que garantam a generalização adequada para a previsão do TOGG nas condições operacionais atuais.

No resumo da classificação, Figura 51, a primeira imagem é um resumo de classificação sobre os resultados obtidos pelo modelo (MLP 18-14-2) em um conjunto de dados de teste. Os dados foram divididos em duas classes, 0 e 1, conforme definido anteriormente. Foram 223 amostras ao total. O modelo conseguiu classificar corretamente 217 amostras, resultando em uma precisão global de 97,3094%. Essa métrica de precisão é subdividida entre as duas classes, com a classe 0 (enquadrada) obtendo uma precisão de 95,0980% e a classe 1 (desenquadrada) atingindo 99,1736%. Esses números refletem a capacidade do modelo de identificar corretamente as amostras pertencentes a cada classe. No entanto, também foram observados erros de classificação, sendo que cinco amostras da classe 0 e uma amostra da classe 1 foram classificadas incorretamente, resultando em uma taxa de erro geral de 2,6906%. Esses resultados são indicativos de um modelo bem treinado e ajustado, capaz de distinguir as classes com alta precisão, mas ainda suscetível a pequenas margens de erro.

|                      |               | classe (Classification summary) (arquivo_oversampled_SMOTE) |          |            |
|----------------------|---------------|---|----------|------------|
|                      |               | Samples: Test   |          |            |
|                      |               | classe-0  | classe-1 | classe-All |
| <b>1.MLP 18-14-2</b> | Total         | 102,0000  | 121,0000 | 223,0000   |
|                      | Correct       | 97,0000   | 120,0000 | 217,0000   |
|                      | Incorrect     | 5,0000  | 1,0000   | 6,0000     |
|                      | Correct (%)   | 95,0980   | 99,1736  | 97,3094    |
|                      | Incorrect (%) | 4,9020  | 0,8264   | 2,6906     |

Figura 51: Resumo da classificação (matriz de confusão)

Na Figura 52, uma parte dos dados, em vermelho um classificado incorretamente encontra-se em destaque.

| classe<br>Target | classe - Output<br>1. MLP 18-14-2 |
|------------------|-----------------------------------|
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 1                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 0                | 0                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |
| 1                | 1                                 |

Figura 52: Tabela comparativa da classificação

A Figura 53 mostra a curva ROC (*Receiver Operating Characteristic*) do modelo, uma ferramenta relevante para avaliar a performance de modelos de classificação. Ela representa a relação entre a sensibilidade (*true positive rate*) e a especificidade (*false positive rate*) para diferentes pontos de corte do modelo. A curva se aproxima do canto superior esquerdo do gráfico, sugerindo um desempenho excelente, onde tanto a sensibilidade quanto a especificidade são altas. A sensibilidade indica a capacidade do modelo de identificar corretamente as amostras positivas, enquanto a especificidade reflete a habilidade de evitar falsos positivos. A proximidade da curva ao eixo superior e ao eixo esquerdo denota que o modelo tem uma alta capacidade discriminativa, minimizando erros em ambos os lados da classificação.

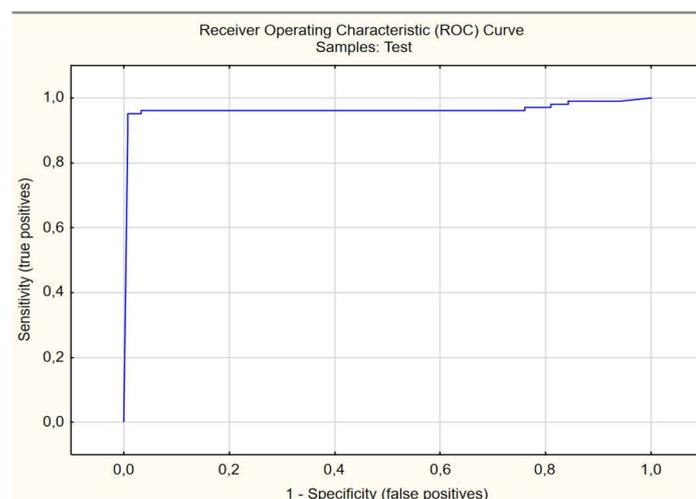


Figura 53: Curva ROC

#### 4.2.5.6. Converter previsor C/C++ para Python

Um código computacional do previsor foi gerado em C/C++ (usando o *Save Networks*) e a sua conversão realizada posteriormente para Python. A Figura 54, destaca esse recurso.

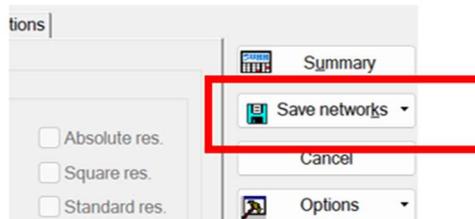


Figura 54: Save networks

```

previsorEnquadramentoTOG.py
1  import math
2  import numpy as np
3  import pandas as pd
4
5  def find_max(vec):
6      max_val = vec[0]
7      max_index = 0
8      for i in range(1, len(vec)):
9          if vec[i] > max_val:
10             max_val = vec[i]
11             max_index = i
12     return max_val, max_index
13
14     def scale_inputs(input_data, minimum, maximum, min_input, max_input):
15         scaled_inputs = []
16         for i in range(len(input_data)):
17             delta = (float(maximum) - float(minimum)) / (max_input[i] - min_input[i])
18             scaled_inputs.append(float(minimum) - delta * min_input[i] + delta * input_data[i])
19         return scaled_inputs
20
21
22     def softmax(vec):
23
24         for i in range(len(vec)):
25             if vec[i] > 200:
26                 max_value, max_index = find_max(vec)
27                 for j in range(len(vec)):
28                     if j == max_index:
29                         vec[j] = 1.0
30                     else:
31                         vec[j] = 0.0
32             else:
33                 vec[i] = math.exp(vec[i])
  
```

Figura 55; Parte do código Python

Os pesos (Details→Weights) da rede selecionada foram obtidos (Figura 56) e convertidos para um arquivo em formato texto. Essa possibilidade torna a atualização em sistemas próprios da empresa possível, com a substituição de apenas um arquivo.

| Network weights (arquivo_oversampled_SMOTE) |  |                                |
|---|--|--------------------------------|
| Weight ID                                   | Connections<br>1.MLP 18-14-2           | Weight values<br>1.MLP 18-14-2 |
| 1   | Vazão Agua Prod --> hidden neuron 1    | 0,0068                         |
| 2   | Temp SG-01 --> hidden neuron 1         | -3,9704                        |
| 3   | Temp TO-01 --> hidden neuron 1         | -5,7518                        |
| 4   | Vazão SG-01 --> hidden neuron 1        | 0,5094                         |
| 5   | LV TO-1 --> hidden neuron 1            | 0,1139                         |
| 6   | Vazão água dessalg --> hidden neuron 1 | 4,2556                         |
| 7   | Pressão FLOTB --> hidden neuron 1      | 0,8103                         |
| 8   | LL-97 --> hidden neuron 1              | -1,5584                        |
| 9   | LL-105 --> hidden neuron 1             | -2,9601                        |
| 10  | LL-100 --> hidden neuron 1             | 4,3659                         |
| 11  | LL-60 --> hidden neuron 1              | 0,9490                         |
| 12  | LL-90 --> hidden neuron 1              | -4,6779                        |
| 13  | LL-69 --> hidden neuron 1              | -4,7363                        |
| 14  | RJS-680 --> hidden neuron 1            | 7,6170                         |
| 15  | LL-102 --> hidden neuron 1             | -5,5552                        |
| 16  | online --> hidden neuron 1             | -1,9294                        |
| 17  | Desemul calc --> hidden neuron 1       | 0,9301                         |
| 18  | Poli2 --> hidden neuron 1              | -4,3076                        |
| 19  | Vazão Agua Prod --> hidden neuron 2    | 1,5457                         |
| 20  | Temp SG-01 --> hidden neuron 2         | 3,9947                         |
| 21  | Temp TO-01 --> hidden neuron 2         | -0,6390                        |
| 22  | Vazão SG-01 --> hidden neuron 2        | 2,1692                         |
| 23  | LV TO-1 --> hidden neuron 2            | 4,9900                         |
| 24  | Vazão água dessalg --> hidden neuron 2 | -8,8918                        |
| 25  | Pressão FLOTB --> hidden neuron 2      | -8,0738                        |
| 26  | LL-97 --> hidden neuron 2              | 1,6295                         |

Figura 56 - Extração dos pesos para transposição para o código em Python

## 5.Considerações

### 5.1.1. Conclusões gerais

Este estudo alcançou com êxito o objetivo central ao desenvolver um método de previsão capaz de estimar os valores de TOG com precisão satisfatória. Utilizando as variáveis de processo da planta de tratamento de petróleo e as medições online de TOGG foi possível criar um modelo preditivo robusto e confiável. A identificação das variáveis mais pertinentes para o modelo foi feita através de uma análise sistemática, que garantiu que as variáveis selecionadas fossem realmente relevantes para a tarefa de previsão. Essa robustez e assertividade foi evidenciada pela acurácia total de 97,3% alcançada pelo modelo.

Os objetivos específicos definidos no início deste estudo foram também alcançados.

Com a forte interação com os especialistas da área, foi possível realizar a mineração dos dados pondo-se o foco naqueles apontados como relevantes ao processo. A integridade desses dados pôde ser avaliada e suas inconsistências tratadas e reportadas para a equipe de especialistas. Desse modo o objetivo *“realizar a mineração de dados, coletando e compilando dados históricos de medições de TOG, bem como as variáveis de processo relevantes”* foi atendido em sua plenitude.

Na etapa 3 do trabalho, em posse dos dados realizaram-se análises estatísticas para avaliá-los, como análise dos efeitos principais, ANOVA, Pareto e outros. Nessa fase atendeu-se o objetivo *“aplicar técnicas de análise estatística e de correlação para identificar as variáveis de processo mais fortemente associadas aos valores do TOG, a fim de selecionar as mais relevantes para o modelo de previsão”*, como definido inicialmente.

Nas etapas seguintes o objetivo central do trabalho foi atendido, ao desenvolver o modelo preditivo proposto e ao validá-lo, alcançando-se assim dois dos objetivos: *“desenvolver um modelo preditivo utilizando técnicas avançadas de análise de dados, como regressão multivariada ou aprendizado de máquina, incorporando as variáveis de processo identificadas como relevantes”* e *“avaliar a eficácia do modelo de previsão por meio de testes de validação cruzada e análises de erro, garantindo que ele seja capaz de prever os resultados de TOG com precisão significativa.”*

Por fim, ao *“disponibilizar o modelo em código que permita a implementação em ferramenta interna da empresa proprietária dos dados”*, foi atingido o último objetivo

proposto, com a conversão do modelo para a linguagem Python.

### 5.1.2. Contribuições do trabalho

Tão importante quanto o modelo proposto foi o ganho de conhecimento referente às variáveis de processo. Estudar seus comportamentos, sentidos, integridade do banco de dados, desenvolvimento de mecanismos de extração e tratamento desses dados foi de grande relevância.

Através de uma abordagem quantitativa, foram identificadas as variáveis que exercem maior influência no processo estudado. Desse modo percepções até então de natureza conceitual, puderam ser comprovados e medidas.

Além disso, o trabalho explorou uma estratégia alternativa para a empresa, destacando a importância da aplicação de ciências de dados na otimização do monitoramento do TOGG. Essa abordagem reflete a preocupação contínua com os aspectos ambientais, evidenciando o compromisso da empresa em promover práticas sustentáveis e responsáveis. Se por um lado demonstra a preocupação ambiental, esse trabalho reafirma a necessidade de os processos produtivos estarem cada vez mais orientados a dados. Isso reafirma o futuro que se desenha para a indústria, onde processos digitais e analíticos serão essenciais e constituirão diferenciais competitivos.

### 5.1.3. Sugestões para estudos futuros

Avaliar a longo prazo a eficiência do modelo, assim como desenvolver mecanismos automáticos de retreino são sugestões para estudos futuros. Esses dois pontos garantem a acurácia do modelo e uma usabilidade operacional. Além disso, sugere-se a expansão do método para outras plataformas, uma vez que no estudo proposto apenas uma Plataforma foi considerada.

## REFERÊNCIAS

- ABASS, A. O. (2020). Recent advances on the treatment technology of oil and gas produced water for sustainable energy industry-mechanistic aspects and process chemistry perspectives. *Chemical Engineering Journal Advances*, 4: 100049.
- ADEWUNMI, A. A., et al. (2018). "Optimization of Produced Water Management: A Review." *Processes*, 6(11), 222.
- SPEIGHT, J. G. (2019). "Environmental Impact of the Petroleum Industry." In *Environmental Impact of the Petroleum Industry* (pp. 207-223). CRC Press.
- BOEIRA, G., Moysés, P., & Amorim, J. (2015). Emulsões na indústria petrolífera. *Bolsista de Valor*, 4. <https://editoraessentia.iff.edu.br/index.php/BolsistaDeValor/article/view/6751>
- Boletim Mensal da Produção de Petróleo e Gás Natural. ([s.d.]). Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Recuperado 12 de outubro de 2023, de <https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/boletins-anp/boletins/boletim-mensal-da-producao-de-petroleo-e-gas-natural>
- BRAGA, Antônio de Pádua; LUDERMIR, Teresa Bernarda; CARVALHO, André C. P. de Leon Ferreira. *Redes Neurais Artificiais: teoria e aplicações*. 2º edição. Editora LTC: Rio de Janeiro, 2007.
- CARLOS, J. F. P. *Aplicação de Redes Neurais no Processamento Digital de Imagens*. Universidade Federal de Minas Gerais, 1994.
- CONAMA. Resolução no 393, de 8 de agosto de 2007. Publicada no DOU nº 153, de 9 de agosto de 2007, Seção 1. Dispõe sobre o descarte contínuo de água de processo ou de produção em plataformas marítimas de petróleo e gás natural, e dá outras providências.
- COOK, L. (2004). "Laboratory Application Articles: Rotary Evaporation." *Labcompare*. <https://www.labcompare.com/10-Featured-Articles/267254-Laboratory-Application-Articles-Rotary-Evaporation/>
- EBENEZER T. Igunnu, George Z. Chen, Produced water treatment technologies, *International Journal of Low-Carbon Technologies*, Volume 9, Issue 3, September 2014, Pages 157–177,
- FAKHRU'L-RAZI, A.; ALIREZA, P.; LUQMAN, C.A.; DAYANG, R.A.B; SAYED, S.M.; ZURINA, Z.A. (2009). Review of Technologies for oil and gas produced water treatment.

Journal of Hazardous Materials, 170:530-551.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.

HAYKIN, S. Redes Neurais: Princípios e prática. McMaster University, Hamilton, Ontário, Canadá, 2001.

MONTGOMERY, D. C. (2013). *Design and Analysis of Experiments* (8th ed.). John Wiley & Sons.

MOTTA, A. R. P. da, Borges, C. P., Kiperstok, A., Esquerre, K. P., Araujo, P. M., & Branco, L. da P. N. (2013). Tratamento de água produzida de petróleo para remoção de óleo por processos de separação por membranas: revisão. *Engenharia sanitária e ambiental*, 18(1), 15–26. <https://doi.org/10.1590/s1413-41522013000100003>

MURTHY, Y.R.; Bhaskar, K.U.; Parametric CFD studies on hydrocyclone, *Powder Technology*, 2012.

NEVES, Marcus Vinícius Silva. Eficiência exergetica de unidades estacionárias de produção de petróleo. 2008 Dissertação (Mestrado em ciências em engenharia mecânica) Universidade

O'NEIL, M. J. et al. (2006). "The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals." 14th Edition, Merck.

PEREIRA, T. V. Otimização e Operação de Processos Off-Shore de Separação ÓleoGás.2003 103 f. Dissertação (Mestrado em Engenharia Química) Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

PROCESSAMENTO PRIMÁRIO DE PETRÓLEO chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://engenhariaquimica.files.wordpress.com/2010/04/apostila-ppp.pdf

SCHARAMM, L. L. Petroleum emulsions: Basic Principles. In: SCHARAMM, L. L. Emulsions: Fundamentals and applications in the petroleum industry, Advanced Chemistry Series 231. Washington: ACS, p. 79-129, 1992.

SEVERIN HANSEN, D., Jespersen, S., Bram, M. V., & Yang, Z. (2020). Uncertainty analysis of fluorescence-based oil-in-water monitors for oil and gas produced water. *Sensors* (Basel,

Switzerland), 20(16), 4435. <https://doi.org/10.3390/s20164435>

SPEIGHT, J. G. (2019). "Environmental Impact of the Petroleum Industry." In *Environmental Impact of the Petroleum Industry* (pp. 207-223). CRC Press.

STREITENBERGER, S. C. (2022). Modelo preditivo do teor de óleos e graxas em água produzida quantificado pelo método gravimétrico. <https://repositorio.unifei.edu.br/jspui/handle/123456789/3319>

VIEIRA, L. G. M. (2006). Otimização dos processos de separação em hidrociclones filtrantes. <https://repositorio.ufu.br/handle/123456789/15053>

VIEIRA, V. M. Água produzida no segmento onshore de petróleo: Caracterização de cenários na baía e prospecção de soluções para gerenciamento. Dissertação de Mestrado. Pós-Graduação em Geologia. Universidade Federal da Bahia – UFBA. Salvador/BA. 2011.