

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA  
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA  
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA EM  
SISTEMAS PRODUTIVOS

ALLAN DOS ANJOS PESTANA

MODELO PREDITIVO PARA OTIMIZAÇÃO DO TRATAMENTO DE ESGOTO COM O  
USO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

São Paulo

Março/2025

ALLAN DOS ANJOS PESTANA

MODELO PREDITIVO PARA OTIMIZAÇÃO DO TRATAMENTO DE ESGOTO COM O  
USO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

Dissertação apresentada como exigência parcial para a obtenção do título de mestre no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação da Profa. Dra. Márcia Ito.

Área de Concentração: Sistemas Produtivos

São Paulo  
Março/2025

P476m Pestana, Allan dos Anjos  
Modelo preditivo para otimização do tratamento de esgoto com o uso de técnicas de aprendizado de máquina / Allan dos Anjos Pestana. – São Paulo: CPS, 2025.  
159 f. : il.

Orientadora: Profa. Dra. Márcia Ito  
Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos) – Centro Estadual de Educação Tecnológica Paula Souza, 2025.

1. Redes neurais com atraso de tempo. 2. Demanda química de oxigênio. 3. Tratamento de esgoto. 4. Redes neurais artificiais. 5. Saneamento. I. Ito, Márcia. II. Centro Estadual de Educação Tecnológica Paula Souza. III. Título.

ALLAN DOS ANJOS PESTANA

MODELO PREDITIVO PARA OTIMIZAÇÃO DO TRATAMENTO DE ESGOTO COM O  
USO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

Documento assinado digitalmente  
 **MARCIA ITO**  
Data: 25/03/2025 19:23:20-0300  
Verifique em <https://validar.iti.gov.br>

---

Profa. Dra. Márcia Ito  
Orientador – CEETEPS

Documento assinado digitalmente  
 **RAZER ANTHOM NIZER ROJAS MONTANO**  
Data: 24/03/2025 09:55:11-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Razer Anthom Nizer Rojas Montaña  
Examinador Externo – UFPR

Documento assinado digitalmente  
 **MARCELO DUDUCHI FEITOSA**  
Data: 01/04/2025 18:37:48-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Marcelo Duduchi Feitosa  
Examinador Interno - CEETEPS

São Paulo, 20 de Março de 2025

A minha família, Anjos Pestana, pelo incentivo a ser melhor a cada dia. E a Yasmin e Pedro, por me ensinar o que é o amor em sua forma mais pura.

## **AGRADECIMENTOS**

Agradeço a Deus pela oportunidade de estudar e aplicar o que foi aprendido em benefício da população e meio ambiente.

A professora Doutora Márcia Ito, pelo aprendizado, parceria, além do incentivo durante a trajetória.

A companhia de Saneamento Básico do Estado de São Paulo, a SABESP, por toda a oportunidade de crescimento, incentivo acadêmico, e permitir colaborar no cuidado da saúde da população através da transformação do esgoto.

A minha mãe, Mirian, e meu pai, Ricardo, que mesmo com pouco estudo, sempre acreditaram na educação como ferramenta de transformação para os filhos.

Ao Rodrigo Alves, amigo e apoiador de pesquisa.

Agradeço ao Antônio Carlos Batista, Fabio Cosme Rodrigues dos Santos, Helvecio Carvalho de Sena, Oto Elias Pinto, Renato Franzzotti, e Rogerio de Souza Marquezini, companheiros de trajetória da Sabesp, pela colaboração no estudo.

Agradeço as instituições anteriores que me trouxeram até aqui.

*Absque sudore et labore nullum opus  
perfectum est.* “Sem suor e sem trabalho  
nenhuma obra é terminada. ”

(Schrevelius, 1176)

## RESUMO

PESTANA, A. A. **Modelo preditivo para otimização do tratamento de esgoto com o uso de técnicas de aprendizado de máquina**. 159 f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2025.

O presente trabalho tem por objetivo analisar e otimizar os processos de tratamento de esgoto por meio da aplicação de técnicas de aprendizado de máquina, com foco específico nos parâmetros de Demanda Química de Oxigênio (DQO) e amônia. A dissertação justifica-se pela necessidade de explorar o potencial das Redes Neurais Artificiais (RNAs) na modelagem e previsão de parâmetros críticos no tratamento de esgoto. O método de pesquisa foi baseado no *framework Design Science Research (DSR)*, a etapa prévia à aplicação das RNAs incluiu análises de PCA (Análise de Componentes Principais) e clusterização com visualização via *K-means*, identificando padrões nos dados. As modelagens realizadas empregaram além de RNAs, Redes Neurais com Atraso de Tempo (*Time Delay Neural Networks – TDNNs*), incorporando parâmetros como o Tempo de Detenção Hidráulica (TDH) que ocorre no tratamento de esgoto para aprimorar a precisão das previsões. Os resultados obtidos a partir de previsões de utilização de inteligência artificial para prever parâmetros operacionais das Estações de Tratamento de Esgoto (ETEs), podem propiciar melhorias significativas em eficiência, confiabilidade, redução de custos com químicos e energia, e no caso da aplicação prática da RNA para predição em uma das ETEs da região metropolitana de São Paulo, antecipar cargas indesejadas ao tratamento. Nesse estudo, a modelagem da DQO em quatro ETEs da região metropolitana de São Paulo obteve média de 0,95 no coeficiente de determinação ( $R^2$ ) e 3,6 no MAE, e inclui a avaliação do impacto da base de dados, temperatura, estrutura da RN e *features* utilizadas nesses resultados. Além disso, o estudo destaca a importância da automação e da análise de dados para atender às normativas ambientais e promover a sustentabilidade no setor de saneamento. Linha de Pesquisa Sistemas de Informação e Tecnologias Digitais. Projeto de Pesquisa Tecnologias digitais em sistemas produtivos.

**Palavras-chave:** Redes Neurais com Atraso de Tempo. Demanda Química de Oxigênio. Tratamento de Esgoto. Redes Neurais Artificiais. Saneamento.

## ABSTRACT

PESTANA, A. A. **Predictive Model for Wastewater Treatment Optimization Using Machine Learning Techniques**. 159 f. Dissertation (Professional Master's in Management and Technology in Production Systems). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2025.

The present study aims to analyze and optimize wastewater treatment processes through the application of machine learning techniques, with a specific focus on the parameters of Chemical Oxygen Demand (COD) and ammonia. The dissertation is justified by the need to explore the potential of Artificial Neural Networks (ANNs) in modeling and predicting critical parameters in wastewater treatment. The research method was based on the Design Science Research (DSR) framework. The preliminary stage before applying ANNs included Principal Component Analysis (PCA) and clustering with visualization via K-means to identify patterns in the data. In addition to ANNs, the modeling employed Time Delay Neural Networks (TDNNs), incorporating parameters such as Hydraulic Retention Time (HRT), which occurs during wastewater treatment, to improve prediction accuracy. The results obtained from artificial intelligence-based predictions of operational parameters in Wastewater Treatment Plants (WWTPs) can lead to significant improvements in efficiency, reliability, cost reduction in chemicals and energy, and, in the practical application of ANN predictions in a WWTP in the metropolitan region of São Paulo, the anticipation of undesirable loads entering the treatment process. In this study, COD modeling in four WWTPs in the São Paulo metropolitan region achieved an average coefficient of determination ( $R^2$ ) of 0.95 and a Mean Absolute Error (MAE) of 3.6. The study also evaluates the impact of the dataset, temperature, neural network structure, and selected features on these results. Additionally, the study highlights the importance of automation and data analysis in meeting environmental regulations and promoting sustainability in the sanitation sector. Research Line Information Systems and Digital Technologies. Research Project Digital technologies in production systems.

**Keywords:** Time Delay Neural Networks. Chemical Oxygen Demand. Sewage Treatment. Artificial Neural Networks. Sanitation.

## LISTA DE QUADROS

|   |    |
|---|----|
| Quadro 1 - Critérios de inclusão e exclusão .....                   | 48 |
| Quadro 2 - Questões para avaliação de qualidade .....               | 49 |
| Quadro 3 - Pontuação para avaliação de qualidade das questões ..... | 49 |
| Quadro 4 - Contextualização do problema .....                       | 51 |
| Quadro 5 - Estratégia PICO.....                                     | 51 |
| Quadro 6 - Artigos após qualificação .....                          | 58 |
| Quadro 7 - Relação dos modelos utilizados .....                     | 60 |

## LISTA DE TABELAS

|  |     |
|--|-----|
| Tabela 1 - Interpretação do coeficiente de determinação .....                                | 46  |
| Tabela 2 - Quantitativo dos resultados nas bases .....                                       | 52  |
| Tabela 3 - Quantitativo dos resultados nas bases após excluir duplicados .....               | 52  |
| Tabela 4 - Resumo das bases utilizadas.....  | 81  |
| Tabela 5 - Relação de instrumentos por planta - ETE A.....                                   | 82  |
| Tabela 6 - Caracterização dos dados - Julho/2024 - ETE A.....                                | 84  |
| Tabela 7 - Relação de instrumentos por planta - ETE B .....                                  | 90  |
| Tabela 8 - Caracterização dos dados - Julho/2024 - ETE B.....                                | 92  |
| Tabela 9 - Relação de instrumentos por planta - ETE C .....                                  | 97  |
| Tabela 10 - Caracterização dos dados - Julho/2024 - ETE C.....                               | 99  |
| Tabela 11 - Relação de instrumentos por planta - ETE D.....                                  | 104 |
| Tabela 12 - Caracterização dos dados - Julho/2024 - ETE D.....                               | 106 |
| Tabela 13 - Relação de instrumentos por planta - ETE E .....                                 | 111 |
| Tabela 14 - Caracterização dos dados - Julho/2024 - ETE E .....                              | 112 |
| Tabela 15 - Parâmetros de configuração do <i>Grid Search</i> da RNA .....                    | 117 |
| Tabela 16 - Métricas obtidas para DQO efluente nas ETEs.....                                 | 127 |
| Tabela 17 - Desempenho de modelos com variável de interesse DQO efluente.....                | 128 |
| Tabela 18 - Cálculo do TDH em cada etapa.....  | 129 |
| Tabela 19 - ETE A - TDH dos grupos definidos .....   | 131 |
| Tabela 20 - ETE D - TDH dos grupos definidos .....   | 134 |
| Tabela 21 - ETE E - TDH dos grupos definidos .....   | 136 |
| Tabela 22 - Set-points esperados para parâmetros analíticos de acordo com especialista ..... | 137 |
| Tabela 23 - Métricas de desempenho da RNA para dados afluentes .....                         | 138 |
| Tabela 24 - Métricas de desempenho da RNA para dados afluentes com previsibilidade.....      | 139 |

## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura 1 - Pirâmide de automação (exemplos no saneamento) .....  | 26 |
| Figura 2 - Fluxograma do tratamento - ETE .....  | 27 |
| Figura 3 - Tratamento fase líquida .....   | 29 |
| Figura 4 - Tratamento fase sólida .....  | 30 |
| Figura 5 - Relação entre métricas para determinação da matéria orgânica .....                            | 34 |
| Figura 6 - Exemplo de medição de manto de lodo .....   | 36 |
| Figura 7 - Diagrama de Venn das Tecnologias da Inteligência Artificial .....                             | 38 |
| Figura 8 - Exemplo do método do cotovelo .....   | 40 |
| Figura 9 - Redes Neurais Artificiais (Exemplo no saneamento) .....                                       | 40 |
| Figura 10 - Funções de ativação .....  | 42 |
| Figura 11 - <i>Cross validation</i> .....  | 43 |
| Figura 12 - Procedimento metodológico adotado .....  | 50 |
| Figura 13 - Artigos por ano .....  | 53 |
| Figura 14 - Palavras mais encontradas .....  | 54 |
| Figura 15 - Palavras mais encontradas ao longo do tempo .....  | 54 |
| Figura 16 - Nuvem de palavras .....  | 55 |
| Figura 17 - Palavras utilizadas ao longo do tempo nos artigos qualificados .....                         | 56 |
| Figura 18 - Nuvem de palavras dos artigos qualificados .....   | 57 |
| Figura 19 - <i>Clusters</i> de palavras dos artigos qualificados .....                                   | 57 |
| Figura 20 - Procedimento metodológico adotado com resultados .....                                       | 59 |
| Figura 21 - Nuvem de palavras e gráfico de frequência de palavras dos potenciais usos da automação ..... | 65 |
| Figura 22 - Nuvem de palavras tecnologias emergentes .....   | 66 |
| Figura 23 - Nuvem de palavras impacto no tratamento de esgoto .....                                      | 67 |
| Figura 24 - Nuvem de palavras para os desafios enfrentados .....   | 72 |
| Figura 25 - Segmentos das respostas com as palavras "Equipamentos" e "Falta" .....                       | 73 |
| Figura 26 - Nuvem de palavras para as tecnologias desejadas .....  | 74 |
| Figura 27 - Segmentos das respostas com a palavra "dados" .....  | 74 |
| Figura 28 - Nuvem de palavras para os benefícios esperados .....   | 75 |
| Figura 29 - Segmentos das respostas com as palavras "melhor" e "redução" .....                           | 75 |
| Figura 30 - Segmentos das respostas com a palavra "processo" .....                                       | 76 |
| Figura 31 - Matriz de correlação entre as respostas - Corpo operacional .....                            | 77 |

|   |     |
|---|-----|
| Figura 32 - Matriz de correlação entre as respostas - Corpo gerencial .....                                 | 78  |
| Figura 33 - Dados ETE A .....   | 83  |
| Figura 34 - <i>Box-plot</i> dos dados - ETE A – Antes e depois da retirada de <i>outliers</i> .....         | 86  |
| Figura 35 - Dados ETE B .....   | 91  |
| Figura 36 - <i>Box-plot</i> dos dados - ETE B - Antes e depois da retirada de <i>outliers</i> .....         | 94  |
| Figura 37 - Dados ETE C .....   | 98  |
| Figura 38 - <i>Box-plot</i> dos dados - ETE C - Antes e depois da retirada de <i>outliers</i> .....         | 101 |
| Figura 39 - Dados ETE D .....   | 105 |
| Figura 40 - <i>Box-plot</i> dos dados - ETE D - Antes e depois da retirada de <i>outliers</i> .....         | 108 |
| Figura 41 - Dados ETE E .....   | 112 |
| Figura 42 – <i>Box-plot</i> dos dados - ETE E .....   | 114 |
| Figura 43 - ETE A - Correlação da DQO efluente com outros parâmetros incluindo atrasos                      | 130 |
| Figura 44 - Localização das sondas e divisão em grupos de atrasos - ETE A .....                             | 131 |
| Figura 45 - ETE C - Correlação da DQO efluente com outros parâmetros incluindo atrasos                      | 132 |
| Figura 46 - ETE D - Correlação da DQO efluente com outros parâmetros incluindo atrasos                      | 133 |
| Figura 47 - Localização das sondas e divisão em grupos de atrasos - ETE D .....                             | 134 |
| Figura 48 - ETE E - Correlação da DQO efluente com outros parâmetros incluindo atrasos                      | 135 |
| Figura 49 - Localização das sondas e divisão em grupos de atrasos - ETE E.....                              | 135 |
| Figura 50 - Testes da utilização da RNA treinada no supervisor Elipse E3.....                               | 141 |
| Figura 51 - Testes da utilização da RNA treinada no supervisor Elipse E3 com pesos e vieses no script ..... | 142 |
| Figura 52 - Levantamento da automação mínima necessário.....  | 154 |
| Figura 53 - Estatística do levantamento .....   | 155 |
| Figura 54 - Levantamento da automação mínima necessária – Plantas principais .....                          | 156 |
| Figura 55 - <i>Dashboard</i> proposto automação .....   | 157 |

## LISTA DE GRÁFICOS

|  |     |
|--|-----|
| Gráfico 1 - Matriz de correlação - ETE A .....   | 85  |
| Gráfico 2 - <i>K-means</i> e PCA - ETE A .....   | 88  |
| Gráfico 3 - Contribuição das <i>features</i> PCA - ETE A .....   | 89  |
| Gráfico 4 - Matriz de correlação - ETE B .....   | 93  |
| Gráfico 5 - <i>K-means</i> e PCA - ETE B.....  | 95  |
| Gráfico 6 - Contribuição das <i>features</i> PCA - ETE B .....   | 96  |
| Gráfico 7 - Matriz de correlação - ETE C .....   | 100 |
| Gráfico 8 - <i>K-means</i> e PCA - ETE C.....  | 102 |
| Gráfico 9 - Contribuição das <i>features</i> PCA - ETE C .....   | 103 |
| Gráfico 10 - Matriz de correlação - ETE D.....   | 107 |
| Gráfico 11 - <i>K-means</i> e PCA - ETE D .....  | 109 |
| Gráfico 12 - Contribuição das <i>features</i> PCA - ETE D .....  | 110 |
| Gráfico 13 - Matriz de correlação - ETE E .....  | 113 |
| Gráfico 14 - <i>K-means</i> e PCA - ETE E.....   | 115 |
| Gráfico 15 - Contribuição das <i>features</i> PCA - ETE E.....   | 116 |
| Gráfico 16 - DQO ETE A - Real x Predito - Base de Teste .....  | 118 |
| Gráfico 17 - Importância do SHAP para DQO efluente - ETE A.....  | 119 |
| Gráfico 18 - DQO ETE C - Real x Predito - Base de Teste .....  | 120 |
| Gráfico 19 - Importância do SHAP para DQO efluente - ETE C .....   | 121 |
| Gráfico 20 - DQO ETE D - Real x Predito - Base de Teste .....  | 122 |
| Gráfico 21 - Importância do SHAP para DQO efluente - ETE D.....  | 123 |
| Gráfico 22 - DQO ETE E - Real x Predito - Base de Teste .....  | 123 |
| Gráfico 23 - Importância do SHAP para DQO efluente - ETE E.....  | 124 |
| Gráfico 24 - Coeficiente de determinação ( $R^2$ ) - DQO efluente .....  | 125 |
| Gráfico 25 - RMSE - DQO efluente.....  | 126 |
| Gráfico 26 - MAE - DQO efluente.....   | 126 |
| Gráfico 27 - Amônia afluyente modelada vs. real – Jan./2024 .....  | 138 |
| Gráfico 28 - Amônia afluyente modelada vs. real – com previsibilidade de 1 hora – Jan./2024 - Média horária..... | 139 |
| Gráfico 29 - Testes da utilização da RNA treinada no Excel.....  | 140 |
| Gráfico 30 - Período de testes da RNA em campo.....  | 143 |

## LISTA DE SIGLAS

|          |   |
|----------|---|
| ACP      | Análise de Componentes Principais   |
| ANFIS-GA | <i>Adaptive Neuro-Fuzzy Inference System Coupled With Genetic Algorithm</i>   |
| AG       | Algoritmo Genético  |
| BOD      | <i>Biological Oxygen Demand</i>   |
| CLP      | Controlador Lógico Programável  |
| COD      | <i>Chemical Oxygen Demand</i>   |
| CONAMA   | Conselho Nacional do Meio Ambiente  |
| DAE      | <i>Denoising Autoencoder</i>  |
| DBO      | Demanda Bioquímica de Oxigênio  |
| DP       | Decantador Primário   |
| DQO      | Demanda Química de Oxigênio   |
| DS       | Decantador Secundário   |
| ELM      | <i>Extreme Learning Machine</i>   |
| ERP      | <i>Entreprise Resource Planning</i> - Sistema Integrado De Gestão Empresarial |
| FFNN     | <i>Feedforward Neural Network</i>   |
| GRU      | <i>Gated Recurrent Units</i>  |
| IA       | Inteligência Artificial   |
| IAE      | <i>Integrated Absolute Error</i>  |
| IEE      | Índice de Eficiência Energética   |
| IMC      | <i>Internal Model Controller</i>  |
| IoT      | <i>Internet of Things</i>   |
| IQR      | <i>Interquartile Range</i>  |
| ISE      | <i>Integrated Square Error</i>  |
| KNN      | <i>K-nearest neighbor</i>   |
| LSTM     | <i>Long Short-Term Memory</i> - Memória De Curto Longo Prazo                  |
| LSTM-AM  | <i>Long Short-Term Memory Based On Attention Mechanism</i>                    |
| MAE      | <i>Mean Absolute Error</i>  |
| MAPE     | <i>Mean Absolute Percentual Error</i>   |
| MLP      | <i>Multilayer Perceptron</i>  |
| MSE      | <i>Mean Square Error</i>  |
| OD       | Oxigênio Dissolvido   |
| ODS      | Objetivos de Desenvolvimento Sustentável                                      |

|      |  |
|------|--|
| ORP  | <i>Oxidation-Reduction Potential</i>                                       |
| PCA  | <i>Principal Component Analysis</i>  |
| RMSE | <i>Root Mean Square Error</i>  |
| RNA  | Redes Neurais Artificiais  |
| RSL  | Revisão Sistemática da Literatura  |
| SHAP | SHapley Additive exPlanations  |
| SST  | Sólidos Suspensos Totais   |
| SVM  | <i>Support Vector Machine</i>  |
| TA   | Tanque de aeração  |
| TDH  | Tempo de Detenção Hidráulica   |
| TDNN | <i>Time Delay Neural Network</i> – Rede Neural com Atraso de Tempo         |
| TOC  | <i>Total Organic Carbon</i>  |
| WCSS | <i>Within-Cluster Sum of Squares</i> - Soma de quadrados dentro do cluster |

## SUMÁRIO

|  |     |
|--|-----|
| INTRODUÇÃO.....  | 19  |
| 1 FUNDAMENTAÇÃO TEÓRICA .....  | 25  |
| 1.1 Transformação digital no tratamento de esgoto .....  | 25  |
| 1.1.1 Indicadores no tratamento de esgoto .....  | 32  |
| 1.2 Instrumentação no tratamento de esgoto .....   | 33  |
| 1.3 Técnicas de análise de dados e aprendizado de máquina .....                                | 37  |
| 1.4 Redes Neurais Artificiais .....  | 40  |
| 1.4.1 Métricas de avaliação .....  | 44  |
| 1.4.2 Tratamento de dados.....   | 46  |
| 2 REVISÃO DA LITERATURA SOBRE TRANSFORMAÇÃO DIGITAL NO TRATAMENTO DE ESGOTO .....              | 48  |
| 3 ENTREVISTAS – PERSPECTIVAS DOS ESPECIALISTAS.....  | 64  |
| 3.1 Resultados das Entrevistas .....   | 65  |
| 4 UMA SURVEY COM PROFISSIONAIS DO SETOR.....   | 69  |
| 4.1 Resultados da Survey .....   | 70  |
| 5 APLICAÇÃO PRÁTICA DO MODELO .....  | 79  |
| 5.1 Caracterização dos dados.....  | 79  |
| 5.2 Início das análises nas Estações .....   | 81  |
| 5.2.1 ETE A.....   | 81  |
| 5.2.2 ETE B .....  | 89  |
| 5.2.3 ETE C .....  | 97  |
| 5.2.4 ETE D.....   | 104 |
| 5.2.5 ETE E .....  | 110 |
| 5.3 Modelagem da variável Demanda Química de Oxigênio (DQO) utilizando RNA.....                | 116 |
| 5.3.1 Conclusões.....  | 124 |
| 5.4 Avaliando a influência do Tempo de Detenção Hidráulica (TDH) das ETEs nas modelagens ..... | 128 |
| 5.4.1 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE A .             | 129 |
| 5.4.2 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE C..             | 132 |
| 5.4.3 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE D .             | 132 |
| 5.4.4 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE E..             | 134 |
| 5.4.5 Resultados com a inclusão do TDH na modelagem.....                                       | 136 |

|   |     |
|---|-----|
| 5.5 Limites de parâmetros segundo especialista.....   | 136 |
| 5.6 Modelagem com amônia por meio de parâmetros em elevatória afluenta .....                      | 137 |
| 5.7 Modelagem da amônia afluenta com atraso .....   | 138 |
| 5.8 Testes da RNA treinada no sistema supervisorío com novos dados .....                          | 140 |
| 6 CONCLUSÕES E PESQUISAS FUTURAS .....  | 144 |
| REFERÊNCIAS .....   | 146 |
| APÊNDICE I – A – RNA ETE A .....  | 151 |
| APÊNDICE I – B – RNA ETE C .....  | 151 |
| APÊNDICE I – C – RNA ETE D .....  | 151 |
| APÊNDICE I – D – RNA ETE E .....  | 152 |
| APÊNDICE II - <i>DASHBOARD</i> PARA AUTOMAÇÃO NO ESGOTO COM FOCO EM<br>TRANFORMAÇÃO DIGITAL ..... | 153 |
| ANEXO I – AUTORIZAÇÃO DO USO DE DADOS NA PESQUISA .....   | 158 |

## INTRODUÇÃO

O tratamento de esgoto brasileiro tem grande desafio pela frente devido o marco legal do saneamento, lei Nº14.026 de 15 de julho de 2020, que preconiza o atendimento mínimo de fornecimento de água e tratamento de esgoto, e os dados provenientes de seus processos, representa oportunidade visando eficiência operacional e por consequência maior volume e qualidade do esgoto tratado.

Uma técnica frequentemente empregada para o tratamento de águas residuais é o princípio de lodo ativado. Nesse processo, é introduzido oxigênio no esgoto para permitir que as bactérias existentes decomponham a matéria orgânica presente. Com o decorrer do tempo, os procedimentos de tratamento passaram a requerer tecnologias mais aprimoradas que minimizem os impactos ambientais e os custos ao máximo (PERES et al., 2021).

O procedimento de tratamento mediante lodo ativado essencialmente acelera o processo natural de decomposição da matéria orgânica ao oxigenar as bactérias nas águas residuais. Essa substância orgânica é composta por proteínas (40-60%), carboidratos (25-50%) e lipídios (10%) e serve como fonte de energia para esses microrganismos (FERREIRA E CORAIOLA, 2008).

O processo de tratamento inclui a fase preliminar em que as grades removem o material particulado, os tanques de aeração, onde ocorre a depuração da matéria orgânica e a desidratação mecânica do lodo, o final do tratamento fase sólida e envolve a separação do lodo prensado em líquido e sólido. E esses processos e tecnologias produzem uma série de dados, que permitem seu estudo visando eficiência.

No tratamento de águas residuais, é comum utilizar várias análises para orientar os processos em cada etapa. Uma das mais importantes é a Demanda Química de Oxigênio (DQO). A DQO é a concentração em massa de oxigênio que é igual à quantidade de um certo reagente químico consumido em seu ensaio, e isso pode estar ligado à carga orgânica presente nas águas residuais (GEERDINK, 2017). Em outras palavras, a relação entre DQO de entrada e saída pode fornecer informações sobre a eficácia da Estação de Tratamento de Esgoto (ETE), daí a frequente utilização desse ensaio.

Diversas abordagens podem ser utilizadas visando obter direcionamentos através de dados no tratamento de esgoto. Dentre elas é possível citar as Redes Neurais Artificiais (RNA)

(PADHMA, 2023), Algoritmos Genéticos (AG) (ARISMENDY, 2021), entre outras. As RNAs podem ser definidas como algoritmos computacionais com capacidade de aproximar o comportamento de determinado processo com precisão através da simulação (ARISMENDY et al., 2021).

Na arquitetura da rede, os neurônios da camada de entrada são as variáveis explicativas do processo e cada um desses são conectados a cada um dos neurônios da camada oculta, que por fim são conectados ao neurônio da camada de saída, que é a variável dependente ou de interesse. Através de algumas métricas e por meio do algoritmo de retro propagação de Levenberg-Marquardt é possível diminuir a função perda a medida que a rede passa pelos ciclos ou quantidade de épocas definidos (LI; VANROLLEGHEM, 2022).

Sendo assim, o setor de tratamento de esgoto no Brasil enfrenta obstáculos significativos para alcançar os objetivos estipulados pelo marco legal. A falta de infraestrutura moderna e a variabilidade dos processos operacionais resultam em desafios para manter a eficiência do tratamento, controlar custos e atender às normas ambientais vigentes. Além disso, embora os dados gerados pelos processos de tratamento sejam diversos e potencialmente valiosos, a falta de integração e análise contínua, dificulta o uso dessas informações para melhorar o desempenho das ETEs. Portanto, explorar tecnologias, como RNAs e técnicas de otimização, torna-se importante para transformar dados brutos em insights que possam melhorar a eficiência e a qualidade do tratamento, ao mesmo tempo em que podem minimizar os impactos ambientais e os custos operacionais.

## **Objetivo**

O objetivo desta dissertação é analisar e otimizar os processos de tratamento de esgoto por meio da aplicação de técnicas de análise de dados, com foco específico nos parâmetros de Demanda Química de Oxigênio (DQO) e amônia.

## **Objetivos específicos**

- Conduzir uma pesquisa bibliométrica para mapear as principais tendências e avanços na área;
- Realizar uma Revisão Sistemática da Literatura (RSL) para identificar trabalhos análogos relevantes no setor de saneamento mundial;

- Realizar uma *Survey* com profissionais do setor e entrevistas com especialistas para coletar dados sobre as práticas atuais, desafios e ideias a fim de justificar a aplicação de aprendizado de máquina no tratamento de esgoto;
- Desenvolver e avaliar a aplicação de RNA na previsão dos parâmetros DQO e amônia, explorando como essas tecnologias podem melhorar a eficiência e a sustentabilidade das estações de tratamento de esgoto;
- Comparar as métricas da modelagem com RNAs e outras técnicas.

### **Justificativa**

O tratamento de esgoto é um componente importante para a proteção ambiental e a saúde pública, uma vez que o tratamento eficaz, por meio da redução de parâmetros como DQO, DBO e amônia, é essencial para evitar impactos negativos nos corpos d'água e na qualidade do ambiente. Um estudo do Instituto Trata Brasil (2022) mostra os benefícios econômicos e sociais da expansão do saneamento no Brasil, como a valorização imobiliária, expansão do turismo, renda gerada pelo investimento entre outros.

Outro fato importante é que de acordo com estimativas da Organização Mundial da Saúde (OMS), cada R\$ 1 investido em saneamento gera uma economia de até R\$ 9 em saúde, dependendo da região e local (FUNASA, 2017; OMS, 2020). A crescente demanda por soluções mais eficientes e sustentáveis tem impulsionado a adoção de tecnologias digitais e automação no setor. Contudo, a implementação dessas tecnologias ainda enfrenta desafios significativos, principalmente na integração e no uso eficaz das informações geradas.

Esta dissertação justifica-se pela necessidade explorar o potencial das RNAs na modelagem e previsão dos parâmetros críticos, e fornece uma base empírica para a melhoria dos processos de tratamento. O estudo busca, portanto, apresentar o potencial na aplicação prática dessas tecnologias e que podem contribuir para a eficiência operacional e a sustentabilidade das ETEs no Brasil e especificamente na região metropolitana de São Paulo.

### **Contribuição Científica**

A contribuição científica desta dissertação está na integração de métodos de análise de dados e tecnologias emergentes no contexto do tratamento de esgoto. O estudo oferece uma análise detalhada da aplicação de redes neurais artificiais para a previsão e otimização dos parâmetros de DQO e amônia, áreas ainda pouco exploradas no setor. A pesquisa também contribui para o corpo de conhecimento existente ao fazer uma revisão sistemática da literatura,

identificando os trabalhos mais relevantes na área, além de incorporar uma análise qualitativa baseada em entrevistas com especialistas e uma *Survey* com profissionais do setor.

### **Método de pesquisa**

A pesquisa desenvolvida no presente trabalho é caracterizada como aplicada, pois visa resolver problemas específicos do setor de saneamento, otimizando processos operacionais em ETEs na região metropolitana de São Paulo por meio da aplicação de técnicas de aprendizado de máquina, e para isso possui autorização da Companhia de Saneamento Básico do Estado de São Paulo, a SABESP, anexo I. No que diz respeito aos objetivos, ela assume um caráter exploratório ao investigar padrões nos dados e testar abordagens como RNAs no contexto de tratamento de esgoto, além de ser explicativa ao buscar compreender relações causais entre variáveis operacionais, como o Tempo de Detenção Hidráulica (TDH), e resultados preditivos, como a DQO. Quanto à abordagem metodológica, adota-se um modelo quali-quantitativo, combinando a análise quantitativa de dados operacionais, utilizando técnicas como PCA, clusterização e modelagem preditiva, com uma análise qualitativa baseada em entrevistas semiestruturadas e na interpretação das percepções de profissionais do setor, proporcionando uma visão abrangente e integrada dos fenômenos analisados.

O método de pesquisa adotado nesta dissertação, baseado no *Design Science Research* (DSR) foi estruturado em cinco procedimentos complementares: pesquisa bibliométrica, RSL, entrevistas com especialistas, *Survey* com profissionais do setor e aplicação de RNAs em ETEs.

A pesquisa bibliométrica envolveu a coleta de publicações científicas em bases de dados relevantes, com análise da frequência de tópicos e identificação dos principais periódicos e conferências relacionados à área de estudo. Enquanto a RSL foi conduzida com base nos procedimentos descritos por Kitchenham (2007), utilizando a ferramenta Parsifal. Os estudos foram selecionados com base em critérios predefinidos de inclusão e exclusão, garantindo a relevância e qualidade dos trabalhos analisados.

Foram realizadas entrevistas com 6 especialistas do setor de saneamento e automação. Os dados foram analisados com a técnica de análise de conteúdo de Bardin, utilizando o software MAXQDA para visualização dos temas identificados. Já a *Survey* foi realizada com 69 profissionais de ETEs, utilizando perguntas em escala Likert para a coleta de dados quantitativos e qualitativos, complementadas por perguntas abertas.

A aplicação prática das RNAs foi realizada utilizando-se a linguagem de programação Python para a previsão dos parâmetros de DQO e amônia nas principais ETEs de São Paulo. Foram utilizados dados históricos para treinar e validar os modelos, buscando avaliar o impacto dessas técnicas na otimização dos processos operacionais, foi também explorada a utilização de técnicas complementares para entender a relação entre os parâmetros estudados, como a Análise de Componentes Principais (PCA), clusterização e visualização com *K-means*, e a análise SHAP das *features* utilizadas nas RNAs.

### **Organização do trabalho**

Inicialmente, na introdução, foram apresentados o objetivo da pesquisa, a justificativa para o estudo, a contribuição esperada e o método de pesquisa adotado. Esta seção estabelece o contexto e a importância da análise de dados em estações de tratamento de esgoto, focando em parâmetros como DQO e amônia.

O primeiro capítulo oferece uma base teórica para a compreensão do tema. São apresentados conceitos fundamentais relacionados à transformação digital e automação no tratamento de esgoto, abrangendo as fases do tratamento (preliminar, líquida e sólida). Também é abordada a Resolução CONAMA 430 e sua relevância para o setor, além da instrumentação analítica no saneamento, que inclui medições de pH, condutividade, entre outros. O capítulo também explora Redes Neurais Artificiais (RNAs), Análise de Componentes Principais (PCA), *K-means*, modelagem matemática e métricas pertinentes ao tema.

No capítulo 2, é realizada uma revisão da literatura do tema, detalhando o método utilizado para a análise quantitativa dos trabalhos relevantes na área. São examinados os locais e periódicos mais frequentes, oferecendo uma visão geral do estado da pesquisa sobre a transformação digital aplicada ao tratamento de esgoto além de uma revisão crítica dos 7 trabalhos mais relevantes sobre a aplicação de RNAs no tratamento de esgoto é apresentada.

No capítulo 3, é apresentada uma análise qualitativa das entrevistas com especialistas do setor. Utiliza-se a técnica de análise de conteúdo de Bardin para interpretar as informações obtidas, e com a utilização do software MAXQDA são criadas nuvens de palavras para visualizar as recorrências nas respostas obtidas nas entrevistas. Além disso, são discutidos a validação do objetivo da pesquisa, os potenciais usos da automação, o impacto da transformação digital no setor e as perspectivas futuras segundo os especialistas entrevistados.

Já no capítulo 4, *Survey* com Profissionais do Setor, uma *Survey* foi realizada com diversos profissionais do setor para obter uma visão abrangente sobre a temática. O capítulo inclui a caracterização dos respondentes e apresenta as respostas a perguntas em escala Likert, divididas entre o corpo operacional e gerencial, além de perguntas abertas. São geradas nuvens de palavras a partir das respostas obtidas e realizada a interpretação dos dados qualitativos obtidos.

Por fim, no capítulo 5, o último capítulo detalha a aplicação da RNAs no tratamento de esgoto, com foco na previsão de parâmetros importantes. É realizada uma caracterização dos dados e ETEs que são objetos de estudo e são apresentados os resultados obtidos com as RNAs treinadas usando a linguagem de programação Python, fornecendo insights sobre a eficácia e a aplicabilidade das técnicas de modelagem na prática.

## **1 FUNDAMENTAÇÃO TEÓRICA**

Este capítulo oferece uma base teórica para a compreensão do tema. São apresentados conceitos fundamentais acerca de automação, tratamento de esgoto, as análises que serão modeladas mais a frente e também sobre as técnicas de aprendizado de máquina.

### **1.1 Transformação digital no tratamento de esgoto**

A transformação digital é um processo de mudança impulsionado por novas tecnologias nas empresas. Sendo caracterizada, principalmente, pela computação embarcada e a conectividade (EBERT; DUARTE, 2018). Os benefícios listados são aumento nas vendas ou produtividade, inovação nas criações de valor, novas formas de interação com os clientes, entre outros (MATT; HESS; BENLIAN, 2015).

Vial (2019) através de uma revisão sobre assunto e outras definições, identifica como um processo que visa melhorar algo por meio de mudanças em suas propriedades e de combinações de tecnologias de informação, computação, comunicação e conectividade.

No sentido da aplicação da transformação digital em serviços avançados, diversos requisitos podem ser necessários. Entre eles podem ser citados inteligência, conexão e recursos analíticos. Inteligência como sendo o aprimoramento e incorporação de produtos inteligentes para detecção e captura de dados operacionais, enquanto os recursos de conexão envolvem transmissão sem fio e de rede de ativos interconectados e os recursos analíticos é obtenção de insights preditivos por meio de processamento de dados lógicos (SAUL; GEBAUER, 2018).

No tratamento de esgoto, a transformação digital atrelada as análises de qualidade que são realizadas nos processos de tratamento e a conectividade dessas informações, podem permitir a capacidade de antecipar manobras necessárias no processo de tratamento.

Com relação a automação, e por meio do controle de processo, diversos benefícios são possíveis em um processo industrial, como aumento da produtividade, minimizar a entrada manual de registro e medições, simplificar a realização de estudos econômicos, aumentar o acesso a dados para análise do processo e problemas operacionais entre outros (MEHTA;

REDDY, 2014).

A literatura de automação usualmente divide os componentes acerca do tema em 5 principais camadas. Na primeira estão os dispositivos de campo, na segunda os controladores, na terceira os sistemas de supervisão, enquanto na quarta estão os gerenciadores das plantas e na quinta os *Enterprise Resource Planning* - Sistema integrado de gestão empresarial (ERPs) conforme a Figura 1 (MARTINEZ et al, 2021). Importantes ressaltar que níveis cada vez maiores de automação permitem controles mais aprimorados e por consequência melhor eficiência operacional, e para isso todo o volume de dados provenientes dos processos são fundamentais.

Figura 1 - Pirâmide de automação (exemplos no saneamento)



Fonte: Adaptado de MARTINEZ et al (2021)

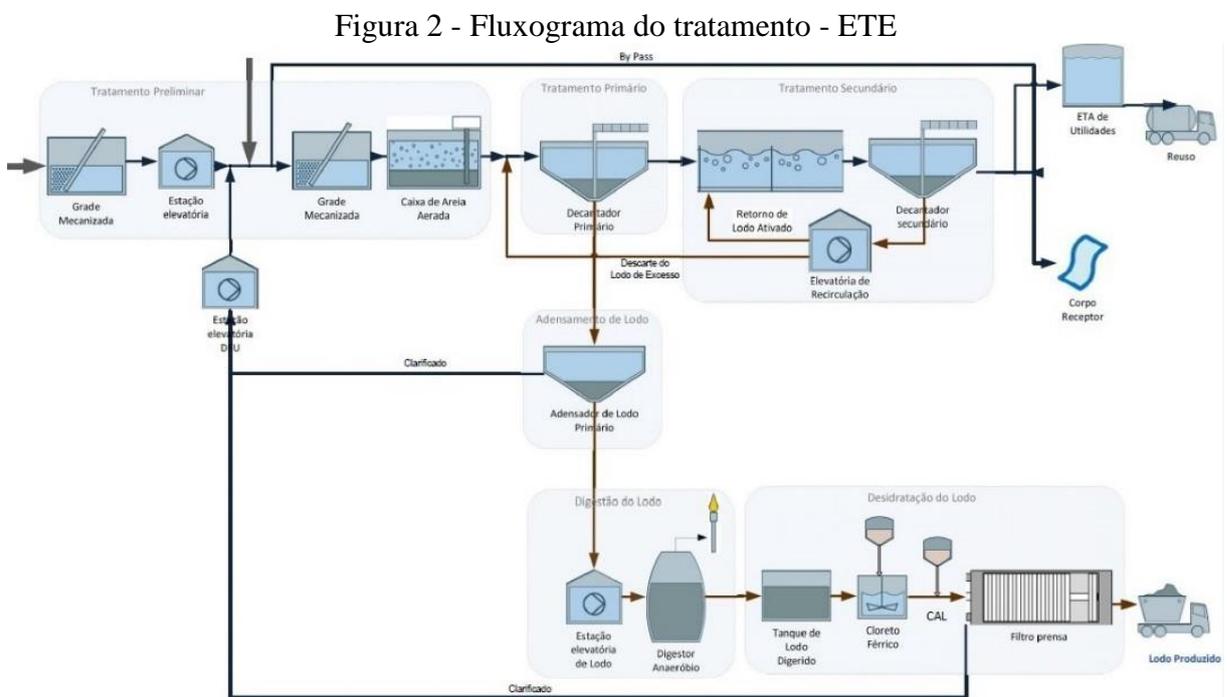
No caso do tratamento de esgoto, os dispositivos de campo da primeira camada usualmente são os elementos de controle como válvulas e bombas enquanto a instrumentação analítica consiste em analisadores como pH, condutividade e afins, e a instrumentação de controle de processo em grandezas como vazão e nível. Já as outras camadas, controle e supervisão, sendo muito parecidas com outros setores, com os Controladores Lógico Programáveis (CLPs) e sistemas supervisórios.

É possível correlacionar a complexidade com o valor dos dados, onde há o caminho de informação para otimização, do *Business Intelligence* para *Business Analytics*, saindo da análise descritiva e diagnóstica, ou seja, o que aconteceu e por que aconteceu, para a análise preditiva e prescritiva, o que acontecerá e como fazer para que algo aconteça, trazendo previsibilidade para os processos produtivos.

A automação em tratamento de esgoto, normalmente é composta por instrumentos que monitoram o processo, enviam essas informações aos controladores, que além de atuar nos elementos finais de controle presentes em cada etapa, fazem a comunicação com o sistema de supervisão da planta, esse último conectado ao banco de dados e historiador quando existente.

A respeito do tratamento de esgoto, a tecnologia mais comum é o processo de lodo ativado, onde basicamente é enviado oxigênio ao esgoto para que as bactérias existentes possam decompor a matéria orgânica ali presente. Cada vez mais os processos de tratamento exigem tecnologias mais eficientes com menor impacto ambiental e custo possível (PERES et al, 2021).

O processo de tratamento por lodo ativado consiste basicamente em acelerar o processo que ocorre na natureza de decomposição da matéria orgânica, oxigenando as bactérias presentes no esgoto (FERREIRA; CORAIOLA, 2008). Na Figura 2 temos um fluxograma que representa o processo de uma ETE com essas características. Uma ênfase para a separação das fases de tratamento (líquida e sólida) e para onde ocorre essa depuração da matéria orgânica, nos tanques de aeração.



Fonte: Do autor (2024)

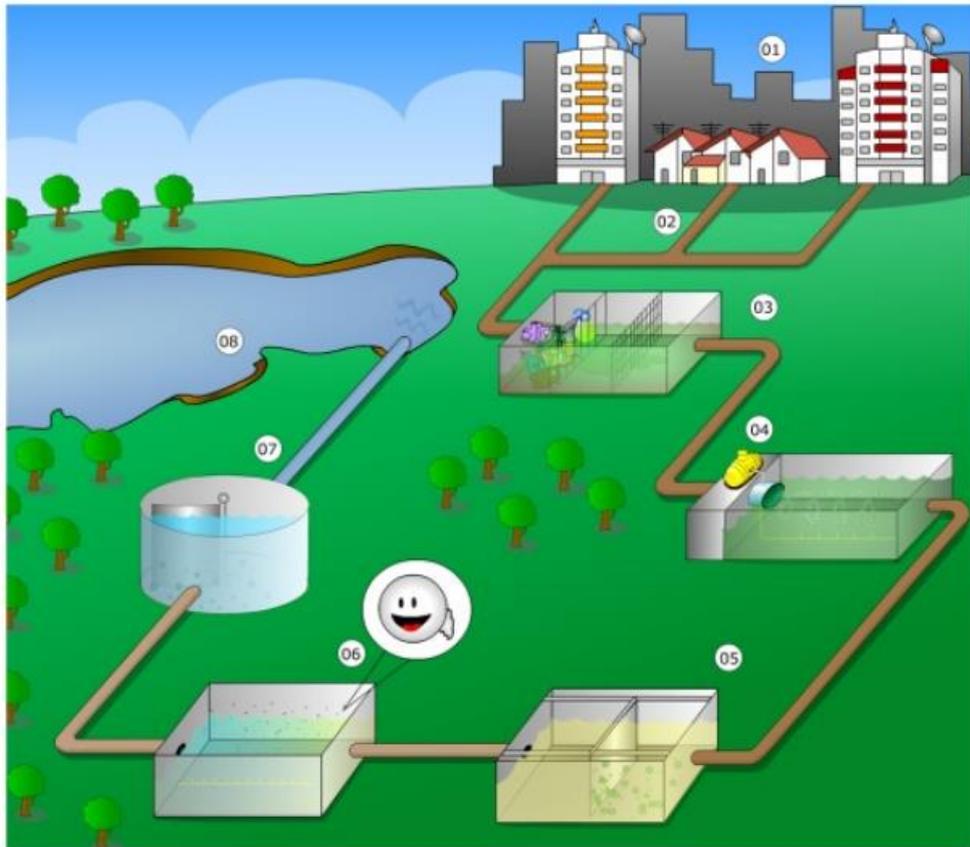
A Figura 2 apresenta cada etapa do processo de tratamento como a fase preliminar onde as grades fazem a primeira retirada de material particulado. Nos tanques de aeração onde ocorre a depuração da matéria orgânica e a desidratação mecânica de lodo, última fase do tratamento

fase sólida onde o lodo é prensado e separa o que é líquido do que é sólido.

A Figura 3 traz uma representação usual da fase líquida no tratamento de esgoto, onde as etapas são numeradas de 01 a 08. A seguir há uma breve descrição de cada uma delas: 01 – Cidade: Após ser utilizada nas residências para higiene pessoal, alimentação e limpeza, a água vira esgoto. Ao deixar as casas, o esgoto é direcionado para as redes coletoras de esgoto. 02 - Rede de Esgotos: Rede de tubulações que transporta o esgoto das residências até as ETEs. Inclui coletores, troncos e interceptores, que são diferentes tipos de tubulações usadas para coletar e direcionar o esgoto para a estação de tratamento. 03 – Grades: Estruturas que removem sólidos grandes do esgoto, como papel, plástico e tampinhas. As grades funcionam como uma peneira, permitindo que o esgoto passe enquanto retêm materiais sólidos maiores, prevenindo obstruções e danos em equipamentos posteriores. 04 - Caixa de Areia: Tanque onde o esgoto é transportado para remover a areia contida nele. Na caixa de areia, a velocidade do esgoto é reduzida, permitindo que partículas de areia e outros sólidos pesados se depositem no fundo. 05 - Decantador Primário: Tanques onde ocorre a sedimentação das partículas mais pesadas. No decantador primário, o esgoto flui lentamente para que partículas sólidas possam se sedimentar e formar um lodo no fundo, separando-se da água residual. 06 - Tanques de Aeração: Tanques onde ar é fornecido para estimular a atividade dos microrganismos que decompõem a matéria orgânica. Os microrganismos presentes nos tanques de aeração se alimentam da matéria orgânica no esgoto, transformando-a em lodo e reduzindo a carga poluidora, importante ressaltar que se espera que areia e outros tipos de material inorgânicos já tenham sido retirados em etapas anteriores, pois a presença desses materiais pode acarretar consumo de energia desnecessário. 07 - Decantador Secundário: Tanques onde os sólidos restantes são separados da parte líquida. Nos decantadores secundários, os sólidos que não foram removidos anteriormente se sedimentam no fundo, enquanto a parte líquida, agora com 90% das impurezas removidas, é separada. 08 - Rio: O ponto final onde a água tratada é lançada. A água, que ainda não pode ser bebida, é lançada nos rios ou reaproveitada para usos não potáveis, como limpeza

de ruas, praças e irrigação de jardins (JORDÃO, 2005).

Figura 3 - Tratamento fase líquida

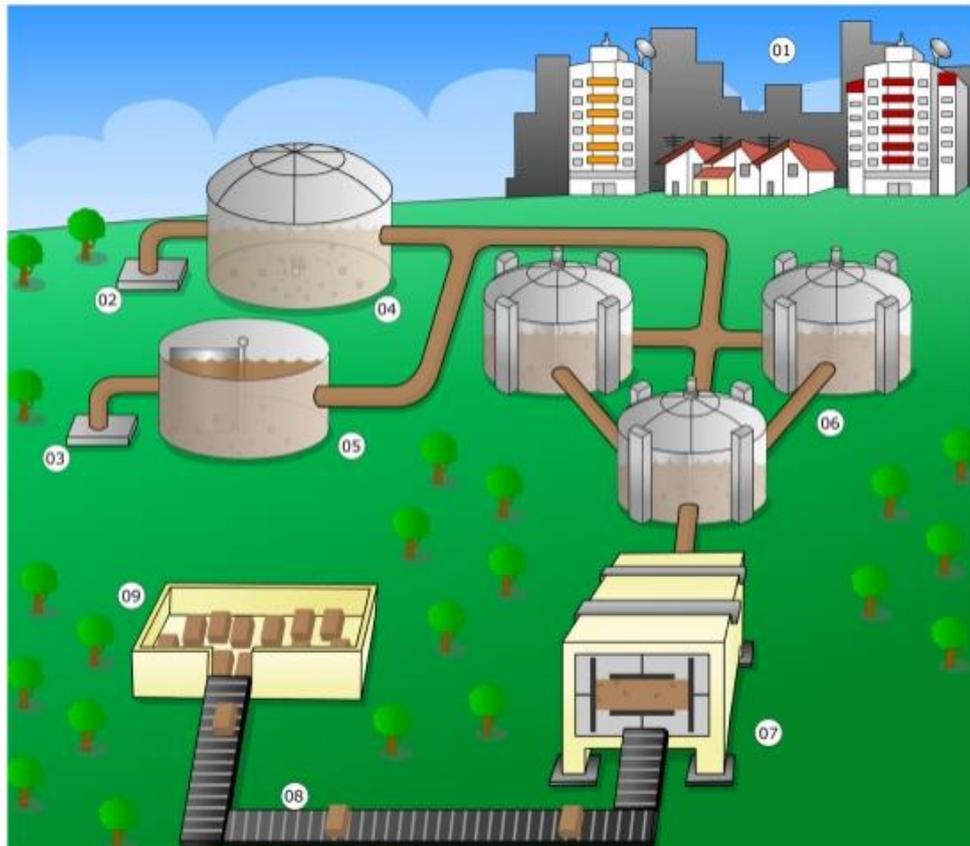


Fonte: SABESP (2024)

Da mesma maneira, a Figura 4 traz uma representação da fase sólida no tratamento de esgoto, onde as etapas são numeradas de 01 a 09. A seguir temos a descrição de cada uma delas: 01 – Cidade: Após ser utilizada nas residências para higiene pessoal, alimentação e limpeza, a água vira esgoto. Assim como na fase líquida, o esgoto sai das casas e segue pela rede de esgoto até a estação de tratamento. Na etapa 02 - Entrada do Lodo Primário, ocorre a separação da água e do sólido através da sedimentação das partículas mais pesadas, semelhante aos decantadores. O lodo do decantador secundário é tratado na etapa 03 - Entrada do Lodo Secundário, pelo processo de adensamento por flotação nos flotadores. Na etapa 04 - Adensadores, o lodo torna-se mais concentrado através da separação de uma parte da água presente. O processo de flotação na etapa 05 – Flotadores, separa a água do sólido através da introdução de água com microbolhas de ar. Na etapa 06 - Digestores, o lodo é estabilizado por meio da ação de microrganismos anaeróbicos, que decompõem a matéria orgânica, gerando biogás (metano) e água, além de reduzir odores desagradáveis. A etapa 07 - Filtros Prensa, desidratam o lodo removendo a maior parte da água e aumentando a concentração de sólidos.

O lodo desidratado é transportado na etapa 08 – Esteira, até a próxima etapa. O lodo desidratado, conhecido como torta, é armazenado e depois enviado para a etapa 09 - Tortas para Aterro Sanitário, onde é disposto em aterros sanitários apropriados para disposição final (JORDÃO, 2005).

Figura 4 - Tratamento fase sólida



Fonte: SABESP (2024)

O sistema de tratamento de esgoto por lodos ativados, desenvolvido por Arden e Lockett em 1914 no Reino Unido, foi inicialmente concebido para operar em tanques de batelada, mas evoluiu para sistemas de fluxo contínuo, amplamente utilizados atualmente. Esse sistema é eficiente no tratamento de despejos domésticos e industriais, especialmente quando há necessidade de alta qualidade do efluente e espaço reduzido. No entanto, o sistema exige uma maior mecanização, o que implica em uma operação mais sofisticada e em maiores custos de energia. O tratamento biológico ocorre em tanques de aeração e decantadores, onde a

biomassa remove matéria orgânica e, em determinadas condições, matéria nitrogenada (VON SPERLING, 1997).

No processo de lodos ativados, a biomassa é recirculada, aumentando a concentração de sólidos em suspensão no tanque de aeração, o que favorece a remoção de matéria orgânica. A sedimentação dos sólidos no decantador secundário permite a clarificação do efluente final. A floculação das bactérias facilita essa sedimentação, garantindo a alta eficiência do sistema. A gestão do lodo biológico excedente, que precisa ser tratado e descartado, é fundamental para o equilíbrio do sistema. Além da remoção de matéria carbonácea, o sistema pode ser adaptado para a remoção biológica de nitrogênio e fósforo, sendo essas etapas essenciais em estações de tratamento mais avançadas (VON SPERLING, 1997).

Um conceito importante no tratamento de esgoto é o Tempo De Detenção Hidráulica (TDH), e nesse trabalho ele se relaciona com as modelagens que serão realizadas nos passos futuros.

O Tempo de Detenção Hidráulica (TDH) é um parâmetro importante no entendimento dos processos e operação de ETEs. Ele é definido como o tempo necessário para que uma unidade de volume de esgoto percorra determinado sistema de tratamento, desde a entrada até a saída. O TDH é calculado pela relação entre o volume do tanque de tratamento e a vazão de esgoto que passa por ele, ou a própria Estação caso o interesse seja o TDH da ETE. De acordo com NBR 12209 e as recomendações das condições para os projetos de ETEs, esse parâmetro é fundamental no dimensionamento das Unidades de tratamento e a relação de seus volumes com a vazão a ser tratada, por exemplo, no caso do decantador primário o TDH deve ser inferior a 3 horas para vazão média e superior a 1 hora para vazão máxima (MIKI, 2022). A eq. (1) apresenta a maneira de calculá-lo.

$$TDH = \frac{\text{Volume da Unidade (m}^3\text{)}}{\text{Volume total da planta}(\frac{\text{m}^3}{\text{dia}})} \times 24 \quad (1)$$

No contexto desse trabalho, o TDH entre etapas do processo e sobretudo onde ocorrem algumas análises, podem fornecer informações importantes para modelagens onde se deseja encontrar correlações entre dados, não somente no mesmo instante de tempo, mas também considerando um determinado “*delay*”, “*lag*” ou atraso entre as variáveis, que no contexto do tratamento de esgoto, é definido justamente pelo TDH entre os tanques.

### *1.1.1 Indicadores no tratamento de esgoto*

O Índice de Eficiência Energética (IEE) é uma métrica amplamente utilizada para avaliar o consumo de energia em diversos setores, sendo adaptado conforme o segmento e o processo produtivo. No caso das ETEs, o IEE relaciona o consumo de energia elétrica com o volume de esgoto tratado. Ele é calculado como a razão entre o consumo total de energia elétrica (kWh) e o volume de esgoto tratado ( $m^3$ ), sendo normalmente expresso em  $kWh/m^3$ . Um IEE baixo indica maior eficiência energética, ou seja, que a Estação utiliza menos energia para tratar a mesma quantidade de esgoto (YAMAKAWA, 2014).

No entanto, é importante ressaltar que o IEE avalia apenas a eficiência energética e não considera, isoladamente, indicadores de qualidade do efluente. Por isso, é descrito como um indicador "do ponto de vista energético", sendo necessário complementá-lo com outras métricas para uma análise mais abrangente.

Esse indicador é particularmente relevante no contexto das ETEs porque etapas como a aeração e o funcionamento das elevatórias de esgoto consomem muita energia elétrica. Estratégias como controle avançado de oxigênio dissolvido, otimização da recirculação de lodo e ajustes no controle de nível e vazão nas elevatórias podem reduzir significativamente o IEE. Além de promover maior eficiência energética, essas ações também contribuem para a redução dos custos operacionais associados ao tratamento de esgoto.

A eficiência de remoção de DBO mede a quantidade de matéria orgânica biodegradável efetivamente retirada do esgoto ao longo do processo de tratamento. É calculada a partir da relação entre a DBO efluente e do afluente. Esse indicador é expresso em % e reflete a capacidade do sistema de remover carga poluente, garantindo que o efluente final atenda aos padrões ambientais (MIKI, 2022).

O indicador de custo de produto químico é essencial para monitorar a eficiência econômica das ETEs em relação aos coagulantes, floculantes, agentes desinfetantes e oxidantes utilizados no tratamento. Ele mede o gasto financeiro com produtos químicos em relação ao volume de esgoto tratado ou massa de lodo retirada do sistema, e pode ser expresso em  $R\$/m^3$  ou  $R\$/Tonelada$  de lodo seco.

Esses produtos desempenham um papel importante no processo de tratamento de esgoto, garantindo a remoção de poluentes e o cumprimento dos padrões de qualidade para descarte ou

reuso do efluente. No entanto, representam uma parcela significativa dos custos operacionais das ETEs. Monitorar esse indicador permite identificar oportunidades de otimização e controle de custos, levando em consideração que esse percentual pode representar de 3,1% a 4,2% do custo total de uma empresa de saneamento (JORDÃO, 2008).

Diversos fatores podem influenciar o custo relacionado a químicos. Como por exemplo: a qualidade do esgoto afluente e maior carga orgânica, ajustes nos processos e o nível de automação e monitoramento presentes, entre outros fatores.

## **1.2 Instrumentação no tratamento de esgoto**

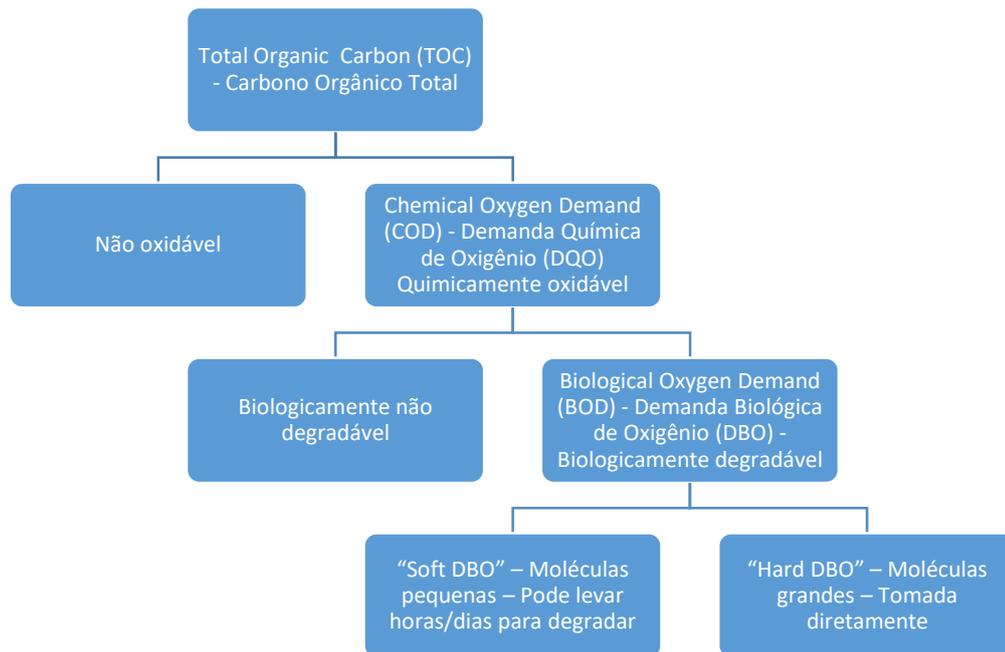
Instrumentação pode ser definido como um sistema que detecta uma variável de processo, quantifica-a e converte essa informação, geralmente para um sinal elétrico, para que posteriormente esse dado possa ser trabalhado, armazenado e acessado (PADMANABHAN, 2012).

A variável medida ou variável de processo é a medida da variável física de interesse no sistema, por exemplo temperatura ou pressão. O elemento primário converte o valor medido em uma forma adequada para posterior conversão, enquanto o elemento secundário apresenta esse valor já em unidade de engenharia ao usuário (PADMANABHAN, 2012). No caso das ETEs, diversas são as grandezas medidas e enviadas aos sistemas de controle através de sinais elétricos, como vazão, nível e pressão para instrumentação de controle de processo e pH, condutividade e OD para instrumentação analítica.

A seguir temos uma descrição dos parâmetros citados e sua relação com o processo de tratamento de esgoto:

Para controle de processos em ETEs usualmente são utilizadas medidas para determinação da matéria orgânica existente. O Carbono Orgânico Total, do inglês *Total Organic Carbon* (TOC), a DQO e a Demanda Bioquímica de Oxigênio (DBO), do inglês *Biological Oxygen Demand* (BOD), (DAVIES, 2005). A Figura 5 traz a relação entre esses ensaios.

Figura 5 - Relação entre métricas para determinação da matéria orgânica



Fonte: Adaptado de DAVIES (2005)

Conforme observado, a DQO traz o dado do que é quimicamente oxidável na amostra, enquanto a DBO daquilo que é biologicamente degradável. É comum que se utilize essas medidas para identificar a eficiência do processo de tratamento, considerando que se tenham esses dados para entrada e saída da Estação, sendo esses, inclusive, parâmetros para lançamento de efluente tratado nos corpos hídricos a depender de sua classe.

O pH é o potencial hidrogeniônico de uma solução e é definido como o logaritmo negativo da concentração de íons de hidrogênio, serve para medir o grau de acidez, neutralidade ou alcalinidade, e no tratamento de esgoto normalmente varia entre 6,5 e 7,5 (JORDÃO, 2005). Para determiná-lo segue eq. (2):

$$pH = \log_{10} \frac{1}{H^+} \quad (2)$$

Onde:  $H^+$ : é a concentração do Hidrogênio em íons-grama por Litro de solução.

No tratamento de esgoto, o valor de pH pode influenciar as atividades biológicas e químicas nos tanques e quando inadequado pode prejudicar os microrganismos comprometendo a remoção da matéria orgânica (METCALF; EDDY, 2014).

A condutividade é uma medida utilizada para determinação de capacidade de conduzir corrente elétrica por determinada solução, eq. (3) (APHA, 1917). Normalmente, moléculas de compostos orgânicos não são bons condutores elétricos, por isso a condutividade é um parâmetro utilizado em algumas etapas do tratamento de esgoto.

$$G = \frac{1}{R} \quad (3)$$

Onde, G: é a condutância, e R: é a resistência.

Valores altos de condutividade no tratamento de esgoto, podem indicar altas concentrações de sais que afetam a saúde dos microrganismos, por isso, muitas vezes ele é utilizado e relacionado a qualidade do efluente final em Estações (DROSTE, 1997).

Em águas residuais, os Sólidos Suspensos Totais (SST) são compostos por substâncias orgânicas e inorgânicas, tais como partículas de carboidratos, proteínas e lipídios para o primeiro caso e partículas minerais e areia para o segundo. O SST é um parâmetro importante pois pode afetar as características óticas de absorção e dispersão de luz solar. Além disso, valores elevados de SST no efluente final podem representar a perda de biomassa e redução da eficiência de remoção da matéria orgânica (DU et al., 2022).

O OD em águas residuais depende das atividades físicas, químicas e bioquímicas do corpo d'água (APHA, 1917). No tratamento de esgoto, é usualmente medido nos tanques de aeração, sendo de suma importância o controle dessa variável nessa etapa, devido custo com energia elétrica elevado.

A análise do Potencial de Oxido-Redução (ORP), é um indicador da capacidade do sistema de realizar reações de oxidação e redução, e essas reações mensuram o comportamento de muitos componentes químicos presentes na água no tratamento de esgoto, sendo possível a partir dessa análise, ajustar as condições afim de evitar a formação de subprodutos tóxicos ao tratamento. As reatividades e mobilidade de componentes químicos em sistemas biológicos, como Fe, S, N e C (APHA, 1917; CHERNICHARO, 2007).

Baixas concentrações de amônia normalmente estão associadas a água potável, águas superficiais ou efluentes de águas residuais nitrificadas de boa qualificada (APHA, 1917).

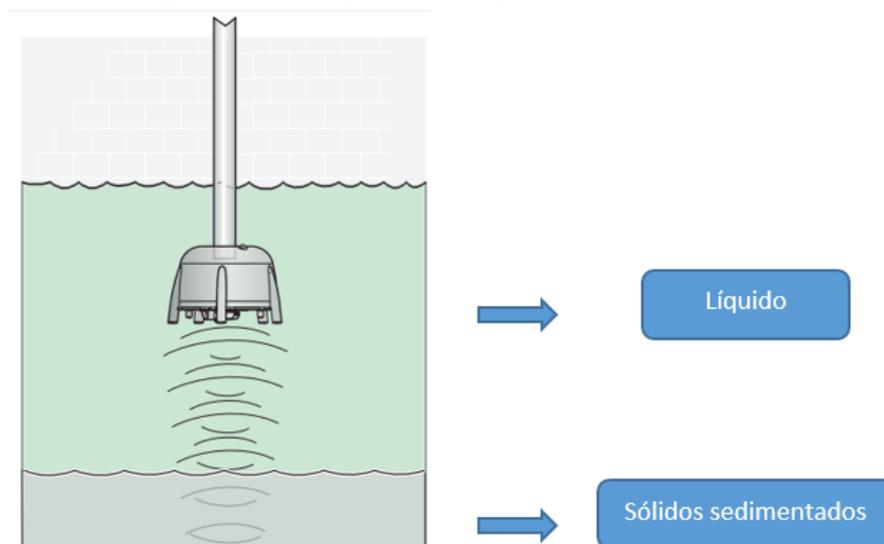
A utilização do cloro no abastecimento de água e tratamento de águas residuais tem o

objetivo principal de desativar ou destruir microrganismos produtores de doenças (APHA, 1917). Para o fornecimento de água de reuso, por exemplo, é necessária uma concentração mínima do mesmo visando a segurança na utilização da água.

A turbidez é uma propriedade ótica que faz com que a luz seja espalhada e absorvida. A clareza da água é um parâmetro muito utilizado em diversos processos produtivos afim de determinação da qualidade da água utilizada. Nos processos de tratamento de esgoto indica a eficiência dos processos de sedimentação e filtração além de ser um parâmetro legal para disposição do efluente nos córregos. Na água, a turbidez é causada por matéria suspensa e coloidal, como lodo, argila, matéria orgânica e inorgânica (APHA, 1917).

A medição do manto de lodo é um parâmetro importante nos processos de tratamento de esgoto, especialmente em unidades como decantadores (primários e secundários). O manto de lodo, ou nível da lama, é a camada formada pelos sólidos sedimentados no fundo do tanque, resultante do processo de separação entre a fase líquida e os sólidos. Essa medição é realizada por meio de equipamentos específicos, como medidores ultrassônicos ou régua graduada, que detectam a altura em que o lodo sedimentado se encontra. Nos medidores ultrassônicos, por exemplo, um feixe de ondas sonoras é emitido e retorna ao sensor ao encontrar a interface entre o lodo e a água, permitindo determinar a profundidade do manto, exemplo na Figura 6.

Figura 6 - Exemplo de medição de manto de lodo



Fonte: Manual SONATAX - Hexis (2024)

A importância dessa medição está diretamente relacionada à eficiência operacional e à estabilidade do sistema de tratamento. Manter o manto de lodo em uma altura adequada é fundamental para evitar problemas como arraste de sólidos para o efluente final no caso dos

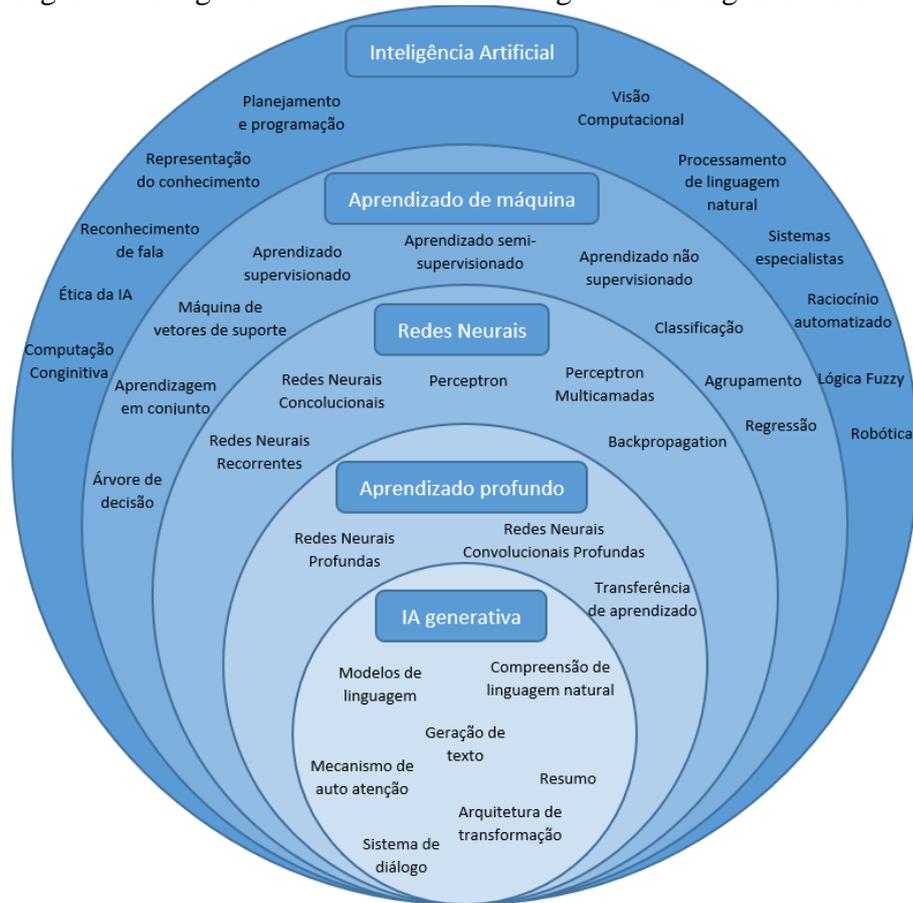
decantadores secundários ou para o tanque de aeração, no caso dos decantadores primários, que pode comprometer a qualidade do esgoto tratado e gerar não conformidades ambientais, além de consumo energético desnecessário. Outro ponto, é que o controle do manto de lodo permite otimizar a recirculação de lodo para os tanques de aeração, garantindo um equilíbrio na biomassa ativa e promovendo uma remoção eficiente de matéria orgânica e nutrientes. Portanto, a medição regular do manto de lodo é essencial para um controle eficiente do processo e para a sustentabilidade operacional das ETEs. A Figura 6 mostra um exemplo dessa medição feita por um medidor ultrassônico.

A partir do próximo item serão tratadas as técnicas de aprendizado de máquina utilizadas nessa dissertação, suas características, métricas de avaliação e por fim o tratamento de dados necessários para a modelagem.

### **1.3 Técnicas de análise de dados e aprendizado de máquina**

A Inteligência Artificial é uma das áreas da ciência da computação que visa criar sistemas que executem serviços que normalmente são realizados por seres humanos. Já o *machine learning* (aprendizado de máquina) é uma das subáreas que estuda o uso de algoritmos na identificação de padrões e melhoria desses mesmos algoritmos ao longo do tempo. Enquanto as RNAs, como visto anteriormente, são um tipo específico de aprendizado de máquina baseado no cérebro humano e seus neurônios, muito eficazes em problemas complexos. A Figura 7 traz um diagrama de Venn das relações dessas tecnologias (PESATI, 2024).

Figura 7 - Diagrama de Venn das Tecnologias da Inteligência Artificial



Fonte: Adaptado de PESATI (2024).

A Análise de Componentes Principais (PCA) é uma técnica de redução de dimensionalidade muito utilizada em ciência de dados, que tem como objetivo transformar um conjunto de variáveis possivelmente correlacionadas em um conjunto de variáveis linearmente não correlacionadas, chamadas de componentes principais. No contexto da pesquisa, onde diversos parâmetros de qualidade são analisados simultaneamente, o PCA pode ser extremamente útil para identificar as variáveis que mais contribuem para a variabilidade dos dados. Isso não só simplifica a análise, mas também facilita a visualização e interpretação dos dados, permitindo focar nas características mais importantes sem perder informações cruciais (JOLLIFFE, 2002).

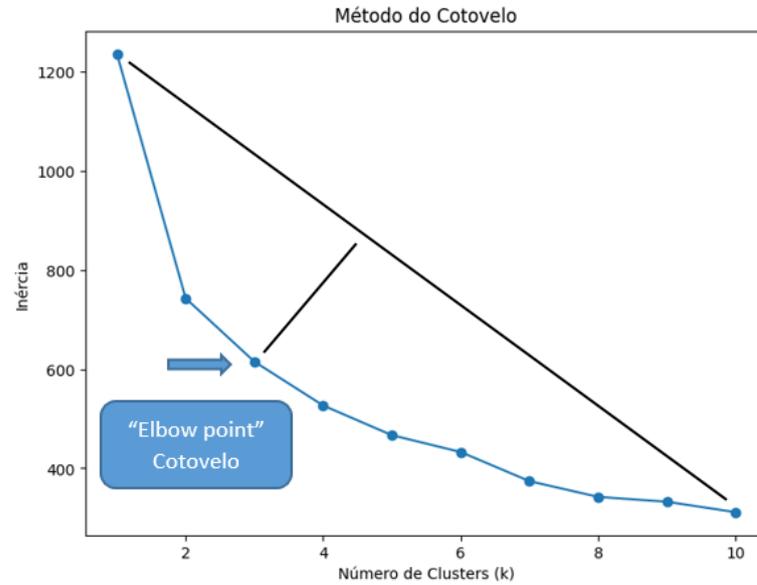
Ao identificar os componentes principais, o PCA ajuda a entender quais parâmetros têm maior impacto na qualidade de uma determinada etapa e como eles interagem entre si. Além disso, essa técnica pode ser integrada ao processo de modelagem preditiva, como uma etapa de pré-processamento, para melhorar a precisão e eficiência dos modelos desenvolvidos.

O *K-means* é um algoritmo de *clustering* que agrupa dados em  $k$  *clusters* ou grupos baseados em características semelhantes, minimizando a variabilidade dentro dos *clusters* enquanto maximiza a variabilidade entre os *clusters*. No contexto do tratamento de esgoto, o *K-means* pode ser usado para segmentar as ETEs com base em características operacionais e de qualidade, permitindo uma análise mais detalhada e orientada à necessidade. Por exemplo, ao aplicar *K-means* aos dados coletados de várias estações, é possível identificar padrões de desempenho, segmentando as estações em grupos com comportamentos semelhantes, ou por etapas do tratamento, o que pode ser útil para implementar estratégias de otimização específicas para cada grupo. Além disso, o *K-means* pode ser utilizado para detectar padrões sazonais ou operacionais nos dados, ajudando a identificar condições que levam a desvios na qualidade no processo (HARTIGAN, 1979).

A técnica do cotovelo normalmente é utilizada em problemas de clusterização, como no método *K-means*, para determinar o número ideal de *clusters* ( $k$ ). Esse método consiste em ter o agrupamento para diferentes valores de  $k$  e a avaliação da soma dos erros quadráticos dentro dos *clusters* (WCSS - *Within-Cluster Sum of Squares*). Ao aumentar o número de *clusters* no *K-means*, é comum que as distinções entre os *clusters* fiquem cada vez menores, enquanto as diferenças dentro de cada *clusters* tendem a aumentar. Por isso, é essencial encontrar um equilíbrio que maximize a homogeneidade das observações dentro de cada agrupamento, ao mesmo tempo em que garante que os *clusters* sejam os mais distintos possível entre si. Ou seja, à medida que  $k$  aumenta, o WCSS tende a diminuir, pois mais *clusters* implicam em uma melhor definição dos grupos, mas isso também pode levar à perda de generalização.

O "cotovelo" é identificado no gráfico que relaciona  $k$  ao WCSS, como o ponto em que a redução no WCSS começa a desacelerar e passa a não compensar. Antes desse ponto, adicionar *clusters* resulta em uma grande melhora na explicação dos dados. Após o cotovelo, os ganhos são marginais, indicando que o aumento de  $k$  pode ser desnecessário e levar ao *overfitting*, ou seja, o sobre ajuste do modelo, impedindo a generalização. Assim, o valor de  $k$  correspondente ao cotovelo é considerado o mais adequado, equilibrando simplicidade e precisão na modelagem dos dados. Essa técnica, apesar de visual e subjetiva, é uma ferramenta prática para decidir a quantidade de *clusters* em diversas aplicações de análise de dados. Na prática e de maneira matemática, também pode ser encontrado pelo ponto mais distante na curva de uma reta traçada entre o primeiro e o último ponto da mesma, exemplo na Figura 8 (BHOLOWALIA, 2014).

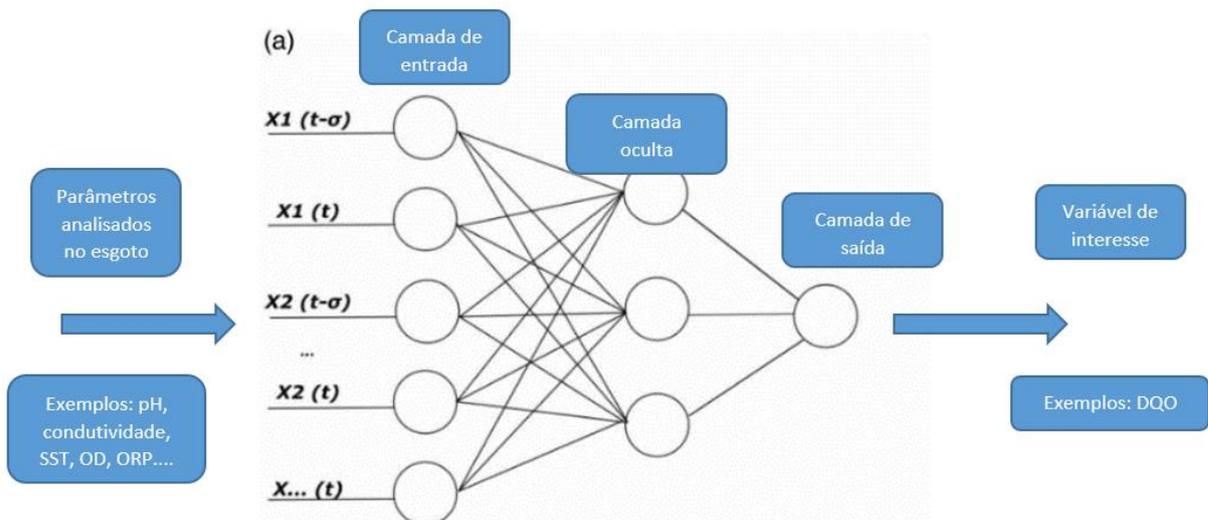
Figura 8 - Exemplo do método do cotovelo



## 1.4 Redes Neurais Artificiais

As Redes Neurais Artificiais podem ser definidas como algoritmos computacionais com capacidade de aproximar o comportamento de determinado processo com precisão através da simulação (ARISMENDY et al, 2021). Na Figura 9 é possível observar um exemplo para arquitetura de rede neural artificial e a representação de seus neurônios.

Figura 9 - Redes Neurais Artificiais (Exemplo no saneamento)



Fonte: Adaptado de LI; VANROLLEGHEM, (2022).

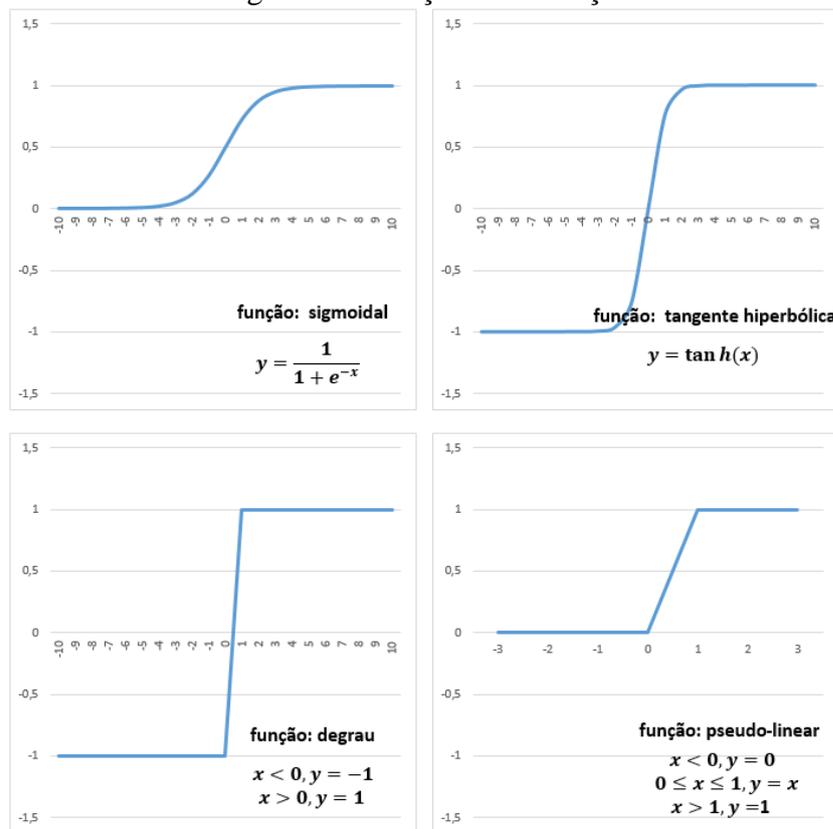
Na arquitetura da rede, os neurônios da camada de entrada são as variáveis explicativas do processo e cada um desses são conectados a cada um dos neurônios da camada oculta, que

por fim são conectados ao neurônio da camada de saída, que é a variável dependente. Através de algumas métricas e por meio do algoritmo de retro propagação de Levenberg-Marquardt é possível diminuir a função perda a medida que a rede passa pelos ciclos ou quantidade de épocas definidos (LI; VANROLLEGHEM, 2022).

Nas redes neurais do tipo *feed-forward*, a informação flui em apenas um sentido, e não há memória daquilo que já foi processado. Já nas redes recorrente, existe uma estrutura que permite reter informações anteriores, permitindo a inclusão do fator temporal na rede neural. Uma vertente das redes neurais recorrentes é o *Long Short-Term Memory* (LSTM) que surgiu como solução para os problemas com o gradiente das redes neurais comuns, e funciona incluindo “*gates*” ou chaves afim de permitir armazenamento ou não de memórias de longo prazo no algoritmo (VAN HOUDT, 2020).

A função de ativação em uma rede neural é um componente crucial que determina a saída de um neurônio com base no valor do sinal de excitação. Após o cálculo do sinal de excitação, este se torna um parâmetro em uma outra equação que, com base no valor desse sinal, determina a saída do neurônio, caracterizando-se como a função de ativação. Esta função, também conhecida como limiar lógico (*threshold*), pode assumir muitas formas ou métodos, sendo simples ou complexa. A função de ativação é responsável por enviar para a saída do neurônio o valor repassado pelo sinal de excitação. Na Figura 10, são apresentadas algumas típicas funções de ativação: a) degrau; b) pseudo-linear; c) sigmoidal; d) tangente hiperbólica. A saída de um neurônio, mapeada por sua função de ativação, pode servir como entrada para outros neurônios ou como resposta final na saída da rede. Em algumas redes, o sinal de saída pode, além de ir para outro neurônio ou para fora da rede, retornar à entrada do próprio neurônio que gerou a saída, servindo como uma realimentação (ALEXANDRIA, 2021).

Figura 10 - Funções de ativação



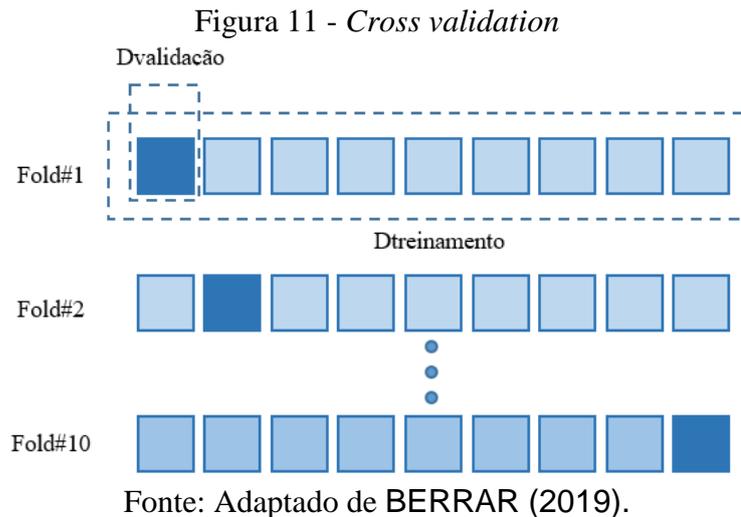
Fonte: Adaptado de ALEXANDRIA, (2021).

A Figura 10 traz exemplos de funções típicas de ativação de neurônios: sigmoidal, tangente hiperbólica, degrau e pseudo-linear.

A divisão dos dados em conjuntos de treino e teste é um passo usual em modelagens preditivas, como as RNAs. Essa etapa ocorre visando garantir a generalização do modelo, evitando o fenômeno de *overfitting* (superajuste), no qual o modelo se ajusta excessivamente aos dados de treino, mas não serve em novos dados. Ou seja, caso as métricas utilizadas na avaliação do modelo permaneçam satisfatórias no conjunto de teste, isso significa que a relação esperada entre as variáveis é verdadeira, e não somente ocorreu devido ao treinamento exaustivo na base de treino.

Para uma divisão adequada, é comum adotar proporções como 75%-25% ou 70%-30% para treino e teste, respectivamente (GOODFELLOW et al., 2016). Em modelos de *machine learning*, é importante que a separação seja feita de manter proporções estatísticas visando garantir a representatividade dos subconjuntos e evitar viés de algum momento específico que por ventura possa ocorrer na base de dados.

A escolha do método de divisão depende da natureza do problema. Em muitos casos, também é utilizada a técnica de *cross-validation* (validação cruzada), que cria diferentes combinações dos subconjuntos de treino e teste para aumentar a robustez da avaliação, conforme visto na Figura 11 (BISHOP; NASRABADI, 2006). Na prática, a validação cruzada pode evitar que um único *split* aleatório influencie excessivamente o resultado final, especialmente em conjuntos de dados menores.



As Redes Neurais com Atraso de Tempo (TDNNs) são um tipo de rede neural projetada para lidar com casos onde a relação temporal entre os dados é fundamental no sistema modelado. Ao contrário das redes neurais *feedforward* tradicionais, as TDNNs possuem um mecanismo que permite levar em consideração as interdependências entre valores em diferentes momentos, tornando-as ideais para aplicações onde o histórico de dados influencia os valores futuros. (PEDDINTI; POVEY; KHUDANPUR, 2015)

A principal característica dessas TDNNs é a capacidade de incorporar atrasos de tempo “lags”, “delay”, ou atrasos em seus neurônios de entrada, criando um deslocamento. Essa janela captura o contexto histórico dos dados, permitindo à rede aprender padrões temporais. No contexto das ETEs, as TDNNs podem ser utilizadas para modelar parâmetros analíticos, levando em consideração o TDH dos reatores e local onde ocorrem as análises, podendo inclusive trazer o caráter preditivo de uma análise efluente, com variáveis do início do processo de tratamento e considerando esse período entre elas.

Como foi descrito até então, as RNAs possuem algumas configurações que podem fazer diferença nos resultados, como a função de ativação, a quantidade de camadas ocultas e de

neurônios, taxa de aprendizagem, entre outros. A seguir será tratada uma técnica que testa diversas dessas configurações e tem como resultado a indicação de qual é a melhor.

O *Grid Search* é uma técnica utilizada para otimizar hiperparâmetros em RNAs. Hiperparâmetros são valores definidos antes do treinamento e que controlam o funcionamento do modelo, como a estrutura da rede, funções de ativação e métodos de otimização. No *Grid Search*, um conjunto de valores é especificado para cada hiperparâmetro e o método testa todas as combinações possíveis, identificando aquelas que produzem os melhores resultados para uma métrica de desempenho escolhida, como precisão ou *Mean Absolute Error* (MAE). Essa abordagem, embora computacionalmente intensiva, garante que o espaço definido de combinações seja totalmente explorado, proporcionando resultados confiáveis (LIASHCHYNSKYI, 2019).

Por fim, já com os resultados obtidos das RNAs, uma necessidade que muitas vezes ocorre, é a do entendimento da contribuição de cada *feature* no resultado da *target*. A técnica SHAP (*SHapley Additive exPlanations*) é utilizada para interpretar o impacto de cada *feature* na predição da variável de interesse. O método SHAP é uma ferramenta para explicar modelos de *machine learning*, pois ele decompõe a predição de um modelo em contribuições aditivas de cada *feature*, com base na teoria dos valores de Shapley da teoria dos jogos cooperativos. Isso permite não apenas observar quais variáveis influenciam mais a DQO efluente, mas também como elas impactam, fornecendo uma visão detalhada da relação entre as variáveis independentes e o resultado predito (MARCÍLIO; ELER, 2020).

#### 1.4.1 Métricas de avaliação

A eq. (4) mostra o cálculo para o Erro Médio Quadrático (do inglês, *Mean Square Error*) (MSE), usualmente utilizado para determinar a diferença entre o valor predito e esperado em um modelo (PADHMA, 2023).

$$MSE = \sum_{i=1}^n \frac{(d_i - y_i)^2}{n} \quad (4)$$

Onde,  $d_i$ : é o valor da variável,  $y_i$ : é o valor estimado pelo modelo, com  $i$  variando de 1 a  $n$ , onde  $n$  é valor de amostras da base de dados modelada.

A eq. (5), a raiz do erro quadrático médio ou *Root Mean Square Error* (RMSE), análogo ao que ocorre na métrica MSE, porém adicionado a raiz quadrada ao cálculo:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(d_i - y_i)^2}{n}} \quad (5)$$

Onde,  $d_i$ : é o valor da variável,  $y_i$ : é o valor estimado pelo modelo, com  $i$  variando de 1 a  $n$ , onde  $n$  é valor de amostras da base de dados modelada.

Outra métrica normalmente utilizada é o erro médio absoluto ou *Mean Absolute Error* (MAE), eq. (6):

$$MAE = \frac{1}{n} \sum_{i=1}^n |d_i - y_i| \quad (6)$$

Onde,  $d_i$ : é o valor da variável,  $y_i$ : é o valor estimado pelo modelo, com  $i$  variando de 1 a  $n$ , onde  $n$  é valor de amostras da base de dados modelada.

Novamente, da mesma forma que ocorre com as métricas MSE e RMSE, o MAPE é uma métrica muito parecida com o MAE, com o acréscimo da divisão, trazendo a porcentagem média da diferença entre o valor predito e real, eq. (7):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{d_i - y_i}{n} \right| \times 100 \quad (7)$$

Onde,  $d_i$ : é o valor da variável,  $y_i$ : é o valor estimado pelo modelo, com  $i$  variando de 1 a  $n$ , onde  $n$  é valor de amostras da base de dados modelada.

Por fim, o coeficiente de determinação ou  $R^2$ , demonstra a quão ajustada está a equação obtida ao conjunto de dados, eq. (8):

$$R^2 = 1 - \frac{\sum_{i=1}^n (d_i - \hat{y}_i)^2}{\sum_{i=1}^n (d_i - \bar{y})^2} \quad (8)$$

Onde,  $d_i$ : é o valor da variável,  $\hat{y}_i$ : é o valor estimado pelo modelo,  $\bar{y}$ : é a média dos dados reais, com  $i$  variando de 1 a  $n$ , onde  $n$  é valor de amostras da base de dados modelada.

Enquanto nos quatro primeiros, MSE, RMSE, MAE e MAPE, esperam-se valores baixos afim de se atestar a qualidade do modelo, no  $R^2$ , que varia entre 0 e 1, valores que se aproximam de 1, aumentam o poder de predição do modelo. A Tabela 1 traz as faixas de valores para o coeficiente de determinação que podem ser interpretados como uma forte correlação ou não:

Tabela 1 - Interpretação do coeficiente de determinação

| Coeficiente de determinação | Interpretação |
|-----------------------------|---------------|
| 0,9 a 1 (-0,9 a -1)         | Muito alta    |
| 0,7 a 0,9 (-0,7 a -0,9)     | Alta          |
| 0,5 a 0,7 (-0,5 a -0,7)     | Moderada      |
| 0,3 a 0,5 (-0,3 a -0,5)     | Baixa         |
| 0,0 a 0,3 (-0,0 a -0,3)     | Inexistente   |

Fonte: MUKAKA (2012).

#### 1.4.2 Tratamento de dados

É fundamental o processo de tratamento de dados quando existe a intenção de trabalhar com modelagem, pois a qualidade dos dados influencia diretamente a qualidade do modelo. É muito comum que os dados contenham erros, dados faltantes, *outliers*. A seguir serão descritas algumas técnicas usualmente utilizadas.

A normalização em ciência de dados é uma técnica importante para garantir que variáveis de diferentes escalas possam ser comparadas e usadas de forma eficaz em algoritmos de aprendizado de máquina. O processo de normalização ajusta os valores dos dados para que eles se situem em uma mesma escala, geralmente entre 0 e 1. Esse ajuste é particularmente importante em modelos que são sensíveis às magnitudes das variáveis, como redes neurais. Sem normalização, variáveis de maior escala podem dominar a análise e distorcer os resultados do modelo, levando a previsões imprecisas (JAMES, 2013).

Diferentes métodos de normalização podem ser aplicados, como a *Min-Max Scaling*, que transforma os dados para um intervalo específico, de acordo com os valores mínimos e máximos encontrados em cada uma das variáveis e a *Z-score Normalization*, que ajusta os valores com base na média e no desvio padrão da variável. Esses métodos ajudam a estabilizar os cálculos e a garantir que todos os parâmetros tenham a mesma importância relativa no processo de modelagem. A escolha do método depende do contexto da análise e do algoritmo

a ser utilizado (JAMES, 2013).

Diferentes técnicas podem ser utilizadas com o interesse de remover *outliers* em determinados dados. A técnica que será aqui demonstrada e utilizada é chamada de intervalo interquartil ou *Interquartile Range* (IQR).

É uma abordagem comum para a detecção e remoção de *outliers* em conjuntos de dados. O IQR é a diferença entre o terceiro quartil (Q3, o valor que delimita os 75% superiores dos dados) e o primeiro quartil (Q1, o valor que delimita os 25% inferiores dos dados). Esse intervalo captura a amplitude do meio dos dados, que é considerada a parte "central" e menos influenciada por valores extremos (MONTGOMERY; RUNGER, 2010).

Para identificar *outliers* usando o IQR, calcula-se o limite inferior e superior do intervalo, definidos por:

- Limite inferior:  $Q1 - 1.5 * IQR$
- Limite superior:  $Q3 + 1.5 * IQR$

Valores abaixo do limite inferior ou acima do limite superior é considerado um *outlier* e pode ser removido ou tratado. Essa técnica é eficaz porque se baseia na distribuição dos dados e evita que os valores extremos afetem a análise e posteriormente até valores preditos (MONTGOMERY; RUNGER, 2010).

## 2 REVISÃO DA LITERATURA SOBRE TRANSFORMAÇÃO DIGITAL NO TRATAMENTO DE ESGOTO

Neste capítulo é realizada uma análise bibliométrica do tema, detalhando o método utilizado para a análise quantitativa dos trabalhos relevantes na área. São examinados os locais e periódicos mais frequentes, oferecendo uma visão geral do estado da pesquisa sobre a transformação digital aplicada ao tratamento de esgoto.

A estratégia escolhida para definição das palavras-chaves foi a PICO, que é um acrônimo para População, Intervenção, Comparação e *Outcome* (desfecho). A estratégia que tem origem na medicina auxilia o pesquisador, através de perguntas sobre cada uma dessas dimensões, a encontrar as palavras mais adequadas que irão compor a *string* de pesquisa. Dois grupos de palavras foram resultantes, um deles descrevendo o campo, ou seja, tratamento de esgoto, lodo ativado e águas residuais e outro caracterizando a evolução tecnológica esperada, chegando em: Indústria 4.0, transformação digital, digitalização e controle e automação.

Na busca dos artigos que mais respondem à questão de pesquisa, os artigos encontrados foram selecionados, segundo critérios de inclusão e exclusão (Quadro 1).

Quadro 1 - Critérios de inclusão e exclusão

| Critérios de inclusão            | Critérios de exclusão   |
|----------------------------------|---|
| Aplicado no tratamento de esgoto | Não está em inglês  |
| Artigos em inglês ou português   | Não responde a questão de pesquisa                                |
|                                  | Tecnologia não caracteriza transformação digital ou Indústria 4.0 |
|                                  | É uma revisão sistemática   |

Fonte: Do autor (2025)

Para avaliação da qualidade dos artigos selecionados, foram elaboradas algumas perguntas visando identificar: (1) se o artigo é relacionado ao tratamento de esgoto; (2) se o uso da tecnologia caracteriza uma transformação digital; (3) se é um dos pilares da indústria 4.0; e (4) se o resultado do artigo tem relação com a eficiência do processo de tratamento de esgoto. O Quadro 2 traz uma síntese das questões aplicadas para esta finalidade.

Quadro 2 - Questões para avaliação de qualidade

|   |
|---|
| Questões  |
| É relacionado ao tratamento de esgoto?  |
| Caracteriza transformação digital?  |
| É um pilar da indústria 4.0?  |
| O resultado da pesquisa tem relação com a eficiência do processo de tratamento de esgoto? |

Fonte: Do autor (2025)

Para cada um dos artigos selecionados e acerca de cada uma das questões, foi verificado se artigo respondia à questão de pesquisa: “Quais e como os conhecimentos de automação já consolidados no tratamento de esgoto são utilizados? ”. Uma pontuação para cada nível de resposta foi atribuída e por fim uma nota de corte foi adotado visando ter como resultado, os artigos mais aderentes a questão de pesquisa. O Quadro 3 apresenta a pontuação utilizada na análise.

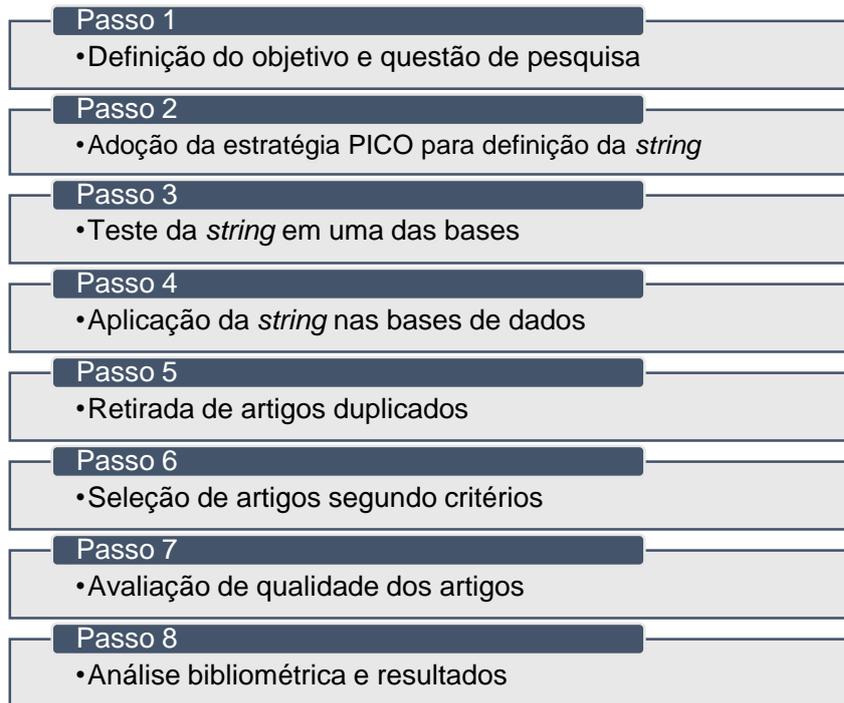
Quadro 3 - Pontuação para avaliação de qualidade das questões

|  |     |
|--|-----|
| O artigo responde ao critério avaliado ou o critério não se aplica | 1.0 |
| O artigo não deixa claro se atende ou não o critério               | 0.5 |
| Não existe nada no artigo que atende ao critério                   | 0.0 |

Fonte: Do autor (2025)

Assim, a Figura 12 traz uma síntese do processo metodológico que foi baseado em Kitchenham (2007), utilizando-se a ferramenta Parsifal.

Figura 12 - Procedimento metodológico adotado



Fonte: Do autor (2025)

No passo 1 são definidos o objetivo e a questão de pesquisa, seguido pelo passo 2 que por meio de um método, nesse caso a estratégia PICO, e atrelado ao objetivo definido permite encontrar uma *string* mais adequada para pesquisa nas bases de dados. Antes da aplicação de fato da *string* na base de dados no passo 4, testes com as *strings* em pelo menos uma das bases se faz necessário visando verificar a abrangência dos resultados obtidos. Muitas das bases de dados acabam incluindo os mesmos artigos em suas bases, por isso a utilização de alguma ferramenta para retirada dos itens duplicados ocorre no passo 5. A seleção dos artigos mais aderentes ao objetivo de pesquisa é feita nos passos 6 e 7, em um primeiro momento por critérios de inclusão e exclusão e posteriormente por uma avaliação de qualidade, respondendo algumas questões de modo a entender o quanto cada artigo responde à questão. Por fim a análise bibliométrica e discussão dos resultados pode ser realizada com o produto obtido.

Para definição das palavras-chave, inicialmente foi realizada uma contextualização com as questões de pesquisa e a problematização que em conjunto com a estratégia PICO deu origem as palavras que foram utilizadas neste trabalho. O quadro 4 apresenta a contextualização do problema e que foi usado para encontrar as palavras chaves.

No Quadro 4 foi possível observar a reincidência de alguns termos chaves em mais de um momento, como por exemplo, transformação digital, tratamento de esgoto e indústria 4.0.

Quadro 4 - Contextualização do problema

|                      | IDENTIFICAÇÃO DA IDEIA   | PALAVRAS-CHAVE(DESCRITORES)  |
|----------------------|--|--|
| CONTEXTO             | Indústria 4.0 e Transformação digital no Tratamento de esgoto  | Indústria 4.0; Transformação digital; Tratamento de esgoto   |
| PROBLEMATIZAÇÃO      | Como automação, controle e transformação digital podem tornar os processos de tratamento de esgoto mais eficientes e por consequência aproximar a companhia do desafio da universalização dos serviços   | Automação, Controle, Eficiência, Universalização; Indústria 4.0; Transformação digital; Tratamento de esgoto                       |
| QUESTÕES DE PESQUISA | Quais os conhecimentos de automação já consolidados no tratamento de esgoto (controles e tecnologias) ?<br><br>Dentro do conceito de indústria 4.0, quais tecnologias tem potencial de contribuir para eficiência de seus processos ?<br><br>Qual o nível de preparação (equipamentos e pessoas) atual das Estações de Tratamento de Esgoto da região metropolitana para receber essas novas tecnologias?<br><br>Como essas tecnologias podem contribuir nos custos de tratamento (Energia, químicos, HH) e eficiência ? | Tecnologias, conhecimentos consolidados<br><br>Tecnologias habilitadoras<br><br>Maturidade; Região metropolitana<br><br>Benefícios |
| PROBLEMA             | A universalização como meta dos serviços de tratamento, e a impossibilidade de atendimento sem a evolução tecnológica.   | Universalização; Tratamento de esgoto; Evolução tecnológica  |
| OBJETIVOS            | GERAL: identificar o conhecimento de transformação digital consolidado e o potencial dentro do tratamento de esgoto.<br>ESPECIFICOS: Realizar bibliometria, e de tecnologias potenciais.   | Conhecimentos; Transformação digital<br>Bibliometria; Tecnologias potenciais   |

Fonte: Do autor (2025)

A estratégia PICO, conforme a Quadro 5, foi utilizada visando delinear a pesquisa e obter as palavras-chave mais adequadas. Com o uso dessa estratégia e a partir de quatro dimensões (População, Intervenção, Comparação e *Outcome* ou o desfecho esperado) é possível encontrar as palavras mais adequadas para a busca.

Quadro 5 - Estratégia PICO

| População   | Intervenção   | Comparação  | <i>Outcome</i> (desfecho)   |
|---|---|---|---|
| Tornar os processos de Tratamento de Esgoto mais eficientes na Unidade de Tratamento de Esgoto da Região Metropolitana de São Paulo | Por meio da automação, controle, transformação digital e uso das tecnologias habilitadoras da Indústria 4.0 no Saneamento | Estações que operam hoje de forma totalmente manual | Espera-se com isso redução de custos com químicos, energia elétrica, tornar os processos mais estáveis e em etapa futura, poder prever situações no Tratamento. |

Fonte: Do autor (2025)

Com as palavras-chave sugeridas pela técnica adotada, fez-se necessário avaliar em pelo menos uma das bases o conjunto de algumas *strings*, com o objetivo de visualizar o resultado

que a busca geraria, e em caso de poucas respostas poder analisar quais palavras usar.

A partir dos testes foi possível observar que pesquisas no título restringiriam muito os resultados, provavelmente excluindo artigos com conteúdo aderente a pesquisa, assim como pesquisas somente no assunto, portanto ficou definido que as pesquisas seriam realizadas de forma geral. Outro ponto observado, foi que a inclusão de uma tecnologia específica na *string*, além de enviesar a pesquisa, poderia dar a falsa impressão de uma pesquisa com muitos resultados acerca daquela tecnologia, enquanto não englobaria pesquisas que até a utilizassem, porém se tratasse de um tipo específico. Por exemplo, a inclusão de *machine learning*, enquanto algum pesquisador pudesse descrever em seu material diretamente qual técnica de *machine learning* utilizou, como árvore de decisão, *K-means* ou *Support Vector Machine* (SVM) (PESTANA, 2023).

Com os resultados, foi possível definir a *string* que seria usada para responder à questão de pesquisa:

***("Sewage Treatment" OR "Activated Sludge" OR "Wastewater") AND ("Industry 4.0" OR "Digital Transformation" OR "Digitization" OR "Control Automation")***

Após seleção dos artigos e aplicados os critérios de inclusão (artigo na língua inglesa e revisados por pares) chegou-se ao resultado final (Tabela 2).

Tabela 2 - Quantitativo dos resultados nas bases

| Base de dados  | Dados      | Resultados | Resultados (Artigos) | Resultados (Artigos) & (Inglês) | Resultados (Artigos) & (Inglês) & (Revisado por par) |
|----------------|------------|------------|----------------------|---------------------------------|--|
| CAPES          | 18/04/2023 | 210        | 161                  | 157                             | 133  |
| SCOPUS         | 18/04/2023 | 99         | 44                   | 37                              | 37   |
| WEB OF SCIENCE | 18/04/2023 | 62         | 40                   | 40                              | 40   |
| Total          |            | 371        | 245                  | 234                             | 210  |

Fonte: Do autor (2025)

Do produto de cada uma das bases e utilizando o software Parsifal, foi identificado que 56 artigos estavam presentes em mais de uma das bases, sendo assim duplicados e retirados nessa etapa. A Tabela 3 apresenta o resultado após esse procedimento.

Tabela 3 - Quantitativo dos resultados nas bases após excluir duplicados

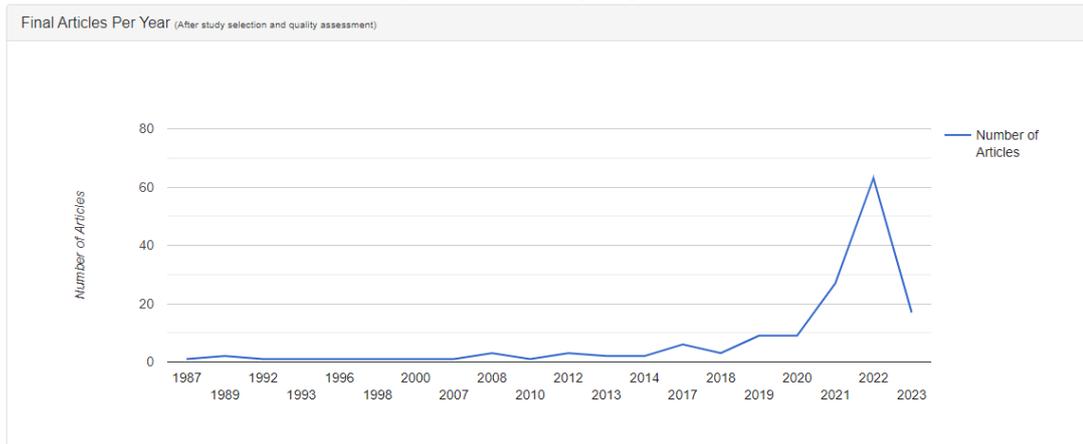
| Base de dados | Após retirar duplicados |
|---------------|-------------------------|
| CAPES         | 107                     |
| SCOPUS        | 30                      |

|                |     |
|----------------|-----|
| WEB OF SCIENCE | 17  |
| Total          | 154 |

Fonte: Do autor (2025)

Nesse passo foi possível observar o crescente número de pesquisas relacionado ao tema, com pico em 2022 (Figura 13).

Figura 13 - Artigos por ano

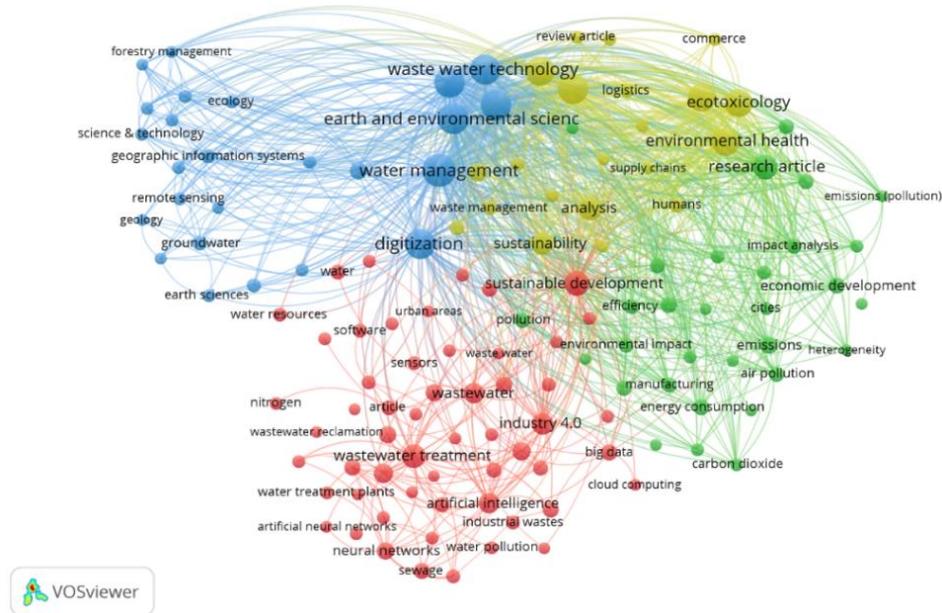


Fonte: Do autor (2025)

Outras análises foram realizadas por meio do software VosViewer, e pelo uso da biblioteca bibliometrix para o RStudio. Foi possível observar que o periódico com maior produção no tema foi a “*Environmental Science and Pollution Research Inter*”.

Quatro *clusters* foram observados nessa etapa e representados na Figura 14, sendo que um dos *clusters* está relacionado a estratégias e tecnologias. Nesse *cluster* algumas tecnologias como *neural network*, *cloud computing*, *big data* e *artificial intelligence* aparecem como temas em comum entre os autores.

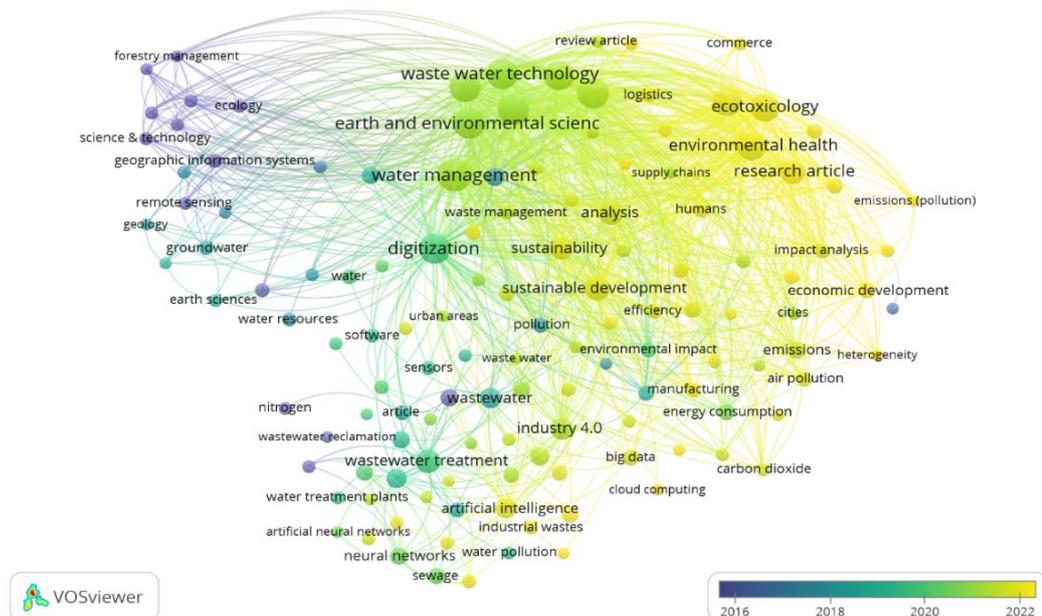
Figura 14 - Palavras mais encontradas



Fonte: Do autor através do software VOSViewer (2025)

Essa mesma análise realizada, permite observar o uso dessas palavras ao longo do tempo, trazendo principalmente as palavras inteligência artificial, redes neurais e *big data* a partir de 2020 (Figura 15).

Figura 15 - Palavras mais encontradas ao longo do tempo



Fonte: Do autor através do software VOSViewer (2025)

Uma nuvem de palavras mostra as palavras mais utilizadas por esses pesquisadores, com ênfase principal para gestão da água, poluição aquática e ciência da terra e meio ambiente (Figura 16). Esse achado sugere que análise bibliométrica carece de refinamento visando aderência dos assuntos abordados pelos autores aos objetivos da pesquisa.

Figura 16 - Nuvem de palavras



Fonte: Do autor através do software RStudio e a biblioteca Bibliometrix (2025)

Após a etapa de qualificação, foram obtidos 16 artigos no total. Como foram selecionados de acordo com as questões de pesquisa, é possível ter maior aderência das pesquisas as ETEs e as tecnologias atualmente utilizadas visando eficiência dos processos de tratamento, relacionados principalmente a digitalização e transformação digital (Figura 17).



Figura 18 - Nuvem de palavras dos artigos qualificados



Fonte: Do autor através do software RStudio e a biblioteca Bibliometrix (2025)

Alguns grupos de pesquisas convergentes foram observados, sendo um deles relacionado ao esgoto, e a aplicação de redes neurais, enquanto outros deles trazia o termo “tratamento de esgoto” e a uso de inteligência artificial (Figura 19).

Figura 19 - Clusters de palavras dos artigos qualificados



Fonte: Do autor através do software RStudio e a biblioteca Bibliometrix (2025)

Para os artigos qualificados iniciou-se então uma análise sobre seu conteúdo, pontos em comum e diferenças sobre o conteúdo da pesquisa, dos modelos e tecnologias utilizadas, que podem ser vistos no Quadro 6.

Quadro 6 - Artigos após qualificação

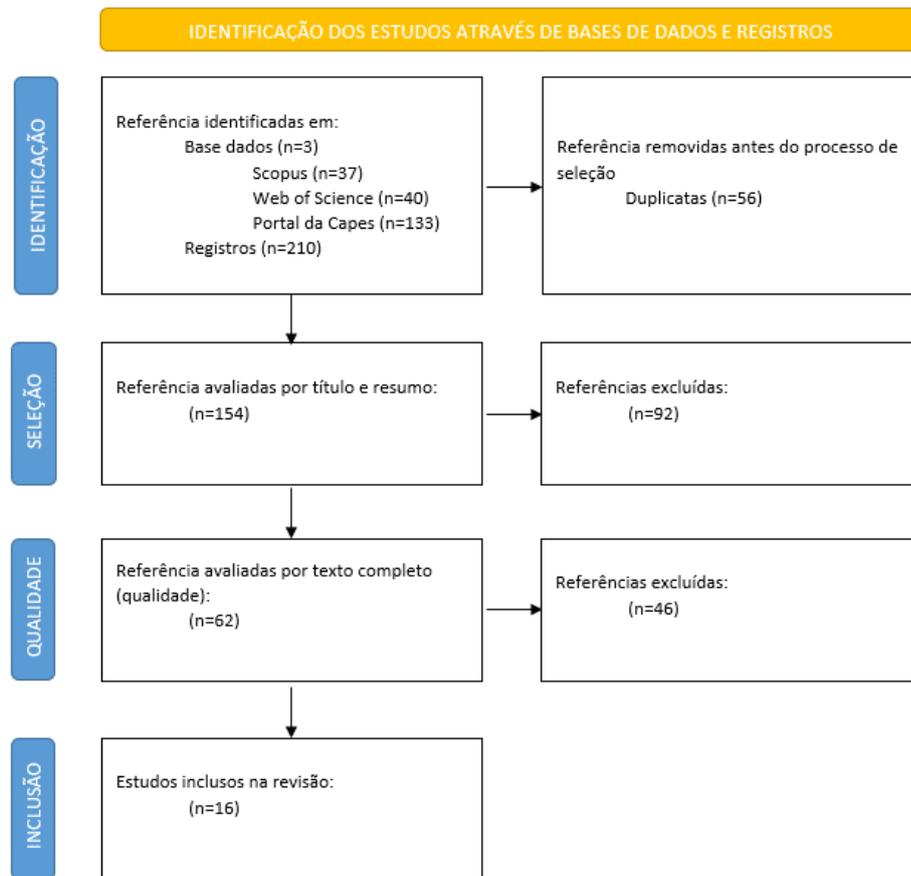
| AU   | TI  | PY   |
|--|---|------|
| SINGH G.;JOSHI P.;RAGHUVANSHI AS                                   | A NOVEL DUTY CYCLE BASED CROSS LAYER MODEL FOR ENERGY EFFICIENT ROUTING IN IWSN BASED IOT APPLICATION   | 2022 |
| ARISMENDY L;CÁRDENAS C;GÓMEZ D;MATURANA A;MEJÍA R;QUINTERO M. CG   | A PRESCRIPTIVE INTELLIGENT SYSTEM FOR AN INDUSTRIAL WASTEWATER TREATMENT PROCESS: ANALYZING PH AS A FIRST APPROACH  | 2021 |
| SHEN Y;LI H;ZHANG B;CAO Y;GUO Z;GAO X;CHEN Y                       | AN ARTIFICIAL NEURAL NETWORK-BASED DATA FILLING APPROACH FOR SMART OPERATION OF DIGITAL WASTEWATER TREATMENT PLANTS   | 2023 |
| HERNANDEZ-DEL-OLMO F;LLANES FH;GAUDIOSO                            | AN EMERGENT APPROACH FOR THE CONTROL OF WASTEWATER TREATMENT PLANTS BY MEANS OF REINFORCEMENT LEARNING TECHNIQUES   | 2012 |
| LI F;VANROLLEGHEM PA   | AN INFLUENT GENERATOR FOR WRRF DESIGN AND OPERATION BASED ON A RECURRENT NEURAL NETWORK WITH MULTI-OBJECTIVE OPTIMIZATION USING A GENETIC ALGORITHM   | 2022 |
| HEO S;NAM K;LOY-BENITEZ J;YOO C                                    | DATA-DRIVEN HYBRID MODEL FOR FORECASTING WASTEWATER INFLUENT LOADS BASED ON MULTIMODAL AND ENSEMBLE DEEP LEARNING   | 2021 |
| PISA I;MORELL A;VICARIO JL;VILANOVA R                              | DENOISING AUTOENCODERS AND LSTM-BASED ARTIFICIAL NEURAL NETWORKS DATA PROCESSING FOR ITS APPLICATION TO INTERNAL MODEL CONTROL IN INDUSTRIAL ENVIRONMENTS—THE WASTEWATER TREATMENT PLANT CONTROL CASE | 2020 |
| VIET ND;JANG D;YOON Y;JANG A                                       | ENHANCEMENT OF MEMBRANE SYSTEM PERFORMANCE USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR SUSTAINABLE WATER AND WASTEWATER TREATMENT: A CRITICAL REVIEW   | 2022 |
| MATHERI AN;NTULI F;NGILA JC;SEODIGENG T;ZVINOWANDA C               | PERFORMANCE PREDICTION OF TRACE METALS AND COD IN WASTEWATER TREATMENT USING ARTIFICIAL NEURAL NETWORK  | 2021 |
| PADHMA PRIYA P;JENIT A;PEARLIN D DJ;SHARMA NK                      | PREDICTIVE BIODEGRADATION OF MULTIPLE TOXIC POLLUTANTS IN BIOREACTORS TREATING REAL WASTEWATER USING ANN AND GP   | 2023 |
| HAO X;LI B   | RESEARCH ON INTELLIGENT SYSTEM ARCHITECTURE OF URBAN DOMESTIC SEWAGE TREATMENT  | 2020 |
| SHARMA N;CHANDRASEKAR S;SUNDAR K                                   | SUITABILITY OF ANN AND GP FOR PREDICTING SOAK PIT TANK EFFICIENCY UNDER LIMITED DATA CONDITIONS   | 2018 |
| MATHERI AN;MOHAMED B;NTULI F;NABADDA E;NGILA JC                    | SUSTAINABLE CIRCULARITY AND INTELLIGENT DATA-DRIVEN OPERATIONS AND CONTROL OF THE WASTEWATER TREATMENT PLANT  | 2022 |
| PISA I;MORELL A;VICARIO JL;VILANOVA R                              | TRANSFER LEARNING IN WASTEWATER TREATMENT PLANTS CONTROL: MEASURING THE TRANSFER SUITABILITY  | 2023 |
| NARDELLI P;GATTI G;CECCHI F;BATTISTONI EM                          | UPGRADING SMALL WWTPS IN THE AUTONOMOUS PROVINCE OF TRENTO (ITALY) BY ALTERNATING OXIC/ANOXIC PROCESS: A DEMONSTRATION STUDY  | 2008 |
| PANDEY S;TWALA B;SINGH R;GEHLOT A;SINGH A;MONTERO EC;PRIYADARSHI N | WASTEWATER TREATMENT WITH TECHNICAL INTERVENTION INCLINATION TOWARDS SMART CITIES   | 2022 |

Fonte: Do autor (2025)

Um ponto observado que há apenas um artigo de 2012 e outros artigos somente após 2018, sugerindo que o tema apresentou maior relevância a partir desta data. Na Figura 20, é possível observar o mesmo passo a passo metodológico apresentado na seção de materiais e métodos, porém agora mostrando os resultados obtidos em cada uma das etapas e as ferramentas utilizadas.

O primeiro produto obtido da *string* trouxe um total 210 artigos que após a retirada dos artigos duplicados resulta em 154 artigos. Após a seleção de artigos em critérios pré-estabelecido chega-se ao total de 62 publicações, agora já bem mais aderentes ao objetivo do estudo. Por fim a análise de qualidade refina esse material a um total de 16 publicações, agora trazendo modelagens utilizadas em ETEs, com métricas e objetivos semelhantes.

Figura 20 - Procedimento metodológico adotado com resultados



Fonte: Do autor (2025)

A partir destes 16 artigos foi feito uma leitura com a finalidade de identificar as tecnologias utilizadas e uma matriz foi criada (Quadro 7).

Quadro 7 - Relação dos modelos utilizados

| Autor | Modelo  |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  |                                     |  |                  |              |  |         |
|-------|---|--|---|---|--|--------------------------------------|------------------------|------------------------|--|---------------------------------|---|--|---------|--|--|-------------------------------------|--|------------------|--------------|--|---------|
|       |   | 1  | 2   | 3   | 4  | 5                                    | 6                      | 7                      | 8  | 9                               | 10  | 11   | 12      | 13   | 14   | 15                                  | 16   | 17               | 18           | 19   | 20      |
|       |   | ALGORITMO GENÉTICO - GENETIC ALGORITHMS (GA) | ANÁLISE POR COMPONENTE PRINCIPAL - PRINCIPAL COMPONENT ANALYSIS (PCA) | APRENDIZADO DE MÁQUINA - MACHINE LEARNING | APRENDIZADO POR REFORÇO - REINFORCEMENT LEARNING | APRENDIZADO PROFUNDO - DEEP LEARNING | AUTOMAÇÃO - AUTOMATION | DENOISING AUTOENCODERS | ESTRATÉGIAS DE LONGO PRAZO MEMÓRIA - LONG SHORT-TERM MEMORY (LSTM) | GEMÊOS DIGITAIS - DIGITAL TWINS | GERADOR INFLUENTE - INFLUENT GENERATOR (IG) | INTELIGÊNCIA ARTIFICIAL - ARTIFICIAL INTELLIGENCE (IA) | K-MEANS | K-VIZINHOS MAIS PRÓXIMOS - K-NEAREST NEIGHBORS (KNN) | APRENDIZADO MULTIMODAL - MULTIMODAL LEARNING | PROGRAMA GENÉTICO - GENETIC PROGRAM | REDES NEURAIS ARTIFICIAIS - ARTIFICIAL NEURAL NETWORKS (ANN) | SÉRIES TEMPORAIS | SIMULAÇÃO 3D | TRANSFERÊNCIA DE APRENDIZADO - TRANSFER LERNING (TL) | REVISÃO |
| A     | SINGH G;JOSHI P;RAGHUVANSHI AS                                      |  |   |   |  |                                      |                        |                        |  |                                 |   |  | X       |  |  |                                     |  |                  |              |  |         |
| B     | ARISMENDY L;CÁRDENAS C;GÓMEZ D;MATURANA A;MEJÍA R;QUINTERO M. CG    | X  |   | X   |  |                                      |                        |                        | X  |                                 |   |  |         |  |  |                                     | X  |                  |              |  |         |
| C     | SHEN Y;LI H;ZHANG B;CAO Y;GUO Z;GAO X;CHEN Y                        |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         | X  |  |                                     | X  |                  |              |  |         |
| D     | HERNANDEZ-DEL-OLMO F;LLANES FH;GAUDIO SO                            |  |   |   |  | X                                    |                        |                        |  |                                 |   | X  |         |  |  |                                     |  |                  |              |  |         |
| E     | LI F;VANROLLEGHEM PA  | X  |   |   |  |                                      |                        |                        | X  |                                 | X   |  |         |  |  |                                     | X  | X                |              |  |         |
| F     | HEO S;NAM K;LOY-BENITEZ J;YOO C                                     |  |   |   |  | X                                    |                        |                        |  |                                 |   |  |         |  | X  |                                     |  |                  |              |  |         |
| G     | PISA I;MORELL A;VICARIO JL;VILANOVA R                               |  | X   |   |  |                                      |                        | X                      | X  |                                 |   |  |         |  |  |                                     | X  |                  |              |  |         |
| H     | VIET ND;JANG D;YOON Y;JANG A  |  |   |   |  |                                      |                        |                        |  |                                 |   | X  |         |  |  |                                     |  |                  |              |  | X       |
| I     | MATHERI AN;NTULI F;NGILA JC;SEODIGENG T;ZVINOWANDA C                |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  |                                     |  |                  |              |  | X       |
| J     | PADHMA PRIYA P;JENIT A;PEARLIN D DJ;SHARMA NK                       | X  |   |   |  |                                      |                        |                        |  |                                 |   | X  |         |  |  | X                                   | X  |                  |              |  |         |
| K     | HAO X;LI B  |  |   |   |  |                                      |                        |                        |  | X                               |   |  |         |  |  |                                     |  |                  | X            |  | X       |
| L     | SHARMA N;CHANDRASEKAR S;SUNDAR K                                    |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  | X                                   | X  |                  |              |  |         |
| M     | MATHERI AN;MOHAMED B;NTULI F;NABADDA E;NGILA JC                     |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  |                                     |  |                  |              |  | X       |
| N     | PISA I;MORELL A;VICARIO JL;VILANOVA R                               |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  |                                     | X  |                  |              | X  |         |
| O     | NARDELLI P;GATTI G;CECCHI F;BATTISTONI EM                           |  |   |   |  |                                      | X                      |                        |  |                                 |   |  |         |  |  |                                     |  |                  |              |  |         |
| P     | PANDEY S;T WALA B;SINGH R;GEHLOT A;SINGH A;MONTERO EC;PRIYADARSHI N |  |   |   |  |                                      |                        |                        |  |                                 |   |  |         |  |  |                                     |  |                  |              |  | X       |
|       | Total   | 3  | 1   | 1   | 0  | 2                                    | 1                      | 1                      | 3  | 1                               | 1   | 3  | 1       | 1  | 1  | 2                                   | 7  | 1                | 1            | 1  | 4       |

Fonte: Do autor (2025)

O Quadro 7 apresenta os modelos e técnicas utilizadas em diferentes estudos, associando cada autor às abordagens exploradas. Dentre os modelos analisados, destaca-se a predominância de RNAs que aparecem em 7 estudos, evidenciando sua ampla aplicação e relevância em análises de tratamento de esgoto e outros processos relacionados. Modelos como Algoritmos Genéticos (GA), e IA também são notados em mais de uma ocorrência, demonstrando a diversificação nas metodologias aplicadas por diferentes pesquisadores. No entanto, técnicas como *Deep Learning*, Transferência de Aprendizado e *K-means* foram exploradas em menor proporção, sugerindo que, embora promissoras, podem ainda estar em estágios iniciais de investigação neste campo específico.

Ao observar os autores individualmente, percebe-se que alguns pesquisadores utilizam uma combinação de modelos, como o estudo de Pisa et al., que integra Algoritmos Genéticos, *Denosing Autoencoders* e RNAs, refletindo a tendência de hibridização de técnicas para maximizar os resultados analíticos. Em contraste, outros trabalhos, como o de Shen et al., focam em abordagens mais específicas, como Gêmeos Digitais e RNAs. Essa diversidade metodológica reflete a interdisciplinaridade do campo, enquanto a concentração em determinados modelos aponta para áreas de pesquisa bem estabelecidas. A análise do uso dessas ferramentas pode ajudar a identificar lacunas no conhecimento e oportunidades para o desenvolvimento de metodologias inovadoras.

Como observado, as redes neurais artificiais é a abordagem mais encontrada nos artigos selecionados. Uma leitura rápida dos artigos que usam a tecnologia das redes neurais foi identificada que a técnica em comum entre os autores, o LSTM. Enquanto as aplicações, incluem principalmente a previsão de variáveis importantes para determinação da eficiência da estação, como a DQO. Já as métricas utilizadas também apresentam concordância entre as publicações, com o MAPE, MAE e RMSE sendo utilizados para verificar a qualidade da modelagem utilizada e determinação do erro entre o valor predito e esperado.

No próximo capítulo é feita a avaliação do conteúdo dos trabalhos que utilizaram RNAs no tratamento de esgoto, e ainda detalhado o seu objetivo, resultados e principais métricas utilizadas.

Arismendy (2021) através do uso de RNAs, LSTM e AG propõe uma redução de DQO em uma ETE. Inicia o trabalho com uma revisão sistemática da literatura e identifica que para a amostra a programação matemática é a técnica de análise prescritiva mais encontrada, representando 39% das publicações. A aplicação ocorreu na China, em uma Estação que analisa 22 variáveis de processo, com uma base de dados de 12/2017 a 07/2020. A rede foi configurada com 18 variáveis nos neurônios de entrada como as variáveis explicativas, e duas camadas ocultas com 2 e 16 neurônios respectivamente, com objetivo de prever a DQO do dia seguinte. Para configuração da rede o autor cria uma matriz de correlação entre todas as variáveis. A intenção é identificar principalmente, altas correlações de análises com parâmetros a manipular. O autor ainda aplica uma árvore de decisão afim de associar o pH em uma das etapas do processo com o nitrogênio, e uma das métricas utilizadas pelo autor para verificação da qualidade do modelo é o MAPE. Por fim a aplicação do AG e da RNA é realizada e o autor demonstra os resultados positivos entre o valor real e predito através de um gráfico em que sobrepõe esses dados.

Shen (2023) em uma base de dados proveniente de uma ETE, propõe um método para preenchimento de dados ausentes de OD através de LSTM, *Gated Recurrent Units* (GRU) uma variante de LSTM, e uma combinação de redes neurais convolucionais. Modelos Florestas aleatórias e o modelo do *K-nearest neighbor* (KNN) são utilizados como benchmark de modelos para comparação de métricas. As métricas utilizadas foram: o erro médio absoluto MAE, o RMSE, o MAPE e a acurácia. O método GRU teve o melhor desempenho de preenchimento com uma precisão de superior a 80%.

Li (2022) assim como outros autores, também faz uso de LSTM, nesse caso associado a *Influente Generator* (IG) e aprimorado por um AG multiobjeto. As métricas utilizadas foram o MAPE e o KLdivergence que mede a diferença entre duas distribuições de probabilidade. A intenção foi modelar os dados de SST, DQO e nitrogênio amoniacal, a partir de dados de chuva, temperatura e vazão. Uma melhoria na série temporal gerada foi obtida através da utilização do processo de passeio aleatório ao LSTM.

Pisa (2020), diferentes dos outros autores utiliza métricas distintas para avaliação do modelo. As métricas: Erro Quadrado Integrado, do inglês *Integrated Square Error* (ISE) e Erro Absoluto Integrado, do inglês *Integrated Absolute Error* (IAE) para o modelo implementado. É implementado um *Internal Model Controller* (IMC) baseado em RNAs também do tipo LSTM. Para eliminação de ruídos foram adotadas duas estratégias a Análise de Componentes

Principais (ACP) e um *Denoising Autoencoder* (DAE). Para comparação dos resultados gerais também foram utilizadas as métricas RMSE e MAPE associados ao coeficiente de determinação  $R^2$ . Toda essa modelagem com intuito de controlar o OD em um dos reatores de uma ETE.

Pisa (2023) propõe aplicação de *transfer learning* de modelos de redes neurais artificiais no tratamento de esgoto. As mesmas métricas de seu trabalho anterior, ISE e IAE são utilizadas para verificar a qualidade da transferência de modelos nos controles de OD e nitrogênio.

Padhma (2023) utilizando uma RNA associada a AG, propõe uma modelagem para operar sistemas complexos que tratam compostos orgânicos específicos, como fenol, xilenol e cianeto. Para mensurar a qualidade do modelo, assim como em outros trabalhos, o autor utiliza a métrica RMSE e MAPE.

Sharma (2018) aplica RNAs em conjunto com programação genética. A métrica RMSE é utilizada para determinação da qualidade do modelo. A intenção é modelar o dado de sólidos suspensos totais em uma etapa específica do processo em uma Estação de tratamento.

Foi observado uma técnica em comum entre alguns autores que utilizaram redes neurais artificiais, o LSTM. Enquanto as aplicações, incluem principalmente a predição de variáveis importantes para determinação da eficiência da estação, como a DQO. Já as métricas utilizadas também apresentam concordância entre as publicações, com o MAPE, MAE e RMSE sendo utilizados para verificar a qualidade da modelagem utilizada e determinação do erro entre o valor predito e esperado.

A seguir, apresenta-se o levantamento do cenário de automação, instrumentação e controle no tratamento de esgoto, que servirá de base para as modelagens a serem realizadas nos capítulos subsequentes. Esse levantamento é essencial para compreender não apenas o nível das análises qualitativas realizadas na unidade, mas também o nível de controle existente, permitindo inferências futuras no processo, após a conclusão das modelagens. Além disso, o levantamento é acompanhado da proposta de um *dashboard*, cujo objetivo é proporcionar uma visualização mais holística sobre o tema.

### 3 ENTREVISTAS – PERSPECTIVAS DOS ESPECIALISTAS

Uma técnica de coleta de dados qualitativa amplamente utilizada em pesquisas sociais e organizacionais é a entrevista semiestruturada, que permite uma exploração ampla e adaptável das percepções dos entrevistados. A combinação de perguntas pré-estabelecidas com a capacidade dos entrevistadores de explorar assuntos que surgem durante as entrevistas permite que os entrevistadores obtenham tendências sobre os assuntos que não poderiam obter de entrevistas mais rígidas (KVALE, 1996). As entrevistas semiestruturadas são particularmente úteis para entender a complexidade das opiniões dos especialistas, fornecendo um equilíbrio entre a comparabilidade dos dados coletados e a profundidade das respostas individuais (COHEN; CRABTREE, 2006).

Esse tipo de entrevista foi escolhido para a pesquisa sobre a transformação digital no tratamento de esgoto para captar as percepções dos profissionais do setor sobre as mudanças tecnológicas e os problemas específicos que ocorrem no Brasil. Como resultado dessa abordagem, é possível examinar como os profissionais percebem e experimentam a incorporação de tecnologias como automação, *big data* e IA em suas práticas diárias. Isso dá a eles uma compreensão mais sólida dos impactos e possibilidades da transformação digital no saneamento (GALLETTA, 2013).

A análise de conteúdo, segundo o método de Bardin, é uma abordagem sistemática e estruturada para a interpretação de dados qualitativos. Esse método envolve algumas etapas. Inicialmente ocorre a pré-análise, onde os dados são organizados e preparados para análise. Após isso, na exploração do material, são identificadas as unidades de registro, ou os elementos significativos que foram identificados no material. Na próxima etapa, as unidades de registro são categorizadas de acordo com seu conteúdo e significado. É essencial nesse processo a criação de categorias, que são os conceitos ou temas que surgem dos dados. Por fim, os resultados são interpretados, e neste processo, as categorias são examinadas em relação ao contexto da pesquisa para descobrir padrões, tendências e significados fundamentais dos dados. O método de Bardin fornece uma estrutura sólida para a análise de conteúdo, o que permite que os pesquisadores explorem com mais profundidade as nuances e complexidades dos dados qualitativos (BARDIN, 1977).

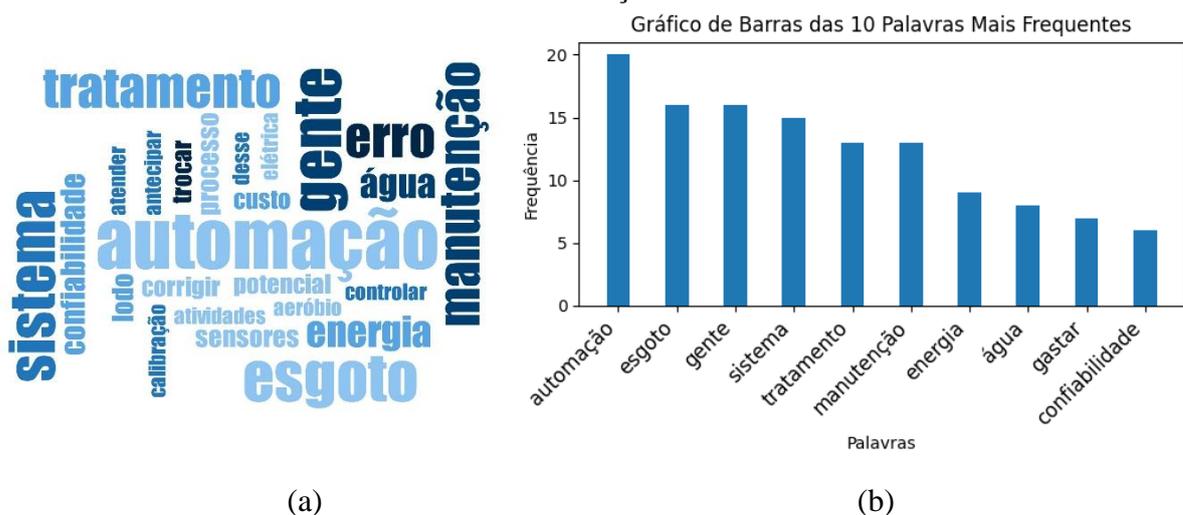
As entrevistas foram realizadas com o uso do software de vídeo conferências zoom. Inicialmente foi realizada breve apresentação do pesquisador e objetivo da entrevista, em seguida, quatro grupos de duas perguntas foram realizadas, com o objetivo de validação do objetivo da pesquisa, identificar potenciais usos da automação no setor, o impacto da transformação digital e as perspectivas futuras para o tratamento de esgoto.

Seis especialistas com carreiras significativas no tratamento de esgoto foram entrevistados. Dois deles são doutores, sendo que um deles possui tese com aplicação de Redes Neurais Artificiais no setor. Outros dois, ex-gerentes operacionais, um deles em Estação de Tratamento de Esgoto, e o outro na área de tecnologia, com experiência sólida em transformação digital. Por fim, temos um assessor técnico da superintendência em automação, e um ex-superintendente, também com implementações significativas em relação a automação, instrumentação e supervisão.

### 3.1 Resultados das Entrevistas

A seguir estão os principais achados em relação as perguntas direcionadas aos entrevistados (PESTANA, 2024). A primeira delas é sobre os potenciais usos da automação no tratamento de esgoto, Figura 21.

Figura 21 - Nuvem de palavras e gráfico de frequência de palavras dos potenciais usos da automação

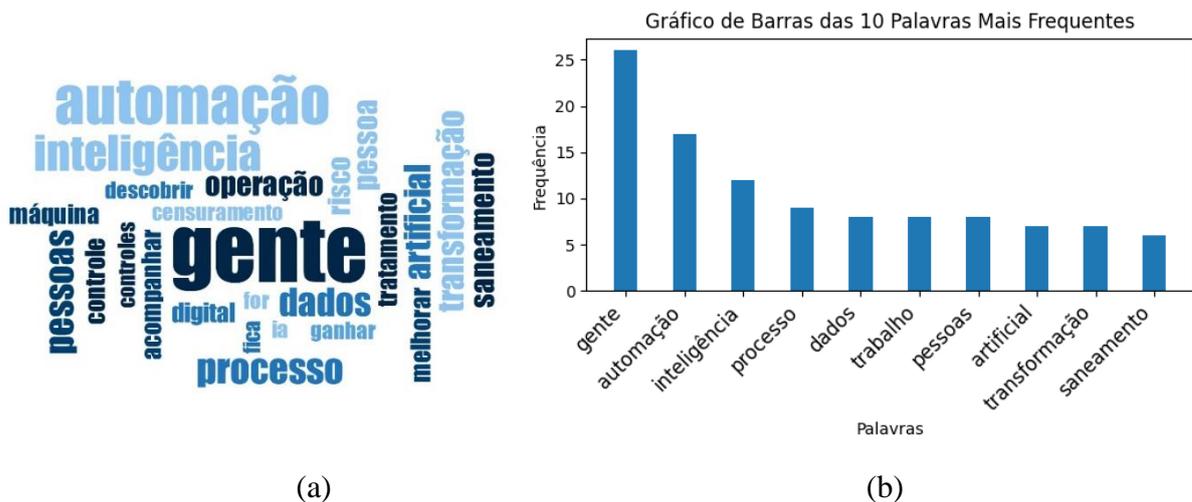


Fonte: Do autor com o uso do software MAXQDA e código Python, 2024.

As palavras-chave que se destacaram na nuvem de palavras, Figura 21 (a), foram "confiabilidade", "atender", "antecipar", "custo", "elétrica", "controlar" e "manutenção", de

acordo com uma análise qualitativa das respostas dos especialistas em automação sobre possíveis usos da automação no tratamento de esgoto. A Figura 21 (b) traz um gráfico de frequência em ordem decrescente das palavras mais encontradas. A automação tem o potencial de trazer uma série de melhorias significativas para o tratamento de esgoto, com foco em aumentar a eficiência operacional, reduzir custos, no caso das Estações com energia e químicos principalmente, e melhorar a confiabilidade do sistema. Um dos principais benefícios destacados pelos especialistas é a capacidade de antecipar problemas e atuar em equipamentos de forma preventiva, o que pode ajudar a evitar falhas inesperadas e garantir a operação contínua das instalações, trazendo um incremento expressivo na confiabilidade geral do tratamento. Outro resultado importante, diz respeito as tecnologias que os respondentes consideram emergentes e promissoras para o setor, Figura 22.

Figura 22 - Nuvem de palavras tecnologias emergentes



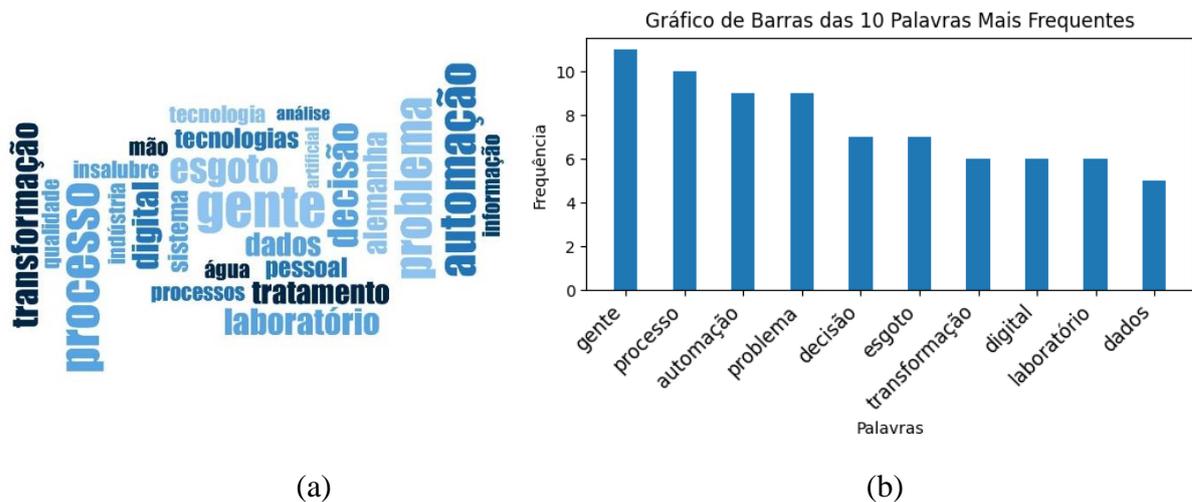
Fonte: Do autor com o uso do software MAXQDA e código Python, 2024.

A análise qualitativa revela um consenso claro sobre as tecnologias emergentes mais promissoras para a transformação digital no tratamento de esgoto: inteligência artificial (IA), tecnologias voltada a obtenção de insights através de dados e a digitalização, Figura 22 (a) e (b).

A IA se destaca como um elemento crucial para otimizar processos, reduzir custos e aprimorar a eficiência do tratamento de esgoto, isso devido a crescente capacidade computacional e volume de dados acerca dos processos de tratamento, cresce também a possibilidade de treinamento de algoritmos complexos com essas componentes. Através de técnicas como aprendizado de máquina e análise preditiva, sistemas inteligentes podem automatizar tarefas repetitivas, monitorar a qualidade do tratamento em tempo real, prever

falhas e otimizar o uso de recursos. Por fim, tem-se a perspectiva desses especialistas acerca de como essas tecnologias podem impactar a eficiência do processo, Figura 23.

Figura 23 - Nuvem de palavras impacto no tratamento de esgoto



Fonte: Do autor com o uso do software MAXQDA e código Python, 2024.

De acordo com a análise qualitativa, as palavras-chave "qualidade", "processo", "decisão", "análise", "digital" e "informação" se destacaram na nuvem de palavras, Figura 23 (a). Esses termos sugerem que os especialistas veem a transformação digital como uma oportunidade para melhorar a qualidade do tratamento de esgoto, otimizando processos por meio de análise de dados e tomada de decisão mais informada. A Figura 23 (b) e a frequência das palavras em ordem decrescente, corroboram com essa análise com os termos "processo", "problema" e "decisão" em destaque.

A transformação digital pode impactar positivamente a eficiência do tratamento de esgoto ao permitir uma análise mais detalhada dos processos envolvidos. Como os instrumentos de voo que auxiliam os pilotos com informações detalhadas, mas sobretudo com o uso de técnicas modernas para transformação desses dados em *insights* que auxiliem a operação. A obtenção de informações em tempo real sobre o funcionamento das instalações de tratamento possibilita uma melhor compreensão dos padrões de operação e a identificação precoce de problemas, seja no âmbito dos equipamentos e manutenção, seja na qualidade dos processos de tratamento. Isso pode levar a uma manutenção mais proativa e à resolução rápida de problemas, contribuindo para a redução de falhas e melhorando a confiabilidade das operações.

As percepções dos especialistas em automação sobre a transformação digital no tratamento de esgoto revelam uma crescente conscientização sobre a importância dessas

inovações tecnológicas. A análise das respostas coletadas por meio de entrevistas qualitativas semiestruturadas destaca a importância de temas como automação, dados e inteligência artificial, demonstrando a relevância desses aspectos para o setor de saneamento, sobretudo com o novo marco legal do saneamento. Os profissionais entrevistados reconhecem os potenciais benefícios da transformação digital, como melhorias na eficiência, qualidade dos processos de tratamento de esgoto, e maior volume tratado por meio da integração de sistemas e do uso de dados para tomada de decisões mais assertivas.

No entanto, também foram identificados desafios significativos associados à adoção de tecnologias digitais. A necessidade de investimentos em infraestrutura e tecnologia, garantia da segurança e proteção dos dados, atualização do parque de equipamentos existentes, bem como a capacitação e conscientização dos usuários, foram alguns dos obstáculos mencionados. A transição para o tratamento de esgoto digitalizado exige uma abordagem cuidadosa e estratégica para superar esses desafios e aproveitar os benefícios potenciais.

As respostas obtidas revelaram que as tecnologias emergentes mais promissoras para o setor incluem inteligência artificial e as técnicas atuais de análise de dados. A implementação de soluções baseadas em IA e análise de dados permitirá uma gestão mais sustentável e eficaz dos recursos hídricos, possibilitando a previsão de problemas, seja operacional ou de manutenção, e a tomada de decisões proativas.

Diante desses resultados, é evidente que a transformação digital tem o potencial de revolucionar o setor de saneamento, melhorando não apenas a eficiência e qualidade do tratamento de esgoto, mas também contribuindo para uma gestão mais sustentável dos recursos hídricos. No entanto, para que isso ocorra de forma eficaz, é necessário enfrentar os desafios identificados e adotar abordagens inovadoras para impulsionar a transformação digital no tratamento de esgoto, incluindo os profissionais envolvidos em cada uma das etapas.

#### 4 UMA SURVEY COM PROFISSIONAIS DO SETOR

Para atender aos objetivos da pesquisa, o método foi utilizar uma *Survey* realizada com profissionais que atuam no tratamento de esgoto. A *Survey* foi dividida em quatro grupos de perguntas. O primeiro grupo consiste em cinco questões que visam caracterizar o perfil dos respondentes, com informações sobre sua formação, tempo experiência e grau de conhecimento sobre transformação digital. Em seguida, foram formuladas oito questões com escala Likert direcionadas ao corpo operacional, buscando avaliar sua percepção sobre diversos aspectos da transformação digital, como melhoria do trabalho, redução de riscos e melhoria do monitoramento de processos. Da mesma forma, foram elaboradas outras oito questões Likert direcionadas ao corpo gerencial, focando em aspectos estratégicos, como tomada de decisão, redução de custos e melhoria da qualidade do tratamento da planta.

A diferenciação das questões para os dois grupos foi feita para captar a percepção da transformação digital em dois contextos diferentes. Enquanto o corpo operacional pode se beneficiar de maneira mais direta com a eficiência operacional, o corpo gerencial pode obter mais dados para melhorar a tomada de decisão.

Finalmente, o último grupo de perguntas consiste em seis questões abertas, que visam identificar benefícios, dificuldades e as principais tecnologias emergentes no tratamento de esgoto. Essas questões abertas permitem uma análise mais aprofundada das percepções dos profissionais.

O método de seleção do público alvo da pesquisa “Transformação digital no tratamento de esgoto” foi feita com base em critérios de experiência e carreira significativas na área de tratamento de esgoto. A pesquisa foi conduzida em uma Unidade de Tratamento de Esgoto com aproximadamente 400 funcionários. Optou-se por uma amostra de conveniência, considerando a disponibilidade e o interesse dos funcionários em participar do estudo. A abordagem envolveu a divulgação do questionário por meio de comunicação interna, como e-mails, incentivando a participação voluntária. No total, aproximadamente 70 funcionários responderam ao questionário, representando diferentes áreas e funções da Unidade, o que proporcionou uma visão diversificada sobre o tema da transformação digital no contexto do tratamento de esgoto.

A *Survey*, ou pesquisa de opinião, é uma metodologia amplamente utilizada para coletar

dados e informações sobre determinado tema por meio de questionários estruturados. Podem também ser descritos como uma investigação a partir de uma determinada amostra e descrita analiticamente (CENDÓN; RIBEIRO; CHAVES, 2014). Esse método de estudo permite ao pesquisador obter a percepção e opiniões de determinado grupo de indivíduos em relação a determinado assunto.

As *Surveys* podem ser conduzidas de diversas maneiras, incluindo entrevistas presenciais e questionários online. Esses questionários podem conter uma combinação de perguntas fechadas (como aquelas que utilizam a escala Likert) e perguntas abertas, onde o participante expressar suas opiniões de maneira mais livre e detalhada. Nesse caso, os *web Surveys*, que são as pesquisas realizadas pelos meios digitais, geralmente são automatizados e facilitam a compilação das respostas obtidas.

A escala Likert, criada por Rensis Likert em 1932, é amplamente utilizada em pesquisas sociais e comportamentais para avaliar as atitudes, percepções e opiniões dos respondentes em relação a um determinado assunto. Esta escala permite que os participantes expressem seu grau de concordância ou discordância com afirmações, geralmente de cinco a sete pontos, variando de "concordo totalmente" a "discordo totalmente". A principal vantagem da escala Likert é sua simplicidade e sua capacidade de quantificar objetivamente percepções subjetivas, o que facilita a análise estatística e a interpretação dos dados coletados (LIKERT, 1932).

#### **4.1 Resultados da Survey**

Com base nos resultados da pesquisa, é evidente que os profissionais a qual a pesquisa foi direcionada reconhecem amplamente os benefícios da automação e da transformação digital. A maioria dos respondentes concorda que a automação pode tornar as operações mais eficientes (71,1%), melhorar a monitorização em tempo real (65,8%) e aumentar a segurança dos trabalhadores (63,2%). Além disso, essas tecnologias também são vistas como uma ferramenta para aumentar a produtividade da equipe operacional (65,8%) e reduzir o tempo necessário para identificar e corrigir problemas (60,5%), (PESTANA,2024).

De maneira menor ou com concordância parcial, existe a percepção sobre a automação reduzir o risco de erros humanos (57,9%). Isso pode se dar ao fato de que mesmo com a automação, os erros humanos ainda serem possíveis e talvez até comuns, sendo possível talvez os diminuir, mas não extinguir a existência.

No entanto, também é importante notar que existem algumas áreas de preocupação e desafios identificados pelos profissionais. Cerca de metade dos respondentes acredita que a automação pode aumentar a capacidade de adaptação a condições variáveis de operação (47,4%), mas uma porcentagem igualmente significativa concorda parcialmente ou não tem certeza sobre isso. Talvez isso se deva ao fato de que a automação pode auxiliar nessa identificação, mas o expertise e know-how de profissionais com longa carreira não deva ser descartado. Da mesma forma, embora a maioria concorde que a automação pode melhorar a capacidade de resposta a eventos de emergência (63,2%), uma proporção considerável não concorda totalmente.

De maneira análoga ao que foi direcionado ao corpo operacional, 8 questões em escala Likert foram formuladas aos respondentes pertencentes ao corpo gerencial, porém visando capturar o ponto de vista de alguém que planeja, lidera e controla recursos dentro de uma ETE.

Com base nos resultados da pesquisa direcionada ao corpo gerencial da Unidade de tratamento de esgoto, novamente fica claro que há uma percepção sobre os benefícios da adoção de tecnologias digitais. A maioria dos respondentes concorda que a transformação digital pode melhorar a tomada de decisão na gestão das estações de tratamento de esgoto (75,9%), reduzir os custos operacionais (69,0%).

Com uma fração um pouco maior, a opinião sobre a transformação digital simplificar o acesso a dados e informações críticas em um único local centralizado (79,3%) e melhorar a eficiência da manutenção preventiva das instalações (79,3%). Isso pode representar o desejo de que a automação e transformação digital auxiliem nesses objetivos.

Além disso, a automação é vista como uma facilitadora para o cumprimento das regulamentações ambientais (69,0%). A transformação digital também é percebida como uma maneira de aumentar a transparência nas operações das estações de tratamento (69,0%), fornecer indicadores-chave de desempenho em tempo real (75,9%). De maneira menor tem-se a percepção que essas tecnologias podem contribuir para melhorar a qualidade do efluente tratado (58,6%), e isso pode representar a experiência dos respondentes de que a qualidade do efluente talvez dependa também de outras variáveis.

Esses resultados destacam a importância atribuída pelos gestores à transformação digital como uma ferramenta para otimizar as operações, reduzir custos, garantir conformidade com as regulamentações e melhorar a qualidade do tratamento de esgoto. No entanto, é importante

notar que alguns respondentes expressaram concordância parcial em relação a algumas questões, indicando que a transformação digital pode ser uma ferramenta importante, mas não deve ser considerada a solução de todos os problemas.

Uma nuvem de palavras foi criada com as respostas obtidas pelo questionário com a utilização do software MAXQDA, conforme Figura 24 (a). Acerca dos desafios encontrados atualmente pelos respondentes nas ETEs objeto de estudo, é possível observar relatos sobre a gestão, volume do esgoto tratado, eficiência, monitoramento e modernização.

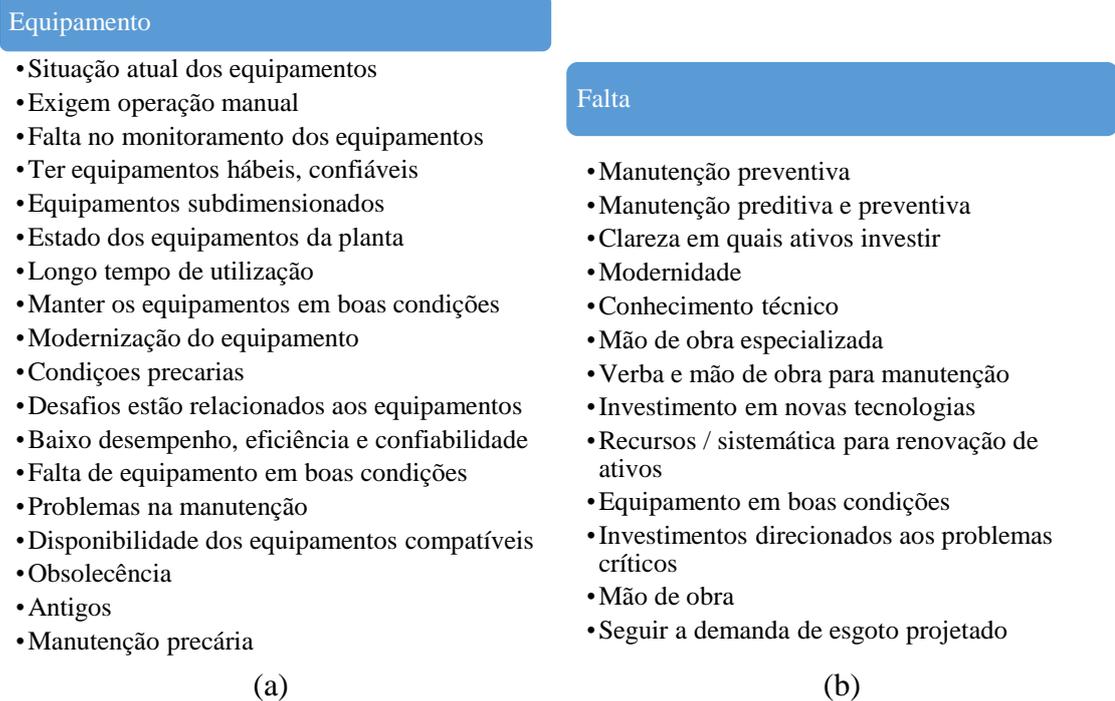
Figura 24 - Nuvem de palavras para os desafios enfrentados



Fonte: Do autor com o uso do software MAXQDA, 2024.

A Figura a direita, Figura 24 (b), traz uma síntese das 5 palavras mais encontradas: Equipamentos, falta, tratamento, manutenção e automação. É possível observar que apesar dos esforços e investimentos em relação a controle, automação, dados e transformação digital, ainda existe uma necessidade relacionada a confiabilidade dos equipamentos responsáveis pelo tratamento, como bombas, válvulas, filtros e afins. A Figura 25 (a) e (b), nos traz os segmentos das respostas obtidas que continham as palavras “Equipamento” e “Falta” reforçam essa apuração.

Figura 25 - Segmentos das respostas com as palavras "Equipamentos" e "Falta"

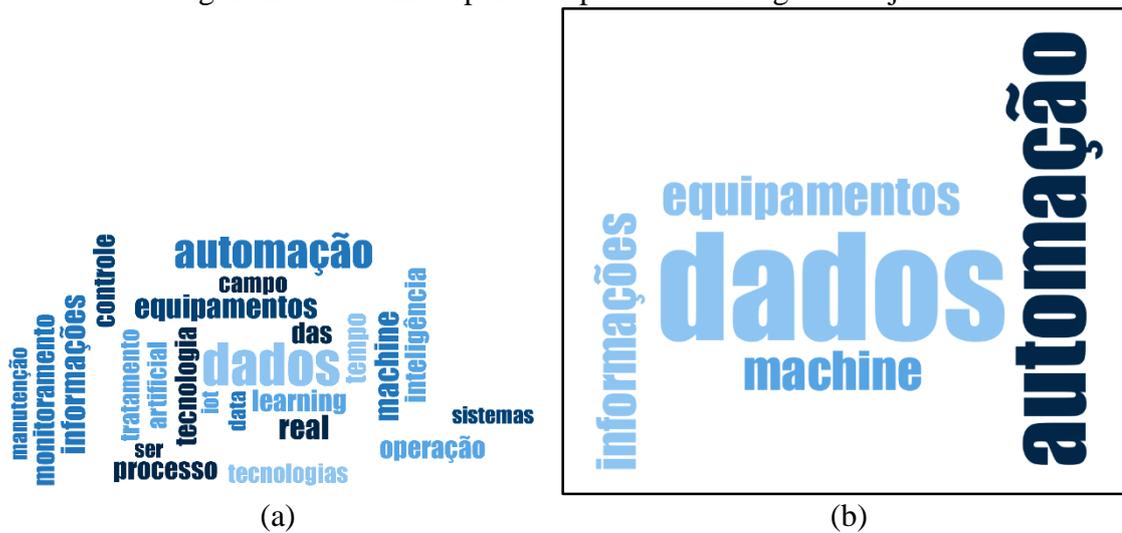


Fonte: Do autor com o uso do software MAXQDA, 2024.

Problemas relacionados a manutenção dos equipamentos, obsolescência dos mesmos e suas condições são listados. Enquanto a percepção de investimentos e mão de obra inadequada também é citada. Outro fato observado, é que dos relatos acerca dos problemas com equipamentos, a maior parte dos relatos é feita pela equipe operacional (56%).

Uma nuvem de palavras também foi criada com as respostas obtidas em relação as tecnologias que os usuários acreditam que poderiam trazer mais benefícios em relação a operação das ETEs, Figura 26 (a). É possível observar que o monitoramento e o uso de inteligência estão entre o desejo dos respondentes.

Figura 26 - Nuvem de palavras para as tecnologias desejadas



Fonte: Do autor com o uso do software MAXQDA, 2024.

A Figura 26 (b) traz uma síntese das 5 palavras mais encontradas: Dados, automação, equipamentos, informações e *machine learning*. É possível observar que com relação aos dados, o interesse dos operadores e gestores está voltado a disponibilidade, comunicação, centralização e análise dos mesmos afim de suportar seus processos de decisão, conforme Figura 27:

Figura 27 - Segmentos das respostas com a palavra "dados"

#### Dados

- Disponibilidade
- Remotas e em tempo real
- Obtenção
- Algoritmo de decisão
- Modelagem de comportamento
- Análise confiável
- Suficientes
- Comunicação
- Captação
- Centralização
- Supervisão
- Análise
- Modelagem

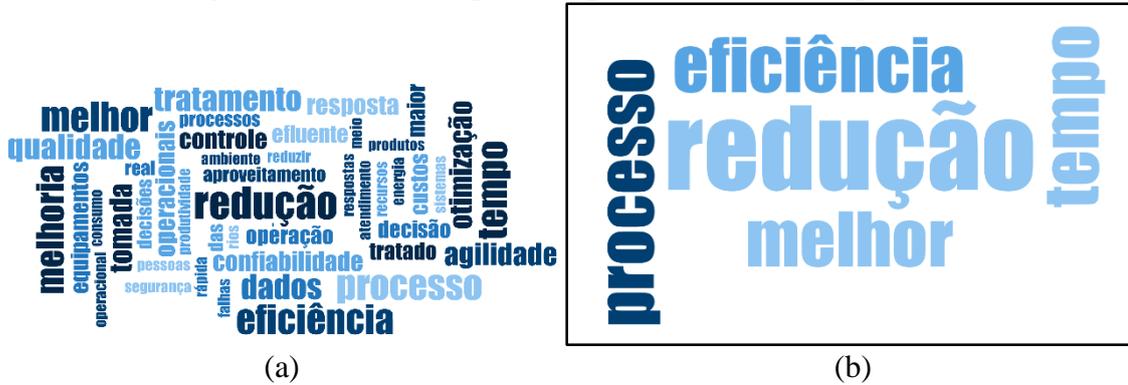
Fonte: Do autor com o uso do software MAXQDA, 2024.

A confiabilidade desses dados, inclusive de plantas remotas também foi observada. Outro fato é que das respostas que são relacionadas melhoria em relação aos dados, a maior parte (58%) é feita por gestores ou universitários.

Diversos benefícios são esperados pelos usuários, entre eles, melhor controle operacional, incremento na qualidade do efluente, segurança, agilidade e resposta a eventos

atípicos ao processo, a nuvem de palavras é vista na Figura 28 (a).

Figura 28 - Nuvem de palavras para os benefícios esperados



Fonte: Do autor com o uso do software MAXQDA, 2024.

A nuvem criada, novamente traz as 5 palavras mais encontradas nas respostas, Figura 28 (b): Redução, eficiência, processo, melhor e tempo. Para o termo redução o corpo operacional e gerencial espera que seja possível para os riscos, impactos ambientais, tempo de deslocamento e os custos inerentes ao processo de tratamento, conforme a Figura 29 (b).

Figura 29 - Segmentos das respostas com as palavras “melhor” e “redução”

#### Melhor

- Resposta para correção de processos
- Dos processos
- Performance operacional
- Eficiência
- Melhorias no nosso trabalho
- Melhoria contínua de todo o processo
- Resposta às tendências no tratamento
- Aproveitamento geral
- Tempo de atendimento as demandas de campo
- Controle do sistema
- Gerenciamento
- Tomadas de decisões
- Resultados
- Qualidade do efluente tratado
- Aproveitamento do corpo técnico
- Recursos melhor empregados
- Logística
- Controle de processo
- Qualidade dos rios
- Controle
- Confiança nos dados

(a)

#### Redução

- Riscos
- Impacto ambiental
- Custo
- Consumo de produtos químicos
- Tempo em deslocamento
- Custos operacionais
- Esforço físico
- Falhas humanas

(b)

Fonte: Do autor com o uso do software MAXQDA, 2024.

Enquanto a melhoria esperada é na eficiência e performance operacional, respostas e

correções as mudanças nas características do esgoto tratado, melhores controles e qualidade do efluente, Figura 29 (a) e Figura 30.

Figura 30 - Segmentos das respostas com a palavra “processo”

#### Processo

- Melhoria na resposta para correção
- Análise de dados e melhoria
- Melhoria contínua
- Excelência
- Aumentar a eficiência e confiabilidade
- Simplificação do controle
- Melhor controle do sistema e aproveitamento
- Tempo real
- Resposta mais rápida e eficiente no controle
- Redução do consumo de energia
- Melhor controle
- Maior eficiência
- Eficiência
- Aumento da confiabilidade

Fonte: Do autor com o uso do software MAXQDA, 2024.

Em resumo, as observações e sugestões dos profissionais enfatizam a importância de uma abordagem holística para a transformação digital no tratamento de esgoto, que envolva não apenas a adoção de tecnologias avançadas, mas também o desenvolvimento de habilidades, a modernização de equipamentos e a gestão eficaz da mudança.

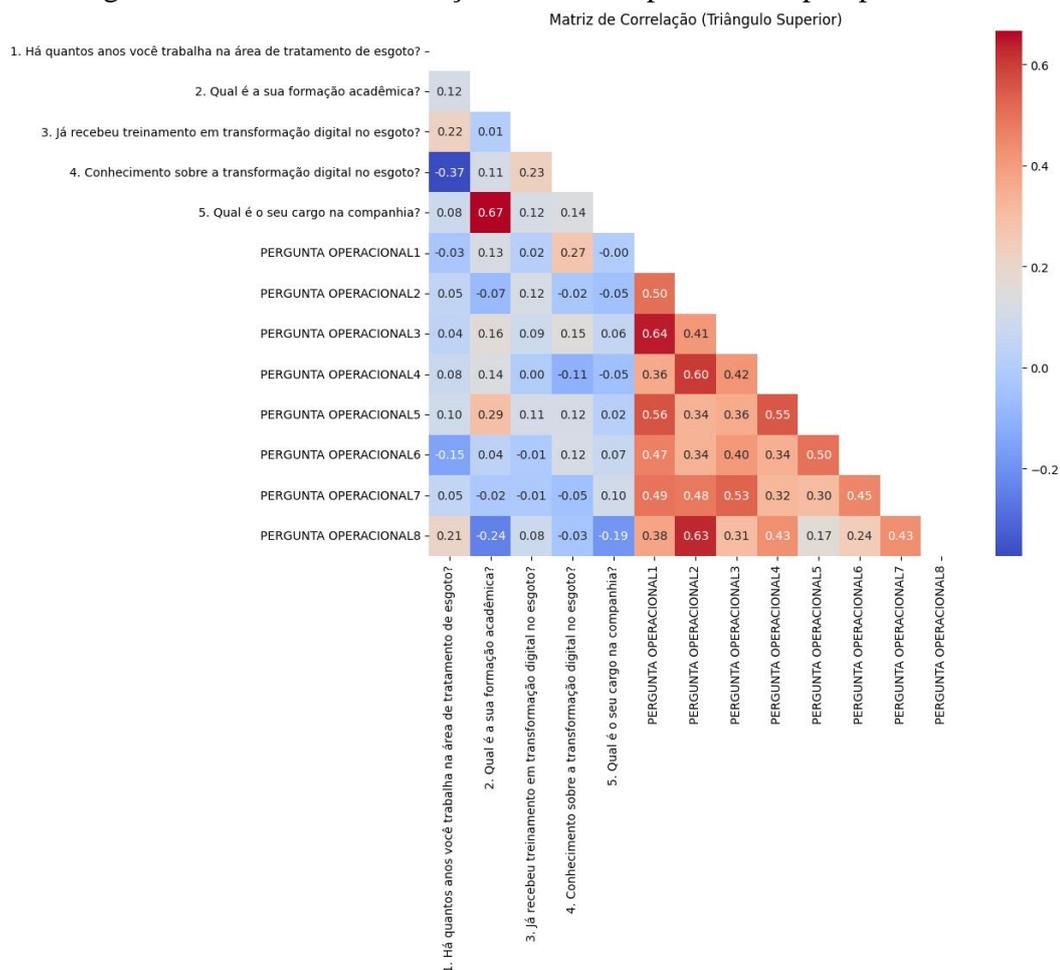
A pesquisa revelou que a maioria dos respondentes vê a automação e a transformação digital como essenciais para o desenvolvimento do setor, especialmente em ETEs. Essa percepção destaca o interesse crescente em integrar tecnologias que tornem os processos mais ágeis e eficientes. No entanto, os entrevistados também identificaram desafios significativos, principalmente relacionados ao estado atual dos equipamentos e à digitalização dos mesmos. Esse cenário evidencia a importância de investimentos não apenas em novas tecnologias, mas também na atualização e adequação dos sistemas já instalados.

Entre as tecnologias mais desejadas pelos operadores e gestores, a inteligência artificial (IA) e a análise de dados se destacam, devido ao potencial de otimização que oferecem. Os benefícios associados ao uso dessas ferramentas incluem uma maior confiabilidade e eficiência operacional, redução de custos com produtos químicos e energia, e o cumprimento das exigências legais. Com essas soluções, as empresas podem melhorar a gestão dos recursos e garantir um serviço mais sustentável e de qualidade para a população, reforçando o papel da

inovação tecnológica na transformação digital para o setor de saneamento.

A Figura 31, traz uma matriz de correlação entre as respostas obtidas para as perguntas direcionadas ao corpo operacional. Por exemplo, quando observado o tempo de atuação no tratamento de esgoto, exista uma correlação sutil (-0,37) inversamente ao conhecimento sobre transformação digital. Ou seja, colaboradores mais antigos podem conhecer menos sobre o tema. Da mesma maneira, maiores cargos estão atrelados a maior formação (+0,67). Outro fato, é que maiores conhecimentos acerca de transformação digital podem estar atrelados a concordância sobre o impacto da mesma na eficiência dos processos (0,27). Já entre as respostas em escala Likert, boa parte delas possui correlação forte positiva entre as outras, e a maior delas (0,64) ocorre entre a concordância da pergunta 1 e 3, que respectivamente avalia o impacto da automação e transformação digital na eficiência e monitorização dos processos.

Figura 31 - Matriz de correlação entre as respostas - Corpo operacional

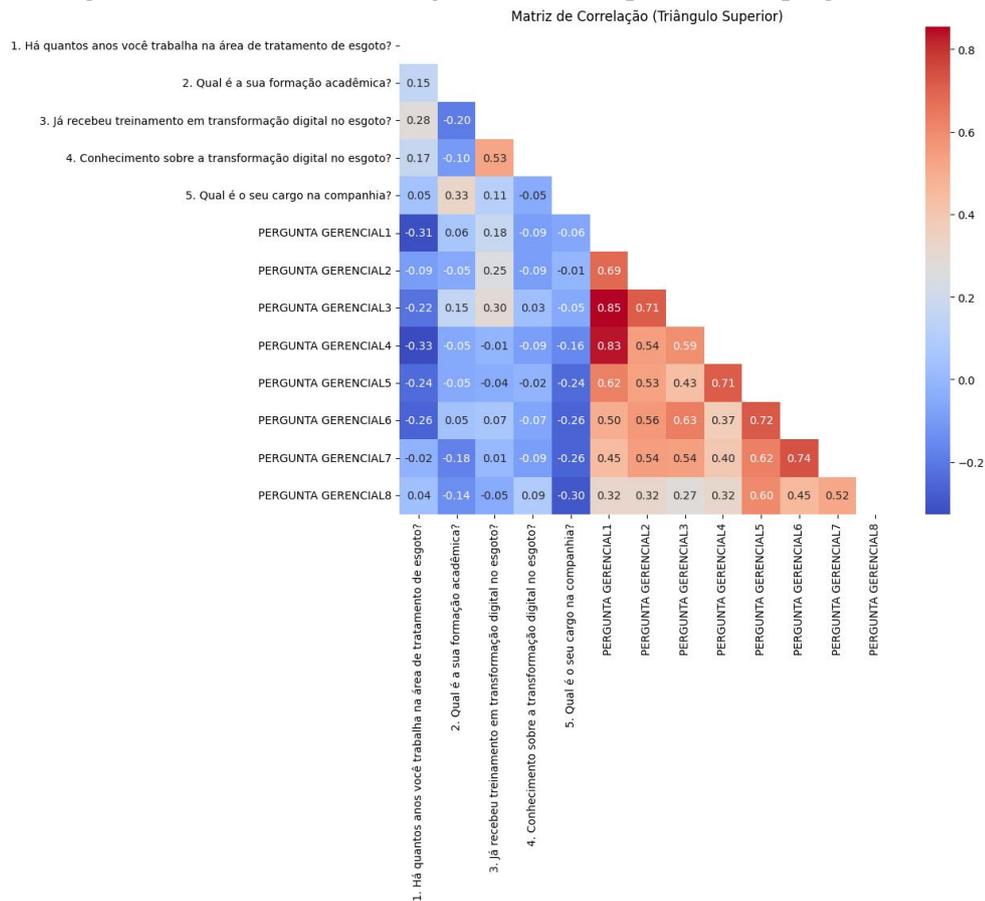


Fonte: Do autor, 2025.

Já a Figura 32, traz uma matriz de correlação entre as respostas obtidas para as perguntas direcionadas ao corpo gerencial. Da mesma forma que na análise feita no grupo operacional,

quando observado o conhecimento sobre transformação digital a correlação é forte com treinamentos realizados sobre o mesmo tema (0,53). Entre as respostas em escala Likert, novamente, boa parte delas possui correlação forte positiva entre as outras, com exceção da pergunta 8, que avalia se a transformação digital simplifica o acesso a dados e informações críticas em um único centralizado. Ou seja, a concordância entre essa resposta ocorre, porém de maneira menor do que entre as outras, e pode não haver uma certeza.

Figura 32 - Matriz de correlação entre as respostas - Corpo gerencial



Fonte: Do autor, 2025.

A partir do capítulo seguinte tem-se de fato a aplicação das RNAs em ETEs da região metropolitana de São Paulo. Algumas análises prévias foram realizadas, como a matriz de correlação, PCA, clueterização com *K-means*, e análise e remoção de *outliers*. Com relação as RNAs, foram testadas em conjunto com a técnica *Grid Search*, SHAP, e também as TDNNs. As principais variáveis de interesse foram a DQO e também amônia.

## 5 APLICAÇÃO PRÁTICA DO MODELO

A companhia de saneamento objeto de estudo, opera diversas ETEs espalhadas pelo seu Estado, sendo responsável por tratar milhões de metros cúbicos de esgoto diariamente. Esses sistemas complexos exigem um monitoramento constante de parâmetros críticos, como DQO, DBO e concentração de amônia, para garantir a eficiência dos processos e o atendimento às normas ambientais.

Historicamente, essa companhia era diretamente responsável pela instalação e manutenção dos instrumentos de medição, bem como pela aquisição dos consumíveis necessários para as análises. No entanto, a partir de 2021, ocorreu uma mudança significativa na gestão desses processos. A companhia passou a adotar um novo modelo de contrato, onde os dados analíticos são adquiridos de fornecedores especializados. Nesse modelo, a empresa contratada é remunerada com base na disponibilidade e qualidade dos dados fornecidos, aliviando a empresa de saneamento de responsabilidades operacionais diretas, mas ao mesmo tempo impondo novos desafios na gestão e utilização desses dados.

Essa mudança trouxe à tona a necessidade de maximizar o valor gerado a partir dos dados analíticos obtidos. Não basta apenas ter acesso a dados precisos e disponíveis, é imperativo que eles sejam utilizados de maneira estratégica para melhorar a eficiência energética e operacional das plantas, além de possibilitar a predição de parâmetros críticos que possam impactar diretamente o desempenho das estações de tratamento. Dessa forma, o foco deste estudo é analisar a qualidade e a utilidade dos dados coletados, com vistas a identificar oportunidades de otimização e melhorias nos processos de tratamento de esgoto na companhia de saneamento objeto de estudo.

### 5.1 Caracterização dos dados

A região objeto de estudo possui cinco grandes ETEs, aqui elas serão descritas como: ETE A, B, C, D e E, com vazões afluente em seus sistemas (ETEs principais e sistemas isolados) de 15588 l/s, 3128 l/s, 19511/s, 1592 l/s e 977 l/s respectivamente para o ano de 2023.

Para a análise dos dados das ETEs e o desenvolvimento dos modelos, foi utilizada a linguagem de programação Python devido à sua ampla gama de bibliotecas específicas para

ciência de dados e aprendizado de máquina. As principais bibliotecas aplicadas neste trabalho incluem: Pandas, para manipulação e transformação de conjuntos de dados (PANDAS, 2024); Scikit-learn (sklearn), para o pré-processamento, divisão de dados e criação de redes neurais (SCIKIT-LEARN, 2024); Numpy, para operações numéricas eficientes (NUMPY, 2024); e Matplotlib, para visualização gráfica (MATPLOTLIB, 2024).

Os conjuntos de dados das cinco ETEs foram apresentados em registros com periodicidade de 5, 10 e 15 minutos, sendo necessária uma padronização para facilitar a análise comparativa e o desenvolvimento dos modelos. As transformações ocorreram em três fases principais:

- A primeira versão do conjunto de dados: Média Horária

Inicialmente, foi calculada a média horária dos registros para padronizar a frequência dos dados. Essa agregação foi necessária para que todas as linhas utilizassem dados consistentes, evitando lacunas nos registros com diferentes periodicidades. A partir desta primeira versão, foram criados gráficos descritivos para cada planta, permitindo uma caracterização visual inicial dos dados, como médias e tendências.

- Segunda versão do conjunto de dados: Remoção de Dados Faltantes

Na segunda etapa, os dados faltantes foram removidos, criando um conjunto de dados limpo. Esse novo conjunto foi utilizado na geração de uma matriz de correlação entre os parâmetros monitorados e na construção dos gráficos *box-plot*, essencial para identificar variações e a distribuição dos dados em cada ETE.

- Terceira versão do conjunto de dados: Remoção de *Outliers*

A última versão do *dataset* incluía a remoção de *outliers* através da técnica do IQR. Esse processo foi essencial para evitar distorções nos resultados das análises. Com o conjunto de dados sem *outliers*, foi possível gerar um segundo conjunto de *box-plots*, realizar a análise PCA e também aplicando a técnica de SHAP para interpretar os resultados das RNAs, fornecendo insights sobre a influência de cada parâmetro no modelo.

Por fim, temos as bases de dados utilizadas nas modelagens conforme a Tabela 4.

Tabela 4 - Resumo das bases utilizadas

|       | Base de dados de 1 mês   |                 |                                       |                 |  |                 | Base de dados (2 meses) | Base de dados (4 meses)<br>Média 15 min |                 |                 |
|-------|--------------------------|-----------------|---------------------------------------|-----------------|--|-----------------|-------------------------|---|-----------------|-----------------|
|       | Versão 1 (Média Horária) |                 | Versão 2 (Remoção de dados faltantes) |                 | Versão 3 (Remoção de <i>Outliers</i> ) |                 | Versão final            |   | Versão final    |                 |
|       | <i>Features</i>          | <i>Amostras</i> | <i>Features</i>                       | <i>Amostras</i> | <i>Features</i>                        | <i>Amostras</i> | <i>Features</i>         | <i>Amostras</i>                         | <i>Features</i> | <i>Amostras</i> |
| ETE A | 34                       | 743             | 32                                    | 650             | 32                                     | 650             | 28                      | 1423                                    | 27              | 5663            |
| ETE B | 57                       | 743             | 52                                    | 394             | 52                                     | 394             | 47                      | 1132                                    |                 |                 |
| ETE C | 37                       | 743             | 29                                    | 633             | 29                                     | 633             | 27                      | 1433                                    | 27              | 7672            |
| ETE D | 40                       | 743             | 39                                    | 272             | 39                                     | 272             | 32                      | 1124                                    | 36              | 10902           |
| ETE E | 18                       | 743             | 15                                    | 344             | 15                                     | 344             | 11                      | 1443                                    | 17              | 11589           |

Fonte: Do autor (2025)

## 5.2 Início das análises nas Estações

### 5.2.1 ETE A

A ETE A beneficia uma população equivalente a 1.400.000 habitantes. Tem uma vazão média de projeto de 3 m<sup>3</sup>/s e uma vazão atual de 3128 l/s. O esgoto é transportado por um sistema de 125 km e tratado através do processo biológico de lodos ativados, com 90% de eficiência na remoção de matéria orgânica (DBO).

Na Tabela 5, a descrição da instrumentação presente na ETE A, o local de instalação e a faixa de trabalho para a análise. A principal variável de interesse é DQO de saída da Estação (efluente final), identificada pela TAG “P1-AIT-506”.

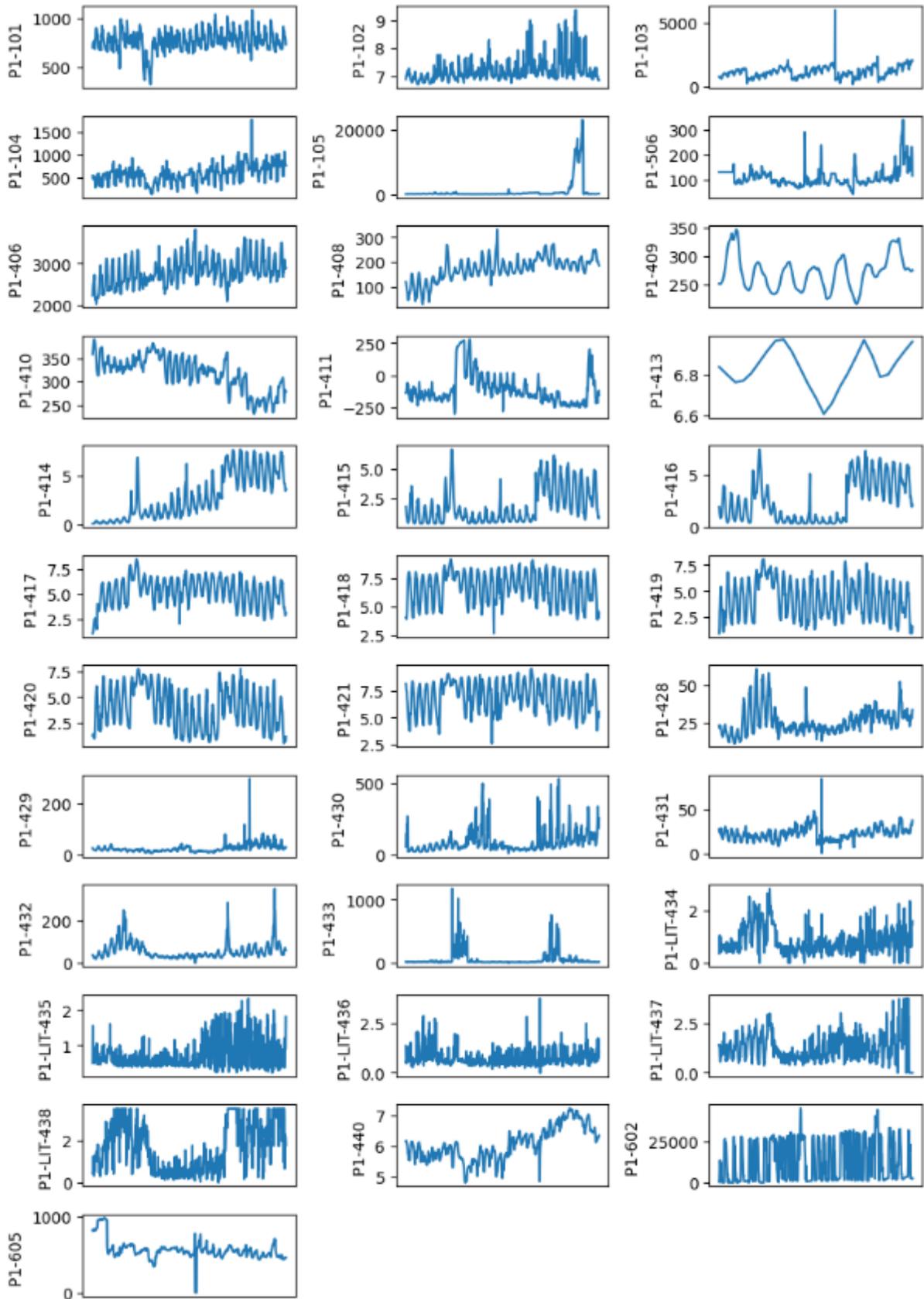
Tabela 5 - Relação de instrumentos por planta - ETE A

|       | TAG        | Parâmetro     | Local                       | Faixa de Trabalho  |
|-------|------------|---------------|-----------------------------|--------------------|
| ETE A | P1-AIT-101 | pH            | Esgoto bruto                | 4 a 12             |
|       | P1-AIT-102 | Condutividade | Esgoto bruto                | 0-10K $\mu$ S/cm   |
|       | P1-AIT-103 | SST (imersão) | Esgoto desarenado           | 100 a 500 mg/L     |
|       | P1-AIT-104 | DQO           | Esgoto desarenado           | 200 a 1.000 mg/L   |
|       | P1-AIT-105 | SST (imersão) | Efluente Primário           | 20 a 200 mg/L      |
|       | P1-AIT-506 | DQO           | Efluente Final              | 10 a 600 mg/L      |
|       | P1-AIT-406 | SST (imersão) | Licor misto                 | 1.000 a 8.000 mg/L |
|       | P1-AIT-407 | SST (imersão) | Licor misto                 | 1.000 a 8.000 mg/L |
|       | P1-AIT-408 | ORP           | Reator Aeróbio 1A           | (-175 a -400 mV)   |
|       | P1-AIT-409 | ORP           | Reator Aeróbio 1B           | (-175 a -400 mV)   |
|       | P1-AIT-410 | ORP           | Reator Aeróbio 2A           | (-175 a -400 mV)   |
|       | P1-AIT-411 | ORP           | Reator Aeróbio 2B           | (-175 a -400 mV)   |
|       | P1-AIT-412 | ORP           | Reator Aeróbio 2A           | (-175 a -400 mV)   |
|       | P1-AIT-413 | ORP           | Reator Aeróbio 2B           | (-175 a -400 mV)   |
|       | P1-AIT-414 | OD            | Reator Aeróbio 1A - Saída   | 0 a 8 PPM          |
|       | P1-AIT-415 | OD            | Reator Aeróbio 1B - Entrada | 0 a 8 PPM          |
|       | P1-AIT-416 | OD            | Reator Aeróbio 2A - Entrada | 0 a 8 PPM          |
|       | P1-AIT-417 | OD            | Reator Aeróbio 2B - Saída   | 0 a 8 PPM          |
|       | P1-AIT-418 | OD            | Reator Aeróbio 2A - Saída   | 0 a 8 PPM          |
|       | P1-AIT-419 | OD            | Reator Aeróbio 2B - Entrada | 0 a 8 PPM          |
|       | P1-AIT-420 | OD            | Reator Aeróbio 1A - Entrada | 0 a 8 PPM          |
|       | P1-AIT-421 | OD            | Reator Aeróbio 1B - Saída   | 0 a 8 PPM          |
|       | P1-AIT-426 | NH3           | Efluente Primário           | 10-150mg/L         |
|       | P1-AIT-427 | NH3           | Licor misto                 | 0,05-20mg/L        |
|       | P1-AIT-428 | Turbidez      | Efluente DS1                | 0-100 NTU          |
|       | P1-AIT-429 | Turbidez      | Efluente DS2                | 0-100 NTU          |
|       | P1-AIT-430 | Turbidez      | Efluente DS3                | 0-100 NTU          |
|       | P1-AIT-431 | Turbidez      | Efluente DS4                | 0-100 NTU          |
|       | P1-AIT-432 | Turbidez      | Efluente DS5                | 0-100 NTU          |
|       | P1-AIT-433 | Turbidez      | Efluente DS6                | 0-100 NTU          |
|       | P1-LIT-434 | Manto de Lodo | Tanque DS1                  | 0-6metros          |
|       | P1-LIT-435 | Manto de Lodo | Tanque DS2                  | 0-6metros          |
|       | P1-LIT-436 | Manto de Lodo | Tanque DS3                  | 0-6metros          |
|       | P1-LIT-437 | Manto de Lodo | Tanque DS4                  | 0-6metros          |
|       | P1-LIT-438 | Manto de Lodo | Tanque DS5                  | 0-6metros          |
|       | P1-LIT-439 | Manto de Lodo | Tanque DS6                  | 0-6metros          |
|       | P1-AIT-440 | pH            | Efluente Final              | 2 a 12             |
|       | P1-AIT-601 | SST (imersão) | Lodo Aflu. Digestor         | 20.000 a 50.00mg/L |
|       | P1-AIT-602 | SST (imersão) | Lodo condicionado           | 20.000 a 50.00mg/L |
|       | P1-AIT-603 | SST (imersão) | Filtrado FPP 1 a 3          | 1.000 a 10.00 mg/L |
|       | P1-AIT-604 | NH3           | Filtrado FPP 1 a 3          | 10-150mg/L         |
|       | P1-AIT-605 | Condutividade | Filtrado FPP 1 a 3          | 0-10K $\mu$ S/cm   |

Fonte: Do autor (2025)

A Figura 33 nos traz uma síntese desses dados em um gráfico de linha para cada uma das análises realizadas na planta para o mês de Julho/2024.

Figura 33 - Dados ETE A



Fonte: Do autor (2025)

A análise prévia desses gráficos já nos permite alguns filtros de possíveis inconsistências em alguma sonda, como por exemplo o dado de ORP da sonda AIT-413 variando em média em 6,8mV durante todo o mês, enquanto as sondas dos outros tanques oscilam entre 200 e 300mV. Na Tabela 6 é possível observar a caracterização desses dados, com valores mínimos, máximos, média, mediana e o desvio padrão dos dados registrados:

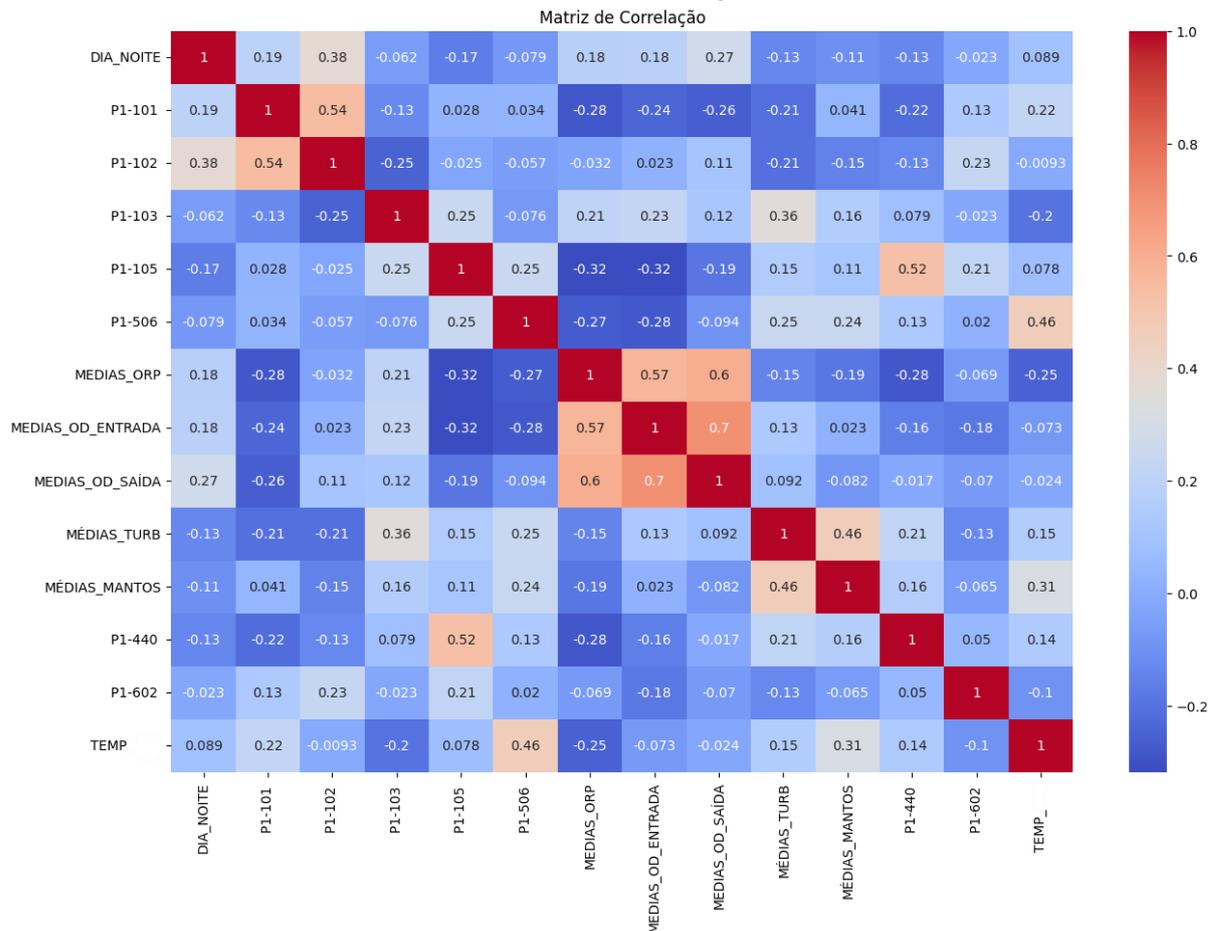
Tabela 6 - Caracterização dos dados - Julho/2024 - ETE A

|        |               | Min     | Max      | Media    | Mediana | Desv Pad |
|--------|---------------|---------|----------|----------|---------|----------|
| P1-101 | Condutividade | 0,00    | 1113,62  | 763,11   | 757,03  | 110,84   |
| P1-102 | pH            | 6,47    | 9,50     | 7,23     | 7,08    | 0,45     |
| P1-103 | SST (imersão) | 111,42  | 22647,02 | 1062,46  | 1046,86 | 550,65   |
| P1-104 | DQO           | 125,70  | 1780,20  | 566,34   | 568,50  | 210,12   |
| P1-105 | SST (imersão) | 53,33   | 23987,26 | 978,27   | 129,71  | 3161,27  |
| P1-506 | DQO           | 42,20   | 340,00   | 108,19   | 98,90   | 37,59    |
| P1-406 | SST (imersão) | 1948,08 | 5740,67  | 2794,05  | 2748,37 | 335,85   |
| P1-408 | ORP           | 23,80   | 349,99   | 169,72   | 171,43  | 47,55    |
| P1-409 | ORP           | 215,86  | 349,33   | 267,14   | 267,95  | 27,50    |
| P1-410 | ORP           | 193,56  | 395,04   | 319,14   | 326,07  | 37,16    |
| P1-411 | ORP           | -426,69 | 310,89   | -106,98  | -141,15 | 119,47   |
| P1-413 | ORP           | 6,82    | 6,85     | 6,84     | 6,84    | 0,01     |
| P1-414 | OD            | 0,14    | 8,83     | 2,77     | 1,96    | 2,25     |
| P1-415 | OD            | 0,32    | 7,98     | 1,75     | 1,14    | 1,52     |
| P1-416 | OD            | 0,33    | 8,36     | 2,40     | 1,72    | 2,00     |
| P1-417 | OD            | 0,53    | 8,60     | 5,23     | 5,38    | 1,37     |
| P1-418 | OD            | 0,67    | 9,30     | 6,69     | 6,89    | 1,38     |
| P1-419 | OD            | 0,51    | 8,67     | 4,29     | 4,33    | 1,79     |
| P1-420 | OD            | 0,42    | 8,53     | 4,00     | 3,93    | 2,03     |
| P1-421 | OD            | 0,25    | 9,65     | 6,97     | 7,15    | 1,45     |
| P1-428 | Turbidez      | 0,24    | 72,99    | 25,31    | 23,02   | 9,15     |
| P1-429 | Turbidez      | 0,67    | 498,74   | 24,78    | 20,37   | 19,69    |
| P1-430 | Turbidez      | 0,83    | 909,36   | 87,68    | 57,85   | 89,74    |
| P1-431 | Turbidez      | 0,24    | 319,09   | 21,83    | 20,88   | 9,40     |
| P1-432 | Turbidez      | 0,05    | 613,73   | 57,24    | 41,99   | 44,28    |
| P1-433 | Turbidez      | 0,21    | 1924,63  | 58,32    | 18,18   | 133,35   |
| P1-434 | Manto de Lodo | 0,00    | 3,54     | 0,85     | 0,70    | 0,57     |
| P1-435 | Manto de Lodo | 0,00    | 3,66     | 0,72     | 0,48    | 0,71     |
| P1-436 | Manto de Lodo | 0,00    | 3,76     | 0,77     | 0,60    | 0,71     |
| P1-437 | Manto de Lodo | 0,00    | 3,75     | 1,28     | 1,13    | 0,84     |
| P1-438 | Manto de Lodo | 0,00    | 3,57     | 1,62     | 1,16    | 1,31     |
| P1-440 | pH            | 1,75    | 7,25     | 6,05     | 5,97    | 0,54     |
| P1-602 | SST (imersão) | 46,69   | 53077,79 | 13441,45 | 8927,46 | 12740,46 |
| P1-605 | Condutividade | 0,00    | 986,81   | 564,34   | 547,06  | 129,40   |

Fonte: Do autor (2025)

Também foi possível construir uma matriz de correlação entre esses dados, conforme Gráfico 1:

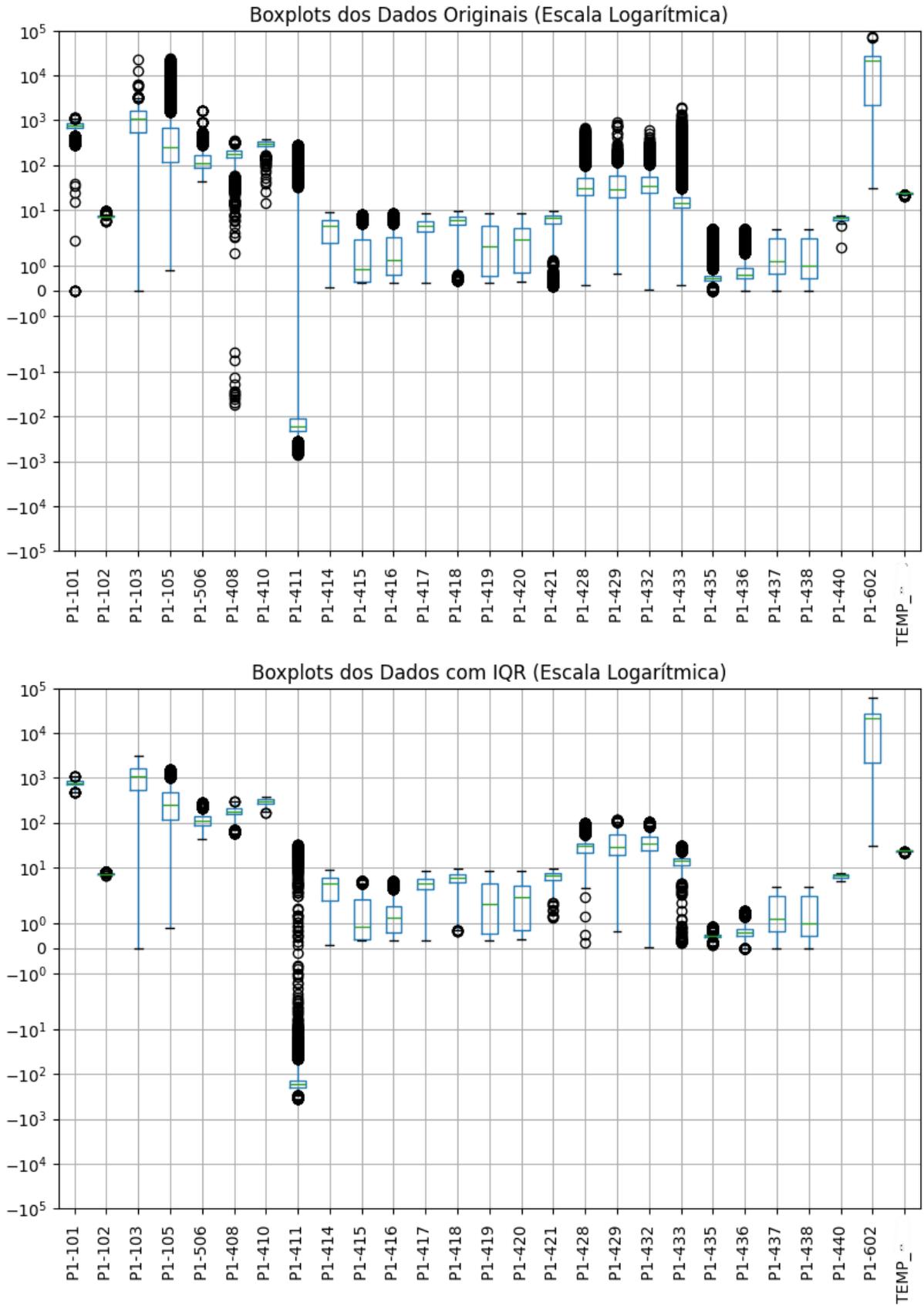
Gráfico 1 - Matriz de correlação - ETE A



Fonte: Do autor (2025)

Na área central da matriz é possível observar uma correlação forte e positiva entre as análises de OD e ORP dos tanques de aeração. Uma correlação coerente considerando que o esgoto presente nos tanques provém de mesmo local. E são esses pontos ou áreas de altas correlações, sejam positivas ou negativas, que merecem atenção e ajudarão a definir as variáveis utilizadas em modelagens mais complexas adiante. Por exemplo, analisando a DQO na saída como uma variável de interesse, é possível observar maior correlação com os dados de temperatura. Outro gráfico importante é o *box-plot* e ele pode ajudar a identificar alguns padrões e a presença de *outliers* nos dados, conforme a Figura 34:

Figura 34 - Box-plot dos dados - ETE A – Antes e depois da retirada de outliers



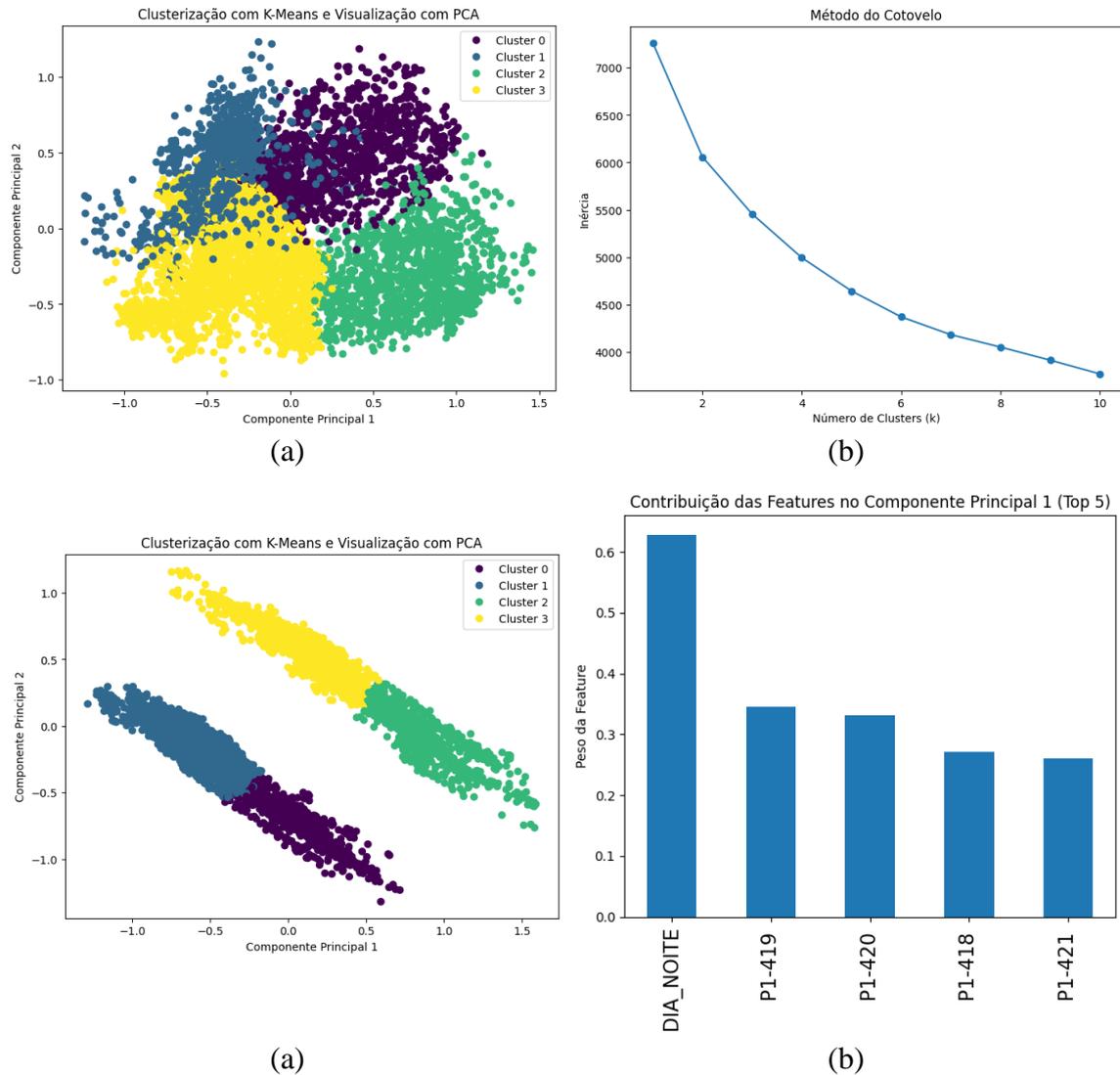
Fonte: Do autor (2025)

É possível observar que algumas análises, possuem uma distribuição de forma que o comprimento da caixa mostra a dispersão dos dados, sem valores fora, que dependendo de quão fora, poderiam ser considerados *outliers*. Essa análise sugere, porém, que outros dados talvez necessitem de um tratamento visando a retirada desses *outliers*. Conforme visto anteriormente, o método IQR que considera o intervalo entre quartis pode ser utilizado visando identificar esses valores. O gráfico a direita, traz um novo *box-plot*, agora com esses dados identificados e substituídos pela mediana da coluna.

A aplicação do algoritmo de *K-means* permite identificar grupos de dados que possuem características similares. Considerando os parâmetros analíticos do tratamento de esgoto, os *clusters* podem indicar diferentes estágios e processos dentro da ETE A, como tratamento do esgoto bruto, reatores aeróbios ou efluente final.

A PCA foi usada para reduzir a dimensionalidade dos dados, facilitando a visualização dos *clusters* formados pelo *K-means*. No gráfico de PCA, cada ponto representa uma observação de um determinado parâmetro de uma TAG específica, colorida de acordo com o *cluster* ao qual pertence. Essa análise pode destacar a variabilidade dos parâmetros em diferentes fases do processo de tratamento, conforme o Gráfico 2. O Gráfico 2 (a) traz a clusterização com os dados já citados, enquanto o Gráfico 2 (b) traz a curva da técnica do cotovelo. O Gráfico 2 (c) inclui uma variável chamada “DIA\_NOITE”, que possui valor 1 das 06:00 às 18:00 horas, e 0 das 18:00 as 06:00. E no Gráfico 2 (d) tem-se a contribuição da Componente Principal 1, com as top 5 componentes do gráfico que inclui essa variável. Uma ênfase justamente para variável “DIA\_NOITE” que traz uma separação clara na clusterização.

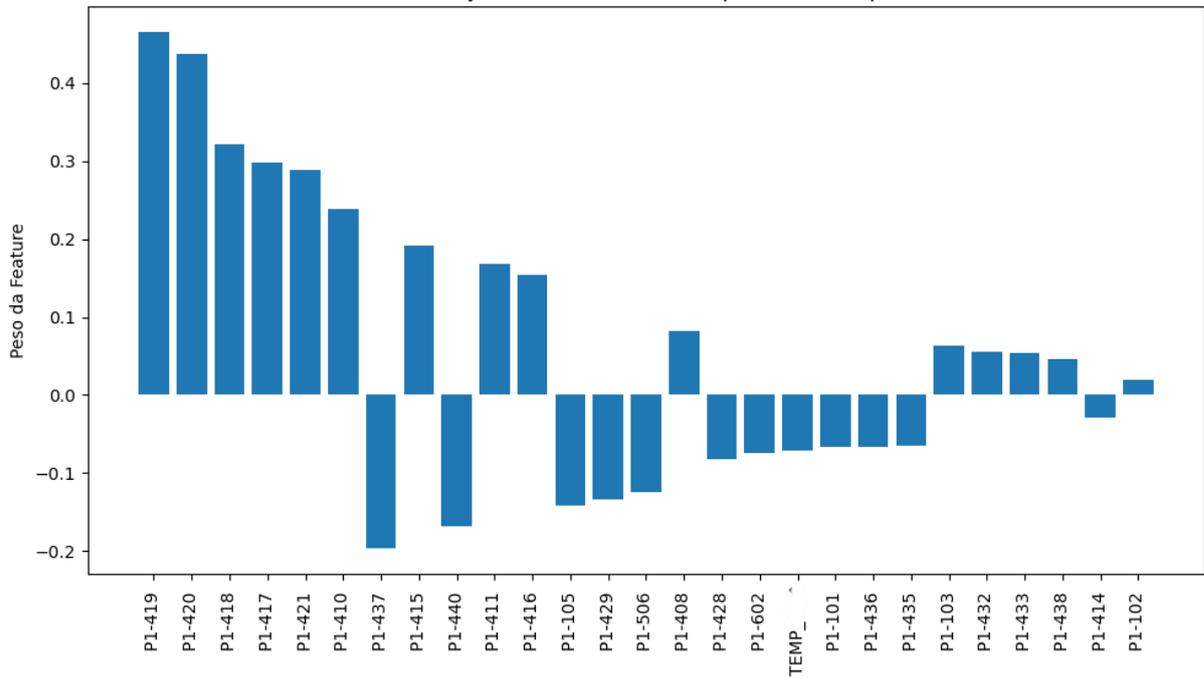
Gráfico 2 - K-means e PCA - ETE A



Fonte: Do autor (2025)

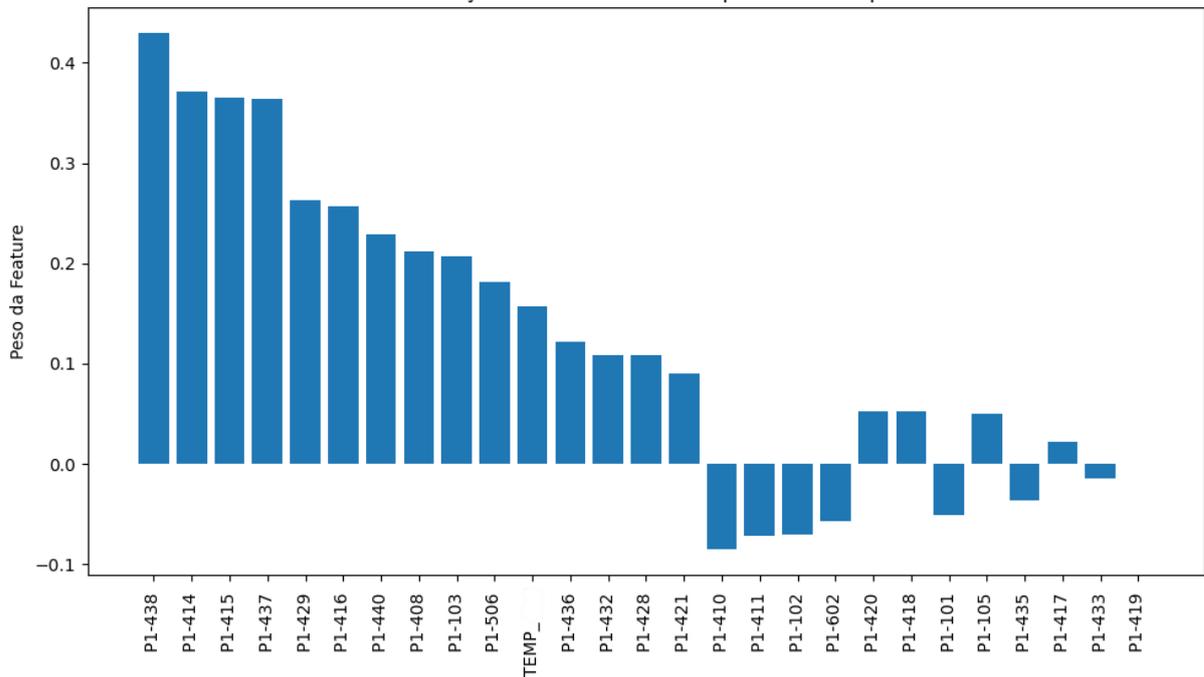
É possível observar que os ODs nos tanques de aeração têm alta contribuição na componente principal 1. Enquanto algumas medições de manto e turbidez, em conjunto com análises de OD também, tem alta influência na segunda componente principal, Gráfico 3:

Gráfico 3 - Contribuição das *features* PCA - ETE A  
Contribuição das Features no Componente Principal 1



(a)

Contribuição das Features no Componente Principal 2



(b)

Fonte: Do autor (2025)

### 5.2.2 ETE B

A ETE B beneficia aproximadamente 7,7 milhões de habitantes. Com uma vazão média de projeto de 16 mil litros por segundo e uma vazão atual de 15.588 litros por segundo (média

de 2023. Os esgotos são transportados para a estação através de um extenso sistema de esgotamento sanitário, composto por coletores-tronco, emissários, interceptores e linhas de recalque, totalizando cerca de 262 quilômetros.

A Tabela 7 relaciona os instrumentos presentes na ETE B, assim como é relacionado para ETE A, com local de instalação e faixa de trabalho. A *target* também é a DQO do efluente na ETE, aqui identificada pela TAG “P2-AIT-106”.

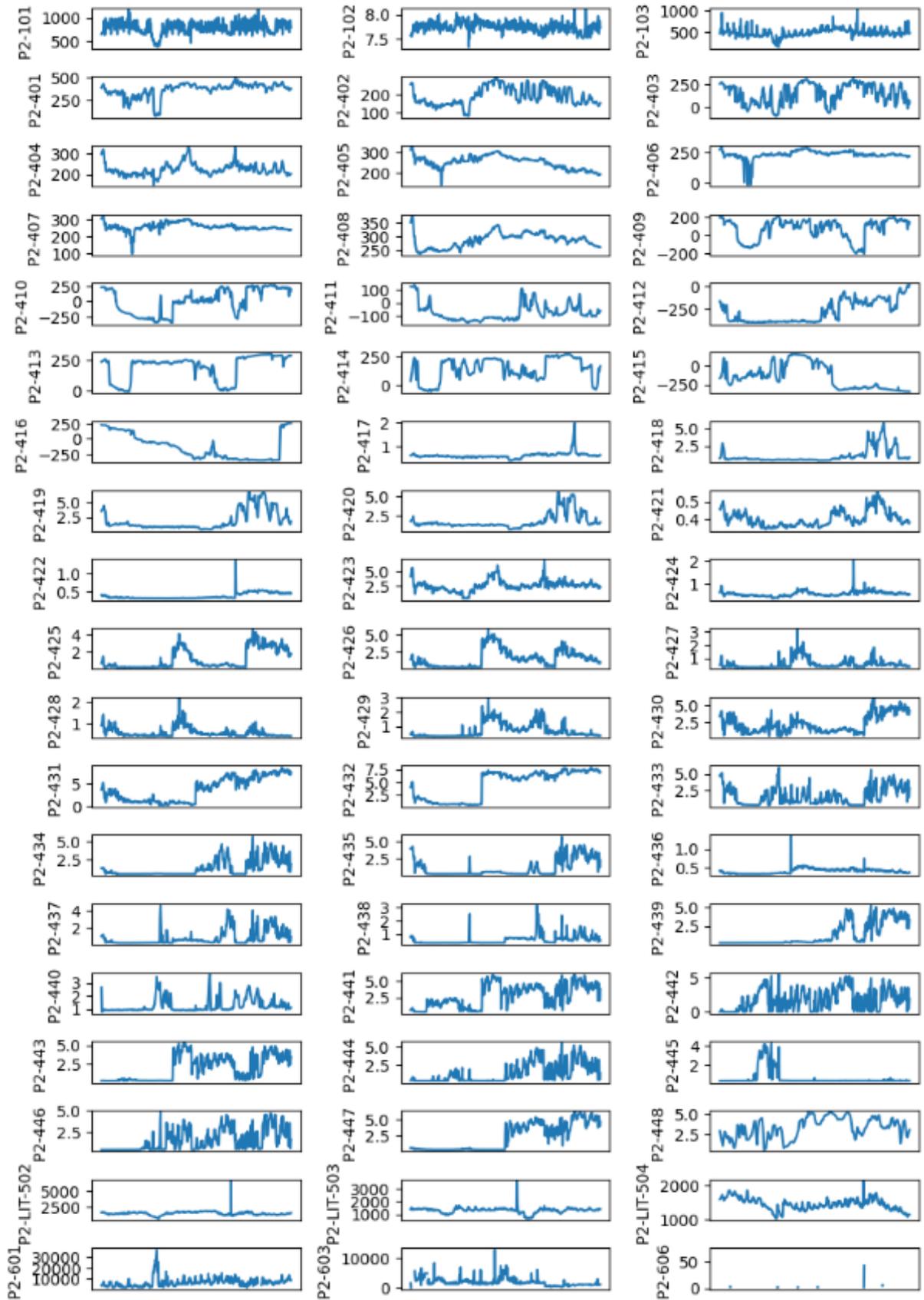
Tabela 7 - Relação de instrumentos por planta - ETE B

|            | TAG              | Parâmetro          | Local                 | Faixa de Trabalho    |
|------------|------------------|--------------------|-----------------------|----------------------|
| ETE B      | P2-AIT-101       | pH                 | Esgoto bruto          | 2 a 12               |
|            | P2-AIT-102       | Condutividade      | Esgoto bruto          | 0-10K $\mu$ S/cm     |
|            | P2-AIT-103       | SST (imersão)      | Esgoto desarenado     | 100 a 500 mg/L       |
|            | P2-AIT-304       | DQO                | Esgoto desarenado     | 200 a 1.000 mg/L     |
|            | P2-AIT-305       | SST (imersão)      | Efluente Primário     | 20 a 200 mg/L        |
|            | P2-AIT-106       | DQO                | Efluente Final        | 10 a 600 mg/L        |
|            | P2-AIT-501       | SST (imersão)      | Licor misto           | 1.000 a 8.000 mg/L   |
|            | P2-AIT-502       | SST (imersão)      | Licor misto           | 1.000 a 8.000 mg/L   |
|            | P2-AIT-503       | SST (imersão)      | Licor misto           | 1.000 a 8.000 mg/L   |
|            | P2-AIT-504       | SST (imersão)      | Licor misto           | 1.000 a 8.000 mg/L   |
|            | P2-AIT-401 a 416 | ORP                | Reator aeróbio 1 a 16 | (-175 a -400 mV)     |
|            | P2-AIT-417 a 448 | OD                 | Reator aeróbio 1 a 16 | 0 a 8 PPM            |
|            | P2-AIT-309       | NH <sub>3</sub>    | Efluente Primário     | 10-150mg/L           |
|            | P2-AIT-450       | NH <sub>3</sub>    | Licor misto           | 0,05-20mg/L          |
|            | P2-AIT-525       | Turbidez           | Efluente Final        | 0-100 NTU            |
|            | P2-LIT-501 a 524 | Manto de Lodo      | Decantador DS1 a DS24 | 0-5 metros           |
|            | P2-AIT-601       | SST (imersão)      | Lodo aflu Adens.      | 20.000 a 50.000 mg/L |
|            | P2-AIT-602       | SST (imersão)      | Lodo aflu C. Adens    | 20.000 a 50.000 mg/L |
|            | P2-AIT-603       | SST (imersão)      | Clarificado Centr.    | 1.000 a 5.000 mg/L   |
|            | P2-AIT-604       | NH <sub>3</sub>    | Clarificado Centr.    | 10-150mg/L           |
| P2-AIT-605 | Condutividade    | Clarificado Centr. | 0-10K $\mu$ S/cm      |                      |
| REÚSO      | P2-AIT-606       | Turbidez           | ETA Reuso             | 0-10 NTU;            |
|            | P2-AIT-607       | Cloro              | ETA Reuso             | 0 a 20 mg/l          |

Fonte: Do autor (2025)

Novamente, um gráfico de cada uma dessas variáveis é plotado e pode ser visto na Figura 35, seguido pela Tabela 8 com a caracterização dos dados:

Figura 35 - Dados ETE B



Fonte: Do autor (2025)

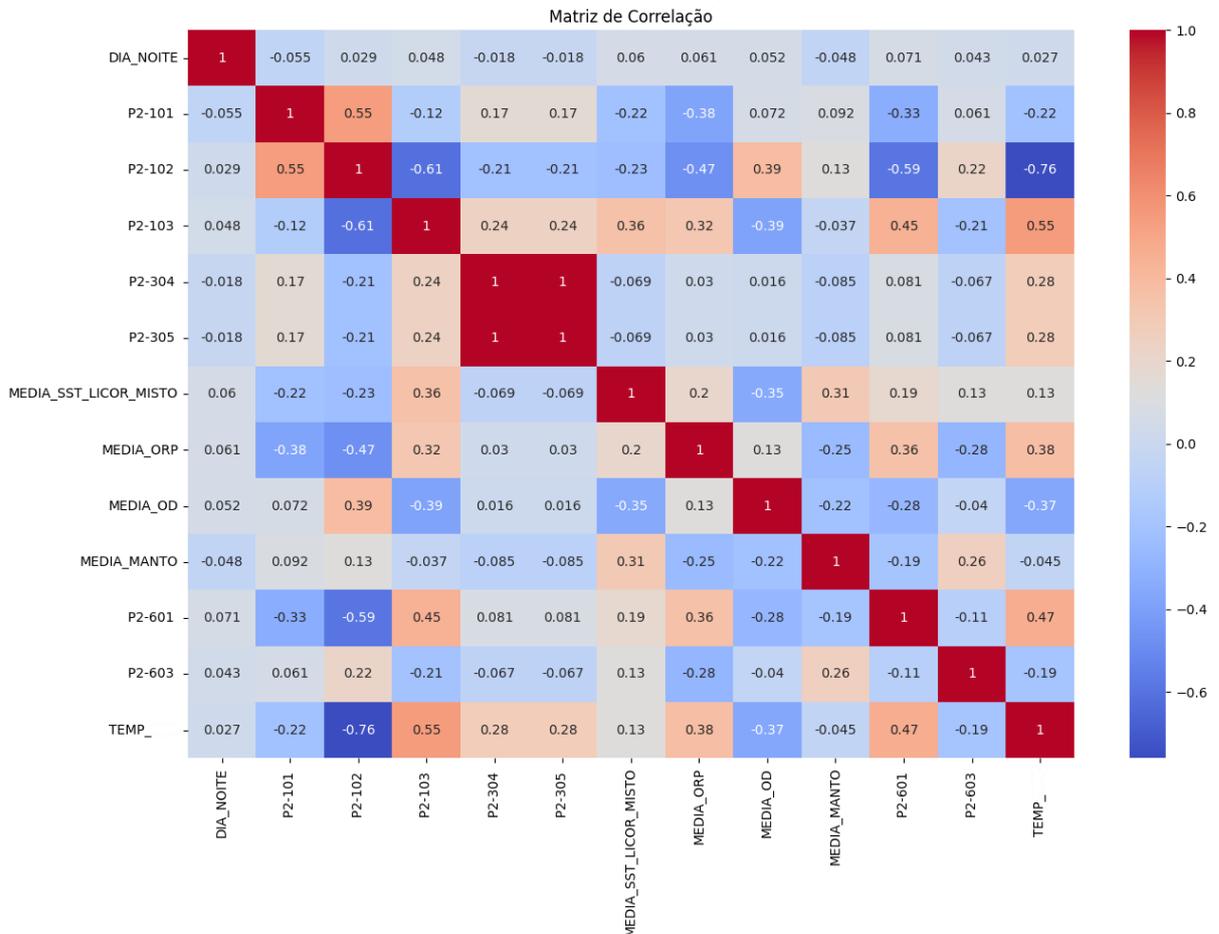
Tabela 8 - Caracterização dos dados - Julho/2024 - ETE B

|        |               | Min     | Max      | Media   | Mediana | Desv Pad |
|--------|---------------|---------|----------|---------|---------|----------|
| P2-101 | pH            | 0,00    | 1405,94  | 777,84  | 776,31  | 136,29   |
| P2-102 | Condutividade | 7,31    | 8,53     | 7,74    | 7,73    | 0,11     |
| P2-103 | SST (imersão) | 147,54  | 1225,55  | 484,02  | 475,87  | 94,15    |
| P2-401 | ORP           | 69,72   | 495,08   | 368,20  | 389,30  | 77,16    |
| P2-402 | ORP           | 55,30   | 287,18   | 191,68  | 180,21  | 48,64    |
| P2-403 | ORP           | -116,51 | 313,36   | 151,41  | 185,31  | 108,59   |
| P2-404 | ORP           | 71,76   | 333,73   | 227,82  | 220,60  | 30,28    |
| P2-405 | ORP           | 126,35  | 329,25   | 253,69  | 257,84  | 30,43    |
| P2-406 | ORP           | -23,13  | 298,77   | 229,93  | 230,42  | 37,40    |
| P2-407 | ORP           | 77,24   | 328,96   | 259,23  | 255,55  | 24,85    |
| P2-408 | ORP           | 236,12  | 376,87   | 289,87  | 293,73  | 29,22    |
| P2-409 | ORP           | -204,36 | 211,27   | 70,87   | 118,89  | 115,07   |
| P2-410 | ORP           | -350,74 | 264,23   | 1,48    | 9,37    | 201,90   |
| P2-411 | ORP           | -183,32 | 138,57   | -69,76  | -93,09  | 71,36    |
| P2-412 | ORP           | -397,77 | 12,41    | -275,90 | -335,45 | 118,85   |
| P2-413 | ORP           | -10,42  | 302,81   | 188,61  | 235,02  | 104,60   |
| P2-414 | ORP           | -58,34  | 273,32   | 133,55  | 145,59  | 101,54   |
| P2-415 | ORP           | -346,99 | 163,00   | -147,81 | -167,60 | 162,79   |
| P2-416 | ORP           | -352,86 | 256,22   | -111,83 | -129,88 | 213,72   |
| P2-417 | OD            | 0,36    | 2,15     | 0,60    | 0,58    | 0,14     |
| P2-418 | OD            | 0,38    | 5,87     | 1,03    | 0,64    | 0,99     |
| P2-419 | OD            | 0,40    | 6,82     | 1,98    | 1,20    | 1,61     |
| P2-420 | OD            | 0,64    | 5,73     | 1,70    | 1,34    | 0,98     |
| P2-421 | OD            | 0,33    | 0,61     | 0,40    | 0,39    | 0,05     |
| P2-422 | OD            | 0,32    | 5,69     | 0,38    | 0,35    | 0,13     |
| P2-423 | OD            | 0,50    | 6,97     | 2,68    | 2,60    | 0,95     |
| P2-424 | OD            | 0,34    | 3,19     | 0,55    | 0,53    | 0,13     |
| P2-425 | OD            | 0,30    | 4,77     | 1,19    | 0,53    | 1,14     |
| P2-426 | OD            | 0,37    | 5,72     | 1,62    | 1,40    | 1,22     |
| P2-427 | OD            | 0,31    | 3,37     | 0,57    | 0,42    | 0,38     |
| P2-428 | OD            | 0,37    | 2,45     | 0,57    | 0,46    | 0,25     |
| P2-429 | OD            | 0,29    | 3,00     | 0,70    | 0,46    | 0,50     |
| P2-430 | OD            | 0,39    | 7,37     | 2,19    | 1,73    | 1,27     |
| P2-431 | OD            | 0,30    | 8,51     | 3,52    | 3,43    | 2,50     |
| P2-432 | OD            | 0,39    | 8,90     | 4,38    | 5,85    | 2,79     |
| P2-433 | OD            | 0,36    | 5,99     | 1,74    | 1,25    | 1,36     |
| P2-434 | OD            | 0,36    | 5,84     | 1,15    | 0,43    | 1,16     |
| P2-435 | OD            | 0,30    | 5,77     | 1,18    | 0,43    | 1,29     |
| P2-436 | OD            | 0,32    | 5,69     | 0,40    | 0,40    | 0,10     |
| P2-437 | OD            | 0,30    | 4,90     | 0,86    | 0,50    | 0,85     |
| P2-438 | OD            | 0,31    | 4,93     | 0,54    | 0,35    | 0,36     |
| P2-439 | OD            | 0,29    | 5,21     | 1,27    | 0,46    | 1,43     |
| P2-440 | OD            | 0,00    | 5,17     | 1,25    | 0,97    | 0,57     |
| P2-441 | OD            | 0,36    | 6,05     | 2,70    | 2,52    | 1,68     |
| P2-442 | OD            | 0,00    | 5,72     | 1,75    | 1,51    | 1,52     |
| P2-443 | OD            | 0,30    | 5,71     | 1,86    | 1,09    | 1,61     |
| P2-444 | OD            | 0,36    | 5,64     | 1,39    | 0,70    | 1,32     |
| P2-445 | OD            | 0,31    | 4,50     | 0,60    | 0,32    | 0,81     |
| P2-446 | OD            | 0,35    | 5,51     | 1,43    | 0,81    | 1,25     |
| P2-447 | OD            | 0,31    | 6,23     | 2,15    | 0,57    | 2,04     |
| P2-448 | OD            | 0,52    | 5,25     | 3,23    | 3,33    | 1,30     |
| P2-502 | SST (imersão) | 752,73  | 13240,29 | 1573,41 | 1572,96 | 341,73   |
| P2-503 | SST (imersão) | 539,77  | 4784,68  | 1307,12 | 1343,40 | 225,91   |
| P2-504 | SST (imersão) | 986,24  | 4167,59  | 1462,52 | 1457,33 | 168,90   |
| P2-601 | SST (imersão) | 1563,03 | 50500,00 | 6511,78 | 5874,78 | 4555,33  |
| P2-603 | SST (imersão) | 0,00    | 31940,60 | 1855,03 | 1494,93 | 1550,90  |
| P2-606 | Turbidez      | 0,84    | 70,19    | 4,18    | 1,88    | 7,75     |

Fonte: Do autor (2025)

A matriz de correlação indica forte correlação positiva entre as análises de condutividade com SST e também a temperatura, conforme o Gráfico 4. Já o dado de DQO efluente da Estação não estava disponível no período e não foi possível observar as maiores influências em relação a esse parâmetro.

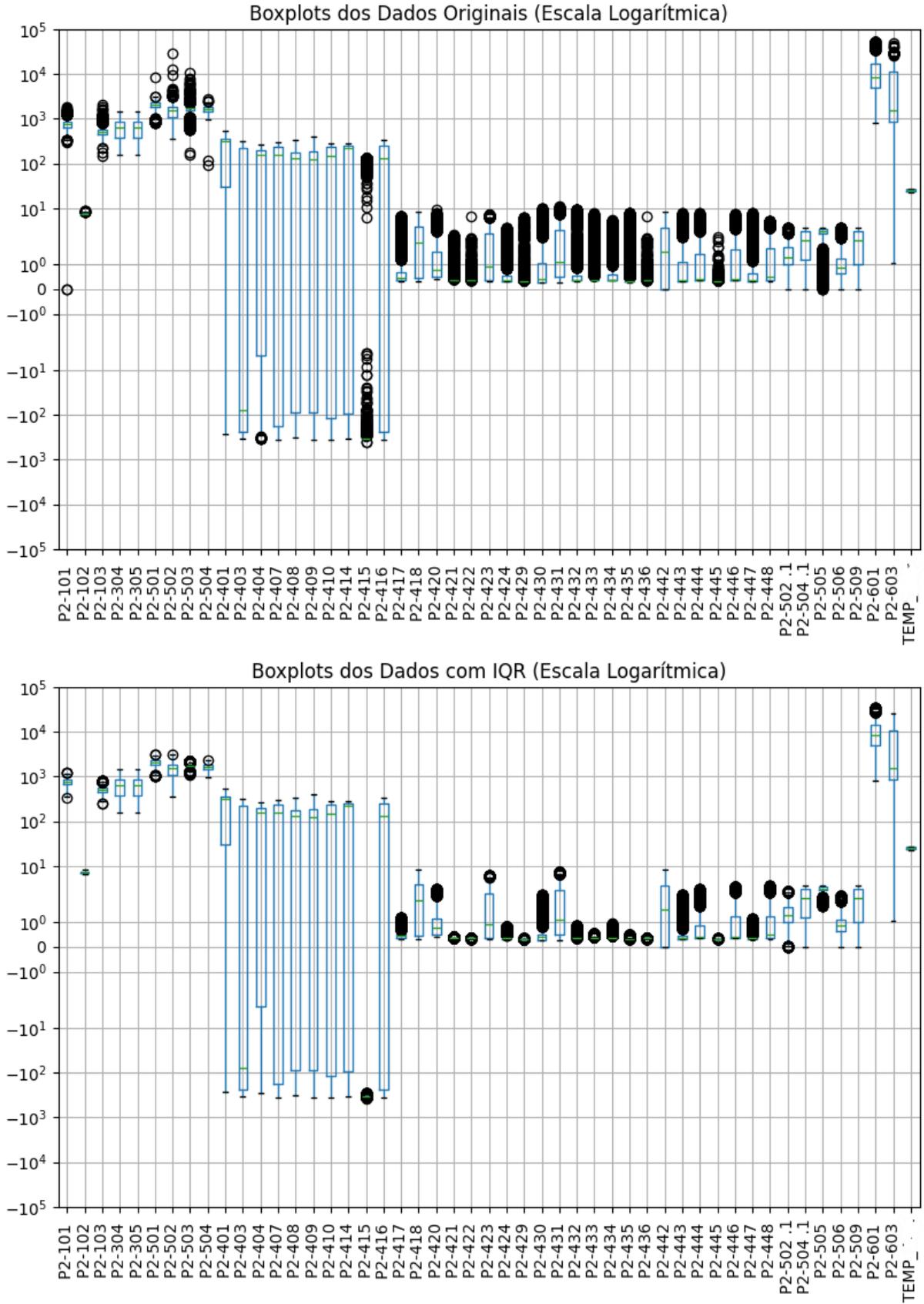
Gráfico 4 - Matriz de correlação - ETE B



Fonte: Do autor (2025)

Também foi possível observar a presença de *outliers* nos dados pertencentes a ETE B. A Figura 36 , traz o gráfico *box-plot*, antes e depois do procedimento de tratamento desses dados que também foi utilizado nos dados da ETE A e serão utilizados nas outras Estações. É possível observar uma redução de dados nas áreas dos “bigodes” (*whiskers*) dos *box-plot*, e quando isso não ocorre, há a diminuição dos valores presentes nessas áreas:

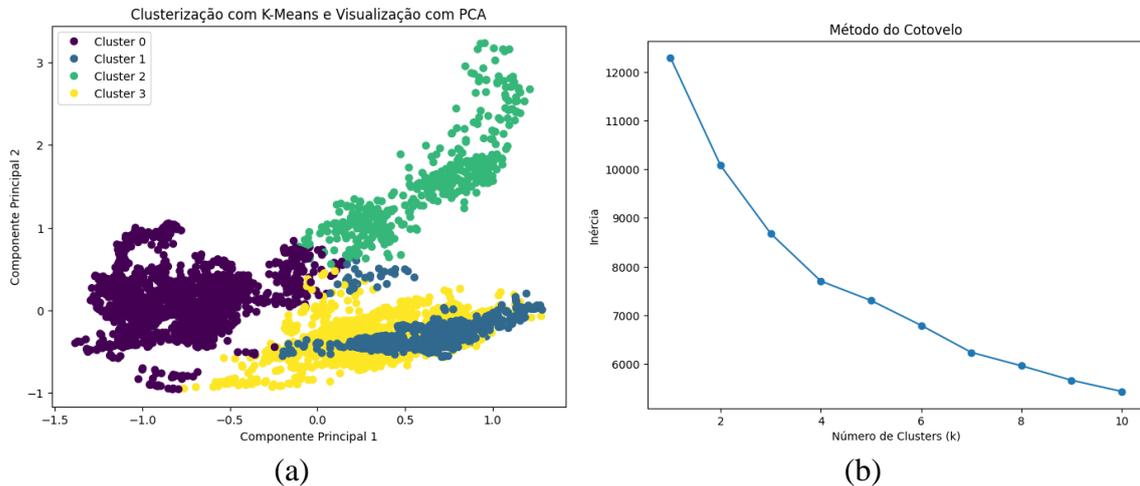
Figura 36 - Box-plot dos dados - ETE B - Antes e depois da retirada de outliers



Fonte: Do autor (2025)

Novamente, a clusterização *K-means* e a visualização com PCA foi utilizada nessa ETE, Gráfico 5 (a), seguido do Gráfico 5 (b) utilizado para determinação da quantidade de *clusters* ideias a partir do método do cotovelo, e do gráfico com a contribuição da *features* nos Componentes Principais, Gráfico 6 (a) e (b):

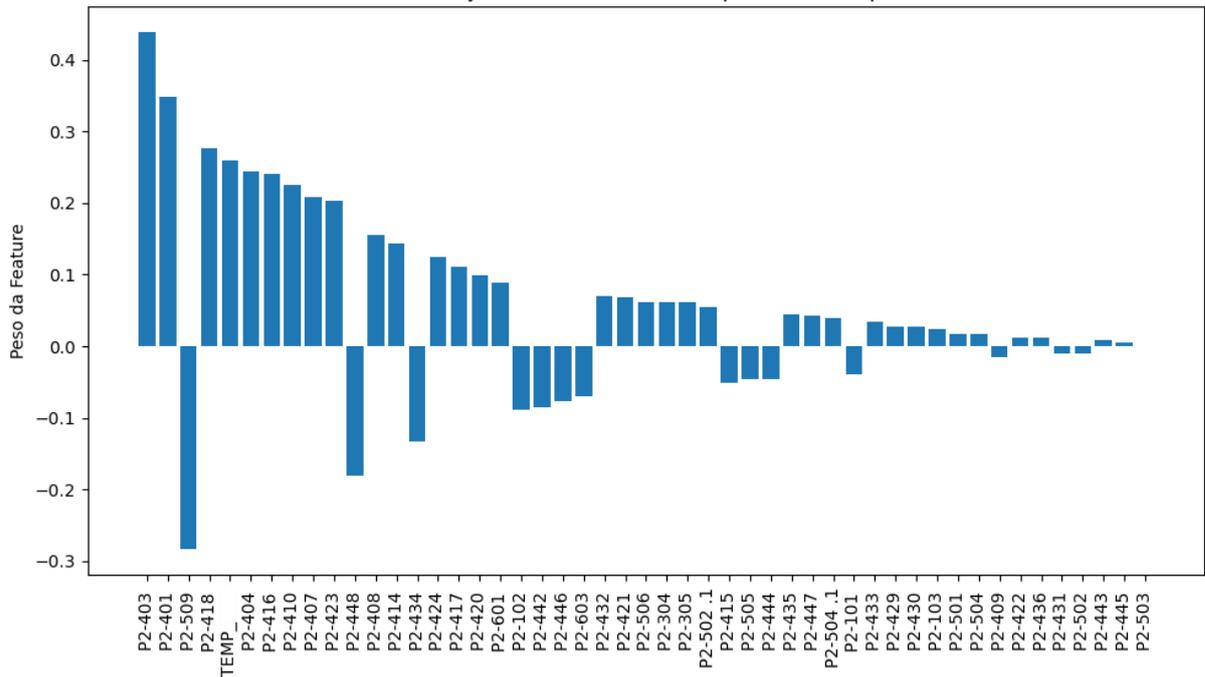
Gráfico 5 - *K-means* e PCA - ETE B



Fonte: Do autor (2025)

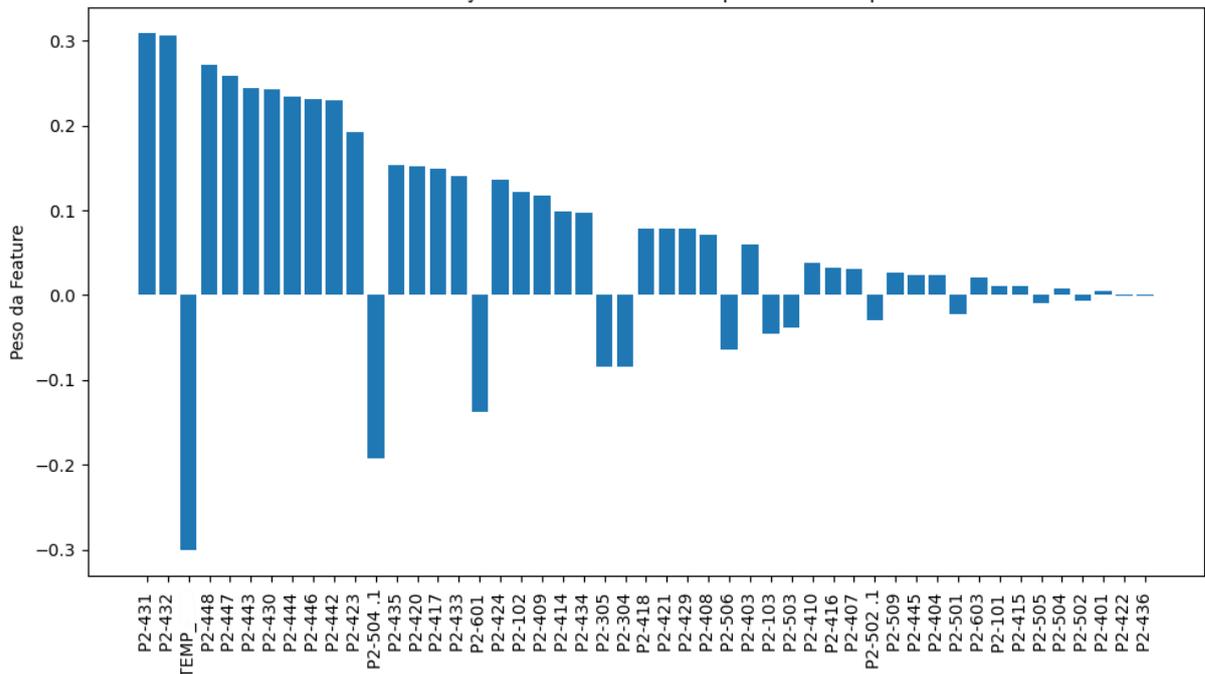
Com relação a inclusão da *feature* "DIA\_NOITE", ficou evidenciado que não é preponderante a participação nas Componentes Principais, como ocorre em outras Estações. A investigação do motivo disso pode ser realizada em estudos futuros, como não é objetivo dessa pesquisa. Contudo, a característica do esgoto nessa região específica, ou o porte da estação, a maior das ETEs avaliadas, pode contribuir para esse resultado.

Gráfico 6 - Contribuição das *features* PCA - ETE B  
 Contribuição das Features no Componente Principal 1



(a)

Contribuição das Features no Componente Principal 2



(b)

Fonte: Do autor (2025)

Para a PCA da ETE B ficou evidenciado a forte contribuição das análises de ORP de TAG “AIT-409” a “AIT-416” na componente principal N°2, enquanto a componente principal N°1 tem influência de diversas análises, mas sobretudo das análises de OD. Porém a análise dessa Estação fica limitada devido ausência de outras análises que existem nas outras plantas, como amônia e DQO.

### 5.2.3 ETE C

A ETE C foi projetada para atender uma população equivalente a 1,2 milhão de habitantes. Com uma capacidade instalada de tratamento de 3.000 litros por segundo e uma vazão atual de 1.951 litros por segundo (média de 2023). O esgoto chega à estação por um sistema de 67 quilômetros de coletores-tronco e interceptores, sendo tratado biologicamente pelo método de lodos ativados, assim como as ETEs A e B.

A Tabela 9 relaciona os instrumentos para ETE C, com os locais de instalação e faixa de trabalho:

Tabela 9 - Relação de instrumentos por planta - ETE C

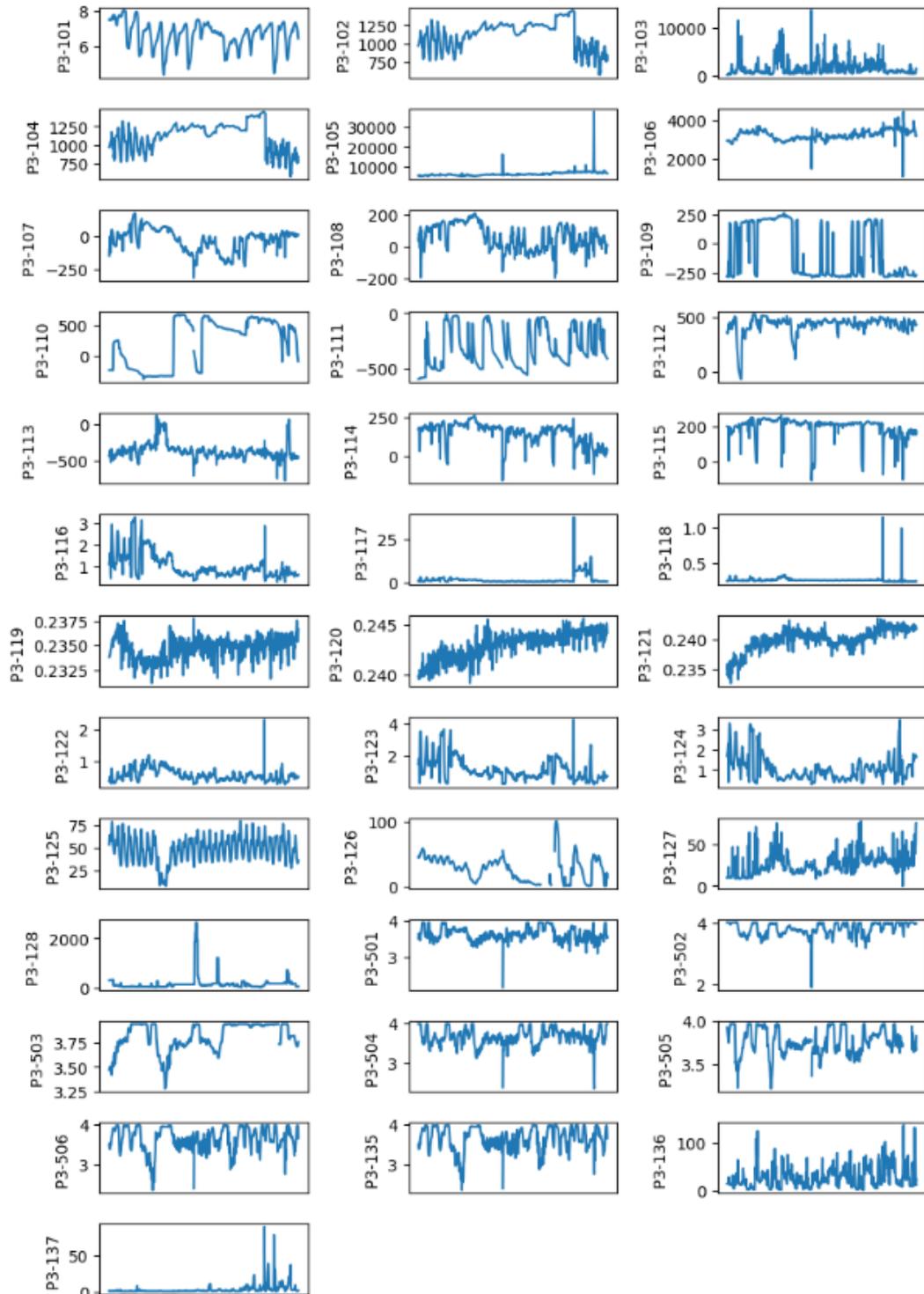
|            | TAG           | Parâmetro      | Local              | Faixa de Trabalho  |
|------------|---------------|----------------|--------------------|--------------------|
| ETE C      | P3-AIT-101    | pH             | Esgoto bruto       | 2 a 12             |
|            | P3-AIT-102    | Condutividade  | Esgoto bruto       | 0-10K $\mu$ S/cm   |
|            | P3-AIT-103    | SST (imersão)  | Esgoto desarenado  | 100 a 500 mg/L     |
|            | P3-AIT-104    | DQO            | Esgoto desarenado  | 200 a 1.000 mg/L   |
|            | P3-AIT-105    | SST (imersão)  | Licor misto        | 1.000 a 8.000 mg/L |
|            | P3-AIT-106    | SST (imersão)  | Licor misto        | 1.000 a 8.000 mg/L |
|            | P3-AIT-107    | ORP            | Reator aeróbio 1   | (-175 a -400 mV)   |
|            | P3-AIT-108    | ORP            | Reator aeróbio 1   | (-175 a -400 mV)   |
|            | P3-AIT-109    | ORP            | Reator aeróbio 1   | (-175 a -400 mV)   |
|            | P3-AIT-110    | ORP            | Reator aeróbio 2   | (-175 a -400 mV)   |
|            | P3-AIT-111    | ORP            | Reator aeróbio 2   | (-175 a -400 mV)   |
|            | P3-AIT-112    | ORP            | Reator aeróbio 2   | (-175 a -400 mV)   |
|            | P3-AIT-113    | ORP            | Reator aeróbio 3   | (-175 a -400 mV)   |
|            | P3-AIT-114    | ORP            | Reator aeróbio 3   | (-175 a -400 mV)   |
|            | P3-AIT-115    | ORP            | Reator aeróbio 3   | (-175 a -400 mV)   |
|            | P3-AIT-116    | OD             | Reator aeróbio 1   | 0 a 8 PPM          |
|            | P3-AIT-117    | OD             | Reator aeróbio 1   | 0 a 8 PPM          |
|            | P3-AIT-118    | OD             | Reator aeróbio 1   | 0 a 8 PPM          |
|            | P3-AIT-119    | OD             | Reator aeróbio 2   | 0 a 8 PPM          |
|            | P3-AIT-120    | OD             | Reator aeróbio 2   | 0 a 8 PPM          |
|            | P3-AIT-121    | OD             | Reator aeróbio 2   | 0 a 8 PPM          |
|            | P3-AIT-122    | OD             | Reator aeróbio 3   | 0 a 8 PPM          |
|            | P3-AIT-123    | OD             | Reator aeróbio 3   | 0 a 8 PPM          |
|            | P3-AIT-124    | OD             | Reator aeróbio 3   | 0 a 8 PPM          |
|            | P3-AIT-125    | NH3            | Efluente Primário  | 10-150mg/L         |
|            | P3-AIT-126    | NH3            | Licor misto        | 0,05-20mg/L        |
|            | P3-AIT-127    | Turbidez       | Efluente DS1 a DS6 | 0-100 NTU          |
|            | P3-AIT-128    | DQO            | Efluente Tratado   | 10 a 100 mg/L      |
|            | P3-LIT-501    | Manto de Lodo  | Decantador DS1     | 0-6metros          |
|            | P3-LIT-502    | Manto de Lodo  | Decantador DS2     | 0-6metros          |
|            | P3-LIT-503    | Manto de Lodo  | Decantador DS3     | 0-6metros          |
|            | P3-LIT-504    | Manto de Lodo  | Decantador DS4     | 0-6metros          |
|            | P3-LIT-505    | Manto de Lodo  | Decantador DS5     | 0-6metros          |
| P3-LIT-506 | Manto de Lodo | Decantador DS6 | 0-6metros          |                    |
| P3-AIT-135 | Cloro         | ETA Reuso      | 0 a 20 mg/l        |                    |
| P3-AIT-136 | Turbidez      | ETA Reuso      | 0-10 NTU           |                    |
| P3-AIT-137 | Turbidez      | ETA Reuso      | 0-10 NTU           |                    |

Fonte: Do autor (2025)

Conforme a Figura 37, fica evidenciado a presença de *outliers* também na série de dados

dessa planta, conforme TAG “P3-117” e “P3-118” e a necessidade do tratamento, conforme será visto posteriormente:

Figura 37 - Dados ETE C



Fonte: Do autor (2025)

A Tabela 10 traz, assim como na outras plantas, a caracterização dos dados:

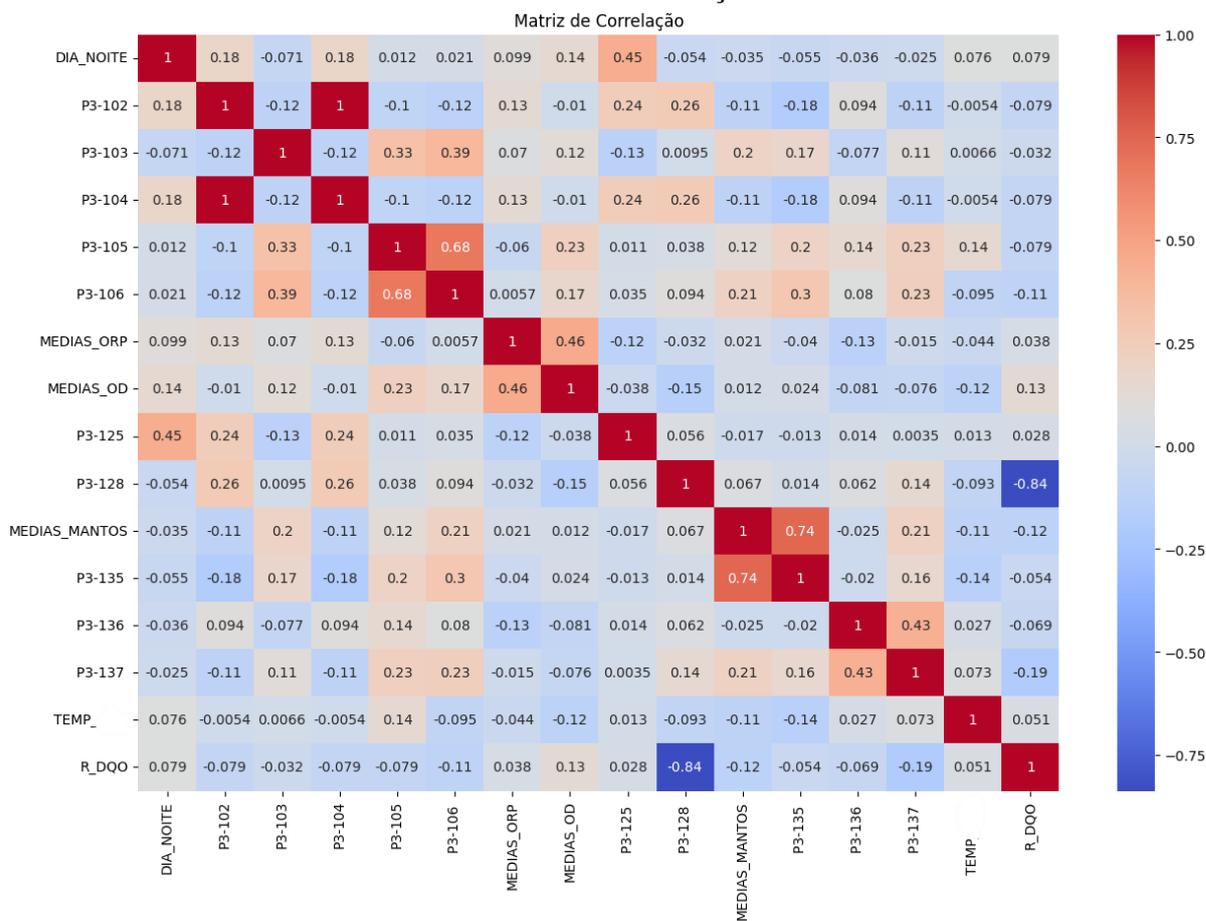
Tabela 10 - Caracterização dos dados - Julho/2024 - ETE C

|        |               | Min     | Max      | Media   | Mediana | Desv Pad |
|--------|---------------|---------|----------|---------|---------|----------|
| P3-101 | pH            | 4,18    | 8,18     | 6,73    | 6,87    | 0,78     |
| P3-102 | Condutividade | 3,24    | 1711,75  | 1126,91 | 1169,43 | 210,79   |
| P3-103 | SST (imersão) | 4,29    | 50500,00 | 1615,55 | 814,24  | 2492,54  |
| P3-104 | DQO           | 3,24    | 1711,75  | 1126,91 | 1169,43 | 210,79   |
| P3-105 | SST (imersão) | 2052,12 | 42919,61 | 6565,95 | 6303,98 | 1613,59  |
| P3-106 | SST (imersão) | 601,79  | 5545,94  | 3209,89 | 3190,79 | 278,59   |
| P3-107 | ORP           | -329,51 | 182,47   | -27,36  | -10,37  | 86,67    |
| P3-108 | ORP           | -380,65 | 480,75   | 62,80   | 58,29   | 79,96    |
| P3-109 | ORP           | -333,56 | 270,58   | -75,33  | -245,04 | 229,13   |
| P3-110 | ORP           | -381,58 | 669,33   | 212,93  | 399,61  | 380,52   |
| P3-111 | ORP           | -601,05 | 62,13    | -315,64 | -369,63 | 178,72   |
| P3-112 | ORP           | -94,70  | 558,51   | 431,06  | 453,47  | 86,07    |
| P3-113 | ORP           | -789,99 | 130,15   | -367,45 | -381,80 | 117,14   |
| P3-114 | ORP           | -189,18 | 482,87   | 153,21  | 179,64  | 70,62    |
| P3-115 | ORP           | -169,40 | 277,52   | 192,82  | 213,61  | 62,38    |
| P3-116 | OD            | 0,28    | 8,30     | 1,05    | 0,84    | 0,63     |
| P3-117 | OD            | 0,26    | 108,64   | 1,57    | 0,72    | 2,92     |
| P3-118 | OD            | 0,24    | 3,26     | 0,26    | 0,26    | 0,08     |
| P3-119 | OD            | 0,23    | 0,24     | 0,23    | 0,23    | 0,00     |
| P3-120 | OD            | 0,24    | 0,25     | 0,24    | 0,24    | 0,00     |
| P3-121 | OD            | 0,23    | 0,24     | 0,24    | 0,24    | 0,00     |
| P3-122 | OD            | 0,28    | 7,57     | 0,58    | 0,53    | 0,23     |
| P3-123 | OD            | 0,26    | 9,92     | 1,21    | 1,00    | 0,75     |
| P3-124 | OD            | 0,27    | 4,78     | 1,06    | 0,89    | 0,67     |
| P3-125 | NH3           | 7,63    | 80,21    | 47,89   | 48,71   | 14,28    |
| P3-126 | NH3           | 1,00    | 100,28   | 28,39   | 30,50   | 18,90    |
| P3-127 | Turbidez      | 0,11    | 100,31   | 25,83   | 23,11   | 13,69    |
| P3-128 | DQO           | 31,30   | 2636,90  | 177,00  | 125,60  | 292,33   |
| P3-501 | Manto de Lodo | 1,40    | 3,95     | 3,62    | 3,70    | 0,28     |
| P3-502 | Manto de Lodo | 1,91    | 4,01     | 3,80    | 3,84    | 0,21     |
| P3-503 | Manto de Lodo | 2,75    | 3,95     | 3,81    | 3,84    | 0,18     |
| P3-504 | Manto de Lodo | 1,55    | 3,99     | 3,66    | 3,72    | 0,27     |
| P3-505 | Manto de Lodo | 2,57    | 3,97     | 3,77    | 3,78    | 0,16     |
| P3-506 | Manto de Lodo | 1,58    | 3,98     | 3,62    | 3,73    | 0,35     |
| P3-135 | Cloro         | 1,58    | 3,98     | 3,62    | 3,73    | 0,35     |
| P3-136 | Turbidez      | 0,88    | 342,50   | 27,34   | 20,85   | 24,92    |
| P3-137 | Turbidez      | 0,08    | 309,82   | 2,60    | 0,64    | 7,81     |

Fonte: Do autor (2025)

Novamente as maiores correlações podem ser observadas entre os dados de OD e ORP presentes nos tanques de aeração. Porém outras relações também podem ser observadas, como por exemplo, as análises de SST que ocorrem no licor misto, o que se mostra coerente, levando em consideração que se trata do esgoto proveniente do mesmo local, conforme Gráfico 7. Outra correlação positiva observada foi entre os mantos nos decantadores secundários e a análise de cloro, o que pode sugerir que quando o manto está maior e pode existir perda de sólidos nessa etapa, existe um reforço normal e necessário na dosagem do cloro, seja pela equipe operacional ou de maneira automática pelos controladores. Já a DQO, as maiores correlações observadas foram com essas mesmas análises de SST no licor misto.

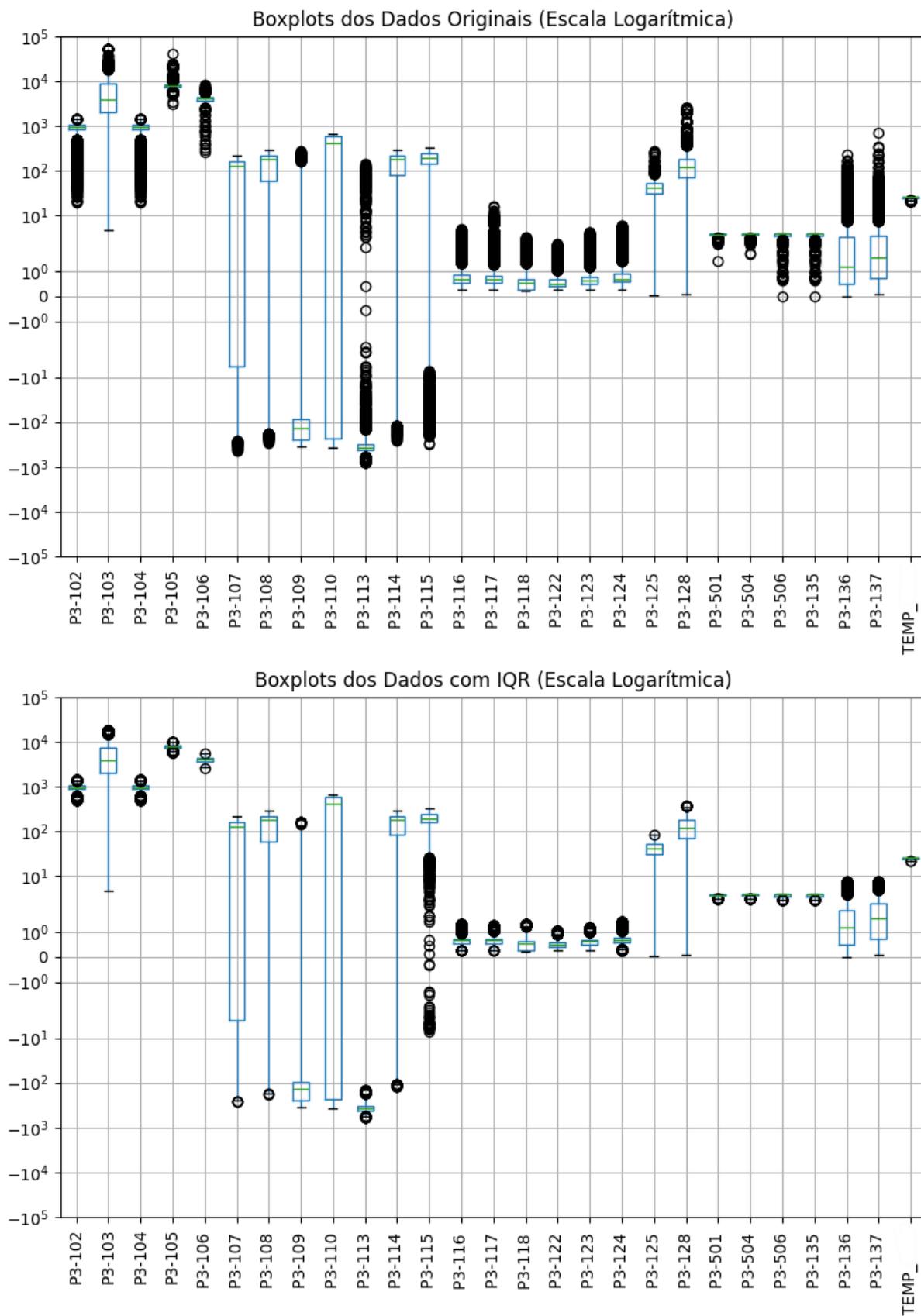
Gráfico 7 - Matriz de correlação - ETE C



Fonte: Do autor (2025)

A Figura 38 nos mostra os gráficos *box-plot*, antes e depois do processo de remoção de *outliers*, evidenciando a redução da presença dos mesmos.

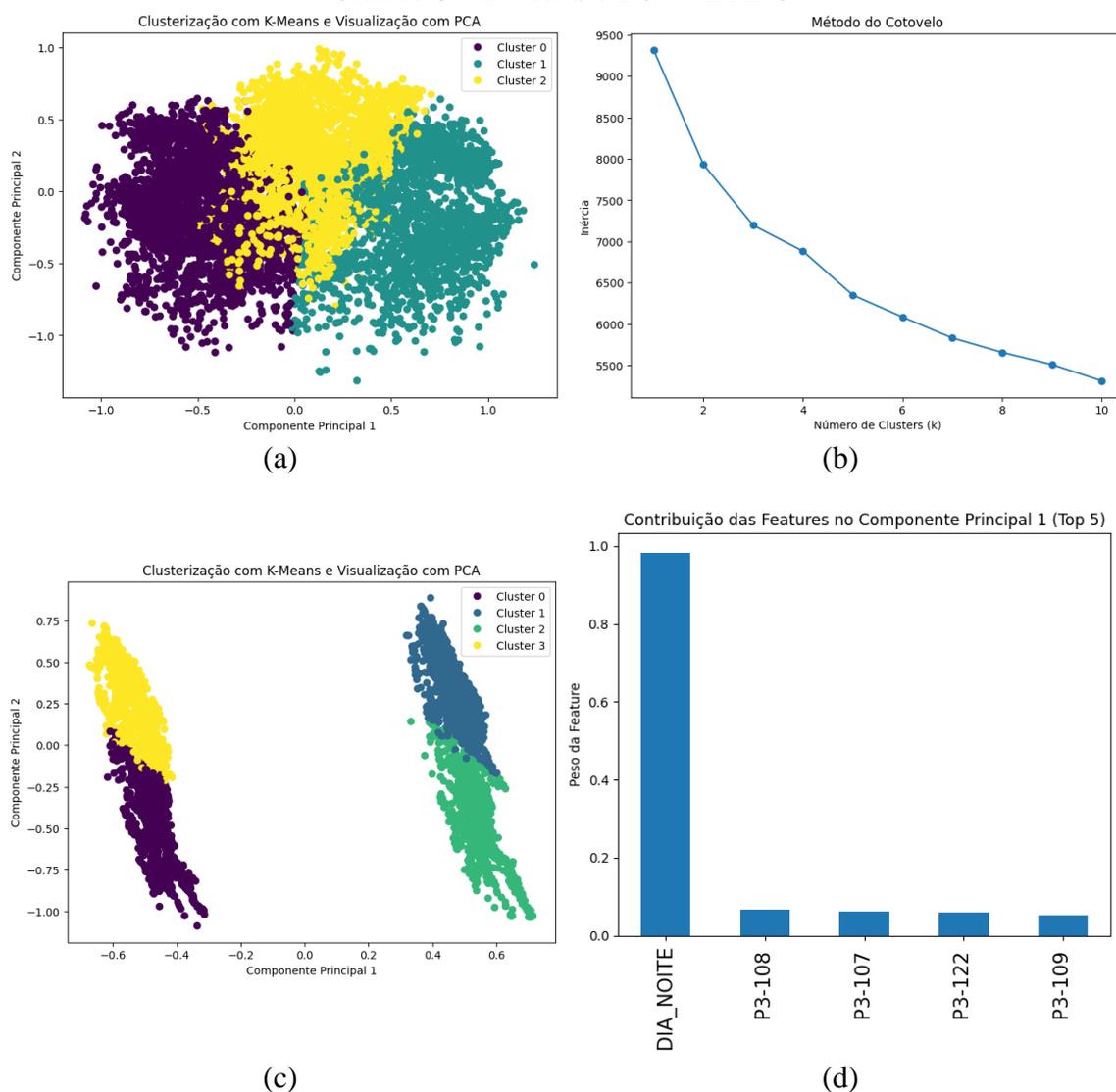
Figura 38 - Box-plot dos dados - ETE C - Antes e depois da retirada de outliers



Fonte: Do autor (2025)

Assim como ocorreu em outras plantas, a clusterização com *K-means* e visualização com PCA é realizada, Gráfico 8 (a), gráfico para determinação do cotovelo, Gráfico 8 (b), e seguido das contribuições de cada uma das *features*, Gráfico 9 (a) e (b). A diferença na quantidade de clusters entre os gráficos com e sem a variável “DIA/NOITE”, Gráfico 8 (a) e Gráfico 8 (c), ocorre devido a utilização do método do cotovelo, nessas duas bases de dados diferentes.

Gráfico 8 - *K-means* e PCA - ETE C



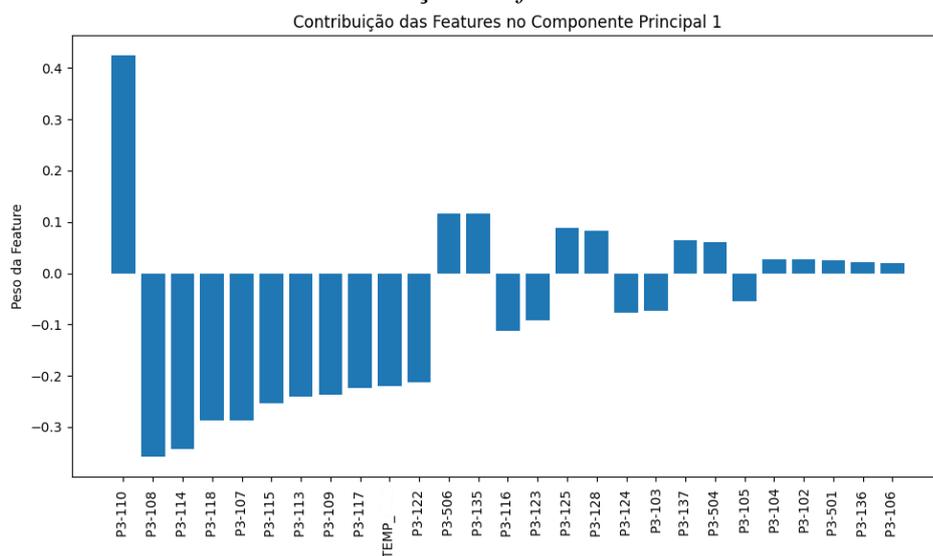
Fonte: Do autor (2025)

No caso das imagens de baixo, do Gráfico 8 (c) e (d), os resultados obtidos na análise de clusterização com *K-means*, visualizados com PCA, mostram uma clara distinção entre dois pares de *clusters* que se relacionam fortemente com a variável categórica "DIA/NOITE", que foi incluída nesse gráfico. Esta variável foi responsável por explicar grande parte da variância

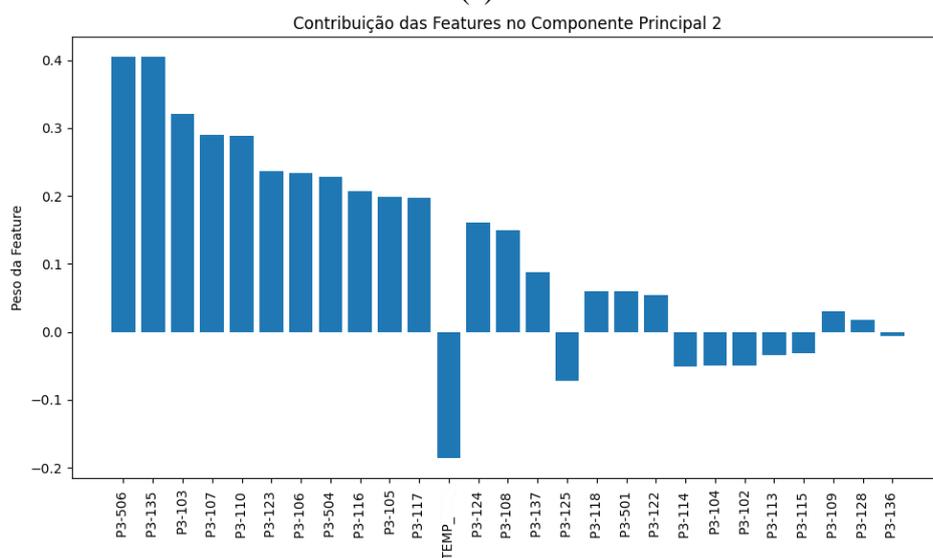
do Componente Principal 1, conforme destacado no gráfico de contribuição das *features*, onde a importância dessa variável é significativamente maior em comparação às demais.

Essa distinção pode ser explicada pelo padrão típico de uso da rede de esgoto ao longo do dia. Durante o período diurno (06:00 às 18:00), e diferença de geração de carga orgânica no período noturno (18:00 às 06:00). Os *clusters* resultantes refletem essas diferenças ao agrupar amostras que apresentam comportamentos similares para cada faixa horária. Essa separação é fundamental para entender a dinâmica da rede de esgoto e otimizar processos de tratamento, permitindo adaptações no manejo operacional de acordo com a carga prevista para cada período.

Gráfico 9 - Contribuição das *features* PCA - ETE C



(a)



(b)

Fonte: Do autor (2025)

As principais análises associadas a Componente Principal 1, são OD e ORP, enquanto para a Componente Principal 2 incluem além de algumas análises de ORP, também análises de SST e manto de lodo.

#### 5.2.4 ETE D

A ETE D tem capacidade para atender uma população equivalente a 720 mil habitantes. Com uma vazão média de projeto de 1.500 litros por segundo e uma vazão atual de 1.189 litros por segundo (média de 2023). O esgoto é transportado até a estação através de um sistema de 47 quilômetros de coletores-tronco e interceptores, sendo tratado antes de o efluente final ser lançado no Rio.

A Tabela 11 relaciona os parâmetros analíticos que são estão disponíveis na ETE D. A planta conta com quatro tanques de aeração, com três zonas de medição de OD e ORP em cada uma delas, totalizando os 12 pontos de cada uma dessas análises. Como usualmente foi visto em outras plantas também há a medição de DQO na entrada e saída da Estação, e medição de manto nos decantadores, nesse caso, nos primários e secundários.

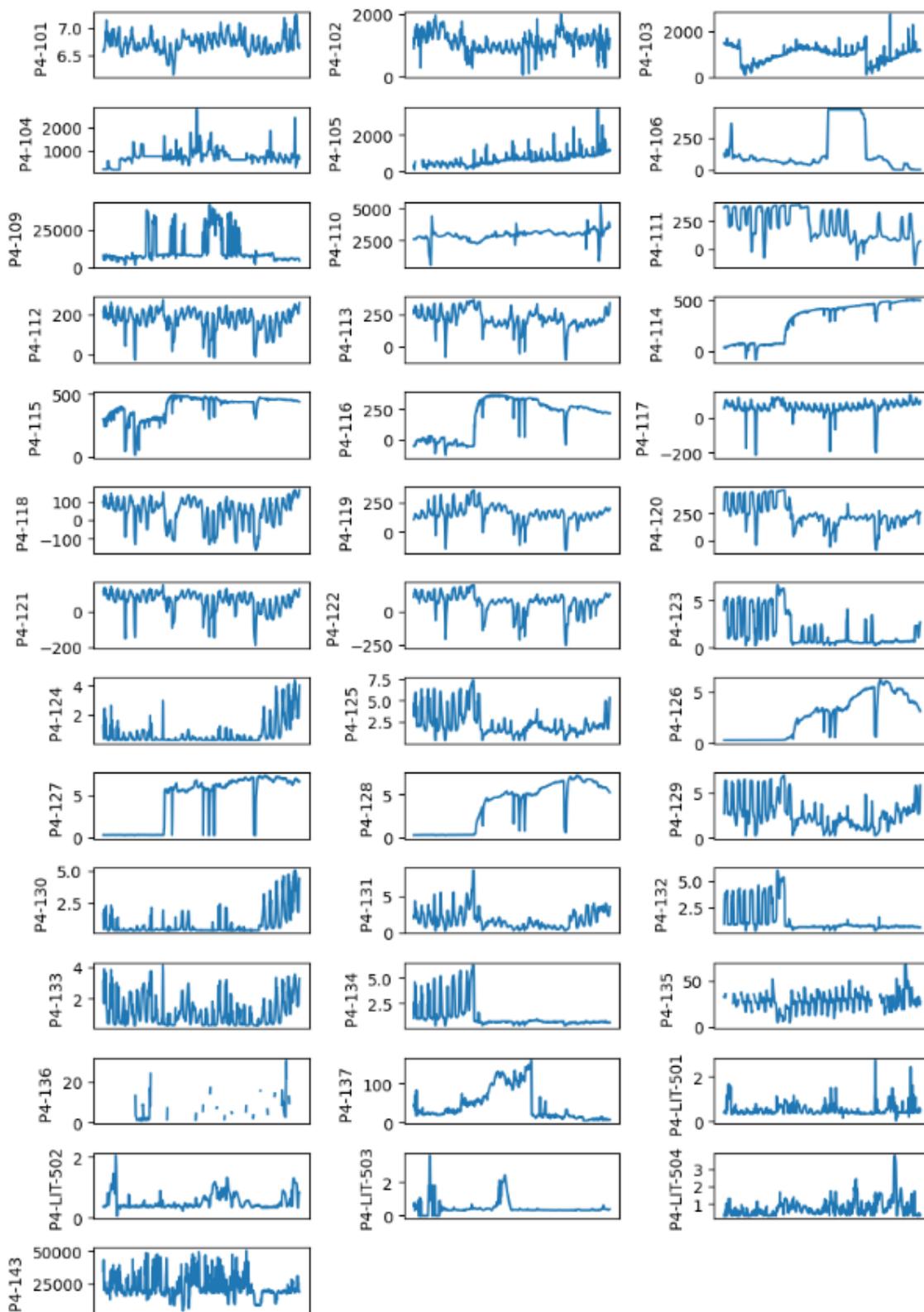
Tabela 11 - Relação de instrumentos por planta - ETE D

|       | TAG              | Parâmetro     | Local                | Faixa de Trabalho   |                      |
|-------|------------------|---------------|----------------------|---------------------|----------------------|
| ETE D | P4-AIT-101       | pH            | Esgoto desarenado    | 2 a 12              |                      |
|       | P4-AIT-102       | Condutividade | Esgoto desarenado    | 0-10K $\mu$ S/cm    |                      |
|       | P4-AIT-103       | SST (imersão) | Esgoto desarenado    | 100 a 500 mg/L      |                      |
|       | P4-AIT-104       | DQO           | Esgoto desarenado    | 200 a 1.000 mg/L    |                      |
|       | P4-AIT-105       | SST (imersão) | Efluente Primário    | 20 a 200 mg/L       |                      |
|       | P4-AIT-106       | DQO           | Efluente Primário    | 100 a 600 mg/L      |                      |
|       | P4-LIT-301       | Manta de Lodo | Decantador Prim.     | 0-6metros           |                      |
|       | P4-LIT-302       | Manta de Lodo | Decantador Prim.     | 0-6metros           |                      |
|       | P4-AIT-109       | SST (imersão) | Licor misto          | 1.000 a 8.000 mg/L  |                      |
|       | P4-AIT-110       | SST (imersão) | Lodo de Retorno      | 2.000 a 15.000 mg/L |                      |
|       | P4-AIT-111 a 122 | ORP           | Reator aeróbio 1 a 4 | (-175 a -400 mV)    |                      |
|       | P4-AIT-123 a 124 | OD            | Reator aeróbio 1 a 4 | 0 a 8 PPM           |                      |
|       | P4-AIT-135       | NH3           | Efluente Primário    | 10-150mg/L          |                      |
|       | P4-AIT-136       | NH3           | Licor misto          | 0,05-20mg/L         |                      |
|       | P4-AIT-137       | Turbidez      | Efluente tratado     | 0-10 NTU            |                      |
|       | P4-LIT-501       | Manto de Lodo | Decantador DS1       | 0-6metros           |                      |
|       | P4-LIT-502       | Manto de Lodo | Decantador DS2       | 0-6metros           |                      |
|       | P4-LIT-503       | Manto de Lodo | Decantador DS3       | 0-6metros           |                      |
|       | P4-LIT-504       | Manto de Lodo | Decantador DS4       | 0-6metros           |                      |
|       | FASE SÓLIDA      | P4-AIT-142    | SST (imersão)        | Lodo Aflu. Digestor | 20.000 a 50.000 mg/L |
|       |                  | P4-AIT-143    | SST (imersão)        | Lodo condicionado   | 20.000 a 50.000 mg/L |
|       | REÚSO            | P4-AIT-147    | Cloro (inserção)     | ETA Reuso           | 0 a 20 mg/l          |
|       |                  | P4-AIT-148    | Turbidez             | ETA Reuso           | 0-10 NTU;            |

Fonte: Do autor (2025)

Os gráficos de linha dos parâmetros novamente permitem uma análise prévia da distribuição dos dados, vide Figura 39:

Figura 39 - Dados ETE D



Fonte: Do autor (2025)

Uma análise conjunta dos gráficos plotados na Figura 39 e dos dados na Tabela 12, permitem identificar a caracterização dos dados analisados na ETE D.

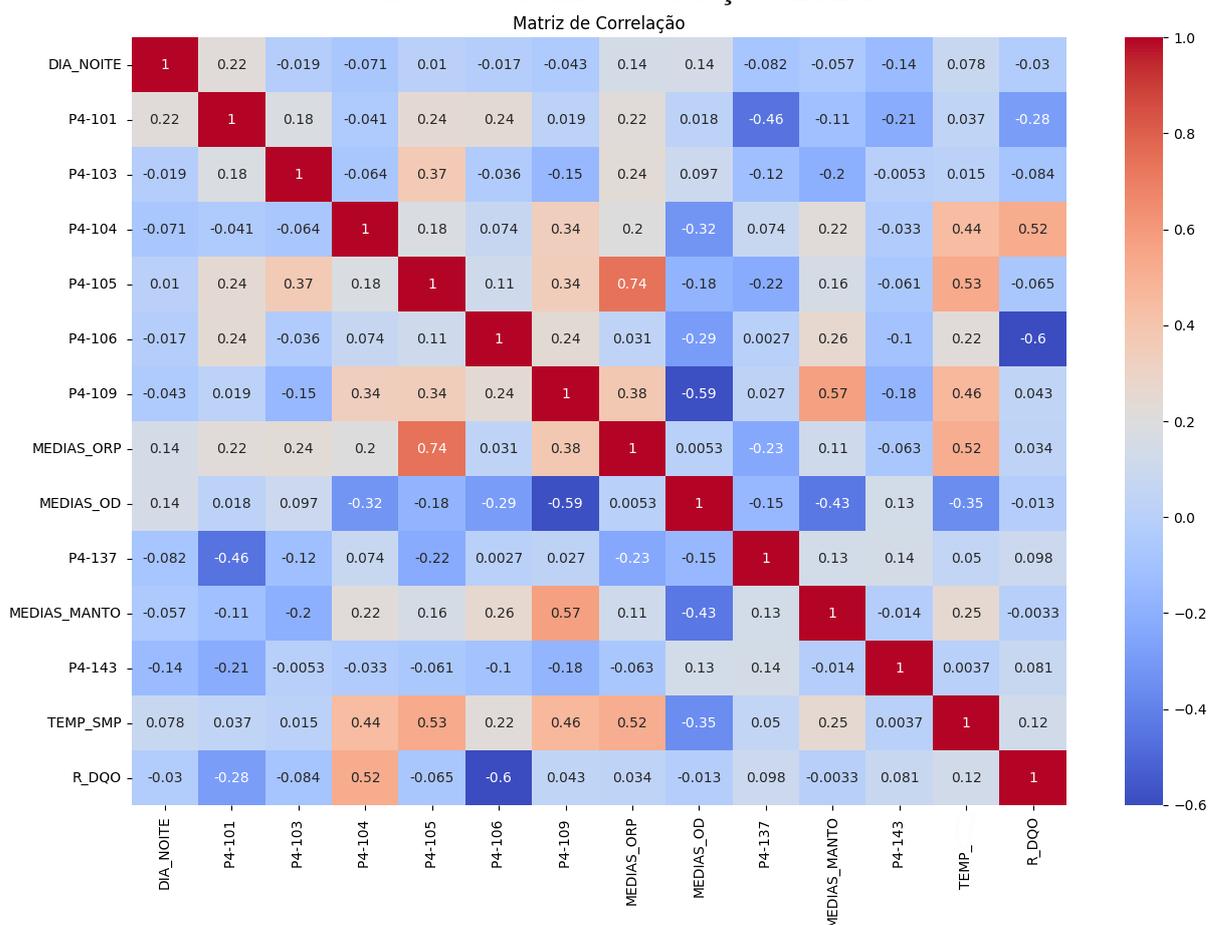
Tabela 12 - Caracterização dos dados - Julho/2024 - ETE D

|            |               | Min     | Max      | Media    | Mediana  | Desv Pad |
|------------|---------------|---------|----------|----------|----------|----------|
| P4-AIT-101 | pH            | 6,13    | 7,29     | 6,71     | 6,70     | 0,15     |
| P4-AIT-102 | Condutividade | 0,01    | 2717,70  | 1086,33  | 1061,04  | 314,41   |
| P4-AIT-103 | SST (imersão) | 71,66   | 5597,64  | 999,72   | 1020,14  | 397,13   |
| P4-AIT-104 | DQO           | 182,20  | 2767,80  | 721,41   | 701,40   | 323,01   |
| P4-AIT-105 | SST (imersão) | 101,34  | 5718,99  | 650,49   | 618,67   | 348,34   |
| P4-AIT-106 | DQO           | 0,10    | 465,30   | 139,73   | 77,90    | 155,87   |
| P4-AIT-109 | SST (imersão) | 1152,33 | 42915,52 | 11192,07 | 7924,31  | 9097,44  |
| P4-AIT-110 | SST (imersão) | 210,75  | 5393,00  | 2914,75  | 2947,68  | 466,61   |
| P4-AIT-111 | ORP           | -147,67 | 395,73   | 213,29   | 191,33   | 132,09   |
| P4-AIT-112 | ORP           | -37,76  | 273,96   | 171,15   | 179,82   | 51,41    |
| P4-AIT-113 | ORP           | -115,47 | 375,04   | 218,27   | 213,11   | 73,18    |
| P4-AIT-114 | ORP           | -163,34 | 502,35   | 294,82   | 393,26   | 183,50   |
| P4-AIT-115 | ORP           | -188,99 | 499,18   | 397,16   | 439,69   | 91,82    |
| P4-AIT-116 | ORP           | -189,61 | 367,50   | 176,82   | 246,07   | 161,60   |
| P4-AIT-117 | ORP           | -224,31 | 128,45   | 57,21    | 61,00    | 42,26    |
| P4-AIT-118 | ORP           | -162,81 | 152,50   | 41,49    | 62,11    | 70,49    |
| P4-AIT-119 | ORP           | -157,72 | 356,73   | 152,03   | 152,19   | 75,69    |
| P4-AIT-120 | ORP           | -87,30  | 461,65   | 242,55   | 221,36   | 111,73   |
| P4-AIT-121 | ORP           | -192,29 | 150,30   | 58,76    | 72,11    | 55,05    |
| P4-AIT-122 | ORP           | -273,04 | 197,55   | 65,84    | 80,67    | 76,12    |
| P4-AIT-123 | OD            | 0,28    | 6,67     | 1,80     | 0,71     | 1,81     |
| P4-AIT-124 | OD            | 0,28    | 4,52     | 0,64     | 0,35     | 0,66     |
| P4-AIT-125 | OD            | 0,36    | 7,65     | 2,29     | 1,80     | 1,46     |
| P4-AIT-126 | OD            | 0,26    | 6,48     | 2,51     | 2,64     | 2,05     |
| P4-AIT-127 | OD            | 0,25    | 7,50     | 4,12     | 5,71     | 2,95     |
| P4-AIT-128 | OD            | 0,28    | 7,26     | 3,57     | 4,69     | 2,64     |
| P4-AIT-129 | OD            | 0,31    | 6,99     | 2,64     | 2,26     | 1,66     |
| P4-AIT-130 | OD            | 0,32    | 5,48     | 0,70     | 0,35     | 0,77     |
| P4-AIT-131 | OD            | 0,33    | 8,61     | 1,75     | 1,45     | 1,24     |
| P4-AIT-132 | OD            | 0,34    | 6,03     | 1,30     | 0,76     | 1,28     |
| P4-AIT-133 | OD            | 0,28    | 4,96     | 1,11     | 0,77     | 0,90     |
| P4-AIT-134 | OD            | 0,31    | 6,36     | 1,18     | 0,70     | 1,23     |
| P4-AIT-135 | NH3           | 1,41    | 85,60    | 26,90    | 27,25    | 9,00     |
| P4-AIT-136 | NH3           | 1,01    | 39,47    | 7,22     | 4,61     | 6,62     |
| P4-AIT-137 | Turbidez      | 4,44    | 171,61   | 45,32    | 24,72    | 40,34    |
| P4-LIT-501 | Manto de Lodo | 0,00    | 3,42     | 0,54     | 0,42     | 0,31     |
| P4-LIT-502 | Manto de Lodo | 0,00    | 2,62     | 0,53     | 0,41     | 0,27     |
| P4-LIT-503 | Manto de Lodo | 0,00    | 4,30     | 0,48     | 0,34     | 0,55     |
| P4-LIT-504 | Manto de Lodo | 0,01    | 3,75     | 0,80     | 0,69     | 0,54     |
| P4-AIT-143 | SST (imersão) | 3470,49 | 50500,00 | 22602,49 | 19371,77 | 10260,42 |

Fonte: Do autor (2025)

No caso da ETE D é possível observar que mais dados possuem maior correlação entre si, positiva e negativa, quando comparada a outras plantas. Como por exemplo o SST do licor primário em relação ao ORP dos tanques de aeração, Gráfico 10.

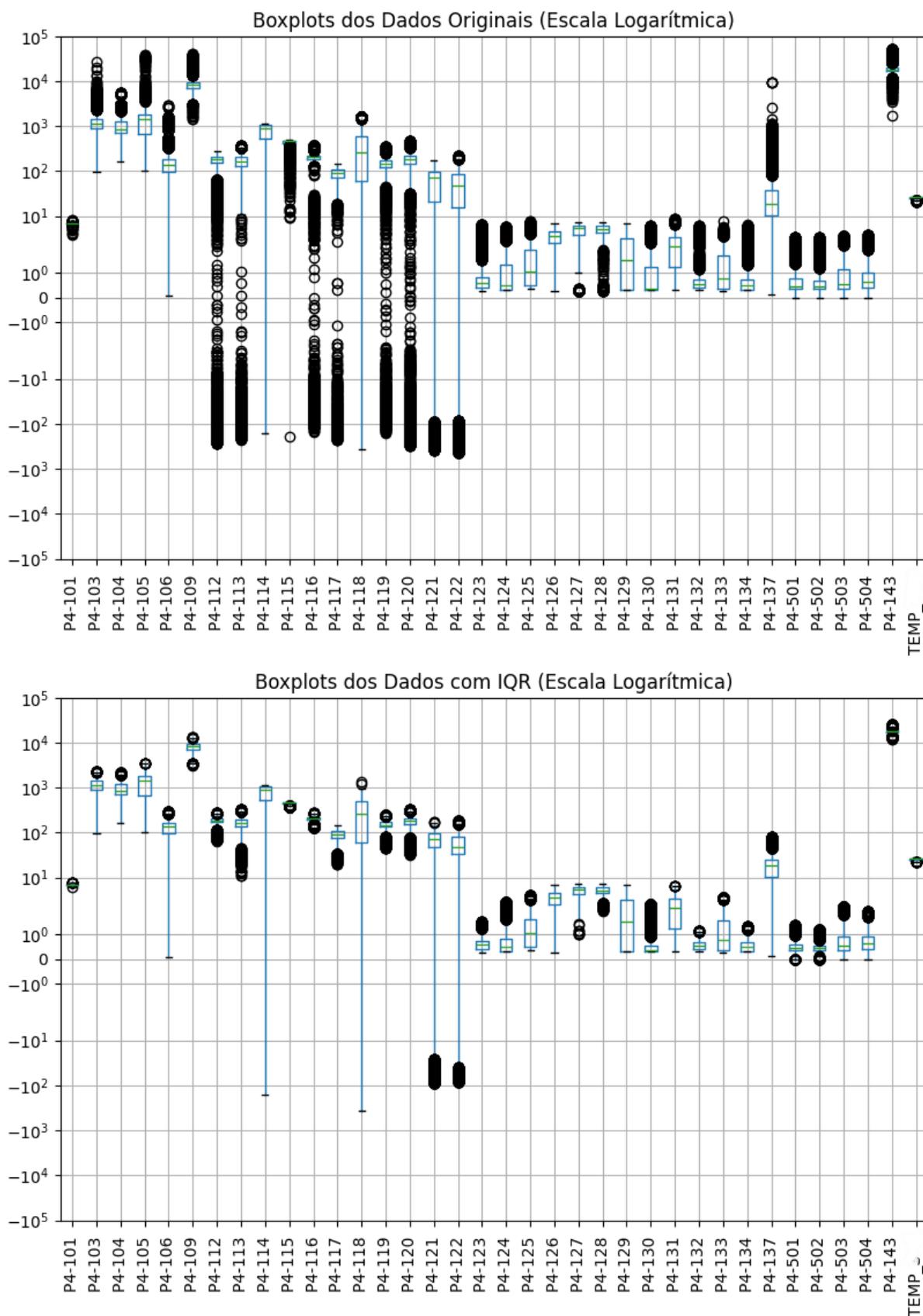
Gráfico 10 - Matriz de correlação - ETE D



Fonte: Do autor (2025)

Assim como foi realizado em outras plantas, o processo de remoção de identificação e remoção de *outliers* na ETE D, também foi feito e o produziu uma diferença expressiva nos dados, com a redução de gráficos com dados nas áreas periféricas do *box-plot*, e quando não ocorreu a ausência dos dados nessas áreas, foi possível observar a redução da diferença entre um dado na área central e fora da área da caixa, conforme a Figura 40.

Figura 40 - Box-plot dos dados - ETE D - Antes e depois da retirada de outliers

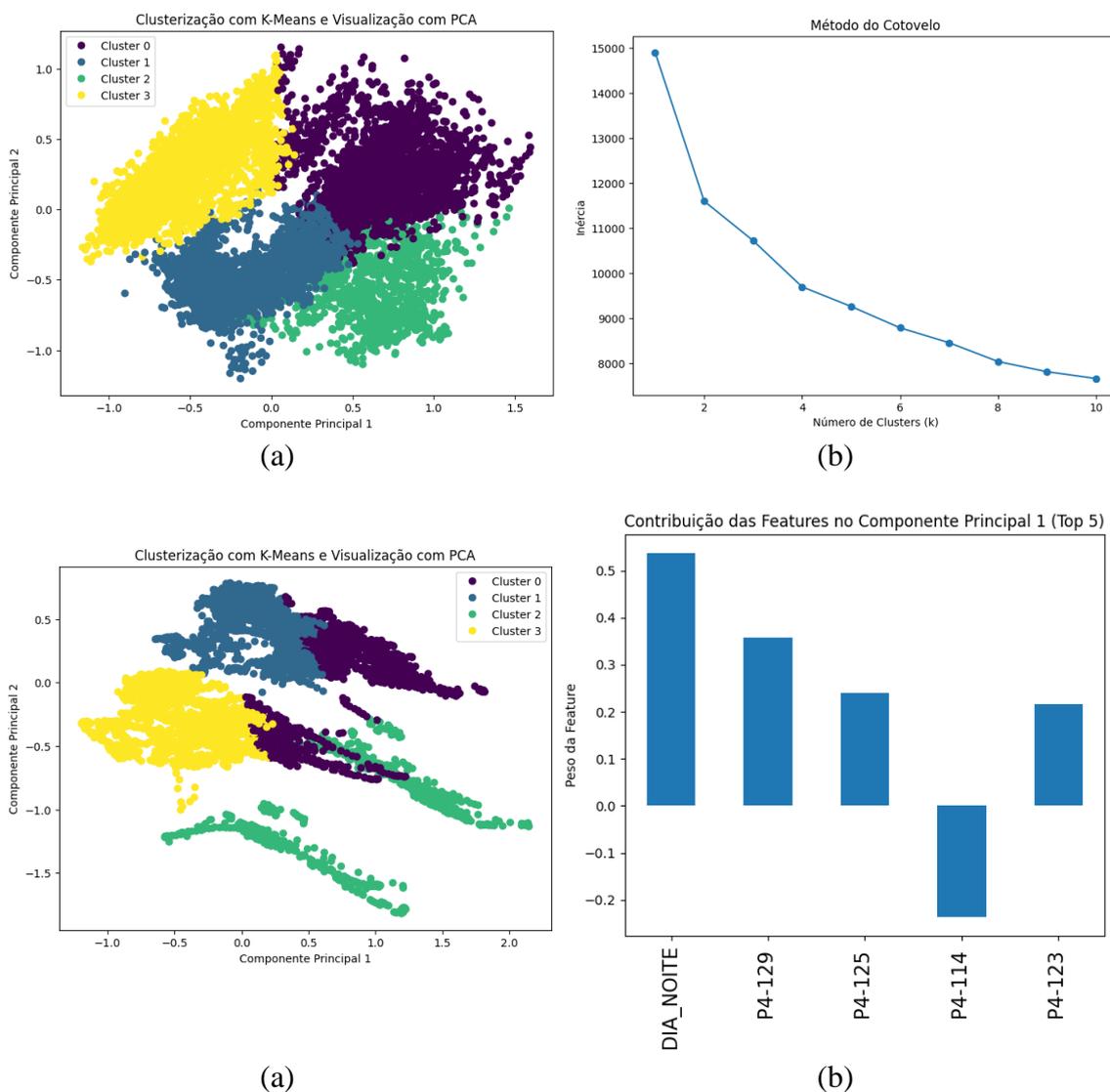


Fonte: Do autor (2025)

O Gráfico 11 (a) traz a visualização da clusterização com *K-means* e visualização com

PCA, já o Gráfico 11 (b), o gráfico para determinação do cotovelo, e os Gráfico 11 (c), assim como feito em outras ETEs, a inclusão da variável “DIA\_NOITE” e a contribuição das principais variáveis nas componentes principais.

Gráfico 11 - *K-means* e PCA - ETE D

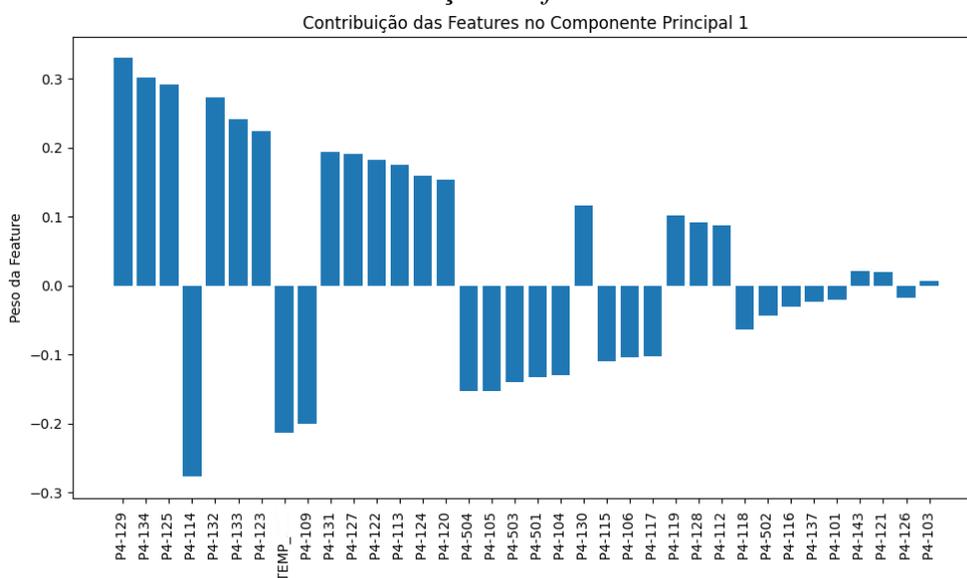


Fonte: Do autor (2025)

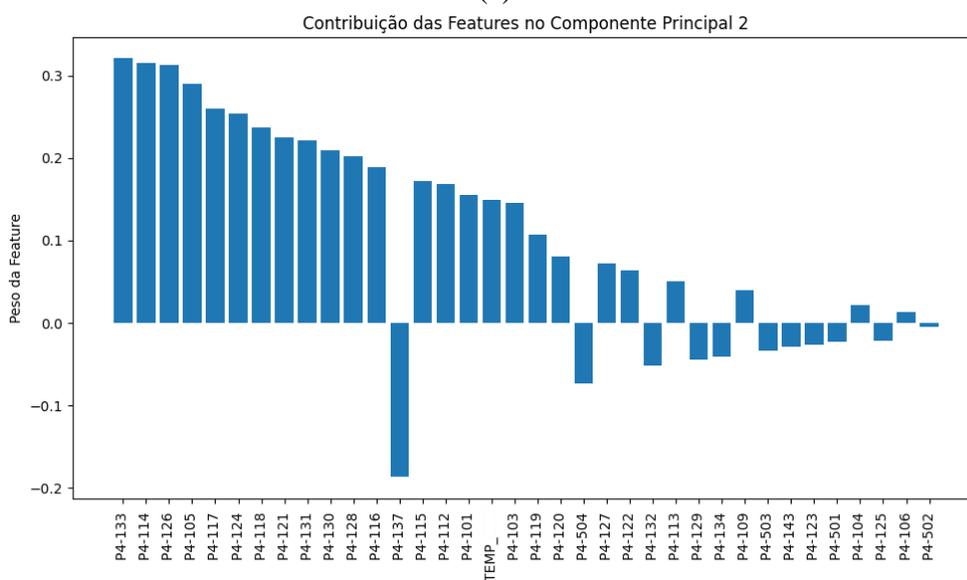
Na ETE D, a influência da variável "DIA/NOITE" também é perceptível na formação dos *clusters*, embora de maneira menos pronunciada em comparação às outras Estações, conforme as imagens inferiores do Gráfico 11. Ainda assim, os grupos apresentam uma leve tendência de separação de acordo com os períodos diurnos e noturnos, indicando que as diferenças nos padrões de geração de esgoto entre esses horários impactam parcialmente os dados. Essa distinção mais sutil pode estar relacionada às características específicas da rede de esgoto local, como menor variação na carga orgânica ao longo do dia ou padrões de uso mais

homogêneos entre os períodos.

Gráfico 12 - Contribuição das *features* PCA - ETE D



(a)



(b)

Fonte: Do autor (2025)

### 5.2.5 ETE E

A ETE E tem uma capacidade instalada para atender uma população equivalente a 720 mil habitantes e tratar 1,5 m<sup>3</sup>/s de esgoto. Em 2023, a vazão média foi de 900 l/s. O esgoto tratado na estação é transportado por um sistema de 30 quilômetros de coletores-tronco e interceptores antes de ser lançado no rio.

A Tabela 13, assim como nas outras plantas, traz a relação dos instrumentos, local de

instalação e faixa de trabalho, por ser uma planta menor, com menor vazão de tratamento e somente um dado de OD e ORP por tanque, a relação é menor:

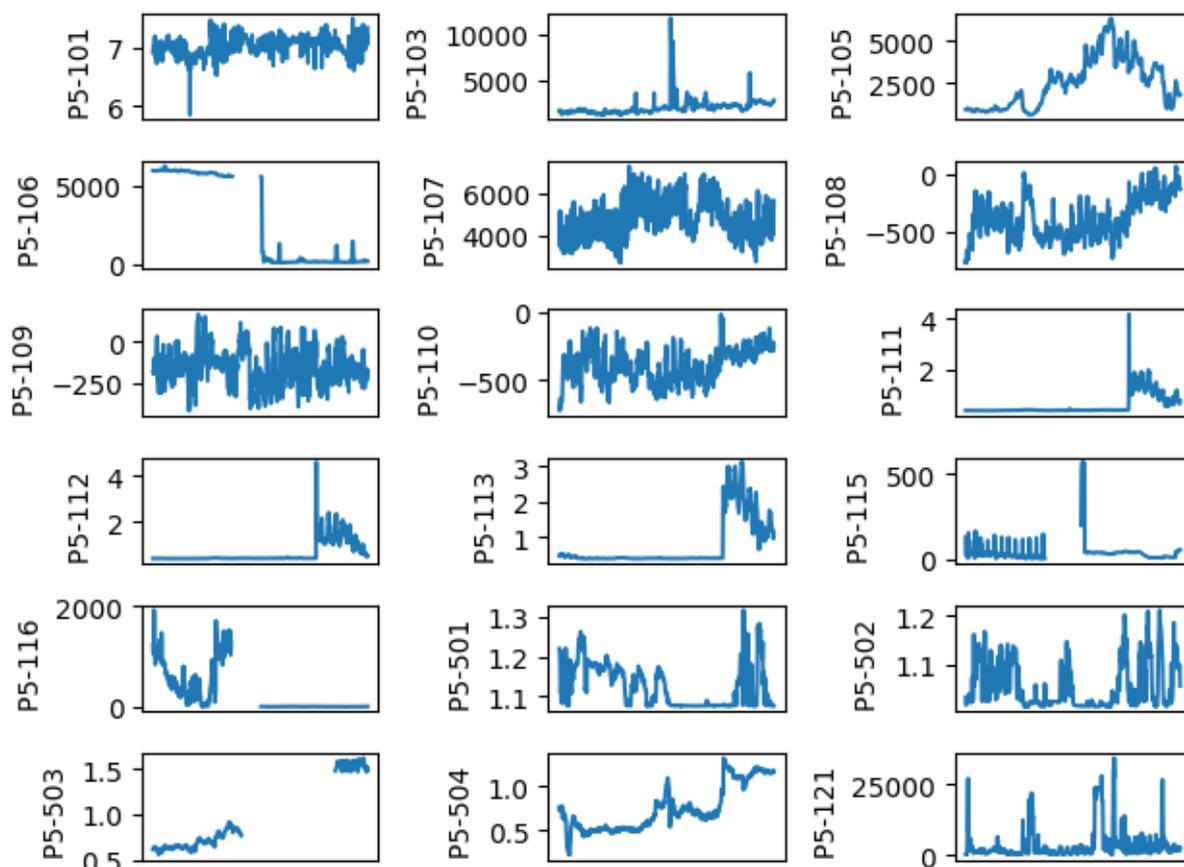
Tabela 13 - Relação de instrumentos por planta - ETE E

|       | TAG              | Parâmetro     | Local                | Faixa de Trabalho   |                      |
|-------|------------------|---------------|----------------------|---------------------|----------------------|
| ETE E | P5-AIT-101       | pH            | Esgoto desarenado    | 2 a 12              |                      |
|       | P5-AIT-102       | Condutividade | Esgoto desarenado    | 0-10K $\mu$ S/cm    |                      |
|       | P5-AIT-103       | SST (imersão) | Esgoto desarenado    | 100 a 500 mg/L      |                      |
|       | P5-AIT-104       | DQO           | Esgoto desarenado    | 200 a 1.000 mg/L    |                      |
|       | P5-AIT-105       | SST (imersão) | Efluente Primário    | 20 a 200 mg/L       |                      |
|       | P5-AIT-106       | DQO           | Efluente Final       | 100 a 600 mg/L      |                      |
|       | P5-AIT-107       | SST (imersão) | Licor misto          | 1.000 a 8.000 mg/L  |                      |
|       | P5-AIT-108 a 110 | ORP           | Reator aeróbio 1 a 3 | (-175 a -400 mV)    |                      |
|       | P5-AIT-111 a 113 | OD            | Reator aeróbio 1 a 3 | 0 a 8 PPM           |                      |
|       | P5-AIT-114       | NH3           | Efluente Primário    | 10-150mg/L          |                      |
|       | P5-AIT-115       | NH3           | Licor misto          | 0,05-20mg/L         |                      |
|       | P5-AIT-116       | Turbidez      | Efluente tratado     | 0-100 NTU           |                      |
|       | P5-LIT-501       | Manto de Lodo | Decantador DS1       | 0-4metros           |                      |
|       | P5-LIT-502       | Manto de Lodo | Decantador DS2       | 0-4metros           |                      |
|       | P5-LIT-503       | Manto de Lodo | Decantador DS3       | 0-4metros           |                      |
|       | P5-LIT-504       | Manto de Lodo | Decantador DS4       | 0-4metros           |                      |
|       | FASE LÍQUIDA     | P5-AIT-121    | SST (imersão)        | Lodo Aflu. Digestor | 20.000 a 50.000 mg/L |
|       |                  | P5-AIT-122    | SST (imersão)        | Filtrado FPP 1 a 3  | 1.000 a 10.000 mg/L  |
|       |                  | P5-AIT-123    | NH3                  | Filtrado FPP 1 a 3  | 10-150mg/L           |
|       |                  | P5-AIT-124    | Condutividade        | Filtrado FPP 1 a 3  | 0-10K $\mu$ S/cm     |
|       | FASE SÓLIDA      | P5-AIT-125    | Cloro (inserção)     | ETA Reuso           | 0 a 20 mg/l          |
|       |                  | P5-AIT-126    | Turbidez             | ETA Reuso           | 0-10 NTU;            |
|       | REÚSO            |               |                      |                     |                      |

Fonte: Do autor (2025)

Os gráficos que permitem a visualização prévia dos dados nos mostram algumas dificuldades em se trabalhar com modelagens de dados no cenário citado, como a ausência de dados em alguns períodos, ou a confiabilidade do registro, como ocorre com os dados de OD na primeira quinzena do mês analisado, conforme Figura 41:

Figura 41 - Dados ETE E



Fonte: Do autor (2025)

Assim como ocorre com as outras ETEs, a caracterização dos dados traz informações como valores mínimos, máximos, média, mediana e desvio padrão, Tabela 14:

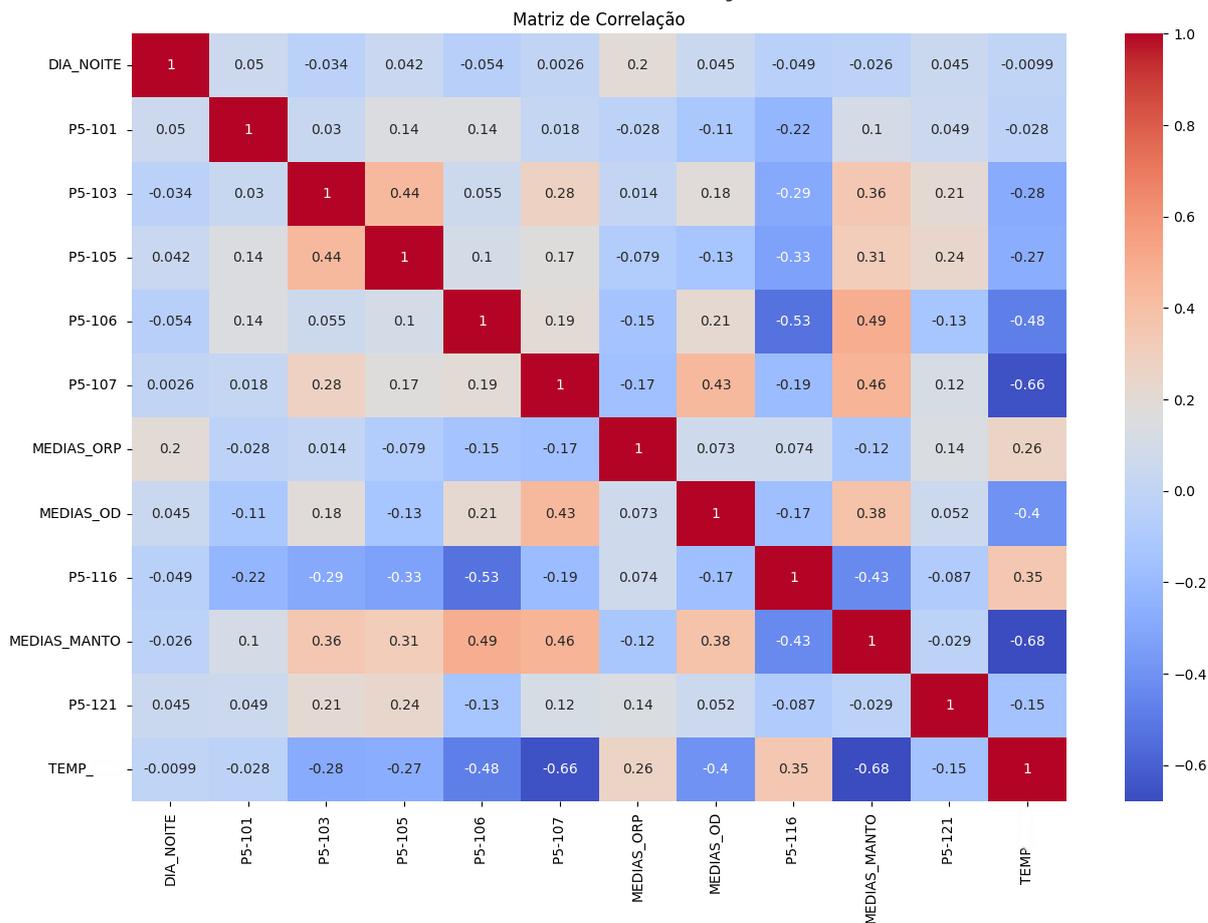
Tabela 14 - Caracterização dos dados - Julho/2024 - ETE E

|        |               | Min     | Max      | Media   | Mediana | Desv Pad |
|--------|---------------|---------|----------|---------|---------|----------|
| P5-101 | pH            | 5,17    | 7,78     | 7,02    | 7,03    | 0,18     |
| P5-103 | SST (imersão) | 1156,46 | 17075,92 | 2015,13 | 1795,59 | 1043,30  |
| P5-104 | DQO           | 1291,30 | 1291,30  | 1291,30 | 1291,30 | 0,00     |
| P5-105 | SST (imersão) | 545,72  | 6578,11  | 2434,01 | 2382,41 | 1536,49  |
| P5-106 | DQO           | 90,70   | 6196,50  | 2848,51 | 405,80  | 2807,90  |
| P5-107 | SST (imersão) | 2397,62 | 7632,55  | 4764,04 | 4746,26 | 1004,21  |
| P5-108 | ORP           | -813,12 | 119,24   | -399,39 | -430,38 | 185,99   |
| P5-109 | ORP           | -518,27 | 259,60   | -149,36 | -147,95 | 127,64   |
| P5-110 | ORP           | -744,30 | 418,90   | -399,65 | -401,18 | 142,22   |
| P5-111 | OD            | 0,35    | 8,89     | 0,54    | 0,38    | 0,41     |
| P5-112 | OD            | 0,37    | 8,29     | 0,60    | 0,39    | 0,48     |
| P5-113 | OD            | 0,34    | 3,23     | 0,72    | 0,40    | 0,67     |
| P5-115 | NH3           | 0,63    | 569,62   | 32,70   | 23,13   | 59,48    |
| P5-116 | Turbidez      | 2,89    | 10098,99 | 352,35  | 9,02    | 674,23   |
| P5-501 | Manto de Lodo | 1,07    | 1,33     | 1,13    | 1,10    | 0,06     |
| P5-502 | Manto de Lodo | 1,01    | 1,27     | 1,06    | 1,03    | 0,05     |
| P5-503 | Manto de Lodo | 0,53    | 1,71     | 0,87    | 0,70    | 0,35     |
| P5-504 | Manto de Lodo | 0,20    | 1,33     | 0,72    | 0,66    | 0,25     |
| P5-121 | SST (imersão) | 105,29  | 47678,79 | 3376,14 | 1390,35 | 5818,31  |

Fonte: Do autor (2025)

A matriz de correlação também apresenta a correlação entre OD e ORP, já esperada, porém outras análises chamam atenção, como a correlação negativa entre alguns parâmetros e os dois pontos de análise de SST “P5-105” e “P5-107” e a correlação positiva de algumas sondas com a análise de manto “P5-504”, conforme o Gráfico 13:

Gráfico 13 - Matriz de correlação - ETE E

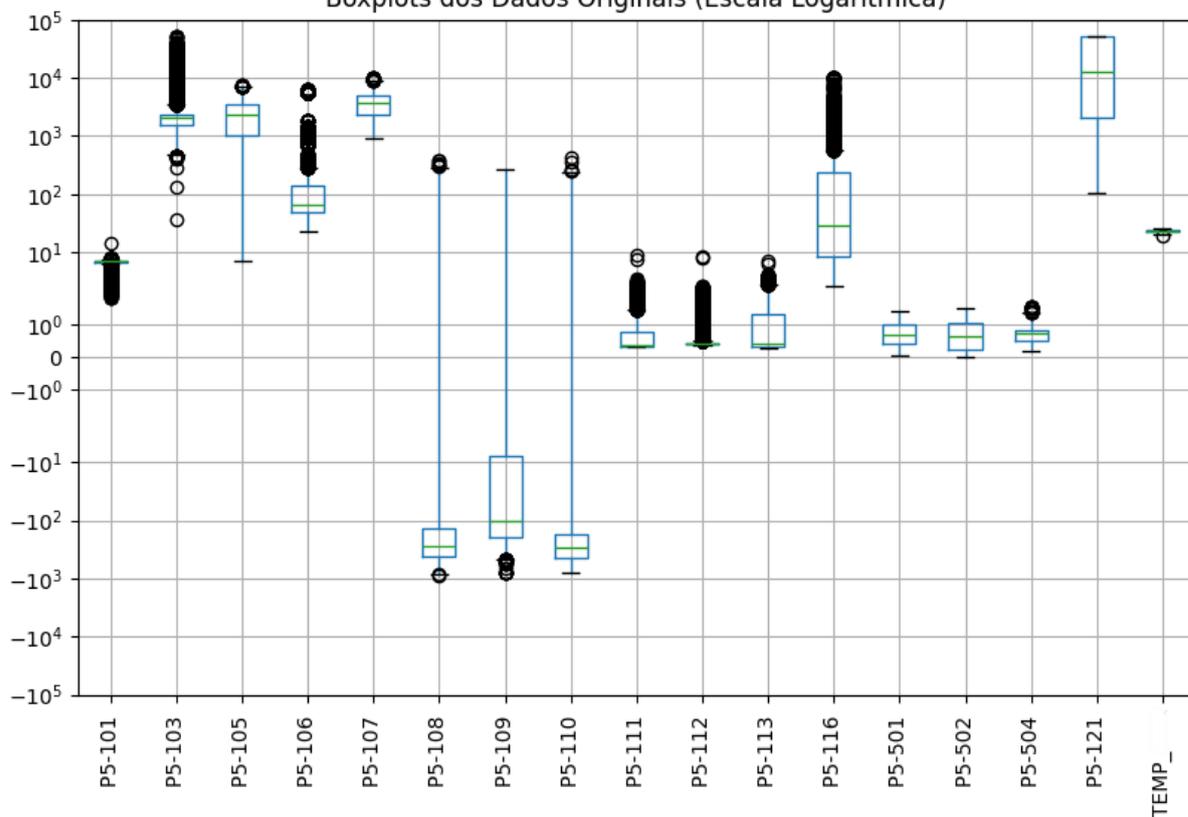


Fonte: Do autor (2025)

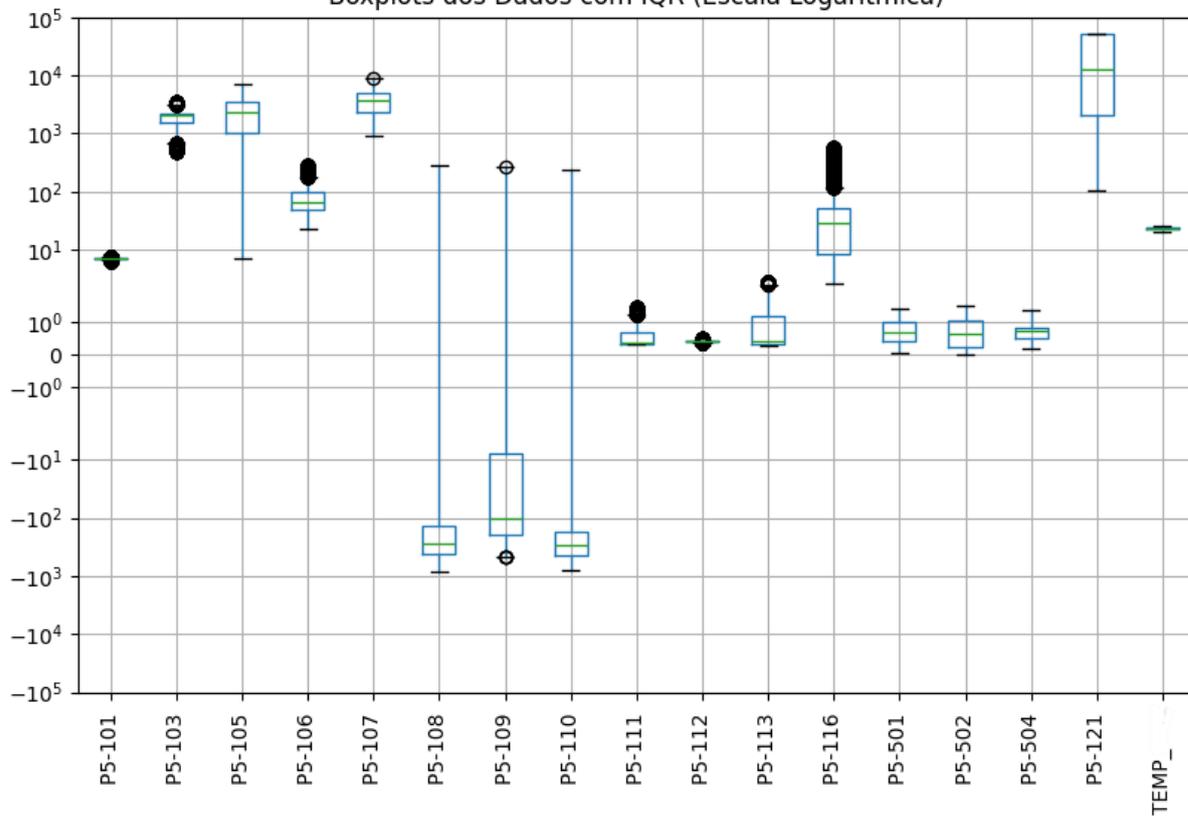
Mais uma vez, a técnica para identificação e redução de *outliers* se mostra eficiente, reduzindo a quantidade de gráficos com amostras fora da área da caixa no *box-plot*, Figura 42.

Figura 42 – *Box-plot* dos dados - ETE E

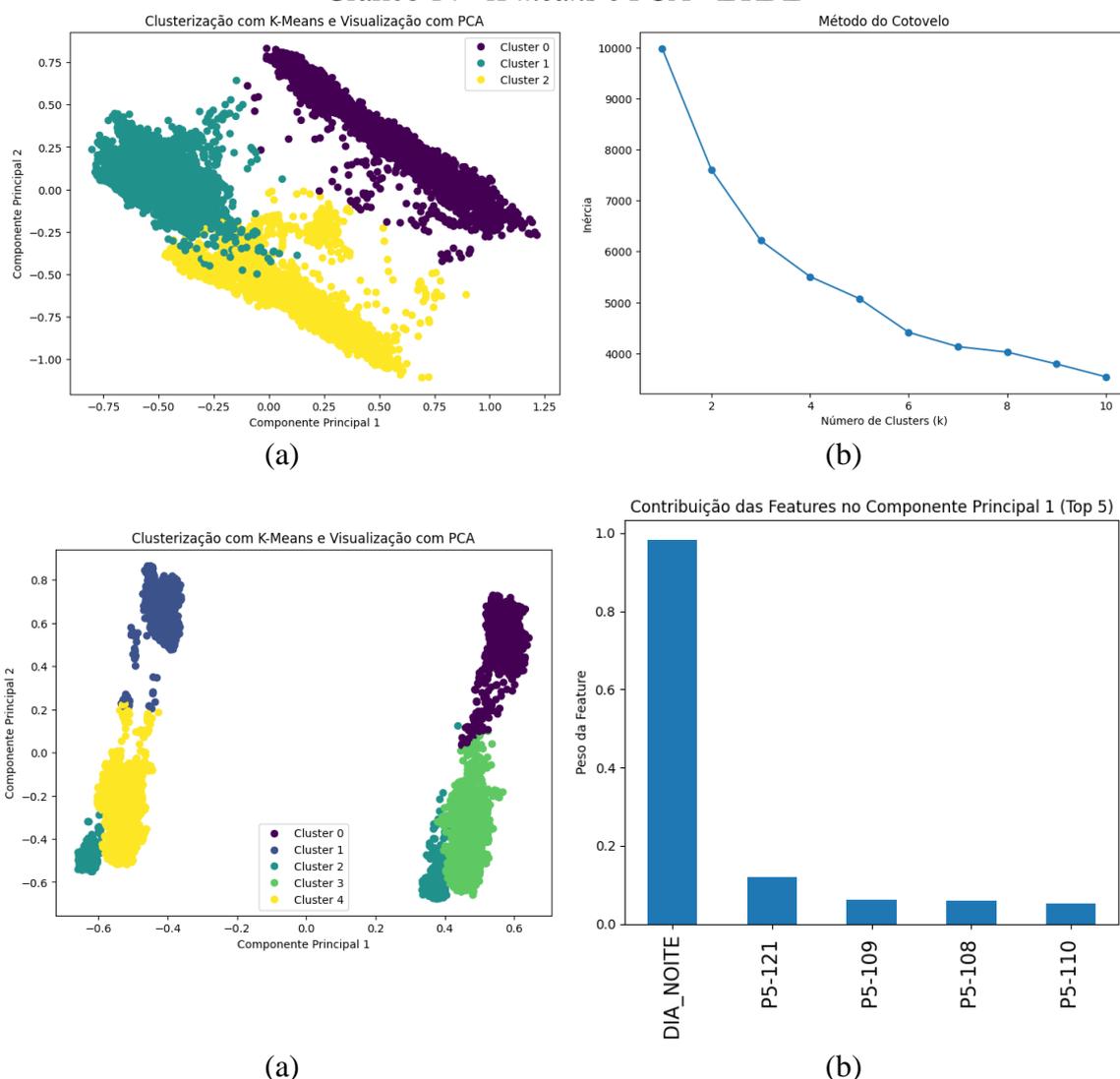
Boxplots dos Dados Originais (Escala Logarítmica)



Boxplots dos Dados com IQR (Escala Logarítmica)

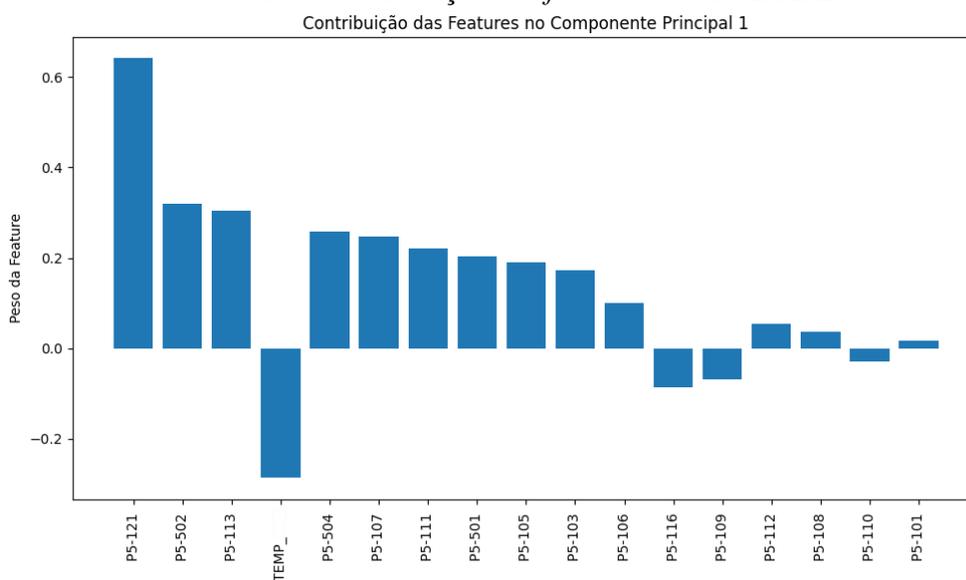


Fonte: Do autor (2025)

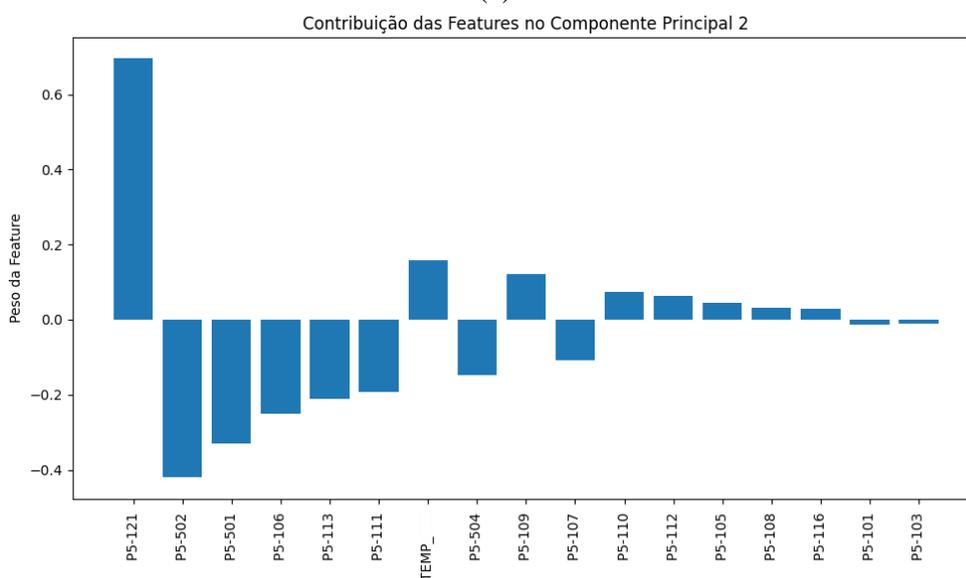
Gráfico 14 - *K-means* e PCA - ETE E

Fonte: Do autor (2025)

Para a ETE E, os resultados obtidos na clusterização também evidenciam uma separação bem definida entre grupos de dados que se alinham ao padrão de comportamento associado ao fator "DIA/NOITE", Gráfico 14 (c) e (d). Essa organização dos *clusters* reflete diferenças características nos períodos analisados, sugerindo uma dinâmica semelhante às ETes observadas anteriormente. A variação nos padrões de geração de esgoto entre os horários diurnos e noturnos também desempenha um papel central, indicando que fatores temporais influenciam de forma consistente as condições operacionais e a carga afluente à estação.

Gráfico 15 - Contribuição das *features* PCA - ETE E

(a)



(b)

Fonte: Do autor (2025)

### 5.3 Modelagem da variável Demanda Química de Oxigênio (DQO) utilizando RNA

Para a base de dados selecionada em cada uma das Estações de Tratamento de Esgoto, adotou-se uma divisão de 70% dos dados para o treinamento da rede neural e 30% para o teste. Essa divisão é comumente utilizada em ciência de dados, pois permite treinar a rede com uma quantidade significativa de informações, enquanto preserva um conjunto independente para avaliação do desempenho do modelo. É crucial destacar que a separação dos dados foi realizada de forma aleatória, a fim de evitar viés no processo de treinamento, assegurando que o modelo possa generalizar adequadamente ao lidar com novos dados. Além disso, é recomendável

realizar o processo de normalização dos dados antes de alimentar a rede neural, de modo a garantir uma convergência mais eficiente e resultados mais precisos.

Na pesquisa realizada, foi aplicada uma modelagem RNA em dados de quatro ETEs, tendo como variável alvo a DQO do efluente tratado, a ETE B não foi incluída nessa modelagem devido a indisponibilidade dessa sonda no período. A análise incluiu a geração de gráficos SHAP (*SHapley Additive exPlanations*) para interpretar o impacto de cada *feature* na predição dessa variável.

Para aplicação da técnica *Grid Search* foram definidos os hiperparâmetros na Tabela 15, com teste para três diferentes funções de ativação, dois algoritmos de *backpropagation*, opções com duas ou uma camada oculta, dois valores para taxa de aprendizagem e cinco estruturas diferentes para teste da RNA.

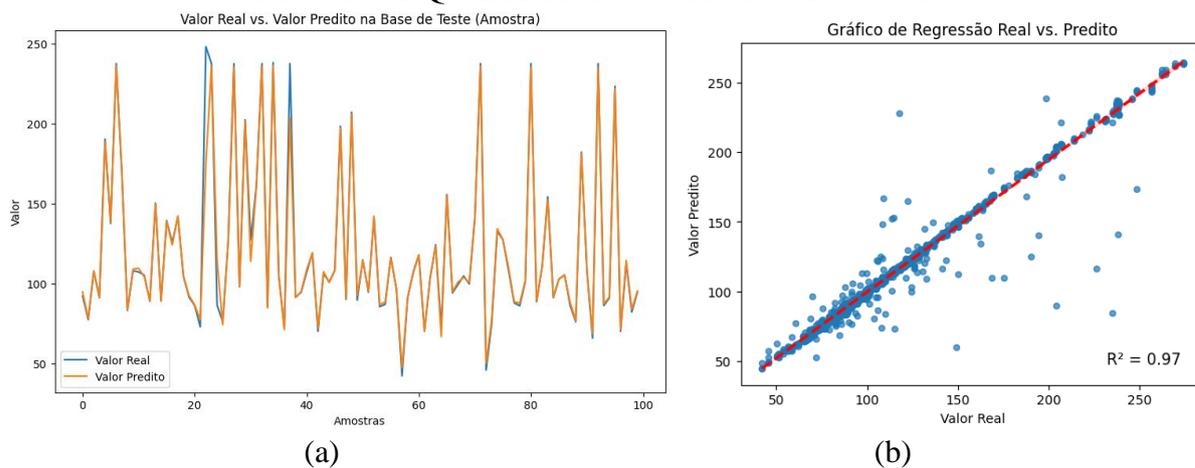
Tabela 15 - Parâmetros de configuração do *Grid Search* da RNA

| Parâmetro da RNA                    | Valores                                 |
|-------------------------------------|---|
| Função de ativação                  | ReLU, Tanh, Logistic                    |
| Épocas                              | 1000                                    |
| Algoritmo de <i>Backpropagation</i> | Adam, Sgd                               |
| Quantidade de camadas ocultas       | 2, 1                                    |
| Quantidade de neurônios por camada  | (30, 15), (20,10), (10,5), (30,), (10,) |
| Taxa de aprendizagem                | (0,1),(0,01)                            |

Fonte: Do autor (2025)

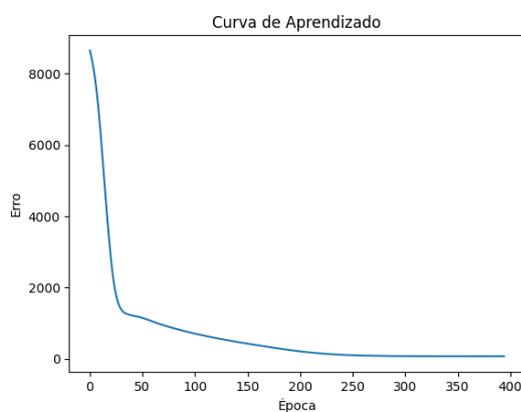
Os melhores hiperparâmetros para o modelo de RNA na DQO efluente na ETE A foram obtidos utilizando o *Grid Search*. Nesse processo, foram testadas múltiplas configurações predefinidas de hiperparâmetros, conforme a Tabela 15, avaliando o desempenho do modelo com cada combinação. A partir dessa análise, a configuração que apresentou os melhores resultados foram: função de ativação 'ReLU', taxa de regularização L2 (alpha) igual a 0.01, estrutura da camada oculta definida como (30,) (uma camada com 30 neurônios), taxa de aprendizado 'constant' e algoritmo de otimização 'sgd'. O Gráfico 16, traz a DQO real versus predita em um gráfico de linha para a base de dados separada para teste.

Gráfico 16 - DQO ETE A - Real x Predito - Base de Teste



(a)

(b)

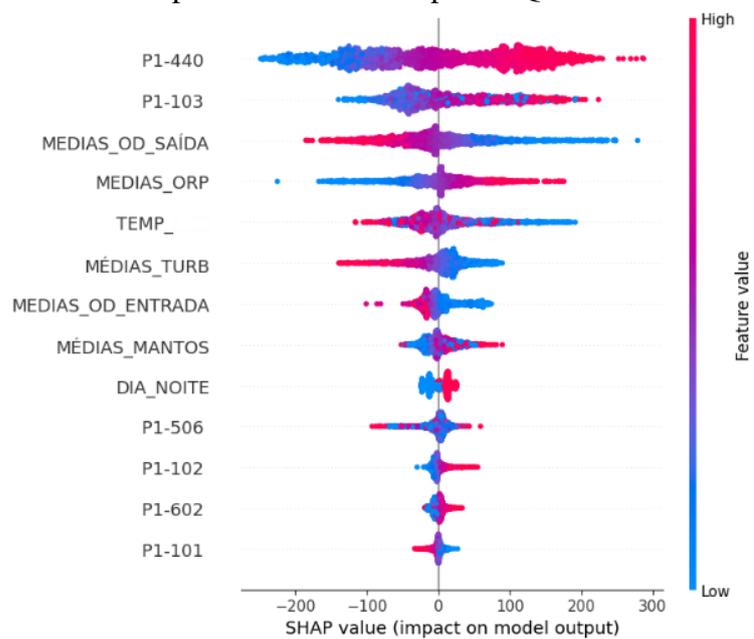


(c)

Fonte: Do autor (2025)

No caso da análise com os valores SHAP para a ETE A, Gráfico 17, os parâmetros mais influentes incluem o SST do esgoto desarenado, o pH do efluente final, e o valor médio de OD de saída no tanque de aeração. Esses fatores desempenham um papel na variável de interesse, que indicam a eficiência do tratamento biológico realizado na estação.

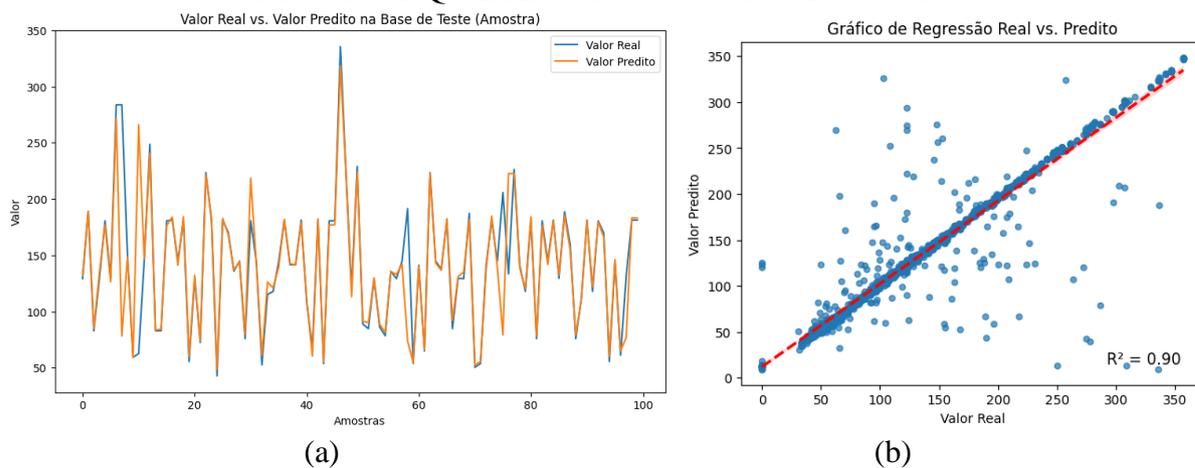
Gráfico 17 - Importância do SHAP para DQO efluente - ETE A



Fonte: Do autor (2025)

Para a ETE C, a seleção dos melhores hiperparâmetros indicou uma configuração distinta, com função de ativação 'ReLU', regularização L2 (alpha) de 0.1, uma arquitetura (10, 5) neurônios, taxa de aprendizado 'constant' e o otimizador 'adam'. O Gráfico 18 apresenta o resultado dessa ETE.

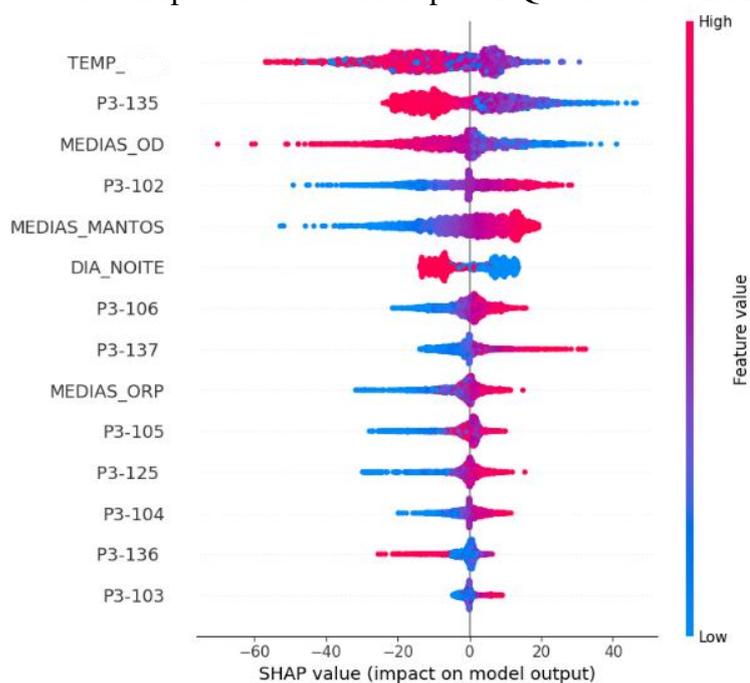
Gráfico 18 - DQO ETE C - Real x Predito - Base de Teste



Fonte: Do autor (2025)

Na ETE C, a análise SHAP revelou que a temperatura o OD no tanque de aeração, e condutividade do esgoto bruto são os principais parâmetros que se relacionam a variável de interesse, Gráfico 19. Esses parâmetros sugerem uma forte dependência dos processos biológicos de aeração e remoção de amônia, que afetam diretamente a qualidade final do efluente.

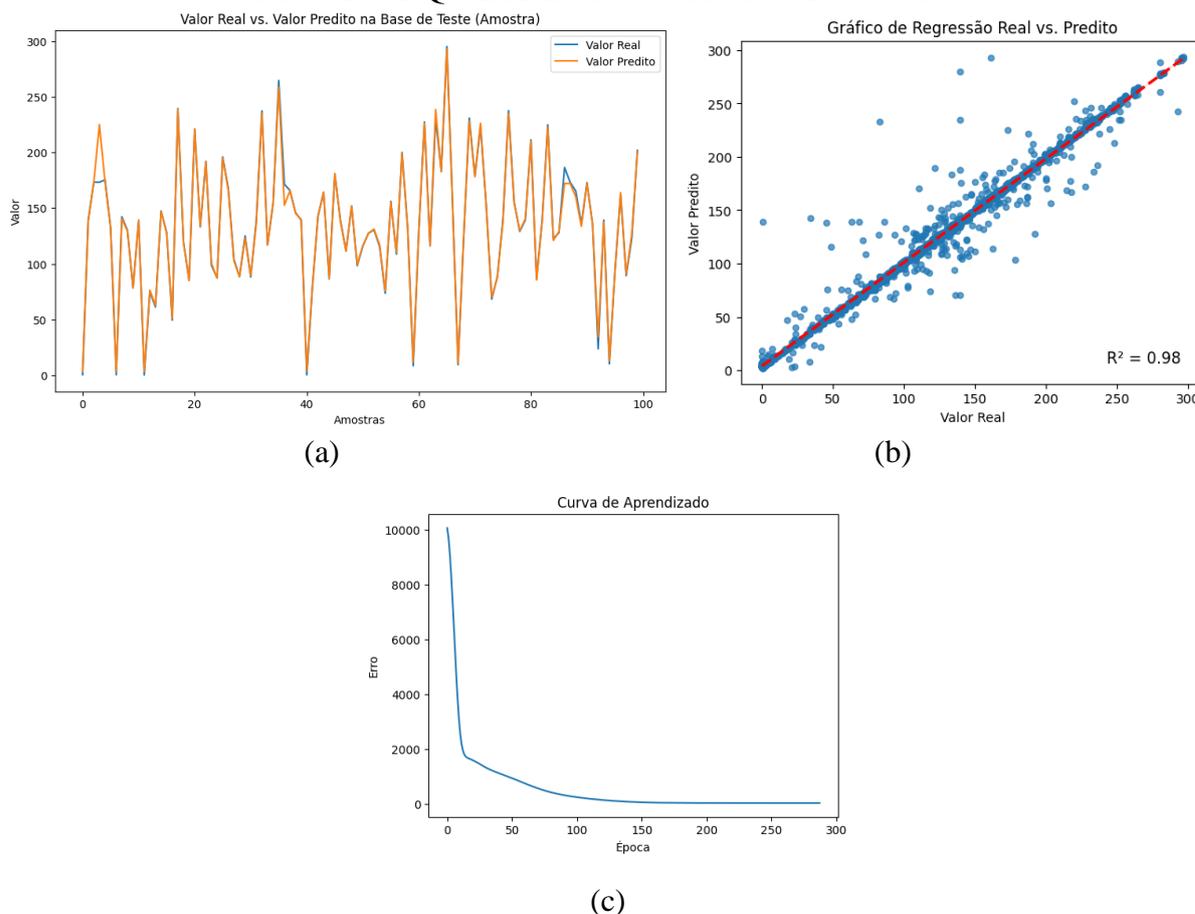
Gráfico 19 - Importância do SHAP para DQO efluente - ETE C



Fonte: Do autor (2025)

Para a ETE D, a configuração de hiperparâmetros que proporcionou os melhores resultados foi composta pela função de ativação 'logistic', regularização L2 (alpha) de 0.1, uma camada oculta com 30 neurônios (30), taxa de aprendizado 'constant' e o otimizador 'sgd', resultados no Gráfico 20.

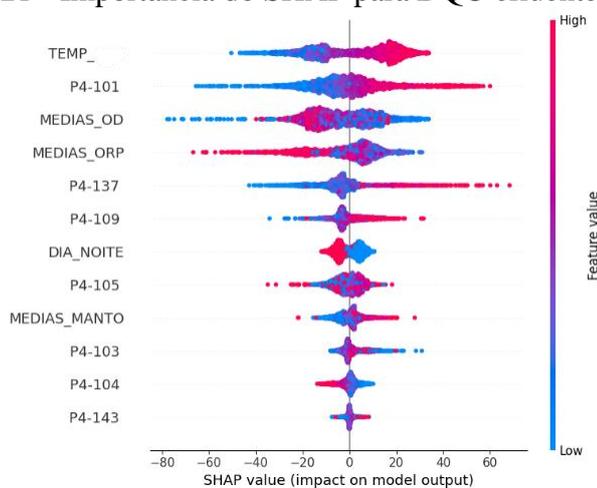
Gráfico 20 - DQO ETE D - Real x Predito - Base de Teste



Fonte: Do autor (2025)

Já na ETE D, novamente a temperatura, o OD e ORP nos tanque de aeração, e o pH no esgoto desarenado são os principais influenciadores, conforme o Gráfico 21. A predominância do OD e ORP indicam a importância de garantir uma boa aeração para sustentar os micro-organismos responsáveis pela degradação da matéria orgânica. Além disso, o controle do manto nos decantadores é fundamental para manter a qualidade do efluente tratado e evitar problemas de sedimentação inadequada.

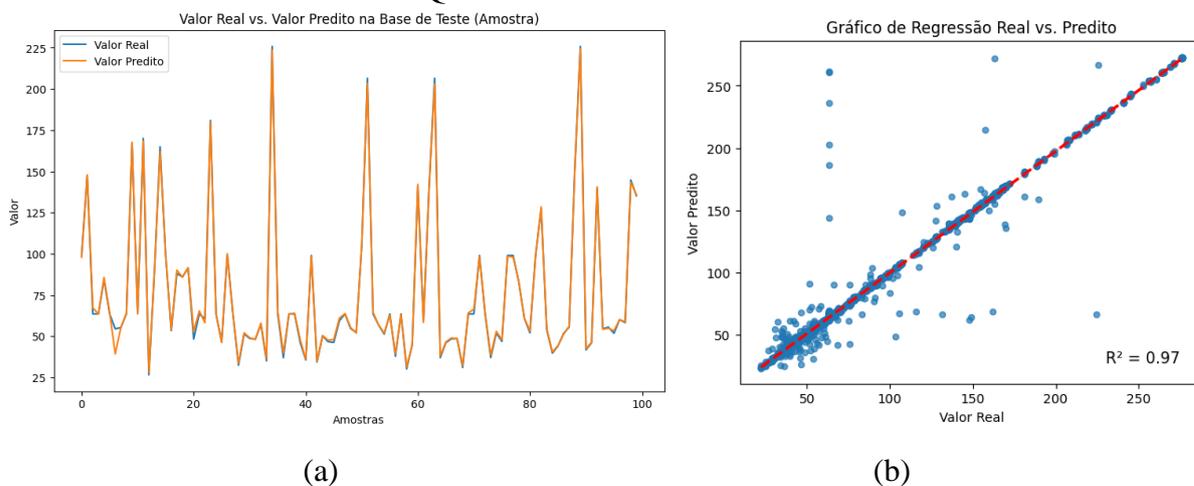
Gráfico 21 - Importância do SHAP para DQO efluente - ETE D



Fonte: Do autor (2025)

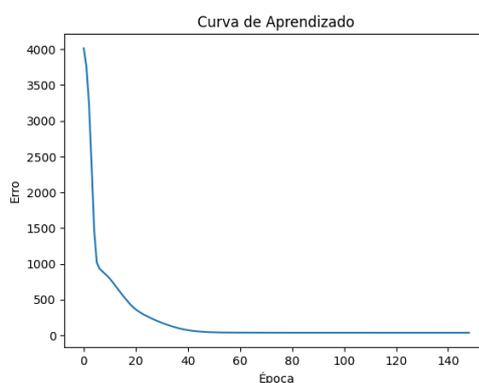
Para a ETE E, a configuração obtida incluiu a função de ativação 'ReLU', regularização L2 (alpha) de 0.1, uma arquitetura de camadas ocultas composta por (30, 15) neurônios, taxa de aprendizado 'constant' e o otimizador 'adam'. Esse resultado pode ser visto no Gráfico 22.

Gráfico 22 - DQO ETE E - Real x Predito - Base de Teste



(a)

(b)

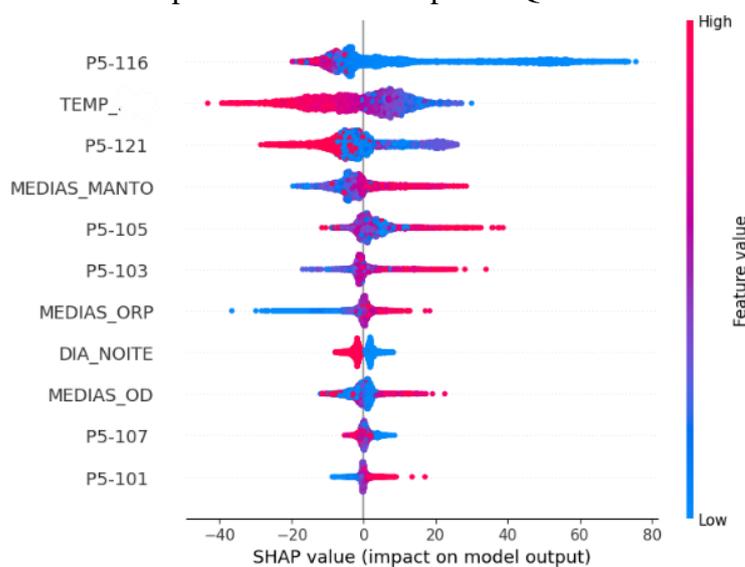


(c)

Fonte: Do autor (2025)

Na ETE E, Gráfico 23, os principais parâmetros identificados foram o Manto no decantador secundário, o SST e a amônia no licor misto, o ORP no tanque de aeração e a turbidez do efluente tratado. A influência desses fatores reforça a necessidade de um controle dos parâmetros sólidos e biológicos, além de destacar a amônia e a turbidez como indicadores importantes para garantir que o efluente final atenda aos padrões ambientais. A combinação desses fatores na ETE E sugere um foco na sedimentação e na eficiência da aeração.

Gráfico 23 - Importância do SHAP para DQO efluente - ETE E



Fonte: Do autor (2025)

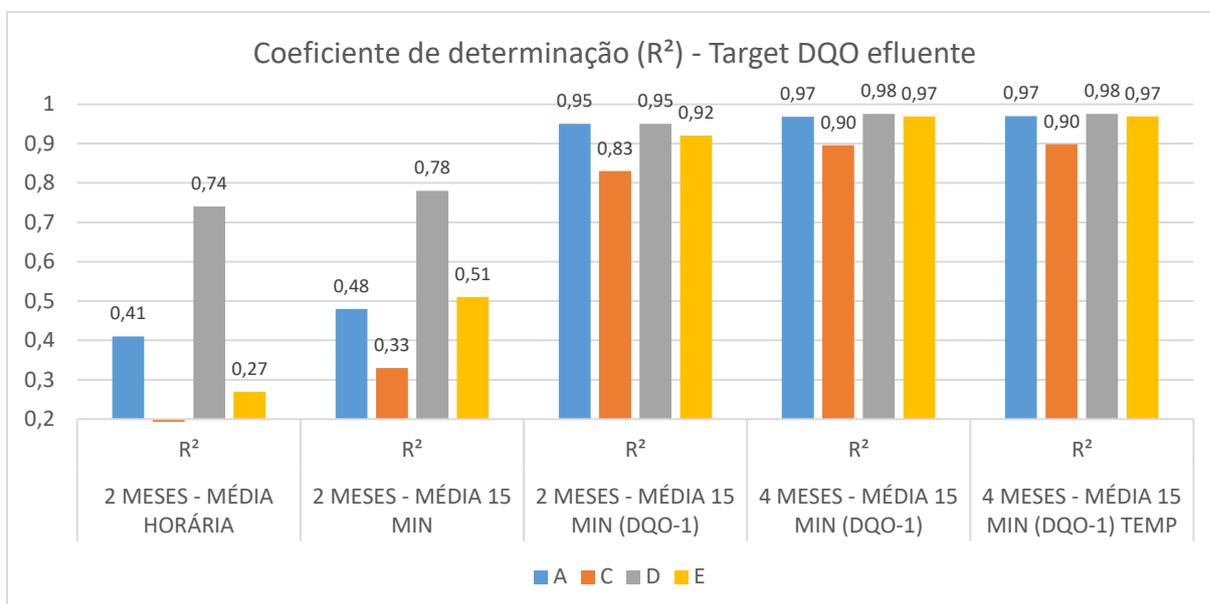
### 5.3.1 Conclusões

As conclusões obtidas a partir das análises realizadas com as RNAs para modelagem da DQO nas ETEs destacam a importância de explorar diferentes configurações de *features* e períodos de dados para otimizar os resultados. A inclusão da própria DQO como uma variável de entrada, com atraso temporal de uma amostra, mostrou-se particularmente relevante, aumentando as métricas gerais em 68,66% e elevando o coeficiente de determinação ( $R^2$ ) em 43,04%. Esse resultado evidencia o impacto positivo do uso de dados históricos para capturar padrões temporais no comportamento da DQO, melhorando significativamente o desempenho preditivo do modelo.

Adicionalmente, foi avaliado o impacto do volume de dados no treinamento. A utilização de médias de quinze minutos ao invés de uma hora, representou um incremento de 23,65% nas métricas gerais das ETEs A, C e D. A comparação entre bases com dois meses e quatro meses de análises revelou um aumento de 4,21% no coeficiente de determinação para

todas as ETEs, e também nas ETEs A e E, um incremento de 39,44% nas métricas em geral. Esses resultados sugerem que bases de dados mais extensas podem enriquecer o modelo, fornecendo maior diversidade de exemplos e melhor representatividade do sistema. Por outro lado, a inclusão da temperatura como *feature* apresentou uma contribuição mais sutil: um aumento de apenas 0,13% no  $R^2$  geral e uma redução de 2,22% nas métricas de erro (MSE, RMSE e MAE). Esses números indicam que, embora a temperatura tenha um efeito marginal no ajuste global do modelo, ela pode ter influência limitada no contexto específico das ETEs analisadas. O Gráfico 24 traz a evolução do  $R^2$  nessas diferentes bases de dados.

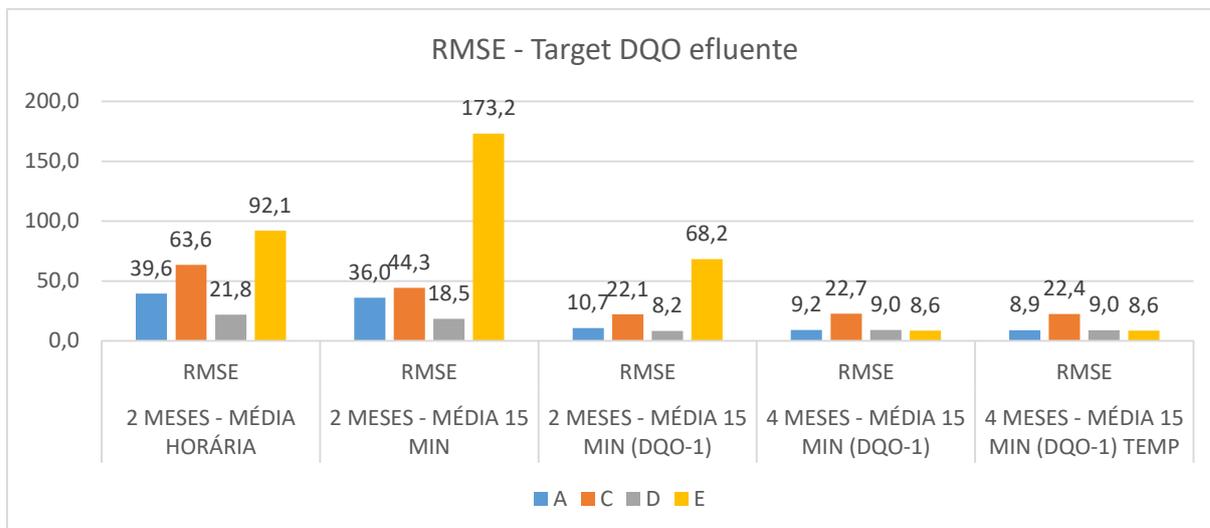
Gráfico 24 - Coeficiente de determinação ( $R^2$ ) - DQO efluente



Fonte: Do autor (2025)

Da mesma maneira, o Gráfico 25 nos mostra a evolução da modelagem em diferentes bases, porém com a informação do RMSE.

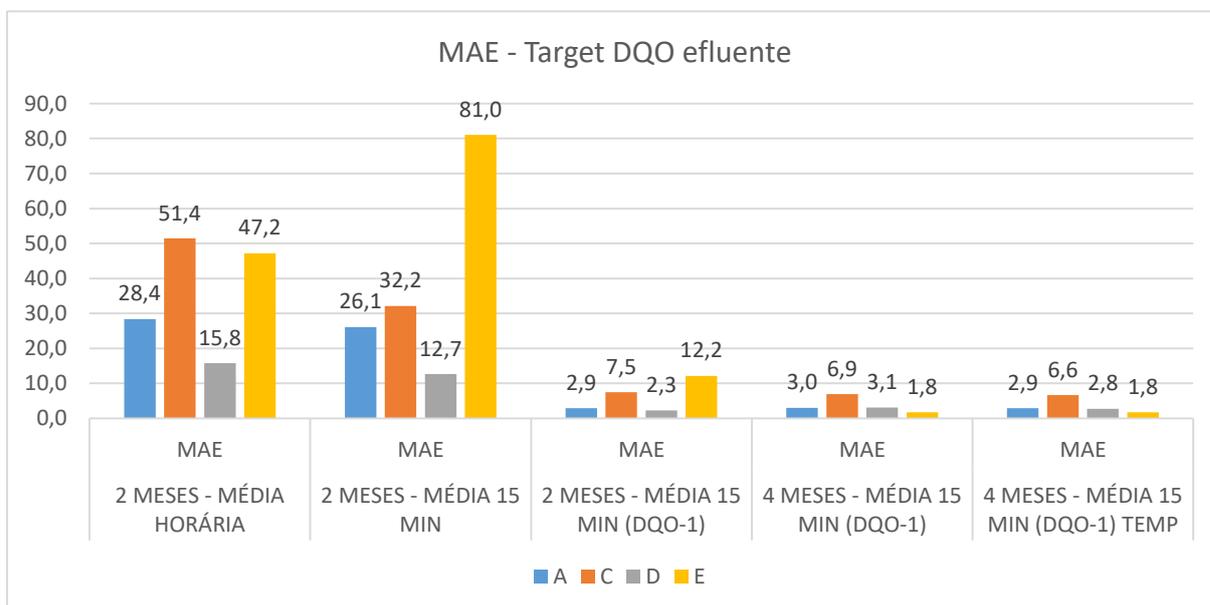
Gráfico 25 - RMSE - DQO efluente



Fonte: Do autor (2025)

E finalmente o MAE entre as bases é representado no Gráfico 26.

Gráfico 26 - MAE - DQO efluente



Fonte: Do autor (2025)

Na Tabela 16 temos as métricas finais obtidas com as bases de dados que reúnem dados de quatro meses, com médias de quinze minutos, inclusão da temperatura, utilização do *Grid Search*, e também da variável de interesse sendo utilizada na entrada com o deslocamento de uma amostra.

Tabela 16 - Métricas obtidas para DQO efluente nas ETEs

|       | MSE    | RMSE  | MAE  | MAPE   | R <sup>2</sup> |
|-------|--------|-------|------|--------|----------------|
| A     | 80,97  | 8,99  | 2,87 | 2,27   | 0,969          |
| C     | 506,24 | 22,49 | 6,75 | 147,65 | 0,8977         |
| D     | 81,59  | 9,03  | 3,09 | 64,39  | 0,9753         |
| E     | 73,51  | 8,57  | 1,7  | 2,48   | 0,9687         |
| Média | 185,58 | 12,27 | 3,6  | 54,20  | 0,95           |

Fonte: Do autor (2025)

Após a finalização das modelagens com variável de interesse a DQO efluente realizadas nesta dissertação, buscou-se contextualizar os resultados obtidos, comparando as métricas de desempenho alcançadas, em especial o coeficiente de determinação (R<sup>2</sup>), com as métricas reportadas na literatura. Para isso, utilizou-se como referência a RSL realizada por Dantas, Christofaro e Oliveira (2023). Este estudo sintetiza diversos trabalhos publicados a partir de 2012 que aplicaram as RNAs no tratamento de esgoto, incluindo a DQO em diferentes contextos e configurações.

A tabela compilada por esses autores apresenta uma visão sobre os métodos utilizados, os parâmetros modelados, o coeficiente de correlação (R) ou coeficiente de determinação (R<sup>2</sup>) alcançadas e os algoritmos de RNA empregados. Tal comparação permite não apenas avaliar a eficiência das modelagens realizadas nesta dissertação, mas também identificar potenciais lacunas ou oportunidades para ajustes nos métodos utilizados. Além disso, a análise comparativa com esses trabalhos busca reforçar a relevância e a aplicabilidade das técnicas empregadas, validando os resultados obtidos em relação ao que já foi pesquisado na literatura.

A seguir, apresenta-se a Tabela 17, extraída e adaptada da RSL de Dantas, Christofaro e Oliveira (2023), que será utilizada como base para a análise comparativa, porém considerando somente as modelagens que tem a DQO efluente como variável de interesse. Alguns dos métodos utilizados para RNA foram: *Extreme Learning Machine* (ELM), *Feedforward Neural Network* (FFNN), *Multilayer Perceptron* (MLP), *Adaptive Neuro-Fuzzy Inference System Coupled With Genetic Algorithm* (ANFIS-GA) e *Long Short-Term Memory Based On Attention Mechanism* (LSTM-AM).

Tabela 17 - Desempenho de modelos com variável de interesse DQO efluente

| Referência                  | Método da RNA | Performance do modelo        |                            |
|-----------------------------|---------------|------------------------------|----------------------------|
|                             |               | Coefficiente de determinação | Coefficiente de correlação |
| Zhao et al. (2012)          | ELM-GA        | $R^2 = 0,7729$               |                            |
| Chen e Lo (2012)            | MLP           |                              | $R = 0,0178$               |
| Jami et al. (2012)          | FFNN          |                              | $R = 0,052622$             |
| Liu et al. (2013)           | ANFIS-GA      | $R^2 = 0,800$                |                            |
| Gaya et al. (2014)          | FFNN          |                              | $R = 0,648$                |
| Bagheri et al. (2015)       | MLP-GA        | $R^2 = 0,98044$              |                            |
| Nourani et al. (2018)       | FFNN          | $R^2 = 0,9363$               |                            |
| Khatri et al. (2019)        | FFNN          |                              | $R = 0,656$                |
| Bekkari e Zeddouri          | FFNN          |                              | $R = 0,8781$               |
| Al-Ghazawi e Alawneh (2021) | FFNN          | $R^2 = 0,45$                 |                            |
| Nourani et al. (2021)       | FFNN          | $R^2 = 0,7178$               |                            |
| Alsulaili e Refaie (2021)   | FFNN          | $R^2 = 0,6115$               |                            |
| Hejabi et al. (2021)        | FFNN          | $R^2 = 0,715$                |                            |
| Liu et al. (2021)           | LSTM-AM       | $R^2 = 0,869$                |                            |
| Saleh (2021)                | FFNN          |                              | $R = 0,77302$              |
| Abba et al. (2021b)         | ELM           | $R^2 = 0,9742$               |                            |

Fonte: Adaptado de DANTAS; CHRISTOFARO; OLIVEIRA, 2023.

Analisando os resultados obtidos pelos autores, é possível observar que o coeficiente de determinação médio foi 0,7827. Foi observado também que maior parte dos trabalhos utilizam o método FFNN. O coeficiente de determinação médio obtido para as quatro ETes modeladas nesta dissertação foi 0,95. Somente 2 autores que utilizaram coeficiente de determinação como métrica de avaliação obtiveram valores maiores. Eles utilizaram os métodos ELM e MLP-GA

#### 5.4 Avaliando a influência do Tempo de Detenção Hidráulica (TDH) das ETes nas modelagens

Na Tabela 18 temos a vazão das cinco plantas que estão sendo estudadas, a vazão total para um dia, a quantidade e volume dos tanques de algumas etapas do tratamento de cada uma dessas plantas, como o decantador primário, tanque de aeração e o decantador secundário. Em posse dessas informações, é possível calcular o TDH entre as etapas e as análises que ocorrem nessas etapas, informação que irá subsidiar o atraso nas TDNNs posteriormente.

Tabela 18 - Cálculo do TDH em cada etapa

| ETE |                             |         |                                   | Decantador primário | Tanque de aeração | Decantador secundário | Total |
|-----|-----------------------------|---------|-----------------------------------|---------------------|-------------------|-----------------------|-------|
| A   | Vazão (l/s)                 | 3129    | Quantidade                        | 4                   | 4                 | 6                     |       |
|     | Vazão (m <sup>3</sup> /dia) | 270346  | Volume unitário (m <sup>3</sup> ) | 4725                | 17595             | 6644                  |       |
|     |                             |         | Volume Total (m <sup>3</sup> )    | 18900               | 70380             | 39864                 |       |
|     |                             |         | TDH (hora)                        | 1,68                | 6,25              | 3,54                  | 11,46 |
| B   | Vazão (l/s)                 | 15588   | Quantidade                        | 16                  | 16                | 24                    |       |
|     | Vazão (m <sup>3</sup> /dia) | 1346803 | Volume unitário (m <sup>3</sup> ) | 5985                | 18500             | 6648                  |       |
|     |                             |         | Volume Total (m <sup>3</sup> )    | 95760               | 296000            | 159552                |       |
|     |                             |         | TDH (hora)                        | 1,71                | 5,27              | 2,84                  | 9,82  |
| C   | Vazão (l/s)                 | 1951    | Quantidade                        |                     | 3                 | 6                     |       |
|     | Vazão (m <sup>3</sup> /dia) | 168566  | Volume unitário (m <sup>3</sup> ) |                     | 18000             | 7500                  |       |
|     |                             |         | Volume Total (m <sup>3</sup> )    | 0                   | 54000             | 45000                 |       |
|     |                             |         | TDH (hora)                        | 0,00                | 7,69              | 6,41                  | 14,10 |
| D   | Vazão (l/s)                 | 1592    | Quantidade                        | 2                   | 4                 | 4                     |       |
|     | Vazão (m <sup>3</sup> /dia) | 137549  | Volume unitário (m <sup>3</sup> ) | 5817                | 8482              | 5817                  |       |
|     |                             |         | Volume Total (m <sup>3</sup> )    | 11634               | 33928             | 23268                 |       |
|     |                             |         | TDH (hora)                        | 2,03                | 5,92              | 4,06                  | 12,01 |
| E   | Vazão (l/s)                 | 977     | Quantidade                        | 4                   | 3                 | 4                     |       |
|     | Vazão (m <sup>3</sup> /dia) | 84413   | Volume unitário (m <sup>3</sup> ) | 1800                | 5760              | 5344                  |       |
|     |                             |         | Volume Total (m <sup>3</sup> )    | 7200                | 17280             | 21376                 |       |
|     |                             |         | TDH (hora)                        | 2,05                | 4,91              | 6,08                  | 13,04 |

Fonte: Do autor (2025)

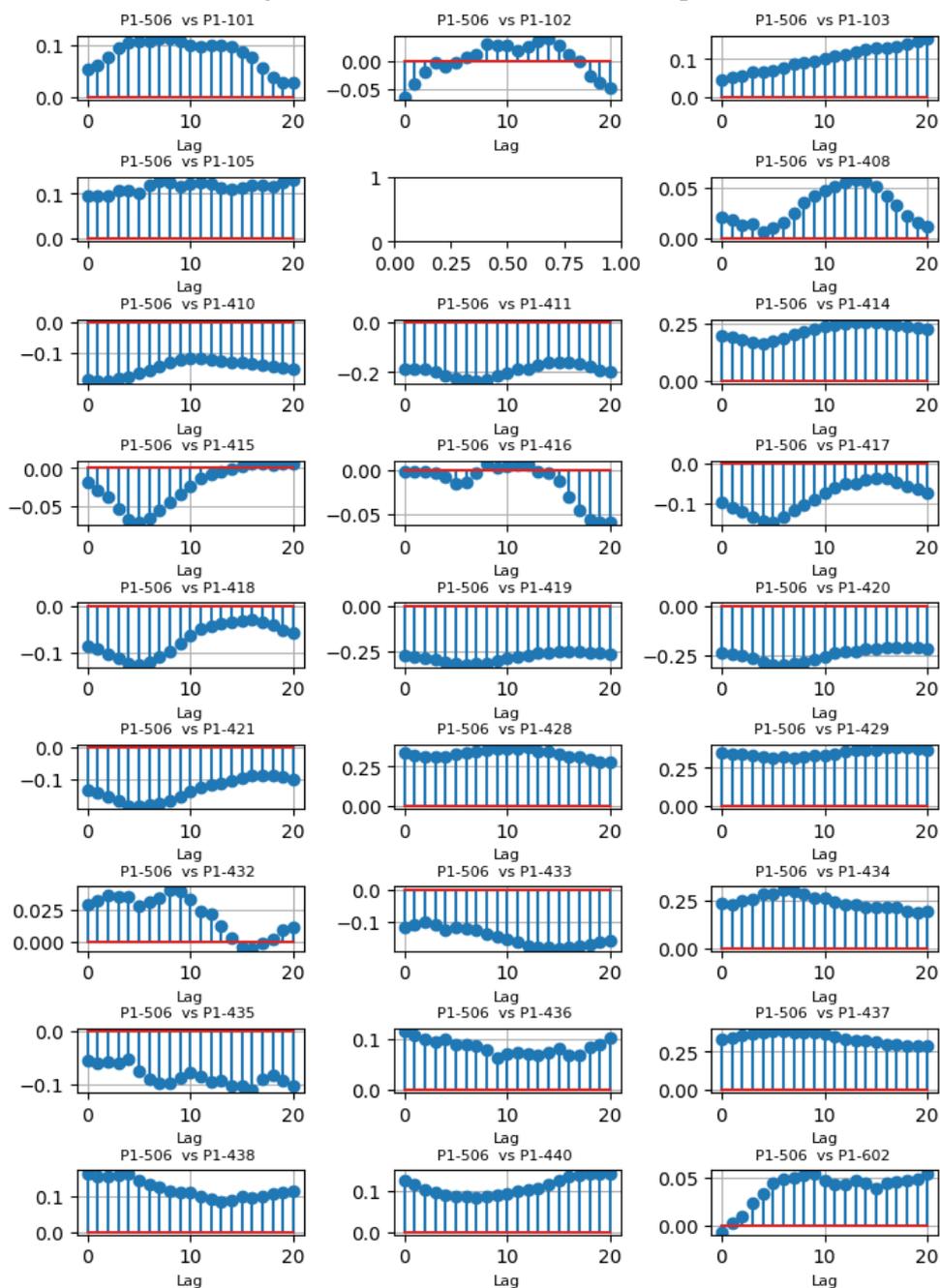
#### 5.4.1 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE A

A seguir, é apresentado um gráfico com a correlação cruzada entre a DQO do efluente final e os demais parâmetros monitorados em cada planta, buscando identificar se há variação na correlação em função do atraso temporal (*delay*) e em qual momento essa correlação atinge seu pico máximo. Com base nos gráficos de correlação cruzada e nos atrasos de pico observados para cada parâmetro em relação à DQO no efluente final (P1-506) da ETE A, é possível notar padrões distintos que sugerem diferentes comportamentos ao longo do tempo, Figura 43.

Os *lags* de pico variam consideravelmente, indicando que alguns parâmetros têm influência imediata na DQO (*lag* = 0), enquanto outros têm correlação maior após alguns períodos. Por exemplo, o pH do esgoto bruto (P1-101) apresenta um *lag* de pico em 8, sugerindo que o efeito desse parâmetro na DQO do efluente final se manifesta de maneira mais significativa após esse intervalo de tempo.

Em resumo, a análise dos *lags* de pico revela uma interdependência temporal complexa entre a DQO no efluente final e os parâmetros ao longo do processo de tratamento. Essas informações são essenciais para ajustar estratégias operacionais e prever com mais precisão a qualidade do efluente, utilizando técnicas como as TDNNs, que podem incorporar esses lags como variáveis temporais para modelar de forma mais robusta a influência de cada parâmetro na DQO.

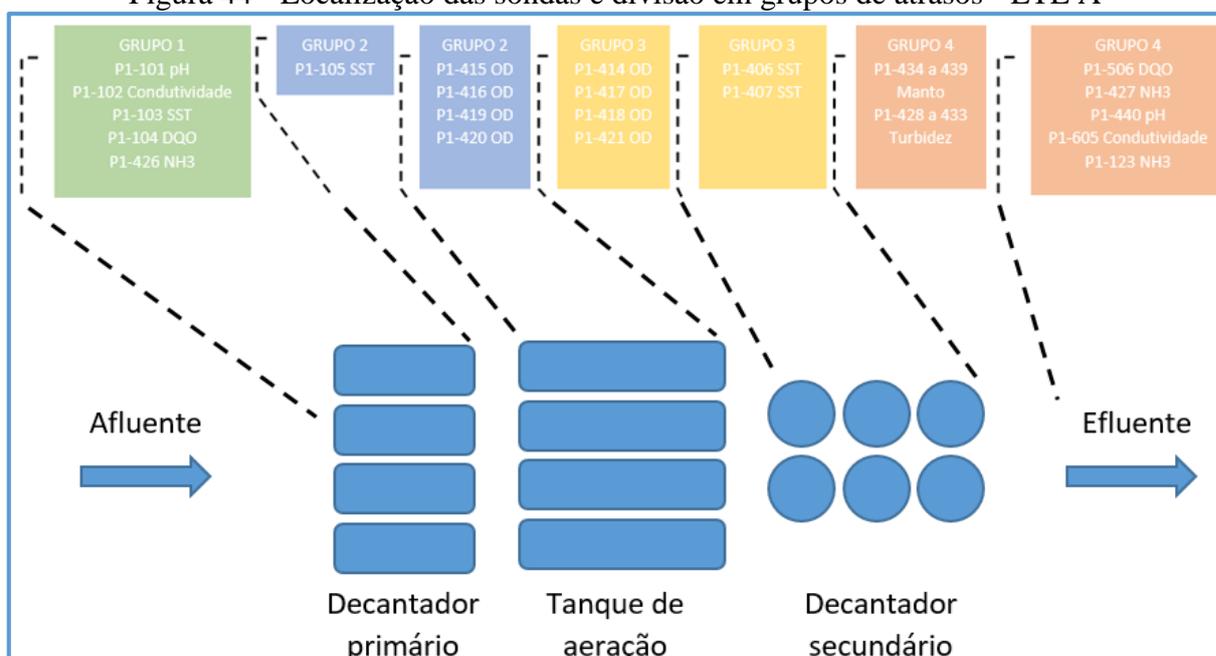
Figura 43 - ETE A - Correlação da DQO efluente com outros parâmetros incluindo atrasos



Fonte: Do autor (2025)

Essa diferença na correlação entre pares de dados e sobretudo a identificação de que o pico dessa correlação ocorre em momentos que não exatamente os dados são analisados na mesma janela de tempo, corrobora o conceito de que o TDH tem relação com as análises que são realizadas na ETE e podem identificar esse momento de maior correlação. Na Figura 44 estão relacionadas as análises que ocorrem na fase líquida da ETE A, e a divisão em grupos de atrasos em relação a DQO do efluente final, ou seja, uma análise que ocorre no afluente ao decantador primário, terá o “lag” ajustado com o TDH total calculado anteriormente para essa ETE. Dessa mesma forma, uma análise que ocorre efluente ao decantador primário, terá o atraso ajustado para o TDH total da ETE menos o TDH calculado para o decantador primário, e assim por diante. A linha tracejada na imagem direciona onde o grupo de instrumentos está localizado na ETE (entrada/saída de determinada etapa do processo).

Figura 44 - Localização das sondas e divisão em grupos de atrasos - ETE A



Fonte: Do autor (2025)

Na Tabela 19 é possível observar o TDH obtido para cada um desses grupos de sondas do tratamento fase líquida que será ajustado nas TDNNs.

Tabela 19 - ETE A - TDH dos grupos definidos  
ETE A

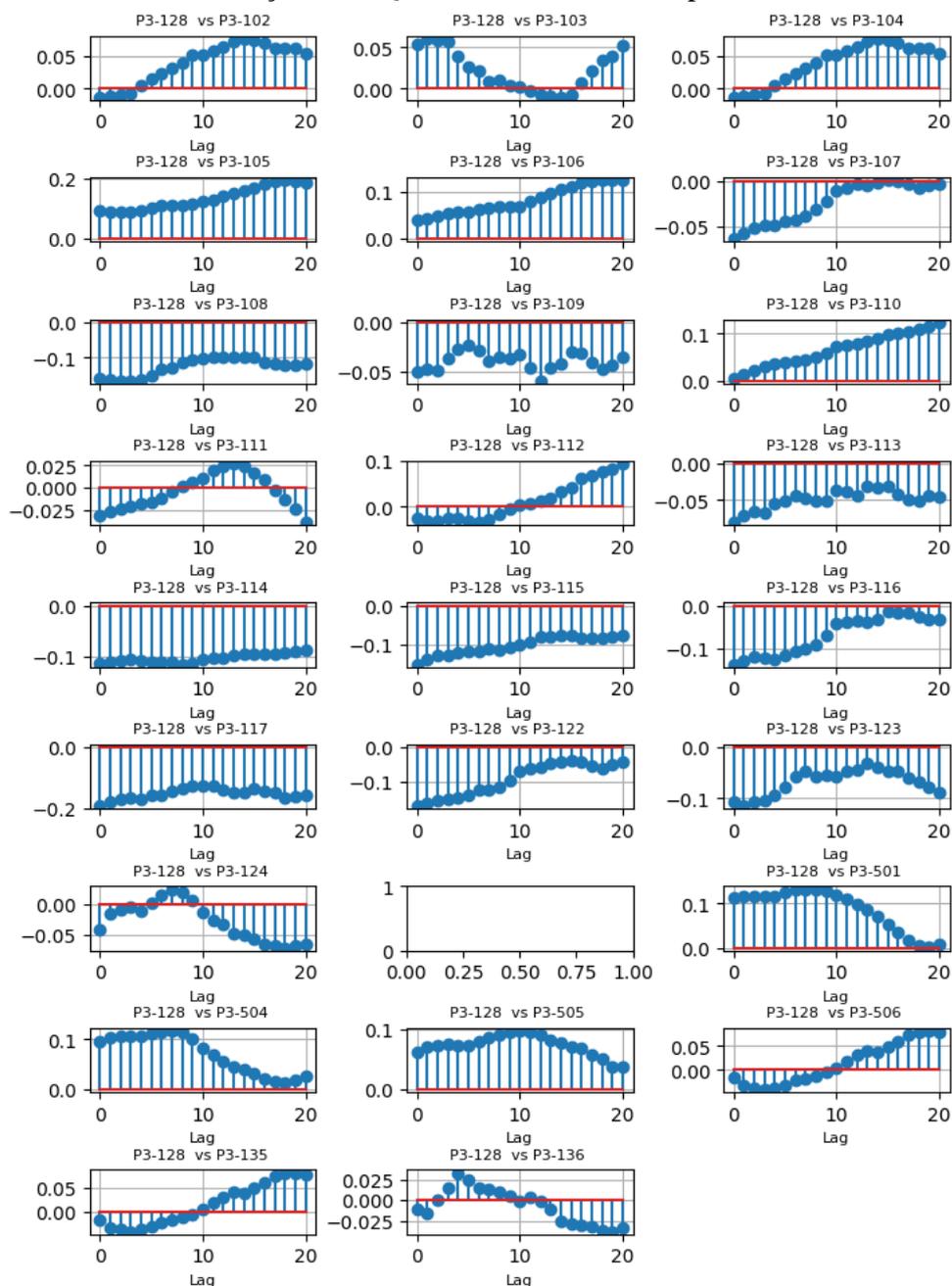
|         | Atraso segundo TDH (horas) |
|---------|----------------------------|
| Grupo 1 | 11,46                      |
| Grupo 2 | 9,79                       |
| Grupo 3 | 3,54                       |
| Grupo 4 | 0                          |

Fonte: Do autor (2025)

#### 5.4.2 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE C

Da mesma maneira, também é possível identificar a diferença do pico de correlação entre a DQO e os outros parâmetros na ETE C a depender do atraso observado, Figura 45:

Figura 45 - ETE C - Correlação da DQO efluente com outros parâmetros incluindo atrasos



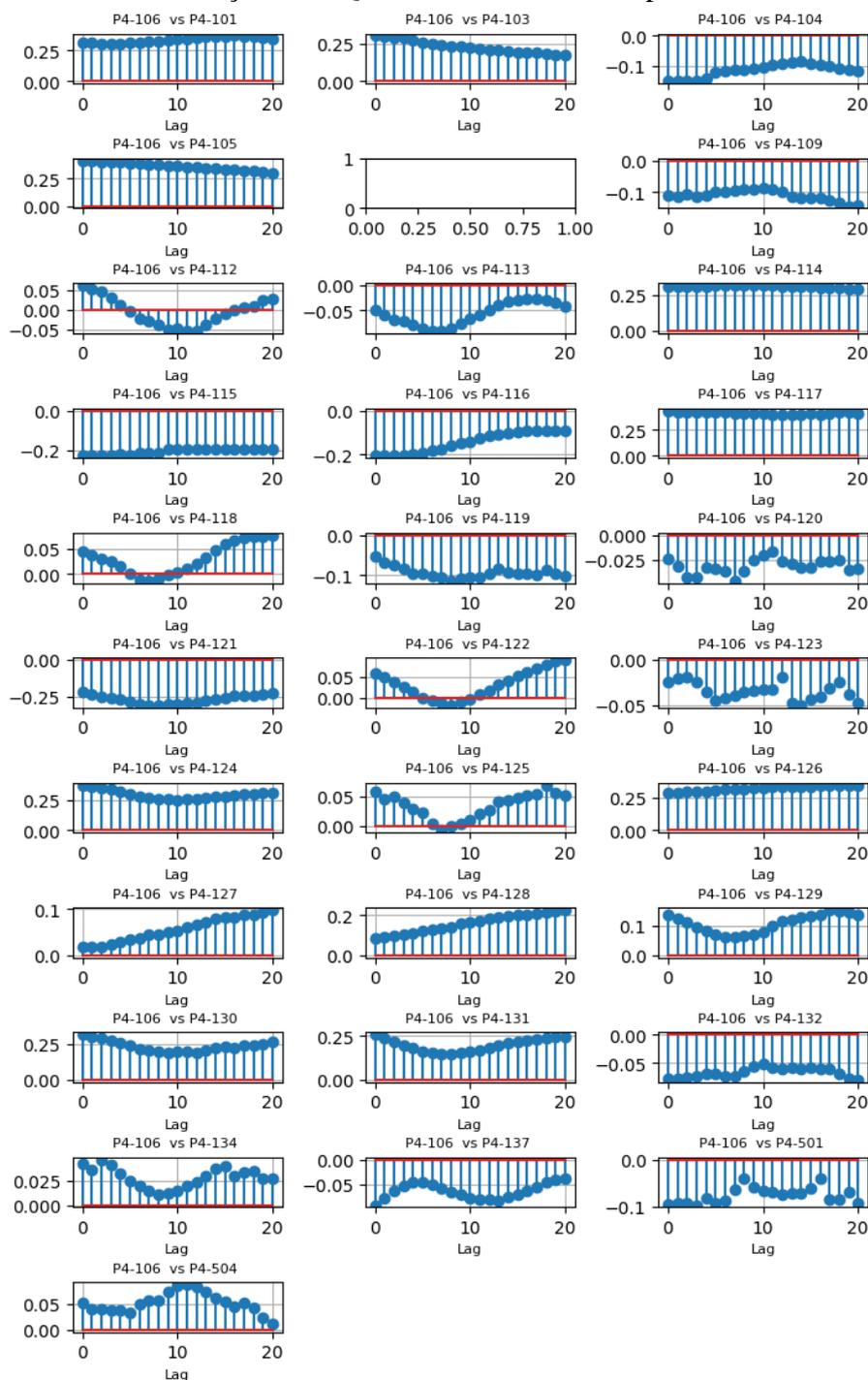
Fonte: Do autor (2025)

#### 5.4.3 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE D

No caso da ETE D também é possível observar diferença entre as correlações nos

atrasos, Figura 46:

Figura 46 - ETE D - Correlação da DQO efluente com outros parâmetros incluindo atrasos

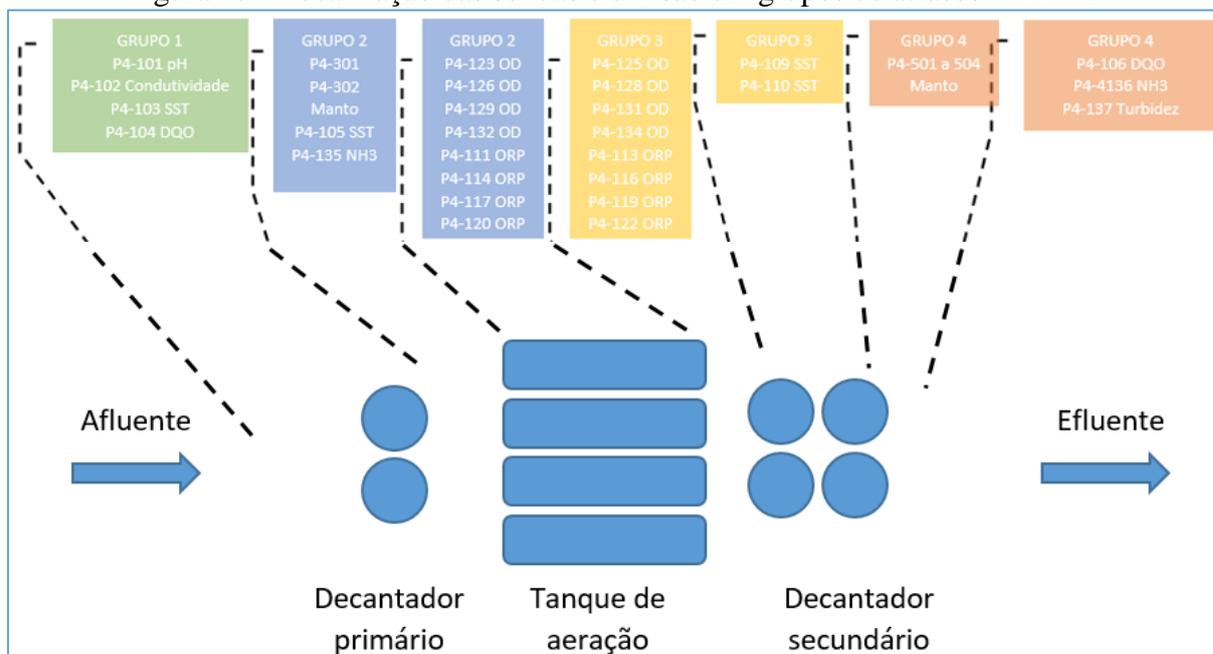


Fonte: Do autor (2025)

Da mesma forma que ocorre com as outras plantas, as análises também foram divididas em grupos para ajuste do TDH entre análises. Importante ressaltar que para determinação do TDH entre essas etapas foi utilizada a vazão média de 2023 e o volume total das Unidades de tratamento, duas informações que provavelmente podem identificar melhores valores de

correlação, mas podem sofrer alteração, no caso da vazão essa alteração é muito comum, e no volume das Unidades no caso da indisponibilidade de alguma delas durante o período, que para essa pesquisa não foi o intuito o levantamento. Esse levantamento e valores definidos podem ser vistos na Figura 47 e Tabela 20:

Figura 47 - Localização das sondas e divisão em grupos de atrasos - ETE D



Fonte: Do autor (2025)

Tabela 20 - ETE D - TDH dos grupos definidos

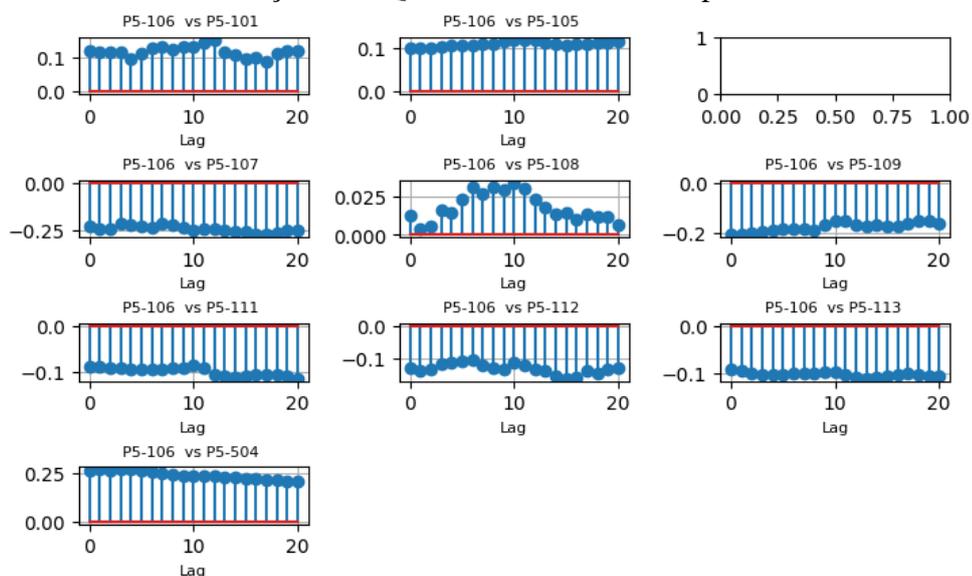
| ETE D                      |       |
|----------------------------|-------|
| Atraso segundo TDH (horas) |       |
| Grupo 1                    | 12,01 |
| Grupo 2                    | 9,98  |
| Grupo 3                    | 4,06  |
| Grupo 4                    | 0     |

Fonte: Do autor (2025)

#### 5.4.4 Tempo de Detenção Hidráulica e relação entre DQO e outras análises - ETE E

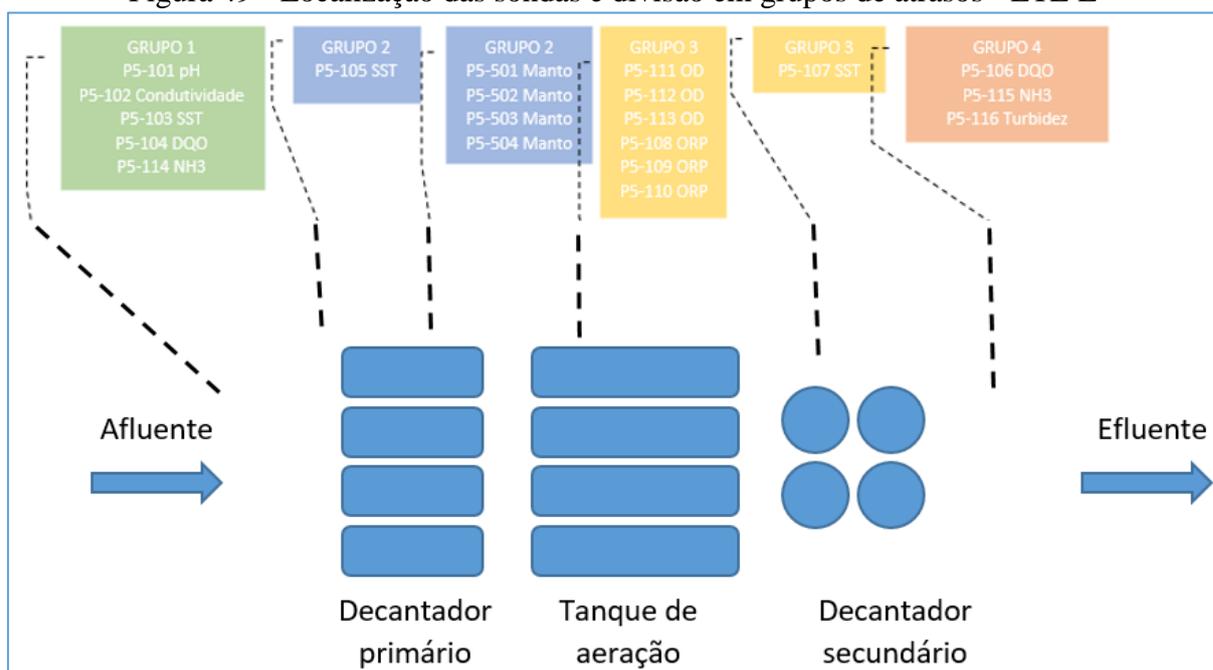
Por fim, temos a análise da influência do TDH na modelagem da DQO também na ETE. A Figura 48 traz a correlação da DQO com as outras análises correlacionando os “lags” temporais, a Figura 49 o agrupamentos das sondas pelas localizações para facilitar a inclusão do TDH e a Tabela 21 com os TDHs para os respectivos grupos. Essas informações visam preparar a modelagem para incluir o TDH assim como ocorreu nas outras ETES.

Figura 48 - ETE E - Correlação da DQO efluente com outros parâmetros incluindo atrasos



Fonte: Do autor (2025)

Figura 49 - Localização das sondas e divisão em grupos de atrasos - ETE E



Fonte: Do autor (2025)

Tabela 21 - ETE E - TDH dos grupos definidos

| ETE E                      |       |
|----------------------------|-------|
| Atraso segundo TDH (horas) |       |
| Grupo 1                    | 13,04 |
| Grupo 2                    | 10,99 |
| Grupo 3                    | 6,08  |
| Grupo 4                    | 0     |

Fonte: Do autor (2025)

#### 5.4.5 Resultados com a inclusão do TDH na modelagem

Após realizar novamente as modelagens, agora incluindo os atrasos calculados entre as análises, foi possível observar alguns avanços, embora com algumas ressalvas. A primeira observação é que a modelagem foi conduzida sem incluir a variável de interesse como uma *feature* com atraso temporal de uma amostra, uma vez que as métricas apresentaram baixo potencial de melhoria nesse caso. O segundo ponto é que, das três ETEs modeladas com TDNNs (ETEs A, D e E), apenas as ETEs D e E apresentaram melhorias. Essas melhorias foram de aproximadamente 2,58% nas métricas MSE, RMSE e MAPE, além de um incremento de cerca de 1,94% no  $R^2$  para as mesmas ETEs.

Com relação ao fato de não ser observado melhorias na ETE A, ou mesmo incrementos maiores na métricas analisadas nas ETE D e E, diversas podem ser as causas prováveis. Primeiramente, pode ser citado que a vazão utilizada para o cálculo do TDH nas plantas, é uma vazão média, que conseqüentemente varia para o período e poderia resultar em um TDH diferente para diferentes períodos analisados. Segundo, que para o processo de preparação dos dados, conforme visto, diversas etapas de limpeza dos dados são necessárias, como a retirada de amostras com dados faltantes, e isso incorre que a retirada dessas linhas no banco de dados, mesmo que em poucos momentos, resulta na diminuição das relações horárias entre os dados de uma sonda e outra.

### 5.5 Limites de parâmetros segundo especialista

Na Tabela 22 é possível observar limites e pontos de ajustes estabelecidos para cada parâmetro obtidos em conjunto com um especialista em tratamento de esgoto, visando nortear etapas futuras do trabalho.

Tabela 22 - Set-points esperados para parâmetros analíticos de acordo com especialista

| Parâmetro        | Local                   | Limites ou set-point desejado? |
|------------------|-------------------------|--------------------------------|
| pH               | Esgoto desarenado       | 6,5 a 7,5                      |
| Condutividade    | Esgoto desarenado       | 450 a 1.600 $\mu$ S/cm         |
| SST (imersão)    | Esgoto desarenado       | 120 a 280 mg/L                 |
| DQO              | Esgoto desarenado       | 400 a 500 mg/L                 |
| SST (imersão)    | Efluente Primário       | 100 a 180 mg/L                 |
| DQO              | Efluente Final          | 24 a 60 mg/L                   |
| Manta de Lodo    | Decantador Prim.        | $\leq$ 1,5 m                   |
| Manta de Lodo    | Decantador Prim.        | $\leq$ 1,5 m                   |
| SST (imersão)    | Licor misto             | 1.500 a 3.500 mg/L             |
| SST (imersão)    | Lodo de Retorno         | 3.000 a 5.000 mg/L             |
| ORP (Zona 1)     | Reator aeróbio 1, 2 e 3 | + 80 a + 180 mV                |
| ORP (Zona 2)     | Reator aeróbio 1, 2 e 3 | + 30 a + 100 mV                |
| ORP (Zona 3)     | Reator aeróbio 1, 2 e 3 | + 100 a + 300 mV               |
| OD (Zona 1)      | Reator aeróbio 1, 2 e 3 | 1 a 3 mg/L                     |
| OD (Zona 2)      | Reator aeróbio 1, 2 e 3 | 0,1 a 2 mg/L                   |
| OD (Zona 3)      | Reator aeróbio 1, 2 e 3 | 1 a 3 mg/L                     |
| NH3              | Efluente Primário       | Até 30 mg/L                    |
| NH3              | Licor misto             | Até 25 mg/L                    |
| Turbidez         | Efluente tratado        | Até 10 NTU                     |
| Manto de Lodo    | Decantador DS1          | $\leq$ 50 cm                   |
| Manto de Lodo    | Decantador DS2          | $\leq$ 50 cm                   |
| Manto de Lodo    | Decantador DS3          | $\leq$ 50 cm                   |
| Manto de Lodo    | Decantador DS4          | $\leq$ 50 cm                   |
| SST (imersão)    | Lodo Aflu. Digestor     | 23.000 a 38.000 mg/L           |
| SST (imersão)    | Lodo condicionado       | 23.000 a 35.000 mg/L           |
| SST (imersão)    | Filtrado FPP 1 a 3      | Até 220 mg/L                   |
| NH3              | Filtrado FPP 1 a 3      | Até 350 mg/L                   |
| Condutividade    | Filtrado FPP 1 a 3      | 200 a 1.500 $\mu$ S/cm         |
| Cloro (inserção) | ETA Reuso               | >1 e <10 mg/L                  |
| Turbidez         | ETA Reuso               | Até 10 NTU                     |

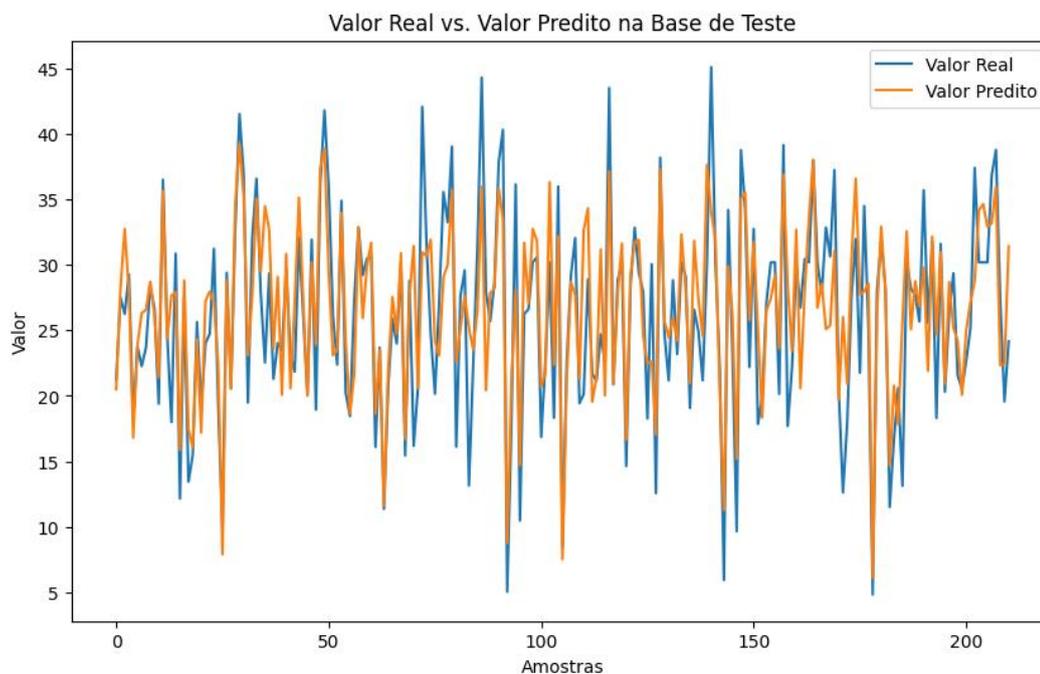
Fonte: Do autor (2025)

## 5.6 Modelagem com amônia por meio de parâmetros em elevatória afluente

Nesse momento, foi realizada a modelagem agora não só considerando as análises realizadas na mesma Estação que a variável de interesse, mas também levando em conta valores em elevatórias de esgoto antes da planta. Por exemplo, com interesse de identificar cargas tóxicas na ETE A, através da análise de Amônia, condutividade e DQO afluente, foi realizada

a modelagem com os dados de pH e condutividade nas elevatórias a montante, somados aos dados de vazão e nível de alguns desses locais. Os resultados podem ser vistos no Gráfico 27:

Gráfico 27 - Amônia afluente modelada vs. real – Jan./2024



Fonte: Do autor (2025)

As métricas utilizadas novamente mostram a correlação entre esses parâmetros, sobretudo para análise de Amônia, conforme Tabela 23:

Tabela 23 - Métricas de desempenho da RNA para dados afluentes

|                | Amônia |
|----------------|--------|
| MSE            | 18,60  |
| RMSE           | 4,31   |
| MAE            | 3,26   |
| MAPE           | 14,79  |
| R <sup>2</sup> | 0,68   |

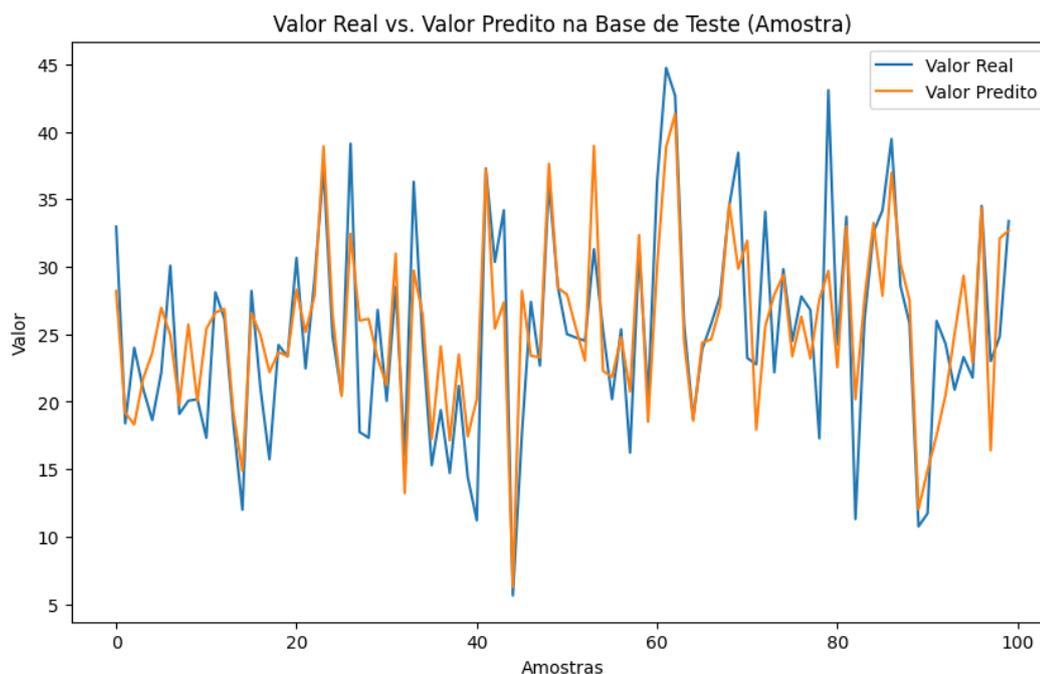
Fonte: Do autor (2025)

### 5.7 Modelagem da amônia afluente com atraso

A partir desse ponto foi considerado também o tempo de deslocamento do esgoto entre esses locais, a fim de poder realizar a modelagem com as variáveis explicativas deslocadas em seu registro horário. Isso aconteceria por exemplo, considerando os dados das elevatórias desse instante, e de uma terceira elevatória de 7h42m atrás e da variável de interesse considerando 42m a frente, devido à distância e velocidade do fluido na rede calculados pelo especialista

local. Essas novas características apresentadas poderiam trazer um caráter preditivo a modelagem e ao ser colocado em produção, estimar o valor de Amônia na Estação com 40 minutos de antecedência. Seguem os resultados no Gráfico 28:

Gráfico 28 - Amônia afluente modelada vs. real – com previsibilidade de 1 hora – Jan./2024 - Média horária



Fonte: Do autor (2025)

Tabela 24 - Métricas de desempenho da RNA para dados afluentes com previsibilidade Amônia (Dados por minuto) – Sem outliers

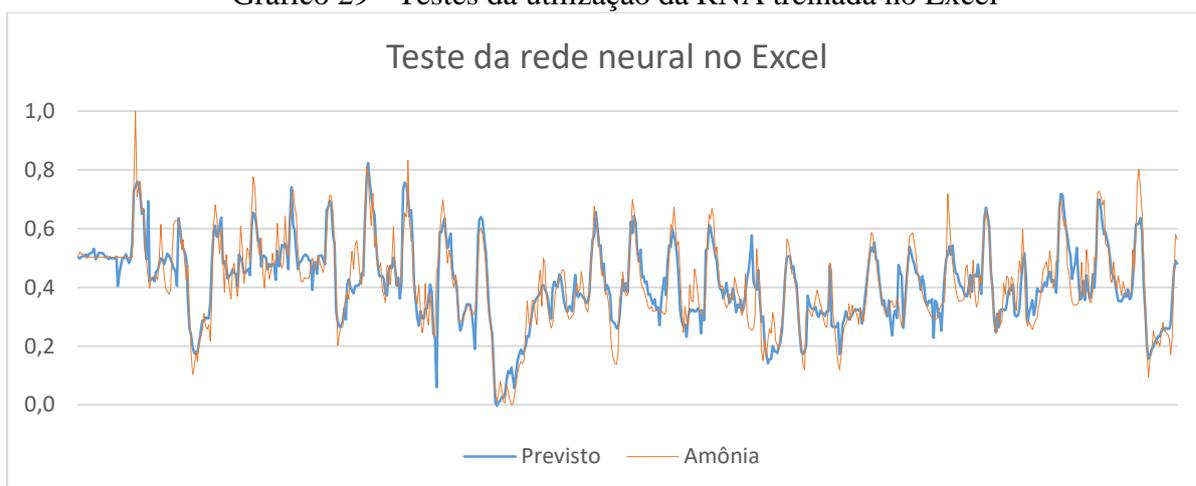
|                |       |
|----------------|-------|
| MSE            | 14,86 |
| RMSE           | 3,85  |
| MAE            | 2,92  |
| MAPE           | 14,67 |
| R <sup>2</sup> | 0,77  |

Fonte: Do autor (2025)

O Gráfico 29 ilustra os resultados obtidos ao testar a RNA previamente treinada, cujos pesos e vieses foram aplicados manualmente em fórmulas no Excel. O processo consistiu em replicar as operações realizadas pela RNA para verificar se o modelo continuava a fazer sentido com os mesmos parâmetros utilizados durante o treinamento. Primeiramente, os pesos das conexões foram multiplicados pelos valores de entrada correspondentes com os dados normalizados, e os resultados somados, incluindo o viés (bias) associado a cada neurônio. Em seguida, o resultado dessa soma foi submetido a uma função de ativação, que no caso foi a

sigmoide, com o objetivo de limitar ou transformar os valores para a próxima camada. Reproduzindo essa sequência para todas as camadas da RNA, foi possível obter as saídas do modelo diretamente no Excel. O gráfico, portanto, compara as saídas obtidas nesse teste com os valores esperados ou reais, permitindo avaliar a consistência do modelo ao replicar manualmente as operações da RNA. Essa abordagem auxilia na validação do modelo e na compreensão detalhada de seu funcionamento interno.

Gráfico 29 - Testes da utilização da RNA treinada no Excel



Fonte: Do autor (2025)

Para o teste no Excel, basta inserir os pesos que foram obtidos pelo treinamento da RNA nas fórmulas, multiplicar pelos parâmetros analíticos normalizados e somar com o viés ou bias conforme a eq. (14), para cada neurônio.

$$\begin{aligned}
 &= ('Pesos e vieses'!C\$3 * \$C3) + ('Pesos e vieses'!C\$4 * \$D3) + \\
 &('Pesos e vieses'!C\$5 * \$E3) + ('Pesos e vieses'!C\$6 * \$F3) + ('Pesos e vieses'!C\$7 * \\
 & \$G3) + ('Pesos e vieses'!C\$8 * \$H3) + ('Pesos e vieses'!C\$9 * \$I3) + \\
 &('Pesos e vieses'!C\$10 * \$J3) + ('Pesos e vieses'!C\$11 * \$K3) + \\
 &('Pesos e vieses'!C\$12 * \$L3) + 'Pesos e vieses'!C\$2
 \end{aligned} \tag{14}$$

## 5.8 Testes da RNA treinada no sistema supervisório com novos dados

Identificado que o modelo respondeu bem após os cálculos matemáticos foi possível então iniciar os testes e implantação no Sistema Supervisório Elipse E3, utilizado na ETE objeto de estudo, conforme é possível observar na Figura 50.

Figura 50 - Testes da utilização da RNA treinada no supervisor Elipse E3

Aplicação E3 Viewer - Screen Title

### Testes com Redes Neurais Artificiais (RNA) no Elipse E3



| NEURÔNIO 1 |            | NEURÔNIO 2 |             | NEURÔNIO 3 |            | NEURÔNIO 4 |             | NEURÔNIO 5 |            | NEURÔNIO 6 |            |
|------------|------------|------------|-------------|------------|------------|------------|-------------|------------|------------|------------|------------|
| BIAS       | -0,9273869 | BIAS       | -3,72792447 | BIAS       | -1,5770182 | BIAS       | -0,7262368  | BIAS       | -1,454445  | BIAS       | -6,26076   |
| w11        | 0,9200179  | w21        | -1,07467879 | w31        | 5,8113062  | w41        | 6,664697    | w51        | -3,8667377 | w61        | -2,553677  |
| w12        | 4,0974762  | w22        | -6,32044141 | w32        | 6,0877358  | w42        | 16,0353445  | w52        | 31,8543794 | w62        | 15,385867  |
| w13        | 0,3719638  | w23        | -0,67296048 | w33        | 4,5294133  | w43        | 0,203535    | w53        | -3,6148446 | w63        | 4,436913   |
| w14        | 1,0491108  | w24        | -0,34823274 | w34        | 5,4145986  | w44        | 1,2340347   | w54        | -5,4081072 | w64        | -23,974697 |
| w15        | -1,5498669 | w25        | -0,14361207 | w35        | -1,071306  | w45        | -14,3529228 | w55        | -0,3016867 | w65        | -8,91391   |
| w16        | 1,1275334  | w26        | -0,48772559 | w36        | -1,3747883 | w46        | -2,8903923  | w56        | 2,9679507  | w66        | 3,536966   |
| w17        | 2,4567398  | w27        | -1,06336451 | w37        | -2,2092813 | w47        | -1,4685416  | w57        | 4,7060548  | w67        | 9,328852   |
| w18        | 2,2162638  | w28        | 3,8940417   | w38        | -1,3164816 | w48        | 2,1688343   | w58        | 1,0412048  | w68        | 6,076109   |
| w19        | 2,5899357  | w29        | 1,02291404  | w39        | 0,2742533  | w49        | 2,1792473   | w59        | 5,0886751  | w69        | 2,322738   |
| w110       | 0,2136627  | w210       | -0,06005336 | w310       | 0,2444713  | w410       | 1,6379652   | w510       | -2,00866   | w610       | -3,974607  |

**Camada 1**

| NEURÔNIO 1 |             | NEURÔNIO 2 |            | NEURÔNIO 3 |            | SAÍDA |            | ENTRADAS |             |
|------------|-------------|------------|------------|------------|------------|-------|------------|----------|-------------|
| BIAS       | -0,5466747  | BIAS       | 0,6205481  | BIAS       | 0,4208446  | BIAS  | -1,7504422 | AIT1     | 0           |
| w71        | 0,769088100 | w81        | -2,5728524 | w91        | 0,5534176  | ws1   | -0,5208933 | AIT2     | 0,546579505 |
| w72        | -2,8984635  | w82        | 4,0457874  | w92        | 13,5236412 | ws2   | 1,4108676  | AIT3     | 7,23078E-06 |
| w73        | -0,4667597  | w83        | 1,6445564  | w93        | 2,0910044  | ws3   | 1,495028   | AIT4     | 0,193508691 |
| w74        | -6,6582836  | w84        | -0,8975276 | w94        | -2,419711  |       |            | AIT5     | 0,613390405 |
| w75        | -2,8708379  | w85        | 0,8088837  | w95        | 2,7552995  |       |            | AIT6     | 0,995397393 |
| w76        | 5,5608949   | w86        | 2,2141014  | w96        | -3,1061401 |       |            | AIT7     | 0,061608825 |
|            |             |            |            |            |            |       |            | AIT8     | 4,98496E-05 |
|            |             |            |            |            |            |       |            | AIT9     | 0,096822595 |
|            |             |            |            |            |            |       |            | AIT10    | 0,072345733 |

**Camada 2**

**SAÍDA DOS NEURÔNIOS**

|    |             |
|----|-------------|
| N1 | 0,89134792  |
| N2 | 0,000411976 |
| N3 | 0,663893152 |
| N4 | 0,04051403  |
| N5 | 0,999999985 |
| N6 | 0,019358586 |

Euler

Min

MaxMin

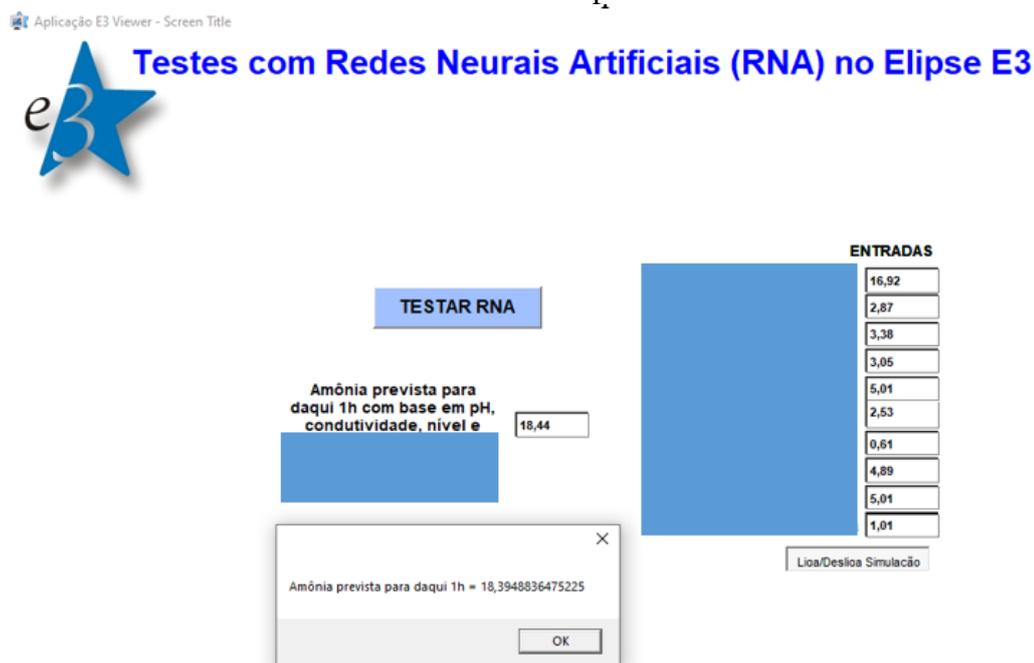
Neuronios = 30,1481844554534

TESTAR RNA

Fonte: Do autor (2025)

Por fim, o cálculo pode ser realizado de maneira automática, sendo necessário somente a troca das variáveis de entrada, que nesse exemplo estão sendo simuladas, por dados reais das Estações Elevatórias, que hoje chegam no Supervisor (Figura 51).

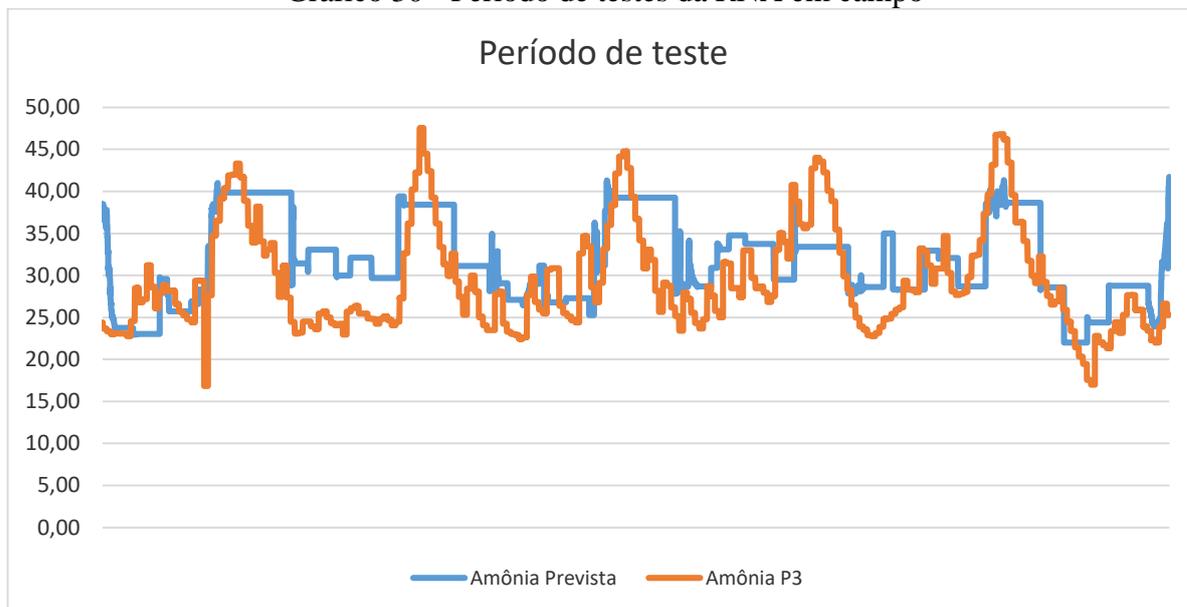
Figura 51 - Testes da utilização da RNA treinada no supervisor Elipse E3 com pesos e vieses no script



Fonte: Do autor (2025)

Após os testes realizados no Excel, onde os pesos e vieses da RNA foram validados manualmente, o modelo foi implementado no supervisor para operar em um ambiente de produção. A partir dessa etapa, a RNA foi submetida a um período de testes contínuos utilizando dados novos, gerados em tempo real pelo processo monitorado. O objetivo foi avaliar o desempenho do modelo frente a cenários dinâmicos, verificando sua capacidade de generalização e consistência ao prever os resultados esperados com base nos padrões dos novos dados.

Gráfico 30 - Período de testes da RNA em campo



Fonte: Do autor (2025)

O Gráfico 30 apresenta os resultados obtidos nesse período de testes, destacando o comportamento da RNA em comparação aos valores reais do processo, evidenciando possíveis ajustes ou melhorias necessários para sua integração definitiva ao sistema operacional.

Um levantamento sobre o cenário de automação e suas componentes (instrumentação, controle, supervisão e comunicação) em cada etapa do processo de diferentes tipos de plantas, foi realizado visando mapear essa temática nas estações objeto de estudo, de acordo com a literatura. Além disso um *dashboard* foi proposto para visualização desse levantamento de maneira holística. Contudo, por não ser objetivo dessa dissertação, e caso tivesse essa intenção necessitar de estudos complementares, esse material acabou não fazendo parte de sua versão final, porém consta como apêndice II. Outro fato importante é que esse levantamento mostrou que a aplicação prática de maneira mais ampla das RNAs no tratamento de esgoto, encontraria gargalos nessas outras camadas da pirâmide de automação (controle, elementos de controle, supervisão).

## 6 CONCLUSÕES E PESQUISAS FUTURAS

A dissertação apresentada teve como objetivo principal a análise e otimização de processos em ETEs, com foco nos parâmetros de DQO e amônia. Através da aplicação de técnicas de análise de dados, como RNA e metodologias estatísticas, foi possível explorar e avaliar a eficiência dos processos em algumas estações.

A análise bibliométrica sobre o tema, ofereceu uma visão ampla do estado atual da pesquisa e apontando tendências e lacunas na aplicação da transformação digital no tratamento de esgoto. Este levantamento foi importante para embasar a análise prática e alinhar a dissertação às pesquisas mais relevantes. Já a revisão da literatura, de sete trabalhos sobre a aplicação de redes neurais no tratamento de esgoto, permitiu a comparação entre diferentes metodologias e resultados, consolidando o uso das RNAs como ferramenta na otimização do tratamento de esgoto. Foi possível identificar que poucos trabalhos fazem uso de maneira prática dessa técnica no setor de saneamento e quando ocorre é fora do país.

Posteriormente, apresentou-se uma análise qualitativa detalhada das entrevistas conduzidas com especialistas do setor. Utilizando a técnica de análise de conteúdo segundo Bardin e softwares como o MAXQDA, foram geradas nuvens de palavras que destacaram as principais convergências nas respostas dos especialistas, fornecendo uma visão prática sobre os desafios e oportunidades da transformação digital no setor. Já a pesquisa realizada com profissionais do setor de saneamento apresenta dados quantitativos e qualitativos sobre suas percepções. As nuvens de palavras geradas a partir das respostas permitiram uma interpretação visual das principais preocupações e expectativas desses profissionais. Ambas pesquisas demonstraram que o setor anseia por dados e ideias geradas por eles, porém identificam desafios relacionados ao estado atual dos equipamentos e infraestrutura para receber esse salto tecnológico.

A aplicação das RNAs para a previsão de parâmetros importantes, como a DQO, nas ETEs, mostraram ser uma ferramenta para otimizar os processos operacionais dessas estações. Os gráficos SHAP, por exemplo, foram essenciais para identificar as variáveis mais influentes na predição da DQO, fornecendo uma visão clara das relações entre as variáveis de entrada e a variável de saída. Através da comparação de algumas métricas, como o erro quadrático médio

(MSE), erro absoluto médio (MAE) e coeficiente de determinação ( $R^2$ ), foi possível realizar o *benchmarking* dos modelos, mostrando variações consideráveis entre as estações. Essas diferenças evidenciam a complexidade dos processos de tratamento de esgoto e a necessidade de ajustes e refinamentos contínuos na modelagem, adaptando-a ao contexto específico de cada ETE. Foi obtida uma média de 0,95 no coeficiente de determinação para a modelagem de DQO na saída de quatro ETEs, métrica maior que 80% dos trabalhos que modelaram a mesma variável de interesse e usaram  $R^2$  para avaliação. E média no MAE de 3,52.

Na presente pesquisa, algumas limitações foram identificadas e merecem destaque para nortear trabalhos futuros. A ausência de dados de vazão afluente, efluente e, especialmente, e etapas intermediárias do processo de tratamento limitou a possibilidade de uma análise mais aprofundada sobre a dinâmica operacional das ETEs. Da mesma forma, a indisponibilidade de dados de consumo de energia elétrica, tanto total da planta quanto segmentado por etapas do tratamento, impossibilitou o cálculo do Índice de Eficiência Energética (IEE), que poderia oferecer *insights* valiosos para otimizar a eficiência energética do sistema. Adicionalmente, a falta de informações sobre o consumo de produtos químicos restringiu as análises sobre a relação entre as variáveis operacionais e o custo operacional associado. Esses dados poderiam, por exemplo, permitir modelagens mais robustas e a identificação de oportunidades para reduzir custos e melhorar a eficiência operacional. A integração desses indicadores aos modelos analíticos desenvolvidos poderia não apenas otimizar os parâmetros de saída, como a DQO efluente, mas também gerar subsídios para decisões mais assertivas relacionadas ao consumo energético e químico, ampliando a abrangência e a aplicabilidade dos resultados obtidos. Por isso, recomenda-se que futuros estudos busquem acessar e incorporar essas informações, de forma a explorar o potencial das técnicas de análise de dados para maximizar a eficiência das ETEs em múltiplas dimensões.

Por fim, a replicação desse trabalho em um contexto internacional, estudando ETEs em regiões com legislações e condições operacionais diferentes, poderia fornecer *insights* valiosos sobre a aplicabilidade dos modelos desenvolvidos.

## REFERÊNCIAS

- ALEXANDRIA, A. R. Sistema de Monitoramento de Chaves Seccionadoras por Visão Computacional e Redes Neurais Artificiais para Suporte ao Sistema de Controle Supervisório da CHESF. In: **IV Electrical Energy Technological Innovation Conference**, ANEEL. 2007.
- AMERICAN PUBLIC HEALTH ASSOCIATION et al. **Standard methods for the examination of water and wastewater**. American Public Health Association., 1917.
- ARISMENDY, L. et al. A Prescriptive Intelligent System for an Industrial Wastewater Treatment Process: Analyzing pH as a First Approach. **Sustainability (Basel, Switzerland)**, v. 13, n. 8, p. 4311, 2021.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 12209. **Elaboração de projetos hidráulico-sanitários de estações de tratamento de esgotos sanitários**. 2011.
- BARDIN, L; **Análise de conteúdo**. Lisboa: Edições 70. Obra original publicada em, 1977.
- BERRAR, D. **Cross-validation**. Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019.
- BHOLOWALIA, P.; KUMAR, A. EB. K-means: A clustering technique based on elbow method and k-means in WSN. **International Journal of Computer Applications**, v. 105, n. 9, 2014.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [s.l.] Springer, 2006. v. 4
- CHERNICHARO, C. DE. Reatores anaeróbios: princípios do tratamento biológico de águas residuárias. **Belo Horizonte: Polytécnica**, 2007.
- COHEN, D.; CRABTREE, B. **Qualitative research guidelines project**, 2006
- DANTAS, M. S.; CHRISTOFARO, C.; OLIVEIRA, S. C. Artificial neural networks for performance prediction of full-scale wastewater treatment plants: a systematic review. **Water Science & Technology**, v. 88, n. 6, p. 1447–1470, 2023.
- DAVIES, P. S. The biological basis of wastewater treatment. **Strathkelvin Instruments Ltd**, v. 3, 2005.
- DECRETO N.º 8.468, DE 8 DE SETEMBRO DE 1976. *Dispõe sobre a prevenção e o controle da poluição do meio ambiente no Estado de São Paulo*. Disponível em: <https://www.al.sp.gov.br/repositorio/legislacao/decreto/1976/decreto-8468-08.09.1976.html>. Acesso em: 22 jul. 2024.
- DROSTE, R. L.; GEHR, R. L. **Theory and Practice of Water and Wastewater Treatment**. [s.l.] Wiley, 2018.
- DU, Y. et al. Total suspended solids characterization and management implications for lakes in East China. **The Science of the total environment**, v. 806, n. Pt 4, p. 151374–151374, 2022.
- EBERT, C.; DUARTE, C. H. C. Digital Transformation. **IEEE software**, v. 35, n. 4, p. 16–21,

2018.

FERREIRA, F. D.; CORAIOLA, M. Eficiência do lodo ativado em fluxo contínuo para tratamento de esgoto. **Revista Acadêmica Ciência Animal**, v. 6, n. 2, p. 259-279, 2008.

GALLETTA, A.; CROSS, W. E. **Mastering the semi-structured interview and beyond: From research design to analysis and publication**. [s.l.] NYU press, 2013. v. 18

GEERDINK, R. B.; SEBASTIAAN VAN DEN HURK, R.; EPEMA, O. J. Chemical oxygen demand: Historical perspectives and future challenges. **Analytica chimica acta**, v. 961, p. 1–11, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [s.l.] MIT Press, 2016.

HARTIGAN, J. A.; WONG, M. A. A K-Means Clustering Algorithm. **Journal of the Royal Statistical Society Series C**, v. 28, n. 1, p. 100–108, 1979.

JAMES, G. **An introduction to statistical learning**. New York: springer , 2013.

JANES, A. et al. Effective dashboard design. **Cutter IT journal**, v. 26, n. 1, p. 17–24, 2013.

JOLLIFFE, I. T. **Principal component analysis for special types of data**. [s.l.] Springer New York, 2002.

JORDÃO, E. P.; PESSÔA, C. A. Tratamento de esgoto doméstico. **ABES**. Ríó de Janeiro, Brasil, 2005.

JORDÃO, E. P. Eficiência energética em tratamento de esgotos Economy of energy in wastewater treatment. **Revista DAE**, 2008.

KVALE, S. **InterViews: an introduction to qualitative research interviewing**. [s.l.] Sage, 1996.

LI, F.; VANROLLEGHEM, P. A. **An influent generator for WRRF design and operation based on a recurrent neural network with multi-objective optimization using a genetic algorithm**. **Water Science and Technology**. IWA Publishing, , 2022.

LIASHCHYNSKYI, P. Grid search, random search, genetic algorithm: a big comparison for NAS. **arXiv preprint arXiv:1912.06059**, 2019.

MARCÍLIO, W. E.; ELER, D. M. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In: **2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)**. IEEE, 2020. p. 340-347.

MARTINEZ, E. M. et al. Automation Pyramid as Constructor for a Complete Digital Twin, Case Study: A Didactic Manufacturing System. **Sensors (Basel, Switzerland)**, v. 21, n. 14, p. 4656, 2021.

MATPLOTLIB. **Matplotlib Documentation**. 2024. Disponível em: <https://matplotlib.org/stable/contents.html>. Acesso em: 21 nov. 2024.

MATT, C.; HESS, T.; BENLIAN, A. Digital Transformation Strategies. **Business &**

**information systems engineering**, v. 57, n. 5, p. 339–343, 2015.

MEHTA, B. R.; REDDY, Y. J. **Industrial Process Automation Systems - Design and Implementation**. Oxford: Elsevier, 2014.

METCALF, L.; EDDY, H. P. **Tratamento de efluentes e recuperação de recursos**. [s.l.] McGraw Hill Brasil, 2015.

MIKI, M. K. et al. **Tratamento Biológico de Esgoto: Princípios, Modelagem e Projeto, 2a edição**. IWA Publishing, , 2022.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**. [s.l.] John wiley & sons, 2010.

MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. **Malawi medical journal**, v. 24, n. 3, p. 69–71, 2012.

NUMPY. **Numpy Documentation**. 2024. Disponível em: <https://numpy.org/doc/stable/>. Acesso em: 21 nov. 2024.

PADHMA PRIYA, P. et al. Predictive Biodegradation of Multiple Toxic Pollutants in Bioreactors Treating Real Wastewater using ANN and GP. **IOP conference series. Earth and environmental science**, v. 1130, n. 1, p. 12040, 2023.

PADMANABHAN, T. R. **Industrial Instrumentation: Principles and Design**. London: Springer London, Limited, 2012.

PANDAS. **Pandas Documentation**. 2024. Disponível em: <https://pandas.pydata.org/docs>. Acesso em: 21 nov. 2024.

PEDDINTI, V.; POVEY, D.; KHUDANPUR, S. **A time delay neural network architecture for efficient modeling of long temporal contexts**. Interspeech. Anais.2015.

PERES, J. et al. Modelagem matemática para sistema de tratamento de esgoto doméstico com remoção simultânea de matéria orgânica e nitrogênio. **Engenharia Sanitaria e Ambiental**, v. 26, n. 4, p. 765–773, 2021.

PESATI, N, Tran., “Security Considerations for Large Language Model Use: Implementation Research in Securing LLM-Integrated Applications”, **IJRTE**, vol. 13, no. 3, pp. 19–27, Sep. 2024.

PESTANA, A. A.; HIRATA, A. Y.; OGAWA, F. I.; COSTA, R. A.; PALADINO, R.; **Automação dos painéis elétricos dos sopradores de ar dos tanques de aeração de estação de tratamento de esgotos, tendo como objetivo melhoria no controle de OD desses tanques e eficiência energética do equipamento**. 29º Congresso Nacional de Saneamento e Meio Ambiente. 2018

PESTANA, A. A.; ITO, M. Análise bibliométrica sobre a transformação digital no tratamento de esgoto. **Anais do XVIII Simpósio dos Programas de Mestrado Profissional – SIMPROFI**. P. 1077-1011, 2023.

PISA, I. et al. Denoising Autoencoders and LSTM-Based Artificial Neural Networks Data

Processing for Its Application to Internal Model Control in Industrial Environments—The Wastewater Treatment Plant Control Case. **Sensors (Basel, Switzerland)**, v. 20, n. 13, p. 3743, 2020.

PISA, I. et al. Transfer Learning in wastewater treatment plants control: Measuring the transfer suitability. **Journal of process control**, v. 124, p. 36–53, 2023.

PESTANA, A. A.; ITO, M. Explorando a transformação digital no tratamento de esgoto: uma survey com profissionais do setor. **Anais do XIX Simpósio dos Programas de Mestrado Profissional – SIMPROFI**, 2024.

PESTANA, A. A.; ITO, M. Transformação digital no tratamento de esgoto: percepções e perspectivas de especialistas. **Anais do XVIII Simpósio dos Programas de Mestrado Profissional – SIMPROFI**, 2024.

RESOLUÇÃO Nº 430, DE 13 DE MAIO DE 2011. **Dispõe sobre as condições e padrões de lançamento de efluentes, complementa e altera a Resolução no 357, de 17 de março de 2005**, do Conselho Nacional do Meio Ambiente-CONAMA. Disponível em: [http://conama.mma.gov.br/?option=com\\_sisconama&task=arquivo.download&id=627](http://conama.mma.gov.br/?option=com_sisconama&task=arquivo.download&id=627). Acesso em: 12 ago. 2023.

SABESP. **Tratamento de esgoto – Fase Líquida**. Disponível em: [https://www.sabesp.com.br/uploads/file/asabesp\\_doctos/Tratamento\\_Esgoto\\_Liquido\\_impresao.pdf](https://www.sabesp.com.br/uploads/file/asabesp_doctos/Tratamento_Esgoto_Liquido_impresao.pdf). Acesso em: 22 jul. 2024.

SABESP. **Tratamento de esgoto – Fase Sólida**. Disponível em: [https://www.sabesp.com.br/uploads/file/asabesp\\_doctos/Tratamento\\_Esgoto\\_Solido\\_imprensa.pdf](https://www.sabesp.com.br/uploads/file/asabesp_doctos/Tratamento_Esgoto_Solido_imprensa.pdf). Acesso em: 22 jul. 2024.

SAUL, C.; GEBAUER, H. Digital Transformation as an Enabler for Advanced Services in the Sanitation Sector. **Sustainability (Basel, Switzerland)**, v. 10, n. 3, p. 752, 2018.

SCIKIT-LEARN. **Scikit-learn Documentation**. 2024. Disponível em: <https://scikit-learn.org/stable/documentation.html>. Acesso em: 21 nov. 2024.

SHARMA, N.; CHANDRASEKAR, S.; SUNDAR, K. Suitability of ANN and GP for Predicting Soak Pit Tank Efficiency under Limited Data Conditions. **MATEC Web of Conferences**, v. 203, p. 3001, 2018.

SHEN, Y. et al. An artificial neural network-based data filling approach for smart operation of digital wastewater treatment plants. **Environmental research**, v. 224, p. 115549–115549, 2023.

TRATA BRASIL. **Benefícios econômicos do saneamento básico**. Instituto Trata Brasil, 2022. Disponível em: <https://tratabrasil.org.br/wp-content/uploads/2022/11/Beneficios-economicos-do-saneamento-no-Brasil.pdf>. Acesso em: 23 set. 2024.

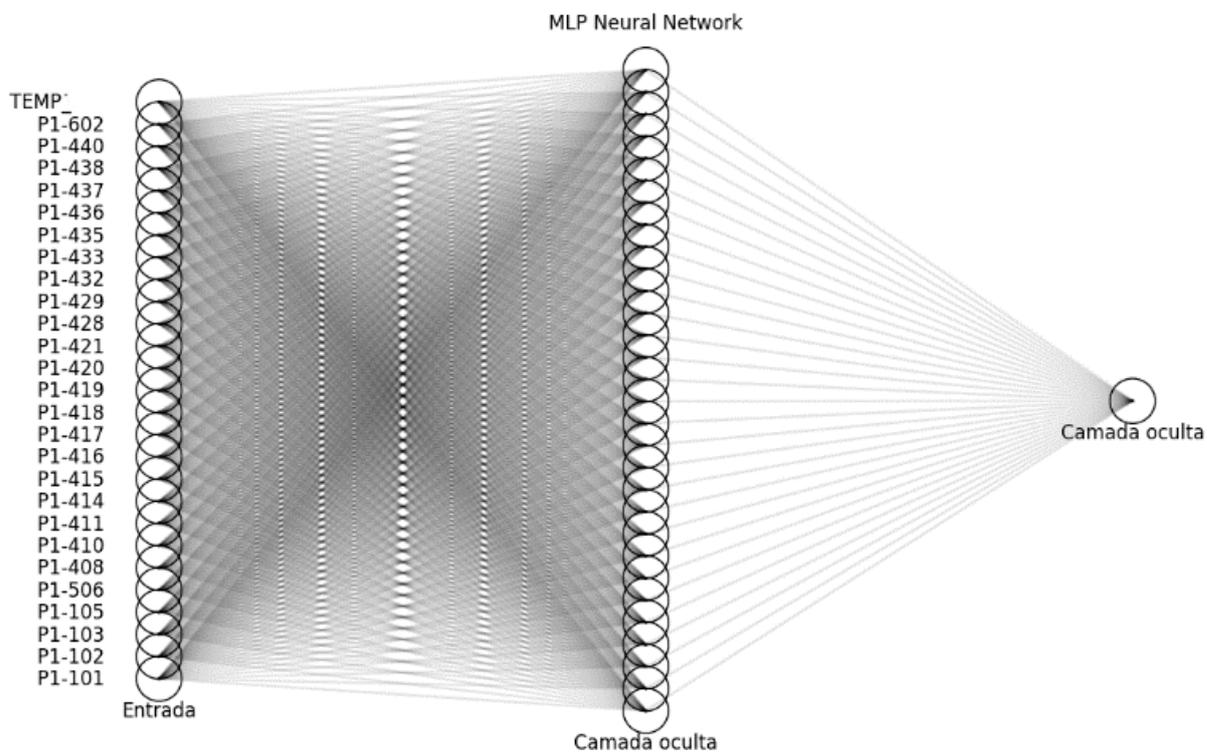
VAN HOUDT, G.; MOSQUERA, C.; NÁPOLES, G. A review on the long short-term memory model. **The Artificial intelligence review**, v. 53, n. 8, p. 5929–5955, 2020.

VIAL, G. Understanding digital transformation: A review and a research agenda. **The journal of strategic information systems**, v. 28, n. 2, p. 118–144, 2019.

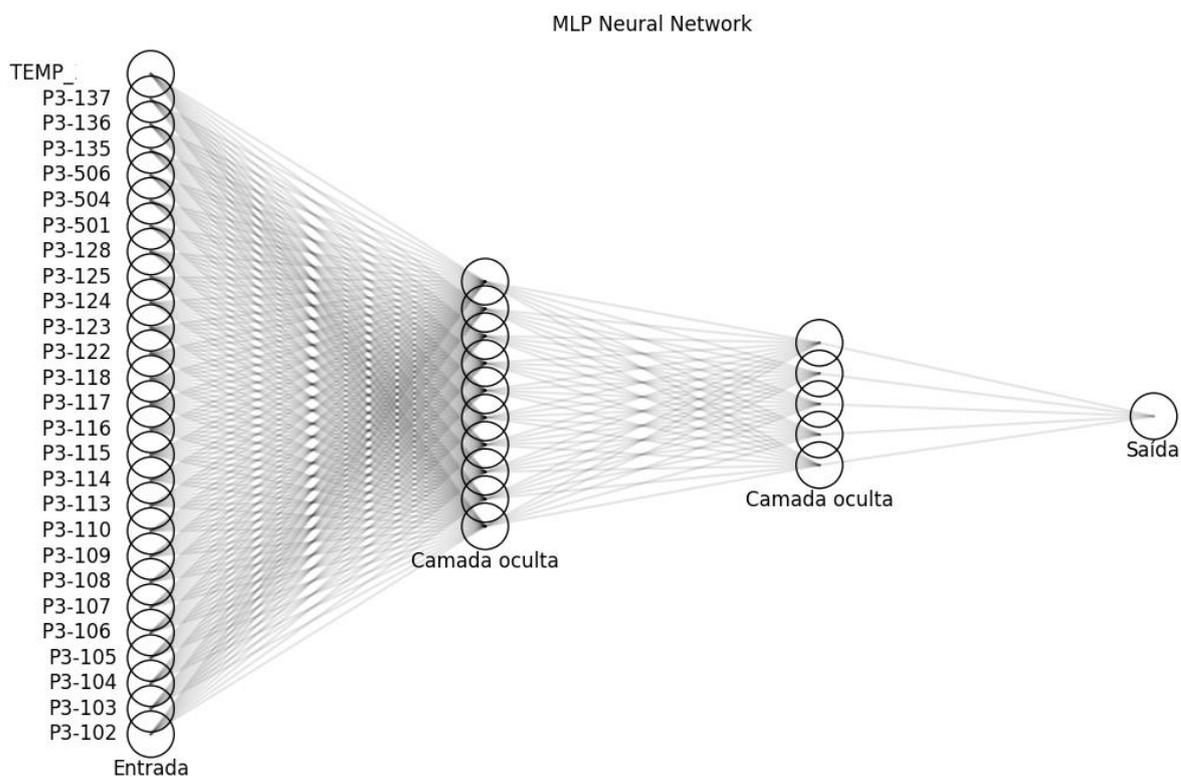
VON SPERLING, M. Lodos Ativados. Belo Horizonte: Departamento de Engenharia Sanitária e Ambiental-Universidade Federal de Minas Gerais, 415 p. **Princípios do Tratamento Biológico de Águas Residuárias**, v. 4, 1997.

YAMAKAWA, E. K.; MIGUEL, P. A. C.; AOKI, A. R. Aplicação de Fuzzy Quality Function Deployment para seleção de indicadores de eficiência energética para utilização em um software de gestão de energia. **Science & Engineering Journal**, v. 23, n. 2, p. 21–31, 2014.

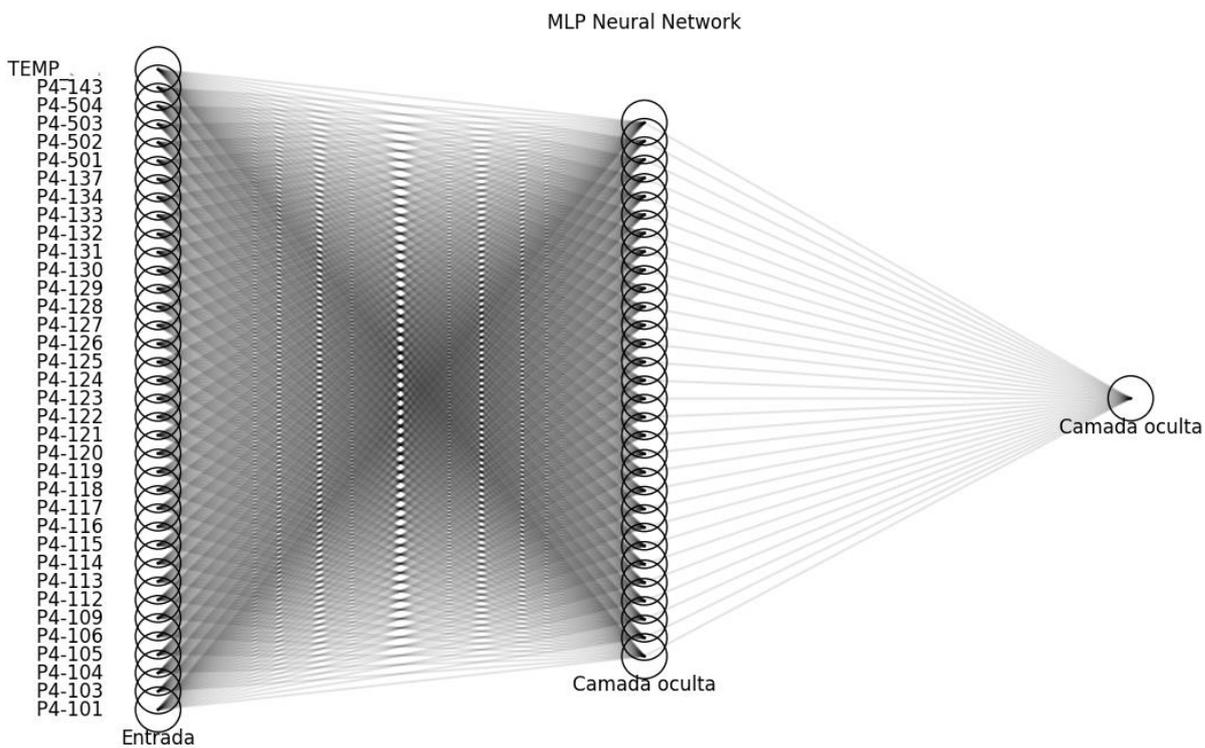
## APÊNDICE I – A – RNA ETE A



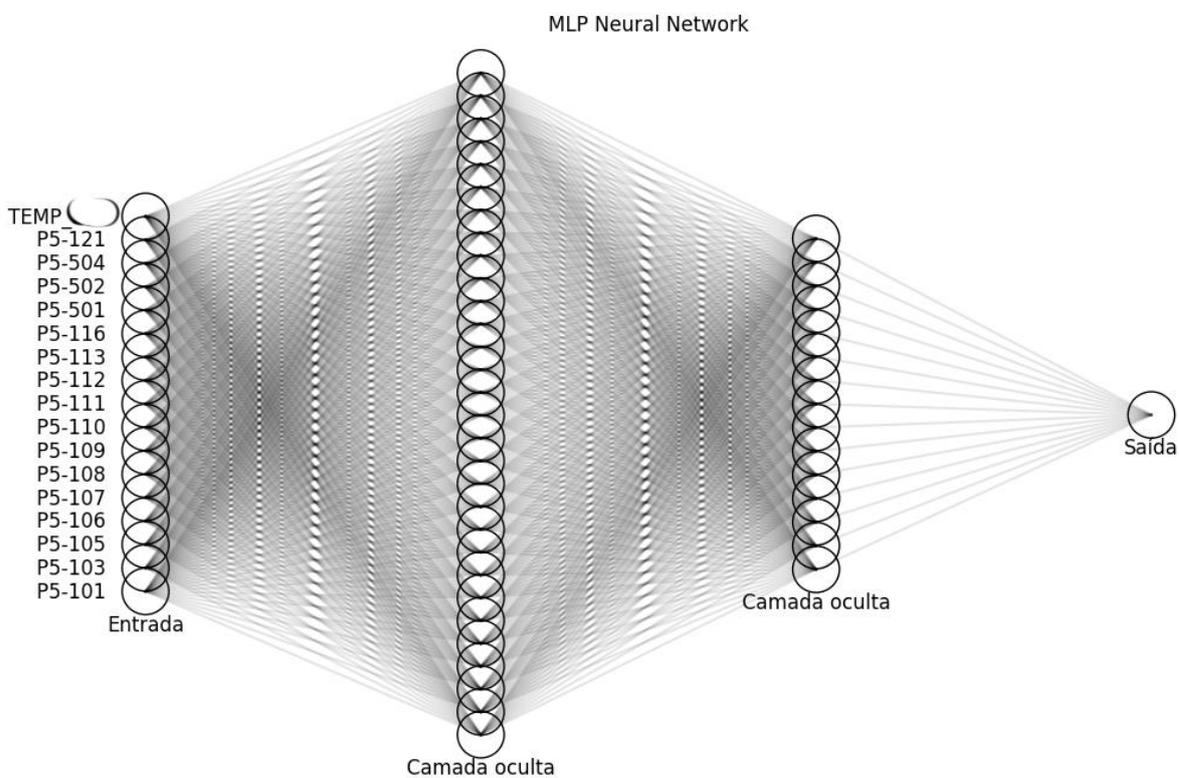
## APÊNDICE I – B – RNA ETE C



## APÊNDICE I – C – RNA ETE D



### APÊNDICE I – D – RNA ETE E



## **APÊNDICE II - DASHBOARD PARA AUTOMAÇÃO NO ESGOTO COM FOCO EM TRANSFORMAÇÃO DIGITAL**

O desenvolvimento de um *dashboard* para automação no esgoto com foco em transformação digital representa um passo fundamental para integrar tecnologia e inovação aos processos de tratamento de esgoto. Em um contexto onde a eficiência operacional e a sustentabilidade são prioridades, o *dashboard* atua como uma ferramenta estratégica para consolidar, monitorar e planejar o cenário nesse contexto, permitindo uma gestão mais ágil e precisa. No escopo desta dissertação, essa proposta conecta-se diretamente com a análise e otimização de dados em estações de tratamento, como um entendimento do cenário.

O primeiro passo foi realizar um levantamento do que há e o que é necessário referente a automação para cada uma das etapas do processo, em cada uma das camadas da pirâmide de automação e em cada tipo de Estação (principal e sistema isolado), e isso foi realizado entre equipe de automação/instrumentação em conjunto com equipe operacional. Por exemplo, foi definido que para os sistemas isolados, o mínimo necessário seria: Primeira camada – proteção de vibração e temperatura e medição de nível e vazão; segunda camada – controlador lógico programável (CLP); terceira camada – comunicação e supervisão; e para etapas futuras a quarta camada – Envio dos dados ao (PIMS), PI System da OSI Software no caso da companhia em questão. Onde o elemento não faz sentido, foi marcado como “N/A” e essa linha não será considerada nos cálculos posteriores. Na Figura 52 é possível ver a planilha Excel com esse levantamento inicial.

Figura 52 - Levantamento da automação mínima necessário

| ITEM | SISTEMA | TIPO | LOCAL | PROTEÇÃO<br>(VIBRAÇÃO/<br>TEMPERATURA) | NÍVEL / COTA | VAZÃO | CLP / CONTROLE | COMUNICAÇÃO /<br>SUPERVISÃO |
|------|---------|------|-------|--|--------------|-------|----------------|-----------------------------|
| 1    |         | EFRF |       | N/A                                    | ✘            | ✔     | ✔              | ✘                           |
| 2    |         | ETE  |       | N/A                                    | ✔            | ✔     | ✔              | ✘                           |
| 3    |         | ETE  |       | N/A                                    | N/A          | ✔     | ✘              | ✘                           |
| 4    |         | ETE  |       | N/A                                    | N/A          | ✔     | ✘              | ✘                           |
| 5    |         | ETE  |       | N/A                                    | N/A          | ✔     | ✘              | ✘                           |
| 6    |         | EEE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 7    |         | EEE  |       | ✘                                      | ✔            | N/A   | ✔              | ✔                           |
| 8    |         | EEE  |       | ✘                                      | ✔            | ✘     | ✔              | ✔                           |
| 9    |         | EEE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 10   |         | ETE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 11   |         | ETE  |       | N/A                                    | ✘            | ✘     | ✘              | ✘                           |
| 12   |         | EEE  |       | ✔                                      | ✔            | ✔     | ✔              | ✔                           |
| 13   |         | ETE  |       | N/A                                    | N/A          | ✔     | ✔              | ✘                           |
| 14   |         | ETE  |       | ✘                                      | ✔            | N/A   | ✔              | ✔                           |
| 15   |         | EEE  |       | ✘                                      | ✔            | ✘     | ✔              | ✔                           |
| 16   |         | EEE  |       | ✔                                      | ✔            | ✔     | ✔              | ✔                           |
| 17   |         | ETE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 18   |         | ETE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 19   |         | ETE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 20   |         | EEE  |       | ✘                                      | ✔            | ✔     | ✔              | ✔                           |
| 21   |         | EEE  |       | ✔                                      | ✔            | ✔     | ✔              | ✔                           |

Fonte: Do autor (2025)

O levantamento do mínimo necessário foi feito em conjunto com a equipe operacional com base em atender legislação, por exemplo, com a importância dos medidores de nível em elevatórias, permitindo-se assim manter a cota operacional da mesma quando em conjunto com o CLP. A existência de comunicação ou não das informações com o sistema supervisor, podem nos dizer a frequência necessária da equipe operacional no local, principalmente em ambientes mais remotos. Após esses levantamentos, estatísticas gerais de plantas e Sistemas de acordo com os elementos desejados já começam a ser possíveis. A eq. (9) a (13) trazem a proposta para o cálculo do nível de automação em cada uma das camadas, de acordo com a pirâmide de automação.

$$\text{Média automação} = \frac{\text{Camada 1} + \text{Camada 2} + \text{Camada 3}}{3} \quad (9)$$

$$\text{Camada 1} = \frac{\text{Média Instrumentação} + \text{Elementos de controle}}{2} \quad (10)$$

$$\text{Instrumentação} = \frac{\text{Proteções} + \text{Analíticos} + \text{Nível} + \text{Vazão}}{4} \quad (11)$$

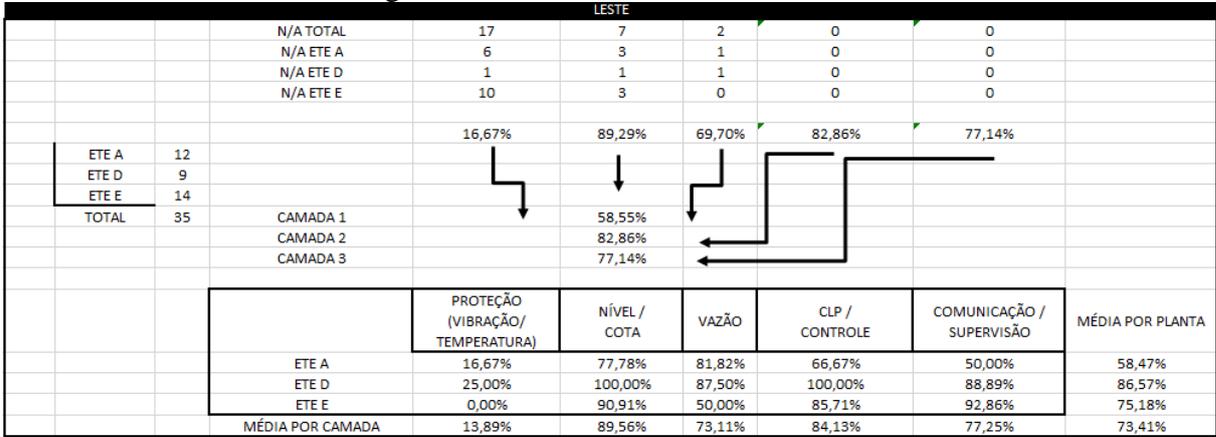
$$\text{Camada 2} = \text{Controlador Lógico Programável} \quad (12)$$

$$\text{Camada 3} = \frac{\text{Comunicação} + \text{Supervisor}}{2} \quad (13)$$

Após o preenchimento e realizados os cálculos, é possível ter uma visão mais holística do cenário de automação na Unidade, com os níveis obtidos em cada uma das camadas, ou por

componente (instrumentação, controle e afins) ou por planta, conforme a Figura 53.

Figura 53 - Estatística do levantamento



Fonte: Do autor (2025)

O levantamento para plantas principais segue o mesmo raciocínio usado nas plantas isoladas, porém com algumas alterações. Na primeira camada, foram inclusos os elementos de controle, usuais em muitas etapas de tratamento, e no campo de instrumentação, os instrumentos analíticos de processo, como pH, condutividade, OD, ST, Turbidez, Amônia, DQO entre outros. Para terceira camada, também se decidiu por separar os itens comunicação/supervisão, isso porque em muitas etapas a comunicação daquele processo de tratamento se dá por um componente diferente de outra etapa, podendo estar disponível em uma delas e indisponível em outras, conforme Figura 54.

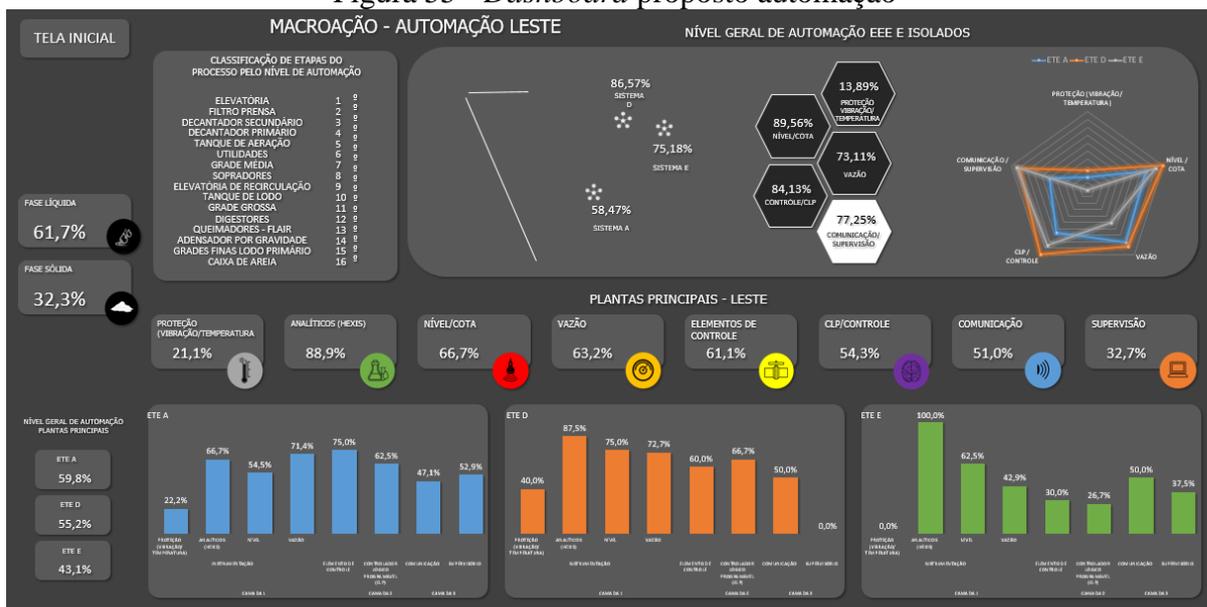
Figura 54 - Levantamento da automação mínima necessária – Plantas principais

| ITEM | ETE | ETAPA DO PROCESSO           | CAMADA 1                        |                    |       |       | ELEMENTO DE CONTROLE | CAMADA 2 | CAMADA 3                             |                           |              |
|------|-----|-----------------------------|---------------------------------|--------------------|-------|-------|----------------------|----------|--------------------------------------|---------------------------|--------------|
|      |     |                             | INSTRUMENTAÇÃO                  |                    |       |       |                      |          | CONTROLADOR LÓGICO PROGRAMÁVEL (CLP) | COMUNICAÇÃO (RÁDIO/FIBRA) | SUPERVISÓRIO |
|      |     |                             | PROTEÇÃO (VIBRAÇÃO/TEMPERATURA) | ANALÍTICOS (HEXIS) | NÍVEL | VAZÃO |                      |          |                                      |                           |              |
| 1    |     | GRADE GROSSA                | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 2    |     | ELEVATÓRIA                  | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 3    |     | GRADE MÉDIA                 | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 4    |     | CAIXA DE AREIA              | N/A                             | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 5    |     | DECANTADOR PRIMÁRIO         | ⊗                               | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 6    |     | TANQUE DE AERAÇÃO           | N/A                             | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 7    |     | SOPRADORES                  | ⊗                               | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 8    |     | DECANTADOR SECUNDÁRIO       | N/A                             | ⊗                  | N/A   | ⊗     | N/A                  | ⊗        | ⊗                                    |                           |              |
| 9    |     | ELEVATÓRIA DE RECIRCULAÇÃO  | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 10   |     | GRADES FINAS LODO PRIMÁRIO  | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 11   |     | ADENSADOR POR GRAVIDADE     | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 12   |     | ADENSADOR POR FLOTAÇÃO      | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 13   |     | DIGESTORES                  | ⊗                               | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 14   |     | QUEIMADORES FLAIR           | N/A                             | N/A                | N/A   | ⊗     | N/A                  | ⊗        | ⊗                                    |                           |              |
| 15   |     | TANQUE DE LODO CONDICIONADO | N/A                             | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 16   |     | FILTRO PRENSA               | ⊗                               | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 17   |     | UTILIDADES                  | ⊗                               | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 18   |     | GRADE GROSSA                | N/A                             | N/A                | ⊗     | N/A   | N/A                  | ⊗        | ⊗                                    |                           |              |
| 19   |     | ELEVATÓRIA                  | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 20   |     | GRADE MÉDIA                 | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 21   |     | CAIXA DE AREIA              | N/A                             | N/A                | N/A   | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 22   |     | DECANTADOR PRIMÁRIO         | N/A                             | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 23   |     | TANQUE DE AERAÇÃO           | N/A                             | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 24   |     | SOPRADORES                  | ⊗                               | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 25   |     | DECANTADOR SECUNDÁRIO       | N/A                             | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 26   |     | ELEVATÓRIA DE RECIRCULAÇÃO  | ⊗                               | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 27   |     | GRADES FINAS LODO PRIMÁRIO  | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 28   |     | ADENSADOR POR GRAVIDADE     | N/A                             | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 29   |     | DIGESTORES                  | N/A                             | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 30   |     | QUEIMADORES FLAIR           | N/A                             | N/A                | N/A   | N/A   | N/A                  | ⊗        | ⊗                                    |                           |              |
| 31   |     | TANQUE DE LODO CONDICIONADO | N/A                             | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 32   |     | FILTRO PRENSA               | ⊗                               | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 33   |     | UTILIDADES                  | ⊗                               | ⊗                  | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 34   |     | GRADE GROSSA                | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 35   |     | ELEVATÓRIA                  | ⊗                               | N/A                | ⊗     | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 36   |     | GRADE MÉDIA                 | N/A                             | N/A                | ⊗     | N/A   | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 37   |     | CAIXA DE AREIA              | N/A                             | N/A                | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 38   |     | DECANTADOR PRIMÁRIO         | N/A                             | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |
| 39   |     | TANQUE DE AERAÇÃO           | ⊗                               | ⊗                  | N/A   | ⊗     | ⊗                    | ⊗        | ⊗                                    |                           |              |

Fonte: Do autor (2025)

Para criação das telas em si, dividiu-se inicialmente as plantas principais dos sistemas isolados. Para os sistemas isolados uma representação do mapa do estado com uma ampliação da região metropolitana da cidade que permitem visualizar as médias da disponibilidade para cada um dos Sistemas e ao lado a média de cada uma das componentes para os Sistemas juntos e separados. Enquanto para os Sistemas principais, um gráfico para cada um deles mostrando o nível para cada elemento está disposto na parte inferior. A esquerda desses é mostrado a média de automação da planta como um todo. Por fim, algumas análises possíveis após o levantamento mostrado até então permitem comparar por exemplo os níveis entre fase líquida e sólida, além de comparar as etapas do processo, por exemplo, elevatória final e tanque de aeração. A proposta desse *dashboard* pode ser vista na Figura 55.

Figura 55 - Dashboard proposto automação



Fonte: Do autor (2025)

Uma das necessidades percebidas, foi que nem todas as pessoas envolvidas no processo de priorização de investimentos de automação, são necessariamente especialistas na área. Por isso, para representação da média de disponibilidade de cada componente, foi criado um símbolo, visando facilitar o entendimento do usuário e criar um link com seu significado, por isso a utilização dos símbolos criados para comunicação, supervisão, vazão, dados analíticos, nível controle, elementos de controle e temperatura.

Por fim, o dashboard mapeia o nível de automação em cada camada da pirâmide de automação, considerando as etapas do processo e os diferentes tipos de Estações. Oferece ainda uma visão holística do nível de controle existente em cada estação, permitindo um entendimento integrado sobre automação e análise de dados no setor. Essa abordagem é fundamental para correlacionar os níveis de automação às capacidades de inferência e predição no processo de tratamento, especialmente a partir das modelagens analíticas propostas na dissertação. Além de ampliar a compreensão sobre a eficiência dos processos, o *dashboard* serve como uma ferramenta estratégica para planejar melhorias e orientar investimentos em automação, promovendo avanços operacionais e alinhamento com a sustentabilidade e inovação no saneamento.

Nos dois capítulos seguintes, são apresentadas as entrevistas com especialistas e *Survey* com profissionais do setor, que foram realizadas visando capturar a perspectiva desses profissionais mais próximos ao tema.

## ANEXO I – AUTORIZAÇÃO DO USO DE DADOS NA PESQUISA



Eu, Ivana Wu Pereira Vidal, na qualidade de Superintendente de Tratamento de Esgoto da Região Metropolitana de São Paulo, autorizo o colaborador Allan dos Anjos Pestana a utilizar os dados analíticos e informações operacionais das estações de tratamento de esgoto da Sabesp como parte integrante de sua pesquisa de mestrado, intitulada "**Modelo preditivo para otimização do tratamento de esgoto utilizando técnicas de aprendizado de máquina**".

Os dados contemplados nesta autorização incluem, medições de parâmetros como pH, condutividade, SST, oxigênio dissolvido, entre outros, bem como informações sobre o nível de automação e instrumentação das unidades.

Esta autorização se dá com base nas seguintes condições:

**Finalidade Exclusivamente Acadêmica:** Os dados serão utilizados exclusivamente para fins da pesquisa acadêmica, conforme o escopo descrito na dissertação.

**Sigilo e Proteção de Dados:** O colaborador se compromete a tratar os dados com confidencialidade, em conformidade com as normas internas da Sabesp e as boas práticas de proteção de dados.

**Não Divulgação de Informações Sensíveis:** A pesquisa não incluirá informações estratégicas ou sensíveis da companhia.

**Compartilhamento de Resultados:** Os resultados e insights provenientes da pesquisa deverão ser compartilhados com a Sabesp, contribuindo para a melhoria contínua dos processos e gestão operacional.

**Não identificação das instalações onde os dados foram extraídos.**

**No compromisso do colaborador em contribuir para o avanço acadêmico e para o desenvolvimento de ferramentas e conhecimentos que possam agregar valor às operações da companhia.**

São Paulo, 03 de janeiro de 2025.

\_\_\_\_\_  
 Ivana Wu Pereira Vidal  
 Superintendente de Tratamento de Esgoto  
 Sabesp – Unidade de Tratamento de Esgoto da Região Metropolitana de São Paulo





## VERIFICAÇÃO DAS ASSINATURAS



Código para verificação: C99C-F0AE-CA7C-6197

Este documento foi assinado digitalmente pelos seguintes signatários nas datas indicadas:

- ✓ IVANA WUO PEREIRA VIDAL (CPF 097.XXX.XXX-00) em 03/01/2025 12:09:36 (GMT-03:00)  
Papel: Parte  
Emitido por: Sub-Autoridade Certificadora 1Doc (Assinatura 1Doc)

Para verificar a validade das assinaturas, acesse a Central de Verificação por meio do link:

<https://assinaturasabesp.1doc.com.br/verificacao/C99C-F0AE-CA7C-6197>