

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CÂMPUS GOIÂNIA
MESTRADO PROFISSIONAL EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

DIEGO TAKASHI SATO

**Modelo computacional preditivo
de regressão supervisionado
guiado por séries temporais para
otimização da cloração e
fluoretação em estação de
tratamento de água**

Goiânia, 2025.

**TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO
NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☒ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☐ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Diego Takashi Sato

Matrícula: 20222011280070

Título do Trabalho: Modelo computacional preditivo de regressão supervisionado guiado por séries temporais para otimização da cloração e fluoretação em estação de tratamento de água

Autorização - Marque uma das opções

1. ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ____/____/____ (Embargo);
3. ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2** ou **3**, marque a justificativa:

- ☐ O documento está sujeito a registro de patente.
☐ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
☐ Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

gov.br
Documento assinado digitalmente
DIEGO TAKASHI SATO
Data: 19/02/2025 18:03:03-0300
Verifique em <https://validar.iti.gov.br>

Local Goiânia - GO, 19 / 02 / 2025.
Data

Assinatura do Autor e/ou Detentor dos Direitos Autorais

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CÂMPUS GOIÂNIA
MESTRADO PROFISSIONAL EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

DIEGO TAKASHI SATO

Modelo computacional preditivo de regressão supervisionado guiado por séries temporais para otimização da cloração e fluoretação em estação de tratamento de água

Programa de Pós-Graduação *Stricto Sensu* em Tecnologia, Gestão e Sustentabilidade do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Dissertação de Mestrado Profissional. Área de Concentração: Tecnologia de Sistemas de Produção Limpa. Linha de Pesquisa: Modelagem de Sistemas.

Orientador: Prof. Wesley Pacheco Calixto, Dr.

Coorientador: Prof. Antônio Pires de Castro Júnior, Dr.

Goiânia, 2025.

S253m Sato, Diego Takashi.

Modelo computacional preditivo de regressão supervisionado guiado por séries temporais para otimização da cloração e fluoretação em estação de tratamento de água / Diego Takashi Sato. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2025.
81f.

Orientação: Prof. Wesley Pacheco Calixto.

Dissertação (mestrado) – Programa de Pós-Graduação *Stricto Sensu* em Tecnologia, Gestão e Sustentabilidade, Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

1. Tratamento de água. 2. Inteligência artificial. 3. Modelagem computacional. I. Calixto, Wesley Pacheco (orientação). II. Instituto Federal de Educação, Ciência e Tecnologia de Goiás. III. Título.

CDD 628.162

Ficha catalográfica elaborada pela bibliotecária Lana Cristina Dias Oliveira CRB1/ 2.631
Biblioteca Professor Jorge Félix de Souza,
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.



INSTITUTO FEDERAL
Goiás

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS GOIÂNIA

PARECER 9/2025 - GYN-CMTGS/GYN-GPPEX/CP-GOIANIA/IFG

SERVIÇO PÚBLICO FEDERAL
INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DE GOIÁS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
IFG CÂMPUS GOIÂNIA
MESTRADO EM TECNOLOGIA, GESTÃO E SUSTENTABILIDADE

DIEGO TAKASHI SATO

MODELO PREDITIVO DE REGRESSÃO SUPERVISIONADO BASEADO EM SÉRIES TEMPORAIS PARA OTIMIZAÇÃO DA CLORAÇÃO E FLUORETAÇÃO EM ESTAÇÕES DE TRATAMENTO DE ÁGUA

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologia, Gestão e Sustentabilidade do Instituto Federal de Educação, Ciência e Tecnologia de Goiás como requisito parcial à obtenção do título de Mestre.

Área de Concentração: Tecnologia de Sistemas de Produção Limpa.

Linha de Pesquisa: Energias Renováveis

Aprovada em 17 de fevereiro de 2025.

Banca examinadora:

Presidente e orientador: Prof. Dr. Wesley Pacheco Calixto - UFG EMC

Coorientador: Prof. Dr. Antônio Pires de Castro Junior - TJ-GO Tribunal de Justiça do Estado de Goiás

Avaliador: Prof. Dr. Antônio Paulo Coimbra - Universidade de Coimbra DEEC

Avaliador: Prof. Dr. Sérgio Botelho de Oliveira - PPGTGS IFG

Avaliador: Prof. Dr. Márcio Rodrigues da Cunha Reis - IFG – Câmpus Senador Canedo

Documento assinado eletronicamente por:

- **ANTONIO PIRES DE CASTRO JUNIOR, ANTONIO PIRES DE CASTRO JUNIOR - 123605 - Diretor de informática - Tribunal de Justiça de Goiás (1)**, em 26/02/2025 18:59:02.
- **Sergio Botelho de Oliveira, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 17/02/2025 17:08:53.
- **Antônio Paulo Mendes Breda Dias Coimbra, Antônio Paulo Mendes Breda Dias Coimbra - Visitante - Ifg - Câmpus Goiânia (10870883000225)**, em 17/02/2025 16:45:45.
- **Marcio Rodrigues da Cunha Reis, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 17/02/2025 16:39:54.
- **Wesley Pacheco Calixto, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 17/02/2025 16:38:31.

Este documento foi emitido pelo SUAP em 17/02/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifg.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 619036

Código de Autenticação: 6a7f795583



Rua 75, nº 46, Centro, GOIÂNIA / GO, CEP 74055-110
Sem Telefones cadastrados

Dedicatória

À minha esposa, cuja paciência, amor e apoio incondicional foram fundamentais para que eu pudesse trilhar este caminho. Nas muitas noites em que estivemos distantes, você me deu força e encorajamento, sempre acreditando, mesmo quando eu duvidei de mim mesmo. Aos meus filhos, que com sua alegria e pureza me lembraram diariamente do verdadeiro significado de perseverar e alcançar nossos sonhos. A todos os amigos que acreditaram e incentivaram.

Agradecimentos

Gostaria de expressar minha mais profunda gratidão a todos que, de alguma forma, contribuíram para a realização deste trabalho. Agradeço especialmente às instituições envolvidas, em particular à Saneamento de Goiás S/A (Saneago), pela disponibilização dos dados indispensáveis para o desenvolvimento deste estudo. Também estendo meu reconhecimento ao Instituto Federal de Goiás (IFG), que proporcionou a oportunidade de expandir meus conhecimentos e consolidar minha trajetória acadêmica.

Agradeço de coração ao meu Orientador, Prof. Dr. Wesley Pacheco Calixto, cuja orientação, paciência e comprometimento foram essenciais em cada etapa deste projeto. Sua expertise e dedicação foram determinantes para enriquecer não apenas este trabalho, mas também minha formação pessoal e profissional.

Também agradeço ao meu coorientador, Prof. Dr. Antônio Pires de Castro Júnior, pela confiança, disponibilidade e contribuições ao trabalho.

Um agradecimento especial é direcionado à minha família, pelo amor, paciência e apoio incondicional. A compreensão e o incentivo constante foram fundamentais, especialmente nos momentos de maior desafio e superação.

Acima de tudo, agradeço a Deus, fonte de força, sabedoria e inspiração, que me guiou ao longo deste percurso, dando-me coragem para enfrentar os obstáculos e determinação para alcançar meus objetivos.

A todos que, direta ou indiretamente, colaboraram para a concretização deste projeto, deixo registrado meu mais sincero agradecimento.

Devíamos ser ensinados a não esperar por inspiração para começar algo. Ação sempre gera inspiração. Inspiração raramente gera ação.

Frank Tibolt,
1897-1989.

Resumo

Título: Modelo computacional preditivo de regressão supervisionado guiado por séries temporais para otimização da cloração e fluoretação em estação de tratamento de água

Autor: Diego Takashi Sato

Orientador: Wesley Pacheco Calixto, Dr.

Coorientador: Prof. Antônio Pires de Castro Júnior, Dr.

Este estudo propõe um modelo preditivo para otimizar os processos de cloração e fluoretação em estações de tratamento de água utilizando algoritmos de aprendizado de máquina, como Redes Neurais Artificiais Multicamadas, Floresta Randômica e XGBoost. Os dados analisados, obtidos de uma estação com captação superficial, passam por técnicas de pré-processamento e otimização de hiperparâmetros para assegurar previsões precisas. Os resultados indicam melhorias significativas nas métricas de desempenho após o processo de otimização, com redução de erros e aumento da capacidade de generalização dos modelos. Os modelos otimizados apresentam $RMSE \approx 0,40$ para a dosagem de cloro, com coeficiente de determinação $R^2 \approx 0,90$, e $RMSE \approx 0,15$ e $R^2 \approx 0,95$ para fluoretação. Os achados indicam que a integração de inteligência artificial em operações de tratamento de água potável proporciona monitoramento eficiente e automatizado, além de garantir a conformidade com os padrões regulatórios de potabilidade. Estes resultados reforçam o potencial da abordagem em contribuir para a sustentabilidade do setor, ao reduzir o uso de insumos químicos e promover operações mais precisas e seguras.

Palavras-chave

Modelagem de sistemas, Tratamento de água, Inteligência artificial, Otimização de processos, Sustentabilidade

Abstract

Title: Supervised regression predictive computer model guided by time series for optimizing chlorination and fluoridation in a water treatment plant

Author: Diego Takashi Sato

Advisor: Wesley Pacheco Calixto, Dr.

Coadvisor: Prof. Antônio Pires de Castro Júnior, Dr.

This study proposes a predictive model to optimize chlorination and fluoridation processes in water treatment plants by applying machine learning algorithms such as multilayer artificial neural networks, random forests, and XGBoost. The data analyzed, obtained from a plant with surface water intake, undergo pre-processing techniques and hyperparameter optimization to ensure accurate predictions. The findings indicate significant performance improvements after the optimization process, with reduced errors and improved generalization capacity of the models. The optimized models present $\text{RMSE} \approx 0.40$ for chlorine dosage, with a coefficient of determination $R^2 \approx 0.90$, and $\text{RMSE} \approx 0.15$ and $R^2 \approx 0.95$ for fluoridation. These observations suggest that the integration of artificial intelligence in drink water treatment operations allows efficient and automated monitoring while ensuring compliance with regulatory drinking water standards. The results highlight the potential of the approach to contribute to the sustainability of the sector by reducing chemical inputs and promoting more precise and safer operations.

Keywords

system modeling, water treatment, artificial intelligence, optimization process, sustainability

Lista de Figuras

Figura 1	– Fluxograma da metodologia proposta.....	39
Figura 2	– Fluxo hipotético dos requisitos da ETA para inclusão no estudo.	39
Figura 3	– Fluxo dos dados no banco de dados tratados.....	42
Figura 4	– Mapa georreferenciado da captação e ETA de Brazabrantas. .	47
Figura 5	– ETA Brazabrantas, Goiás/Brasil.	48
Figura 6	– Equipamentos, ambientes e processo: (a) tanque de solubilização de cloreto férrico em batelada, (b) medidor de turbidez em NTU, (c) floco-decantador, (d) tela de retenção de sólidos flutuantes/grosseiros na saída do floco-decantador, antes de entrar no filtro de fluxo ascendente, (e) calha de coleta de água superficial no floco-decantador e (f) medidor automatizado de cloro, flúor, M_T e M_{pH}	49
Figura 7	– Comparativo de métodos de preenchimento de dados faltantes.	52
Figura 8	– Valores reais do sistema e previsões dos modelos com amostra de 2022: (a) cloração sem otimização de hiperparâmetros, (b) cloração com otimização de hiperparâmetros, (c) fluoretação sem otimização de hiperparâmetros e (d) fluoretação com otimização de hiperparâmetros.	57
Figura 9	– Valores reais do sistema e previsões dos modelos com amostra de 2024: (a) cloração sem otimização de hiperparâmetros, (b) cloração com otimização de hiperparâmetros, (c) fluoretação sem otimização de hiperparâmetros e (d) fluoretação com otimização de hiperparâmetros.	58
Figura 10	– Distribuição dos erros dos modelos preditivos para: (a) cloro e (b) flúor.	60
Figura 11	– Resíduos dos modelos preditivos ao longo das observações: (a) XGBoost sem otimização cloro, (b) XGBoost otimizado cloro, (c) RNA-MLP sem otimização flúor e (d) RNA-MLP otimizado flúor...	61
Figura 12	– Dispersão com valores reais $\times$ previstos: (a) RNA-MLP sem otimização cloro, (b) RNA-MLP otimizado cloro, (c) FR sem otimização flúor e (d) FR otimizado flúor.	62
Figura 13	– Artigo submetido.	68
Figura 14	– Relatório técnico-científico.	70

Lista de Tabelas

Tabela 1 – Síntese de estudos que empregam modelos de inteligência artificial no tratamento de água.....	23
Tabela 2 – Correlação entre algumas variáveis observadas e a dosagem de cloro e flúor.	50
Tabela 3 – Estatística descritiva dos parâmetros observados.	50
Tabela 4 – Critérios de avaliação dos métodos utilizados para preenchimento de lacunas.	51
Tabela 5 – Valores dos hiperparâmetros e das validações internas utilizando os métodos RNA-MLP, FR e XGBoost.	53
Tabela 6 – Valores otimizados de hiperparâmetros e validações para os métodos RNA-MLP, FR e XGBoost.	55
Tabela 7 – Validação externa comparando os resultados dos modelos com os valores reais coletados no sistema.....	56
Tabela 8 – Validação externa comparando os resultados dos modelos otimizados com os valores reais coletados no sistema.	56
Tabela 9 – Resultados das métricas de avaliação para os modelos preditivos de cloro e flúor.	59

Lista de Algoritmos

Algoritmo 1	–	<i>TrataLacunas(dadosBrutos)</i>		41
-------------	---	----------------------------------	-----------	----

Lista de Abreviaturas e Siglas

k -FCV	<i>k-fold cross-validation</i>
R^2	coeficiente de determinação
AAB	adutoras de água bruta
ADN	ácido desoxirribonucleico
AHA	ácido haloacéticos
AM	aprendizado de máquina
ANA	Agência Nacional de Águas
CRL	cloro residual livre
EMA	erro médio absoluto
EMQ	erro médio quadrático
ETA	Estação de Tratamento de Água
FR	Floresta Randômica
GBC	<i>Gradient Boosting Classifier</i>
GBT	<i>Gradient Boost Tree</i>
IA	inteligência artificial
ICF	índice de coliformes fecais
ICT	índice de coliformes totais
IQA	índice de qualidade da água
KNN	<i>kernel nearest neighbor</i>
LSTM	<i>long short-term memory</i>
MLP	<i>multi-layer perceptron</i>
OMS	Organização Mundial de Saúde
ONU	Organização das Nações Unidas
POA	processos de oxidação avançados
REMQ	raiz do erro médio quadrático
RNA	redes neurais artificiais
RNP	redes neurais profundas
RNR	rede neural recorrente
RVS	regressão de vetor de suporte
SAA	Sistema de Abastecimento de Água
SODIS	desinfecção solar da água
THM	trihalometanos
UV	ultravioleta
XGBoost	<i>Extreme Gradient Boosting</i>

Lista de Símbolos

α	paciência
γ	ganho mínimo
Δ_{min}	melhoria mínima na validação
ϵ	número máximo de épocas
η	taxa de aprendizado
μ	média
ρ	correlação
σ	desvio padrão
c_{mf}	crescimento mínimo de folhas
C_{ST}	características amostradas aleatoriamente em cada árvore durante o treinamento
D_{Cl}	vazão volumétrica de produto adicionado para cloração
D_F	vazão volumétrica de produto adicionado para fluoretação
D_M	profundidade máxima das árvores de decisão durante o treinamento
D_S	vazão volumétrica de produto adicionado para coagulação
M_{Cl}	medições de Cl
M_{cw}	peso mínimo de instância
M_F	medições de F
M_{pH}	medições de pH
M_T	temperatura da água tratada
n_a	número máximo de árvores
N_{RA}	nível do reservatório apoiado
N_{rc}	número de variáveis escolhidas aleatoriamente
N_{RE}	nível do reservatório elevado
N_{rn}	número de rodadas do algoritmo de treinamento
pH	potencial hidrogeniônico
S_{sa}	observações amostradas aleatoriamente em cada árvore durante o treinamento
S_z	neurônios na camada oculta
T_{ab}	turbidez da água bruta
t_{maf}	tamanho mínimo da amostra por folha
T	temperatura
V_{max}	valor máximo
V_{min}	valor mínimo
V_{ab}	vazão de água bruta
V_{at}	vazão de água tratada
W_D	decaimento de peso

Sumário

1	Introdução	19
2	Processos de purificação da água	26
2.1	Coleta de água bruta	26
2.2	Purificação em estações de tratamento de água.....	27
2.3	Processo de cloração	28
2.4	Processo de fluoretação	30
2.5	Considerações finais.....	31
3	Inteligência artificial	32
3.1	Ciência de dados	32
3.2	Tecnologia cognitiva	33
3.3	Aprendizado de máquina.....	33
3.4	Sanitização de dados	35
3.5	Processo de otimização	36
3.6	Considerações finais.....	37
4	Metodologia	38
4.1	Contextualização.....	38
4.2	Características desejadas no sistema real	39
4.3	Coleta de dados e pré-processamento	40
4.4	Modelagem computacional e aplicação de aprendizado de máquina	41
4.5	Métricas de avaliação e critérios de convergência	43
4.6	Avaliação estatística dos modelos preditivos	44
4.7	Considerações finais.....	45
5	Resultados	46
5.1	Escolha do sistema para aquisição, tratamento e processamento de dados .	46
5.2	Construção do modelo e validação interna	51
5.3	Otimização de hiperparâmetros	54
5.3.1	Análise da solução com hiperparâmetros otimizados	55
5.3.2	Avaliação de desempenho dos modelos preditivos.....	59
5.4	Discussão.....	63

6	Conclusão	66
6.1	Contribuições diretas do trabalho.....	67
6.1.1	Rotinas desenvolvidas para aplicação na ETA.....	67
6.1.2	Publicação científica	67
6.1.3	Relatório técnico	69
6.2	Trabalhos futuros	69
	Referências	71

Introdução

A importância da água tratada para a vida e saúde da população é indiscutível. Este tratamento envolve a aplicação de processos físico-químicos com o intuito de aprimorar a qualidade da água proveniente de fontes naturais, para que atenda aos padrões regulatórios estabelecidos pela Declaração Universal dos Direitos da Água (Gleick, 1998). Antes de ser distribuída à população, a água passa por um processo de tratamento. Isto se deve ao fato de que a **água bruta**, captada de rios, lagos ou aquíferos, pode estar contaminada por impurezas, microrganismos como bactérias, vírus, fungos, protozoários, além de substâncias poluentes orgânicas e inorgânicas.

O tratamento da água tem como objetivo primário atender às necessidades de consumo humano. Para isto, é necessário melhorar a qualidade microbiológica e controlar a presença de metais e substâncias químicas perigosas. Além disto, existem objetivos secundários, como preservar e proteger os sistemas de distribuição e canalização, bem como gerenciar e monitorar características estéticas da água, como sabor, odor, cor e teor de minerais (Allen, 2008). A água contaminada pode ser veículo de transmissão de diversas doenças, como *cólera*, *febre tifoide*, *giardíase*, *shigelose*, *campilobacteriose*, *salmolenose*, além de infecções virais como *hepatite A*. Estas doenças afetam pessoas de diferentes idades, com elevados índices de morbidade e mortalidade, gerando impactos econômicos na sociedade e sobrecarregando os sistemas de saúde (Momba; Obi; Thompson, 2009).

O processo de desinfecção da água é uma das principais ações realizadas para a preservação da saúde pública, pois tem o objetivo de eliminar agentes patogênicos, bem como impurezas orgânicas e inorgânicas (Galal-Gorchev, 1996). Existem diversas abordagens disponíveis para este fim, abarcando métodos químicos, térmicos, de saturação e físicos. Embora o cloro seja amplamente empregado devido à sua disponibilidade e facilidade de uso, ele gera derivados de subprodutos do cloro, os quais podem ser prejudiciais à saúde (Hrudey, 2009). Entre os derivados mais comuns estão: i) trihalometanos (THM) e ii) ácido haloacéticos (AHA). Para a remoção destes subprodutos, torna-se necessário aplicar métodos adicionais, tais

como aeração, utilização de carvão ativado, emprego de filtros de membrana e adoção de processos de oxidação avançada, entre outros (Ocampo-Rodríguez *et al.*, 2022). Além disto, é necessário administrar o cloro com precisão, visto que o ácido hipocloroso HClO presente neste composto pode ter efeitos citotóxicos se utilizado em excesso. Desta forma, a cloração deve ser realizada com rigor para evitar riscos à saúde e assegurar a eficácia do processo de desinfecção da água (Slaughter *et al.*, 2019).

O emprego de flúor nos sistemas públicos de abastecimento de água teve início a partir da constatação da relação entre o flúor e a saúde bucal, observando que a presença de flúor na água poderia causar manchas nos dentes, mas também proporcionar proteção contra cáries (Murray; World Health Organization *et al.*, 1986). Com base nisto, a prática de fluoretação da água foi implementada em várias regiões, respaldada pela Organização Mundial de Saúde (OMS), e apresentou resultados positivos na redução da necessidade de tratamentos odontológicos (Peres; Fernandes; Peres, 2004). No entanto, o uso de flúor na água tem sido objeto de questionamento devido ao aumento da ocorrência de fluorose dentária em algumas áreas, bem como ao crescimento do consumo de flúor a partir de outras fontes. Por outro lado, a interrupção da fluoretação pode resultar em aumento de casos de cáries, especialmente entre populações menos privilegiadas (Newbrun, 1989). Assim, a fluoretação da água é medida de saúde pública amplamente reconhecida, mas que requer vigilância e equilíbrio para evitar efeitos adversos (Charrois, 2010).

O desempenho das pequenas e médias Estação de Tratamento de Água (ETA) na garantia dos padrões de potabilidade da água para os consumidores finais frequentemente enfrenta desafios, principalmente relacionados à produção inadequada ou fora dos padrões estabelecidos, utilização de tecnologias desatualizadas, operação eficaz, falta de capacitação e desmotivação dos operadores, limitações financeiras e falta de conhecimento das operações básicas de tratamento da água. Aspectos tanto técnicos quanto não técnicos, referentes à operação, manutenção e gestão, têm impacto direto na eficácia e sustentabilidade destes sistemas (Makungo; Odiyo; Tshidzumba, 2011). Os sistemas de tratamento de água devem assegurar o produto em conformidade com os regulamentos de potabilidade estabelecidos, priorizando a satisfação do consumidor. A água deve ter qualidade e pressão adequadas a qualquer momento e para todos os consumidores. No entanto, o desafio para os gestores de ETA reside em equilibrar custos e benefícios de forma sustentável, visando a gestão eficaz e responsável (Kučera; Tuhovčák; Biela, 2017).

A avaliação da qualidade da água destinada ao consumo humano requer a análise de variedade de fatores relevantes (Doria, 2010). Esta avaliação abrange ampla gama de variáveis, como turbidez, cor aparente, pH , cloretos, índice de

coliformes totais (ICT) e índice de coliformes fecais (ICF), além de considerar sabor, odor, temperatura, alcalinidade, acidez, presença de ferro, manganês e nitrogênio. Não há padronização aplicável a todos os sistemas hídricos e o propósito do índice de qualidade da água (IQA) é estabelecer correlações entre algumas destas variáveis, com o intuito de detectar alterações, sejam elas ocasionadas por ações humanas ou por processos naturais, que possam afetar a fonte de água (World Health Organization, 2004; World Health Organization, 2018). Assim, o IQA deve permitir a identificação de condições endêmicas, sendo medida adimensional capaz de representar tanto a realidade anual quanto a sazonal (Giri; Qiu, 2016).

Além disto, a legislação determina que é responsabilidade do gestor do sistema de abastecimento de água para consumo humano assegurar a conformidade com os regulamentos estabelecidos (World Health Organization, 2018; Marin *et al.*, 2022; Silverio *et al.*, 2023). Com o objetivo de aprimorar a qualidade da água, os provedores e concessionárias têm recorrido a modelos de inteligência artificial (IA) para garantir que o maior número possível de amostras receba a quantidade correta de produtos químicos, de forma a atender aos padrões regulatórios, principalmente com o uso de redes neurais artificiais (RNA), Floresta Randômica (FR), redes neurais profundas (RNP), *multi-layer perceptron* (MLP), regressão de vetor de suporte (RVS), rede neural recorrente (RNR) *long short-term memory* (LSTM) multivariada, *kernel nearest neighbor* (KNN), *Gradient Boosting Classifier* (GBC) e *Gradient Boost Tree* (GBT). O estudo de Zhang *et al.* (2019) descreve que regulamentações rigorosas e a deterioração das fontes de água afetam a capacidade das ETA de produzir água de alta qualidade. Os autores empregam modelos para prever o desempenho das ETA sob condições de estresse, fornecendo percepções valiosas para a tomada de decisões e o planejamento futuro. Solaimany-Aminabad, Maleki e Hadi (2013) implementam modelo de previsão em ETA que analisa a qualidade da água de entrada, criando base de dados para orientar a operação do processo. Este modelo utiliza RNA e oferece análises e diagnósticos para compreender e simular o comportamento não linear das características da água de entrada $\times$ água de saída, sendo ferramenta valiosa para operadores e tomadores de decisão nas ETA.

Pinto *et al.* (2009) aborda a problemática dos elevados níveis de magnésio e turbidez encontrados na água tratada da cidade portuguesa de Évora. Para solucionar esta questão, os autores propõem metodologia com base no uso de conjunto de mineração de dados e sistemas de suporte à decisão. O processo envolve a coleta regular de amostras, a realização de análises químicas e a tabulação dos dados obtidos. Em seguida, os dados são tratados, normalizados e aplicados em algoritmos, como: árvore de decisão e o método de clusterização *K*-Means. O conjunto de dados é dividido, reservando 2/3 para o treinamento dos modelos e utilizando 1/3 para teste

a acurácia. Os resultados obtidos pelos autores permitem prever as condições que levam a este problema específico, permitindo a implementação de ações corretivas com menor custo, maior eficácia e a garantia de melhoria na qualidade final do produto tratado.

Librantz, Santos e Dias (2018) propõem a utilização de RNA para controlar a quantidade de produto destinado à cloração em estação de tratamento de água. O problema identificado é a presença de transientes de cloro residual livre (CRL) durante períodos de chuva, os quais são desafiadores de ajustar manualmente. Os autores definem três cenários, com as variáveis de saída sendo o **objetivo pós-cloração** e o **volume de produto para cloração**. São treinadas RNA com base no critério de erro médio quadrático (EMQ), e os cenários são comparados entre si e com a situação atual utilizando o erro médio absoluto (EMA). Os resultados obtidos nos três cenários são passíveis de aplicação prática, possibilitando maior economia de produtos químicos, reduzindo a necessidade de intervenções humanas e, conseqüentemente, diminuindo os custos operacionais.

Khan *et al.* (2021) propõem metodologias para a predição e classificação do IQA. Inicialmente, coletam dados do lago Gulshan, em Bangladesh, calculam o IQA e, em seguida, realizam o pré-processamento com redução de dimensionalidade e segmentação dos dados. Após o pré-processamento, aplicam algoritmos de IA para obter os dados preditos e classificados. Consideram nove parâmetros: *pH*, oxigênio dissolvido, demanda química de oxigênio, total de sólidos dissolvidos, condutividade elétrica, turbidez, cloração, sólidos suspensos e alcalinidade. Os resultados experimentais demonstram que os métodos de predição e classificação utilizando algoritmos de aprendizado de máquina (AM) alcançam os resultados desejados. Em particular, a RVS destaca-se como o melhor algoritmo de regressão, enquanto o GBC apresenta-se o mais eficaz para a tarefa de classificação.

O trabalho de Wang *et al.* (2021) tem como intuito otimizar plantas industriais de tratamento de água por meio do uso de IA. A metodologia adotada considera como processos independentes a captação de água, a vazão volumétrica de reagentes, a lavagem de filtros, o descarte de lodo e a distribuição de água. O emprego de modelo preditivo com base em FR e RVS, aliado à disponibilidade de elevada quantidade de dados históricos empíricos, permite que o sistema ajuste a quantidade de químicos, minimizando a ocorrência de erros operacionais humanos. O modelo proposto demonstra eficácia na gestão dos equipamentos na planta industrial. Isto valida a efetividade da combinação entre teoria e o sistema de controle existente, destacando-se como abordagem eficaz para otimizar o funcionamento das plantas de tratamento de água.

More e Wolkersdorfer (2023) demonstram que diversos modelos tradicionais

falham ao utilizar todos os parâmetros disponíveis para prever outros parâmetros em sistemas de tratamento de água subterrânea. O trabalho se destaca pelo emprego de técnicas estatísticas avançadas associadas ao AM. São comparados o desempenho de modelos como RNR-LTSM, RNA, RNP, FR e GBT. Além destas técnicas, o estudo incorpora tratamentos para estimar dados faltantes, remoção de anomalias de dados e correlação de parâmetros para a seleção dos mais relevantes. Os modelos são avaliados utilizando métricas como EMQ e EMA, com o objetivo de prever a concentração de ferro (Fe) e a acidez com base em dados de alcalinidade, pH e sulfato. O algoritmo com melhor desempenho é o FR, indicando que o uso de AM para a otimização do tratamento de água apresenta resultados desejados. Adicionalmente, o estudo ressalta a necessidade frequente de transformar séries temporais para utilizá-las nos modelos, a fim de obter previsões precisas.

Pai *et al.* (2023) desenvolveram modelos computacionais avançados de inteligência artificial para correlacionar o volume de flúor ingerido com a ocorrência de cáries, utilizando algoritmos de classificação RVS e KNN. Aoun *et al.* (2018) realizaram uma revisão sistemática sobre a eficácia da fluoretação, considerando a adição e ajuste de concentrações em leite, sal e água. Destaca-se que a recomendação da OMS para fluoretar a água de abastecimento público é adotada em países como Estados Unidos, Brasil, Austrália, Canadá, Espanha, Argentina, Coreia do Sul e Nova Zelândia. Além disso, outros estudos empregam inteligência artificial para controle das concentrações de flúor em águas naturalmente fluoretadas, provenientes principalmente de fontes subterrâneas (Nadiri *et al.*, 2013; Khan; Gupta; Sekhri, 2021).

Diversos trabalhos exploram a aplicação de IA no tratamento de água, com o objetivo de aprimorar a qualidade de forma sustentável. A Tabela 1 sintetiza os principais estudos nesta área, destacando a utilização de IA em diferentes contextos. Este trabalho busca preencher a lacuna existente na literatura ao propor o uso de modelos de IA para predição da vazão volumétrica de cloro e flúor adicionada a sistemas de tratamento de água potável provenientes de captação superficial.

Tabela 1 – *Síntese de estudos que empregam modelos de inteligência artificial no tratamento de água*

Autor	Localização	Variáveis	Amostras	Método utilizado
Pinto <i>et al.</i> (2009)	Portugal	35	642	C5.0
Librantz, Santos e Dias (2018)	Brasil	8	18.035	RNA
Khan <i>et al.</i> (2021)	Bangladesh	9	108	RVS e GBC
Wang <i>et al.</i> (2021)	China	5	20	FR e RVS
More e Wolkersdorfer (2023)	África do Sul	9	1.123	RNR-LSTM, RNA, RNP, FR e GBT

A aplicação de AM e inteligência artificial no tratamento de água demonstra resiliência e proporciona redução de custos, viabilizando a universalização do acesso à água potável, especialmente para populações vulneráveis. Embora existam estudos na literatura que abordam o uso de IA em estações de tratamento de água, há a lacuna em pesquisas voltadas a prever a vazão volumétrica de cloro e flúor adicionado em estações de captação superficial, o que reforça a justificativa deste trabalho. A hipótese formulada é que a implementação de modelo preditivo de regressão supervisionado com base em séries temporais melhora a precisão das previsões de vazão volumétrica de cloro e flúor, reduzindo os erros de previsão e garantindo maior segurança, qualidade e eficácia operacional.

O uso de modelos computacionais ajustados com abordagens matemático-estatísticas representam avanço significativo, uma vez que os custos computacionais são substancialmente menores em comparação às alterações operacionais na planta industrial. Além disto, as modificações computacionais são mais ágeis, permitindo prever valores e comparar a acurácia das previsões com o sistema implantado, apresentando a relevância do estudo. Adicionalmente, várias estações de tratamento de água em países em desenvolvimento compartilham características semelhantes de planta industrial, ampliando a aplicabilidade e o impacto potencial deste trabalho.

A originalidade desta pesquisa reside na elaboração de um modelo preditivo capaz de simular o comportamento da ETA de captação superficial com maior precisão, considerando um volume de amostras maior do que o utilizado em estudos anteriores. A inovação do estudo está na aplicação de inteligência artificial para predição no setor de saneamento, um campo que historicamente apresenta pouca adoção destas tecnologias.

O objetivo principal desta pesquisa é apresentar o modelo computacional preditivo de regressão supervisionado, guiado por séries temporais, que possibilite a otimização dos processos de cloração e fluoretação em determinada estação de tratamento de água superficial. Adicionalmente, os objetivos específicos incluem: i) alcançar a redução dos custos operacionais e da qualidade da água, ii) obter valores de dosagens de produtos químicos críticos para o tratamento de água e seguros para a saúde da população, iii) aprimorar o processo para garantir a qualidade do produto final e iv) contribuir para o alcance do Objetivo de Desenvolvimento Sustentável n.º 6 (ODS 6), estabelecido na Agenda 2030 pela Organização das Nações Unidas (ONU).

A motivação ao desenvolvimento deste trabalho surge da crescente importância de assegurar o acesso à água potável de qualidade, recurso necessário para a vida humana e para o alcance dos objetivos de desenvolvimento sustentável (UN-Water, 2021). O propósito é impactar positivamente a sociedade, resultando na

redução de custos operacionais, na preservação da saúde da população e na contribuição para a sustentabilidade ambiental. Desta forma, alinha-se ao compromisso global de assegurar o acesso universal a água limpa e segura.

A estrutura deste trabalho é organizada na forma: no Capítulo 2, são brevemente apresentados os processos de purificação da água, enquanto no Capítulo 3, são abordados conceitos fundamentais sobre IA. Estes capítulos fornecem base necessária para compreender a metodologia proposta e os resultados. No Capítulo 4, detalha-se a metodologia proposta, enquanto o Capítulo 5 apresenta os resultados e as discussões correspondentes. Finalmente, no Capítulo 6, são apresentadas as conclusões, nas quais são resumidas as principais contribuições deste estudo.

Processos de purificação da água

Neste capítulo são abordados brevemente os processos fundamentais de purificação da água, abarcando coleta de água bruta e procedimentos executados nas estações de tratamento de água para garantir a potabilidade e segurança para o consumo humano. São discutidas as etapas de oxidação por meio da cloração, importante na eliminação de microrganismos e impurezas, assim como a fluoretação, procedimento adicional voltado para o fortalecimento da saúde dental da população.

2.1 Coleta de água bruta

A captação da água ocorre predominantemente em mananciais superficiais, meteóricos ou subterrâneos. Os mananciais superficiais incluem córregos, rios, lagos e represas, enquanto os meteóricos são formados por eventos climáticos como precipitações, derretimento de neve e granizo. Os mananciais subterrâneos são obtidos dos aquíferos freáticos e artesianos. Enquanto os mananciais superficiais têm fluxo rápido através de cursos de água, os subterrâneos têm velocidade menor, pois precisam atravessar formações geológicas sedimentares, ígneas ou metamórficas. Nas formações sedimentares, o veio de água transpõe poros e interstícios, enquanto nas formações ígneas e metamórficas, a água flui por planos de fraqueza estrutural (Tsutiya, 2006).

Exceto pelas águas meteóricas, a captação de outras fontes requer autorização governamental intitulada de outorga. Além disto, é necessário considerar vários aspectos relacionados ao Sistema de Abastecimento de Água (SAA), que deve ser guiado pelos princípios de eficácia, eficiência e economicidade durante o período estipulado. O planejamento do SAA envolve a análise de dados populacionais, incluindo o cálculo da população total atendida pelo projeto. Neste cálculo, é necessário distinguir entre a população residente, composta por aqueles que têm a localidade como residência permanente, a população flutuante, que vem de outras comunidades e ocasionalmente se transfere para a área em estudo, consumindo água na taxa equivalente à da população residente e a população temporária, que também vem de

outras comunidades, mas consome água a taxa inferior à da população residente. Assim, é possível determinar a população atendida pelo SAA (FUNASA; UFF; IBAM, 2016).

Estimar a população residente requer análises estatísticas dos dados demográficos disponíveis, enquanto as populações flutuantes e temporárias são avaliadas de acordo com os padrões locais. O consumo *per capita* é determinado levando em conta a demanda projetada, bem como fatores socioeconômicos e ambientais. Diretrizes da Associação Brasileira de Normas Técnicas (ABNT) e da OMS oferecem orientações sobre a demanda mínima de água por habitante. Portanto, é importante equilibrar a demanda e o consumo, considerando as variações anuais e sazonais, com aumento anual de 1% e ajustes sazonais conforme o clima. Encontrar estas variações é necessário para calcular coeficientes importantes no dimensionamento das vazões. A demanda de água potável do SAA deve levar em conta a vazão média diária, a vazão máxima diária e a vazão máxima horária (Guidi, 2017).

Na captação superficial, diversas estruturas são construídas junto ao manancial para viabilizar a extração da água. Para garantir o sucesso, é necessário considerar critérios como variação do nível de água, coerência entre vazão de captação e vazão do manancial, concentração de sólidos em suspensão e seleção do local de captação (Gudina, 2017). Para a captação subterrânea requer análise e autorização da Agência Nacional de Águas (ANA), assim como avaliação físico-química e microbiológica mais detalhada. Em geral, as águas meteóricas não atendem às demandas do SAA e, portanto, não são utilizadas. Em todos os casos, são construídas adutoras de água bruta (AAB) para transportar a água até a ETA, seja por gravidade, recalque ou sistema misto (FUNASA; UFF; IBAM, 2016).

2.2 Purificação em estações de tratamento de água

Desde antes de 500a.C., há registros da preocupação com a qualidade da água, mas é após esta época que os egípcios começam a utilizar produtos químicos, como Sulfato de Ferro FeSO_4 e Sulfato de Alumínio $\text{Al}_2(\text{SO}_4)_3$, para melhorar tanto o sabor quanto a aparência da água (Namor, 2007). Por outro lado, na antiga Índia, eram empregadas plantas, ervas, além de métodos como fervura, exposição à luz solar e até mesmo a utilização de cobre e ferro para purificar a água. Os processos modernos de tratamento de água em estações de tratamento variam conforme as características da água e os padrões de qualidade exigidos.

Estes processos incluem etapas como coagulação, floculação, decantação, fil-

tração, ajuste do pH , aeração, adsorção em carvão ativado e desinfecção com agentes como cloro, ozônio O_3 ou luz ultravioleta. Tais medidas visam remover impurezas e garantir a segurança do consumo humano. Além dos métodos convencionais, existem abordagens avançadas, como a oxidação eletroquímica, ultrassom, fotocatalise, micro-ondas e radiação ionizante, que oferecem alternativas para garantir a potabilidade da água. Estes processos são adaptados conforme as necessidades específicas de cada estação de tratamento, buscando garantir a qualidade e a segurança do abastecimento de água (Cescon; Jiang, 2020; Miklos *et al.*, 2018).

2.3 Processo de cloração

Dentro do conjunto de procedimentos realizados na estação de tratamento de água, a desinfecção da água é uma das fases mais importantes. A remoção de agentes patogênicos da água destinada ao abastecimento da população é o elemento principal para assegurar a preservação da saúde (Hofkes, 1981; Namor, 2007). De acordo com Ocampo-Rodríguez *et al.* (Ocampo-Rodríguez *et al.*, 2022), nos últimos anos têm sido empregados diversos métodos para a desinfecção da água, agrupados predominantemente em quatro categorias: i) métodos químicos que envolvem o uso de oxidantes potentes, ii) métodos térmicos como a pasteurização, iii) saturação utilizando íons de metais não ferrosos como prata, cobre e propriedades bactericidas e iv) métodos físicos que incluem a aplicação de radiação ultravioleta, ultrassom ou radiações ionizantes (Negoitescu; Tokar, 2012). Além destes, existem outros métodos, como o uso de ozônio (Faust; Aly, 1998), e a ultrafiltração utilizando membranas (LAINÉ; VIAL; MOULART, 2000). A desinfecção de águas destinadas ao consumo humano busca reduzir significativamente a presença de microrganismos, em vez de atingir esterilização completa (Faust; Aly, 1998).

Segundo Miklos *et al.* (2018), os processos de oxidação avançados (POA) podem ser categorizados em cinco grupos principais: i) com base em ozônio, ii) com base em radiação ultravioleta, iii) processo avançado de oxidação eletroquímica, iv) processo avançado de oxidação catalítica e v) processo avançado de oxidação física. Todos estes métodos envolvem a geração *in loco* do agente oxidante, que então reage com os contaminantes. Esta característica de produção *in loco* demanda consumo elevado de energia e utiliza ampla gama de processos físico-químicos para a remoção eficaz de matéria orgânica, minimizando a formação de derivados de subprodutos do cloro. Em todos os POA, o objetivo é gerar o oxidante *in loco* e promover a reação com os contaminantes alvo (Kim; Zoh, 2016).

O estudo de Ocampo-Rodríguez *et al.* (2022) descreve o uso de cloro e seus derivados para a desinfecção da água destinada ao abastecimento é amplamente

empregado em escala global, presente em mais de dois terços dos sistemas de tratamento de água do planeta. Esta preferência advém de seu custo financeiro e facilidade de manuseio (Chhetri *et al.*, 2018). A eficácia deste método de tratamento pode ser avaliada pela redução ou eliminação da presença de microrganismos, principalmente *Escherichia coli* e *Enterococcus* spp, que servem como indicadores de contaminação fecal. Entretanto, os desinfetantes à base de cloro e seus derivados reagem com a matéria orgânica e inorgânica presentes na água, gerando derivados de subprodutos do cloro que são prejudiciais à saúde e desafiadores de eliminar devido à sua diversidade química, física e biológica. Estas substâncias, associadas a doenças como o câncer, são principalmente produzidas pela reação do cloro com materiais orgânicos. Em águas com elevada concentração de compostos orgânicos, formam-se nitritos, cloratos, fenacetina, acetanilida, sulfanilamida, nitrobenzeno e quinonas, substâncias especialmente nocivas para crianças, com potencial de causar metemoglobinemia e/ou sulfemoglobinemia, condição conhecida como síndrome do bebê azul (Fernícola; Azevedo, 1981; Lukens, 1987; Sadeq *et al.*, 2008; Johnson, 2019). A remoção destes subprodutos pode ser alcançada por meio de processos como osmose reversa, troca iônica e adsorção. No entanto, o mecanismo mais comumente empregado é a prefiltração com carvão ativado e trocadores iônicos magnéticos (Zou; Tang, 2019).

Para Ocampo-Rodríguez *et al.* (2022), o mecanismo de desinfecção microbiológica do cloro consiste em ocasionar danos às membranas cito-celulares, interrompendo as funções celulares e afetando o ácido desoxirribonucleico (ADN), além de ser absorvido pelos biofilmes e eliminar as bactérias interiores (Silva *et al.*, 2020). Entre as várias formas de oxidantes à base de cloro, o HClO destaca-se por sua capacidade de degradar proteínas e destruir as enzimas envolvidas no metabolismo de carboidratos, oxidando grupos sulfídricos. Por sua vez, o dióxido de cloro ClO₂ atua de maneira seletiva, permeabilizando as membranas celulares e interrompendo a síntese proteica, favorecendo a perda de nutrientes vitais e, conseqüentemente, levando à morte celular. A escolha pelo uso de cloro é justificada pela ampla eficiência na inativação de patógenos, na proteção da saúde humana e pelo seu efeito residual (Destiani; Templeton, 2019).

A pesquisa conduzida por Slaughter *et al.* (2019) investiga o hipoclorito de sódio quando dissolvido em água, observando maior presença de HClO em relação ao íon hipoclorito, sendo este equilíbrio dependente do *pH* da solução. Quando em contato com mucosas, como as da boca, esôfago, estômago e trato respiratório, o HClO se converte rapidamente em ácido clorídrico e íons de oxigênio, resultando na morte de tecidos e induzindo intoxicação por citotoxicidade. A ingestão de hipoclorito de sódio em concentrações diluídas raramente é clinicamente

relevante, embora possa causar desconfortos digestivos, como dores de estômago e vômitos (Rhee; Beaumont, 1990). Na ausência de antídoto específico é necessário o tratamento primário focado na redução dos sintomas observados. Diante destes aspectos, destaca-se a importância de utilizar o cloro como oxidante na quantidade adequada, visto que quantidade insuficiente não atende ao propósito de sanitização, enquanto em excesso pode resultar em envenenamento (Martinez; Long, 1995).

O processo de cloração, fundamental para garantir a qualidade microbiológica da água, apresenta variáveis críticas que precisam ser controladas de forma precisa para evitar riscos à saúde pública e desperdícios operacionais. Assim, torna-se relevante adotar ferramentas capazes de prever e otimizar os níveis de dosagem, permitindo a operação segura e eficaz.

2.4 Processo de fluoretação

O flúor é frequentemente adicionado à água destinada ao consumo humano como medida de saúde pública para prevenir a cárie dentária. A prática de controlar a adição de flúor à água está estabelecida em vários países (Kaur *et al.*, 2021). O primeiro indício da relação entre flúor e saúde dental surge em 1911, quando o dentista McKay observa que crianças em Colorado Springs/EUA, têm dentes manchados de marrom, enquanto as crianças das cidades vizinhas não têm estas manchas. McKay suspeita que a causa das manchas esteja na água que as crianças consomem. Em 1916, McKay e Black confirmam a hipótese ao analisarem a água de abastecimento de Colorado Springs, encontrando níveis elevados de flúor. Em 1928, McKay sugere que, além de causar as manchas, o flúor também pode proteger os dentes contra cáries (Murray; World Health Organization *et al.*, 1986). Em 1931, o químico Petrey corrobora a ideia ao observar que as crianças com dentes mais manchados são as que têm menos cáries, pois bebem água com maior teor de flúor (Mullen, 2005).

Desde então, vários estudos têm surgido, associando as doses seguras de flúor para a promoção da saúde bucal da população. Esta iniciativa é reconhecida e incentivada pela OMS como medida terapêutica cientificamente comprovada, eficiente e segura. A recomendação da fluoretação da água é apresentada na 25^a Assembleia Mundial de Saúde, em 1975, e aprovada por 148 membros (Mullen, 2005; Newbrun, 1989). Atualmente, vários países incorporam a fluoretação das águas destinadas ao abastecimento humano, incluindo Estados Unidos, Brasil, Austrália, Canadá, Espanha, Coreia do Sul e Nova Zelândia (Mullane *et al.*, 2016; Garbin *et al.*, 2017). A prática da fluoretação é mantida, geralmente, com o objetivo de assegurar a conformidade por meio do heterocontrole exercido por entidades/instituições

reguladoras e pela vigilância em saúde pública. Entretanto, a responsabilidade primária recai sobre os operadores dos sistemas de abastecimento público (Peres; Fernandes; Peres, 2004; Charrois, 2010).

O estudo de Roy e Dass (2013), realizado na Índia, evidencia diversos problemas associados ao consumo excessivo de flúor, frequentemente decorrentes do uso de fontes subterrâneas sem tratamento adequado e da ingestão de fontes alternativas, como chás. De modo geral, a exposição crônica ao flúor causa efeitos prejudiciais, incluindo deformações severas e permanentes nos ossos e articulações (osteoclerose). Ressalta-se que a fluorose é uma condição irreversível e não possui tratamento.

O estudo conduzido por Levy *et al.* (2023) compara a incidência de cárie em jovens adultos não expostos ao abastecimento de água fluoretada e que não passaram por tratamentos dentais, com outros jovens adultos expostos à água fluoretada, seja com ou sem tratamentos dentais associados. Os autores observam correlação significativa entre a fluoretação obrigatória da água para consumo humano e a redução da necessidade de tratamentos dentários. De acordo com Charrois (Charrois, 2010), a melhoria das condições de higiene bucal e dentária na população tem levado à diminuição da adição de flúor na água potável em diversos países, mesmo sem considerar as mudanças nos hábitos alimentares. A OMS alerta que concentrações de flúor superiores a 1,5mg por litro podem levar ao desenvolvimento de fluorose dentária. A fluoretação da água de abastecimento público é medida de saúde pública de baixo custo, ampla abrangência populacional, segura e eficaz no controle e redução da doença cárie (Peres; Fernandes; Peres, 2004; Newbrun, 1989; Charrois, 2010).

A fluoretação, embora eficaz na prevenção de doenças bucais, exige controle rigoroso para evitar concentrações inadequadas, que podem ser prejudiciais à saúde. Desta forma, o uso de técnicas avançadas de modelagem e predição é uma alternativa promissora para assegurar a dosagem correta e reforçar a eficiência operacional das estações de tratamento de água.

2.5 Considerações finais

Neste capítulo foram apresentados conceitos relacionados aos pontos de atenção para concepção e manutenção do SAA. Além dos diferentes tipos de mananciais e etapas do processo de tratamento da água para abastecimento público humano, foram enfatizados os processos de cloração e fluoretação, seus benefícios à saúde e os riscos de dosagens em quantidades inadequadas. Estes conceitos são importantes para o correto ajuste dos modelos de IA apresentados no próximo capítulo.

Inteligência artificial

Neste capítulo são explorados os fundamentos da inteligência artificial, com foco nos pilares da ciência de dados, contemplando o aprendizado de máquina e processos de otimização. É descrita brevemente a importância da ciência de dados como base para a construção e análise de modelos inteligentes, destacando a coleta, organização e análise de grandes volumes de dados, bem como os benefícios do uso da inteligência artificial. Por fim, discorre-se sobre o funcionamento do sistema, os processos de otimização aplicados aos modelos de inteligência artificial.

3.1 Ciência de dados

A ciência de dados é área multidisciplinar que dedica-se a extrair conhecimento e percepções de conjuntos de dados que encerra várias atributos ou ideias (Al-Haija, 2022), utilizando variedade de técnicas estatísticas, computacionais e de visualização (Blei; Smyth, 2017). Este enfoque integrado abarca desde a coleta e processamento até a análise de elevados volumes de dados provenientes de diversas fontes. Profissionais de ciência de dados empregam ferramentas como Python, R e bibliotecas especializadas, como scikit-learn, caret and scipy, além de técnicas avançadas de estatística, para identificar padrões, tendências e relações nos dados, contribuindo para a tomada de decisões informadas (Riepl *et al.*, 2010).

O termo ciência de dados tem atraído muita atenção do mercado, do governo e de áreas de estudo acadêmico estatístico. Uma mistura de estatística e ciência de computação focada no contexto e necessidades adequadas a uma análise de dados moderna, não é um conceito novo, tendo referências desde os anos 50 como análise de dados. Como todos os conjuntos de dados envolvem incerteza, a estatística é utilizada para auxiliar na compreensão de dados complexos, a casualidade destes e reduzir a alta dimensionalidade, enquanto a ciência de computação procura testar exaustivamente estes dados para obter modelos preditivos ou classificativos para otimizar o comportamento esperado, evitando *underfit* e *overfit* (Blei; Smyth, 2017).

A geração de dados na área do saneamento sofreu um forte incremento desde 2010, grande parte devido à instalação de sensores automatizados com registro das informações em sistemas de tratamento de água potável e esgoto. Isso possibilitou a criação de diversos conjuntos de dados onde é possível aplicar técnicas de ciência de dados para predição e classificação, sendo estas otimizações fatores de grande relevância, uma vez que a falta de acesso a água segura causou a morte prematura de 1 milhão e 200 mil pessoas em 2017 (Wongburi; Park, 2021).

3.2 Tecnologia cognitiva

A definição de tecnologia cognitiva ou inteligência artificial, pode abarcar sistemas que reproduzem o pensamento e comportamento humano, ou que agem de forma racional, dependendo do autor. O teste de Alan Turing em 1950, é utilizado para avaliar a inteligência de sistemas, considerando aspectos como processamento de linguagem natural, representação do conhecimento, raciocínio automatizado, aprendizado de máquina, visão computacional e robótica. Esta é a área de estudo relativamente recente, que se beneficia do avanço da capacidade computacional, evolução da ciência da computação e disponibilidade de grandes bases de dados para modelagem e treinamento (Russel; Norving, 1995).

Ludermir (2021) categoriza a IA em três tipos: focada, generalizada e superinteligente. A IA focada, ou fraca, tem como base os algoritmos complexos para resolver problemas específicos. A IA generalizada, ou forte, busca igualar a eficiência das habilidades humanas em diversas áreas, como visão computacional. Por fim, a superinteligência é o conceito no qual os algoritmos ultrapassam as habilidades humanas, porém sistemas deste tipo ainda não existem. O uso da IA tem transformado o cotidiano, influenciando várias interações humanas e suscitando questionamentos éticos e morais.

3.3 Aprendizado de máquina

O AM é subcategoria da IA que desenvolve algoritmos capazes de identificar padrões automaticamente a partir de dados e melhorar seu desempenho com o tempo (Fedorko; Král'; Fedorko, 2022). Em algoritmos de aprendizado supervisionado, modelos são treinados com conjuntos de dados rotulados, permitindo que generalizem padrões e façam previsões precisas sobre novos dados (Ghahramani, 2015). Algoritmos não supervisionados exploram dados não rotulados para identificar padrões naturais, enquanto algoritmos por reforço envolvem agentes que tomam decisões sequenciais em ambientes dinâmicos, recebendo opiniões na forma

de recompensas ou penalidades (Morales; Escalante, 2022). Estas abordagens são aplicadas em tarefas como classificação, clusterização, regressão e otimização, respectivamente (Jolly, 2018).

Com a crescente disponibilidade de conjuntos de dados massivos e o avanço das técnicas de AM, incluindo redes neurais profundas, a capacidade de criar modelos preditivos precisos tem se destacado. Dentro do AM, as RNA, especialmente as MLP, são importantes. As MLP são arquiteturas de redes neurais profundas (Chandra *et al.*, 2021), compostas por camadas de neurônios organizados em múltiplas camadas, incluindo: entrada, ocultas e saída (Dayhoff, 1990). Cada conexão entre neurônios possui um peso associado e funções de ativação não lineares são aplicadas para introduzir não linearidades. Esta flexibilidade faz das MLP componentes fundamentais em diversas aplicações de AM. Apesar de sua eficácia, o treinamento de MLP pode ser desafiador em conjuntos de dados extensos, exigindo técnicas avançadas para otimizar seu desempenho em cenários práticos (Lashari *et al.*, 2019).

A aplicação de RNA-MLP na modelagem de tratamento e qualidade da água representa avanço considerável no campo da engenharia ambiental (Aliashrafi *et al.*, 2021). As RNA-MLP são arquiteturas capazes de capturar relações difíceis em conjuntos de dados hidroambientais. Ao integrar dados de parâmetros de tratamento, características hidrológicas e indicadores de qualidade da água na RNA-MLP, é possível desenvolver modelos preditivos precisos para otimizar processos de tratamento, prever concentrações de poluentes e antecipar mudanças na qualidade da água (Tahershamsi; Menhaj; Ahmadian, 2006; Gourmelon *et al.*, 2010). Esta abordagem avançada oferece uma ferramenta valiosa, permitindo a tomada de decisão eficaz e proativa no monitoramento e controle dos sistemas de tratamento de água (Jayaweera; Aziz, 2018; Ghobadi *et al.*, 2022; Safeer *et al.*, 2022).

Em aprendizado de máquina, uma técnica amplamente utilizada devido à sua versatilidade em trabalhar tanto com regressão quanto com classificação é a FR (Breiman, 2001). Nesta abordagem, várias árvores de decisão são combinadas durante o treinamento, e cada árvore na floresta opera de forma independente. A combinação de suas previsões é empregada para aprimorar a precisão e o desempenho do modelo. No contexto do aprendizado supervisionado, a FR é frequentemente escolhida por ser um método de propósito geral que não exige parametrização (Jena *et al.*, 2022). Os dados de treinamento são subdivididos em subconjuntos menores, organizados em árvores, e os melhores são selecionados para formar novos agrupamentos. Este procedimento é realizado exaustivamente, reduzindo erros e minimizando a influência de ruídos e valores atípicos. A adoção desta técnica é apropriada ao lidar com conjuntos de dados não lineares (Jenkins; Kowalski; Alvarenga, 2018).

O *Extreme Gradient Boosting* (XGBoost) também é técnica de aprendizado de máquina da categoria de métodos *ensemble*, usada em classificação e regressão. Utilizando *gradient boosting*, o algoritmo constrói sequencialmente árvores para corrigir erros residuais, otimizando a função de perda. Esta abordagem, guiada por dados, é amplamente adotada na ciência de dados, proporcionando resultados eficientes (Chen; Guestrin, 2016). O XGBoost utiliza paralelismo e otimização de recursos de memória, dividindo o conjunto de treinamento em árvores e ajustando os dados até que os cruzamentos resultem no menor valor da função objetivo. Ao final do treinamento, as pontuações das árvores restantes são calculadas nas folhas, gerando o modelo do valor predito (Li *et al.*, 2022). Em sistemas de tratamento de água, caracterizados por não linearidade dura, técnicas de aprendizado de máquina, como FR (Breiman, 2001) e XGBoost (Mo; Li; Li, 2023; More; Wolkersdorfer, 2023), são apropriadas.

3.4 Sanitização de dados

A maioria dos estudos que utiliza dados provenientes de anotações, sistemas computacionais, fontes informativas e outras origens requer tratamento adequado para identificar valores anômalos e ausentes, os quais podem impactar significativamente os resultados de pesquisas científicas. A sanitização dos conjuntos de dados constitui uma etapa fundamental na ciência de dados, pois afeta diretamente a estatística descritiva, influenciando os modelos e podendo introduzir tendências a comportamentos anômalos.

Scheffer (2002) identifica três categorias principais de ausência de dados: i) ausência completamente aleatória, ii) ausência ao acaso e iii) ausência não aleatória. A ausência completamente aleatória ocorre quando as lacunas são independentes tanto das variáveis de interesse quanto de quaisquer outras variáveis do conjunto de dados. A ausência ao acaso ocorre quando os dados ausentes estão associados a uma variável que não é de interesse. Por fim, a ausência não aleatória refere-se a dados ausentes que dependem diretamente da variável de interesse. Técnicas comuns, como a exclusão de casos com dados ausentes ou a imputação pela média, são consideradas prejudiciais, pois reduzem a quantidade de dados e comprometem a variância da amostra. Para lidar com lacunas, recomenda-se o uso de algoritmos de regressão ajustados à natureza dos dados e a aplicação de imputação múltipla, utilizando modelos de análise adequados.

Blázquez-García *et al.* (2021) aborda a identificação e o tratamento de valores atípicos em séries temporais. Quando uma anomalia é detectada, é necessário determinar se o dado é relevante e merece análise ou se pode ser descartado por sua

irrelevância. Valores anômalos podem ser pontuais, influenciar outras variáveis em uma sequência ou apresentar desvios em relação a variáveis de uma série temporal. A detecção pode ser realizada utilizando histogramas, análises de densidade, modelos de estimativa ou predição e técnicas com base em dissimilaridade¹.

A análise de séries temporais no contexto do tratamento de água enfrenta desafios adicionais devido à elevada complexidade dos sistemas, que resulta de interações não lineares entre variáveis químicas, físicas e operacionais (Bulhoes *et al.*, 2020). Estas interações dificultam o monitoramento contínuo e preciso, especialmente diante de lacunas nos dados, frequentemente ocasionadas por falhas de medição ou interrupções na coleta. Tais problemas podem comprometer a eficiência dos modelos preditivos, exigindo a aplicação de métodos avançados de preenchimento e análise (Wu *et al.*, 2022). Neste cenário, as abordagens ensemble, que combinam múltiplos algoritmos, destacam-se por oferecer previsões mais precisas e resilientes (Rodriguez *et al.*, 2021; Choi; Lim; Ji, 2023).

A interpolação/extrapolação polinomial por funções spline (PFS) é uma das técnica frequentemente aplicada para preencher dados ausentes em séries temporais, sendo comum o uso de funções spline. No entanto, estas funções podem apresentar o efeito de Runge durante extrapolações, o que leva a oscilações nas extremidades da série e compromete a precisão das previsões. Por esta razão, recomenda-se a utilização de polinômios de grau menor para minimizar este problema (Fornberg; Zuev, 2007; Chen *et al.*, 2014). Uma alternativa utilizada é a decomposição de tendência sazonal em séries temporais STL (Cleveland *et al.*, 1990), devido às vantagens em relação aos métodos tradicionais que realizam predições diretas (Huo *et al.*, 2019; Sebastian; Gao; Xing, 2020; Zhao *et al.*, 2023). O processo de decomposição consiste em dividir os dados em três componentes principais: crescimento, sazonalidade e erro residual. Após esta separação, os dados são tratados em segmentos menores, o que reduz a complexidade computacional e o erro envolvido. Por fim, as variáveis interpoladas/extrapoladas individualmente são recombinaadas por meio do processo inverso de decomposição, resultando em previsões com menor margem de erro (Ouyang; Jabloun; Ravier, 2023).

3.5 Processo de otimização

O processo de otimização envolve identificar, entre todas as possíveis soluções, aquela que melhor atende às restrições do problema. Para isto, é necessário

¹Medida que expressa o grau de diferença ou desigualdade entre elementos em um conjunto de dados.

compreender a natureza do problema e definir o objetivo da solução, seja minimizar ou maximizar. Uma vez que o problema está definido e modelado, utilizando a função objetivo ou de avaliação, busca-se atender às restrições e encontrar os melhores valores possíveis para os parâmetros (Ganzaroli *et al.*, 2022). Os desafios atuais na área de otimização são, frequentemente, multidisciplinares e multiobjetivo.

As técnicas de otimização utilizadas para resolver problemas incluem não apenas aquelas com base em gradientes, mas também os métodos heurísticos inspirados em processos naturais. Em problemas de otimização, a função de avaliação deve ser maximizada ou minimizada. Geralmente, prioriza-se a minimização, uma vez que até os problemas de maximização podem ser reformulados como problemas de minimização (ARAÚJO, 2020).

3.6 Considerações finais

O uso da IA possibilita que grandes volumes de dados sejam analisados com agilidade e precisão. Este capítulo apresentou os conceitos de ciência de dados e AM. O aprendizado de máquina oferece um conjunto robusto de ferramentas para análise de séries temporais e previsão de variáveis complexas, permitindo ajustes automáticos com base em padrões históricos. Assim, a aplicação destas técnicas ao contexto de estações de tratamento de água pode melhorar significativamente o controle dos parâmetros de cloro e flúor, trazendo previsões mais precisas e redução de erros operacionais. No próximo capítulo são reunidos todos os conceitos discutidos descrevendo a metodologia.

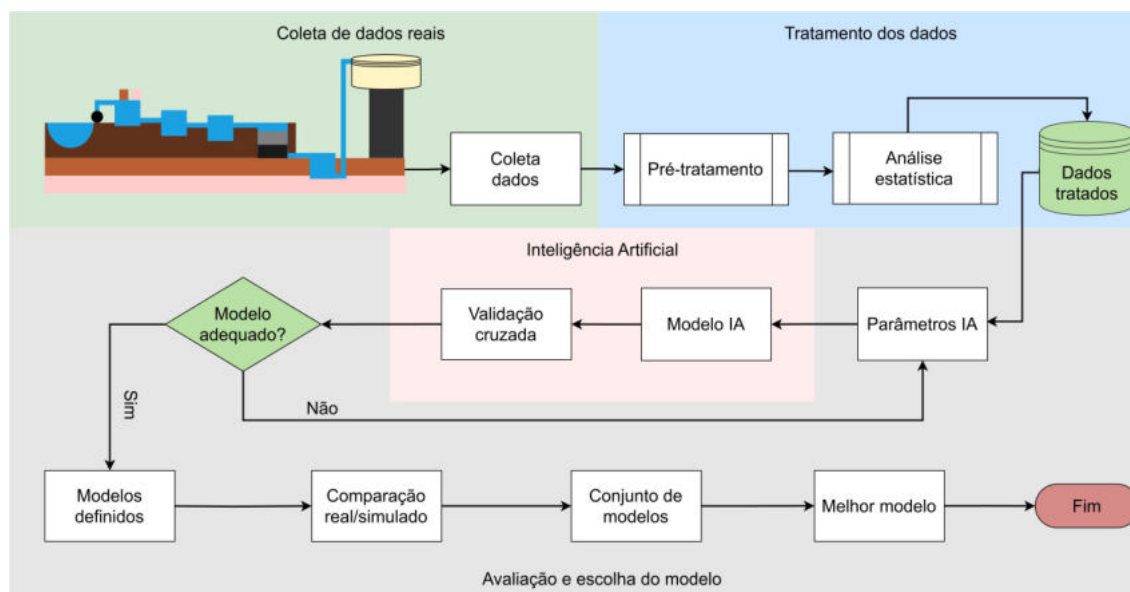
Metodologia

Este capítulo descreve a metodologia proposta, iniciando com a breve contextualização do estudo, que inclui a apresentação da problemática e dos requisitos necessários para o sistema de tratamento de água. Em seguida, detalha-se o processo de coleta de dados e pré-processamento para a aplicação das técnicas computacionais, incluindo aprendizado de máquina. Por fim, são abordados os critérios de convergência e as métricas de avaliação para validar e avaliar os modelos propostos.

4.1 Contextualização

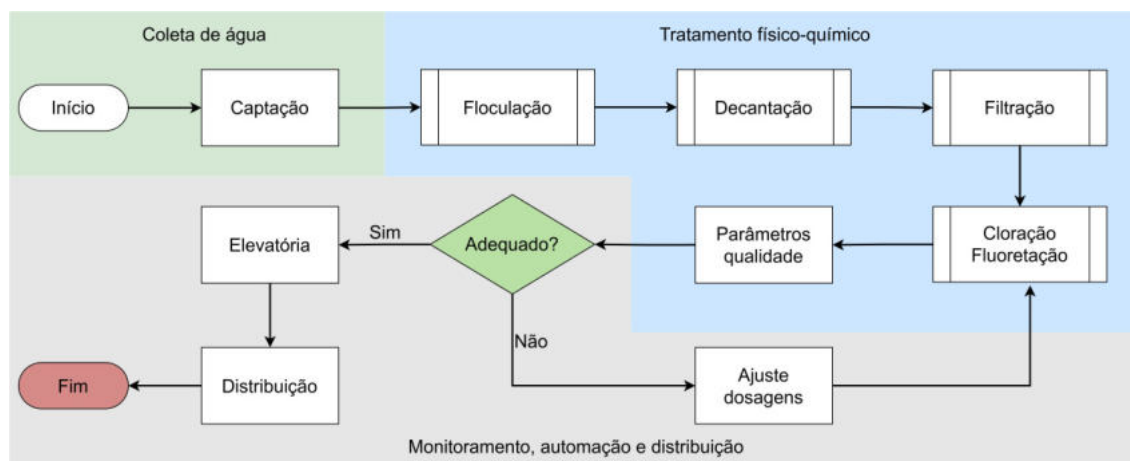
O objetivo principal deste estudo é desenvolver modelo computacional preditivo de regressão supervisionada, orientado por séries temporais, para otimizar os processos de cloração e fluoretação em estação de tratamento de água específica. O intuito é reduzir os custos operacionais associados à determinação das dosagens de produtos químicos, garantindo simultaneamente a segurança da saúde pública. Pretende-se aprimorar o processo, visando assegurar a qualidade do produto final, enquanto contribui para o alcance do Objetivo de Desenvolvimento Sustentável número 6 (ODS 6) da Agenda 2030 da Organização das Nações Unidas (ONU).

Este estudo emprega a abordagem de pesquisa hipotético-dedutiva (Lawson, 2000), concentrando-se na aplicação de IA nos dados obtidos em ETA, para o desenvolvimento do modelo que representa o sistema. Possuindo o modelo, é possível realizar simulações, reproduzindo o comportamento do sistema. Assim, é possível realizar previsões e análises de variáveis definidas no modelo, otimizando a saída do sistema sem a necessidade de alteração física na planta. O processo inicia-se com a captação superficial da água e termina na distribuição da água potável. A Figura 1 ilustra o fluxo da metodologia proposta.

Figura 1 – *Fluxograma da metodologia proposta.*

4.2 Características desejadas no sistema real

As características fundamentais que a ETA deve apresentar para se enquadrar no escopo desta pesquisa são: i) a necessidade de ser dotada de captação superficial de água e ii) conter sistema de tratamento convencional, envolvendo os processos de floculação, decantação, filtração, cloração e fluoretação. Adicionalmente, a ETA deve realizar sistematicamente a aferição e registro dos parâmetros relevantes, promovendo a retroalimentação destes dados para o ajuste das dosagens de produtos químicos, de forma manual ou automatizada. Estas características compõem os requisitos necessários que orientam o desenvolvimento e funcionamento eficiente da ETA, como ilustrado na Figura 2.

Figura 2 – *Fluxo hipotético dos requisitos da ETA para inclusão no estudo.*

A captação superficial da água justifica-se pelo benefício da desinfecção solar da água (SODIS), na qual a incidência de radiação solar, especialmente a luz ultravioleta (UV), tem a capacidade de eliminar bactérias como *Escherichia coli*, *Pseudomonas aeruginosa*, *Shigella flexneri*, *Salmonella typhimurium*, *Salmonella enteritis*, *Salmonella paratyphi* (Meulemans, 1987), vírus como *bacteriophage f2*, *rotavirus* (Zavadova; Gresland; Rosenbergova, 1968), fungos como *Aspergillus niger*, *Aspergillus flavus* (Jun *et al.*, 2003), *Candida*, *Geotrichum*, e protozoários como *Giardia spp.*, *Cryptosporidium spp* (Masjoudi; Mohseni; Bolton, 2021), embora para cada micro-organismo seja exigido um protocolo específico, seja para evitar reativação, seja para inativar os esporos.

A SODIS é capaz de inativar ampla gama de microrganismos patogênicos, associando a radiação e temperatura. No entanto, este método não é suficiente para tratar elevados volumes de água, necessitando de processos auxiliares, como a oxidação por meio da cloração. Captações subterrâneas apresentam maior probabilidade de contaminação química e biológica, exigindo processos desinfetantes mais rigorosos e maior número de análises. Assim, por estes motivos, não são objeto desta pesquisa.

4.3 Coleta de dados e pré-processamento

Após a seleção da ETA, os dados devem ser obtidos no formato de *dataframe*. As informações necessárias para a avaliação incluem: vazão de água bruta V_{ab} [m^3/min], turbidez da água bruta T_{ab} [NTU], dosagem de cloro D_{Cl} [mg/L], dosagem de flúor D_F [mg/L], concentração de cloro na água tratada M_{Cl} [mg/L], concentração de flúor na água tratada M_F [mg/L], temperatura da água tratada M_T [$^{\circ}C$], pH da água tratada M_{pH} e vazão de água tratada V_{at} [m^3/min]. Após a coleta, realiza-se o pré-processamento, assegurando a consistência das informações e a coerência das medições. Em caso de lacunas ou valores atípicos nos dados, aplica-se o método de regressão polinomial ou de decomposição de tendência sazonal em séries temporais, conforme o desempenho avaliado por métricas específicas (Durrleman; Simon, 1989; Wegman; Wright, 1983; Cleveland *et al.*, 1990), como descrito no Algoritmo 1.

As variáveis monitoradas, como D_{Cl} e D_F , T_{ab} e V_{ab} , estão inter-relacionadas no contexto do modelo, influenciando diretamente as decisões de dosagem e a qualidade final da água tratada. A turbidez, por exemplo, afeta a quantidade de coagulante necessária e, indiretamente, a demanda de D_{Cl} para desinfecção. Além disto, parâmetros como M_{pH} e M_T influenciam a estabilidade da solução de cloro e flúor, afetando sua eficácia ao longo da distribuição. Desta forma, o modelo deve considerar estas interações para prever com precisão os ajustes necessários, buscando

minimizar erros de predição e garantir a estabilidade dos níveis de concentração recomendados por normas.

Antes de aplicar os dados nos modelos de IA, realiza-se análise utilizando a estatística descritiva na tentativa de compreender e resumir a natureza dos dados. Avalia-se as médias μ , o desvio padrão σ e os valores máximos V_{max} e mínimos V_{min} dos parâmetros para obtenção do comportamento típico. A análise temporal indica possíveis tendências sazonais, enquanto a investigação de correlações ρ entre diferentes parâmetros fornece percepções sobre suas inter-relações no processo de tratamento da água.

Algoritmo 1: *TrataLacunas(dadosBrutos)*

Entrada: Dataframe *dadosBrutos*, Dataframe *conjuntoValidacao*.

Saída: Dataframe *dadosTratados*.

```

1 declare Dataframe dadosTratadosPFS
2 declare Dataframe dadosTratadosSTL
3 set.seed(1234)
4 para (n de 1 até dadosBrutos.tamanho) faça
5     se dadosBrutos[n] for atípico ou ausente então
6         | dadosTratados[n] ← vazio
7     senão
8         | dadosTratados[n] ← dadosBrutos[n]
9     fim
10 fim
11 dadosTratadosPFS ← interpolação spline de dadosTratados
12 dadosTratadosSTL ← interpolação STL de dadosTratados
13 se (RMSE de dadosTratadosPFS ≤ dadosTratadosSTL) então
14     | dadosTratados ← dadosTratadosPFS
15 senão
16     | dadosTratados ← dadosTratadosSTL
17 fim
18 retorne dadosTratados

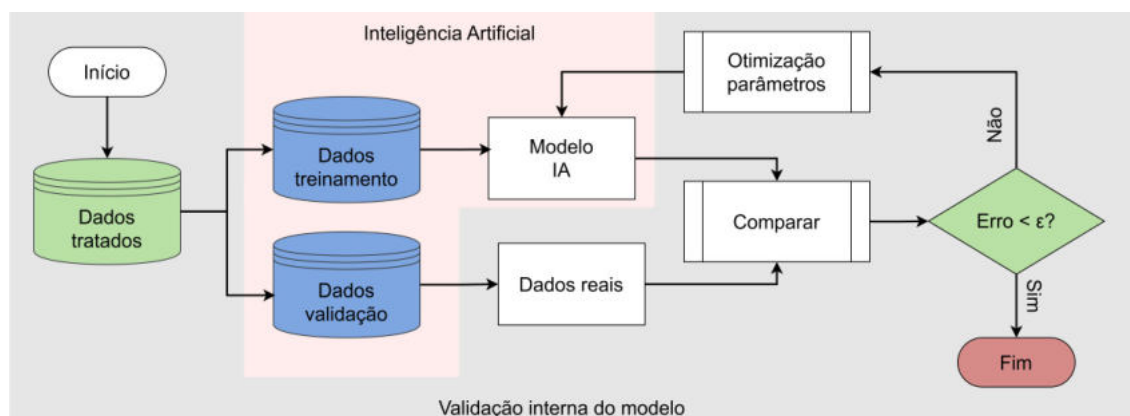
```

4.4 Modelagem computacional e aplicação de aprendizado de máquina

Por se tratar de séries temporais, após o tratamento dos dados e a análise estatística, é necessário segmentar as séries em períodos contíguos, intervalos contínuos e ininterruptos de pelo menos um mês. Este processo garante o reconhecimento

dos padrões de sazonalidade do sistema. Os demais dados são utilizados para treinamento, com foco na segmentação e independência dos dados. Os percentuais de validação representam, no mínimo, 20% dos dados, e caso um mês não atinja este percentual, é adicionado mais um mês sucessivamente até atender a este critério. Os modelos são construídos utilizando os algoritmos FR, XGBoost e RNA-MLP com diferentes quantidades de camadas. A proposta visa otimizar as dosagens de cloro e flúor no sistema. A Figura 3 ilustra o fluxo dos dados no banco de dados tratados da metodologia proposta.

Figura 3 – Fluxo dos dados no banco de dados tratados.



No modelo desenvolvido utilizando o algoritmo RF, inicia-se com o conjunto de dados de treinamento, realizando amostragem aleatória com reposição para construir diversas árvores de decisão. Cada árvore é treinada em conjunto de dados diferente e a previsão final é determinada pela votação ou média das previsões individuais das árvores. Este processo diversificado reduz o sobreajuste, resultando em um modelo eficiente para classificação ou regressão. Para o modelo desenvolvido utilizando o algoritmo XGBoost, inicia-se com a árvore base, calcula os resíduos entre as previsões atuais e os valores reais, construindo árvores adicionais para corrigir deficiências nas previsões anteriores de maneira iterativa. As previsões das árvores são combinadas e o modelo é atualizado para minimizar a função de perda global. Com técnicas de regularização para evitar sobreajuste, o XGBoost produz previsões finais ponderadas pelo desempenho das árvores individuais.

No modelo desenvolvido utilizando o algoritmo RNA-MLP, inicia-se com os dados nas camadas de entrada dos neurônios correspondentes às características do conjunto de dados. A propagação direta passa pelas camadas ocultas, em que cada neurônio realiza a combinação linear ponderada e a função de ativação não linear. As camadas ocultas extraem características complexas dos dados, culminando na camada de saída que gera a previsão final. O cálculo do erro, realizado comparando a saída prevista com os valores reais, desencadeia a retro propagação do erro.

Este processo ajusta iterativamente os pesos entre neurônios para minimizar a função de perda. Nos modelos construídos pelos algoritmos FR e RNA-MLP, a avaliação do desempenho ocorre com dados de teste ou validação, medindo a capacidade do modelo de generalizar para novos dados. No modelo construído pelo algoritmo XGBoost, o processo de impulsionamento por gradientes resulta em modelos eficientes.

4.5 Métricas de avaliação e critérios de convergência

Para os algoritmos FR, XGBoost e RNA-MLP, são empregadas as métricas de avaliação: i) raiz do erro médio quadrático (REM_Q) que calcula a média dos quadrados das diferenças entre os valores observados e os valores previstos, ii) EMA que calcula a média das diferenças absolutas entre os valores observados e os valores previstos e iii) coeficiente de determinação (R^2) que avalia o modelo de regressão observando se este ajusta aos dados observados, representando a proporção da variância da variável dependente explicada pelos regressores. Na escolha do método de avaliação, é necessário considerar a natureza do problema, especialmente em casos de regressão de séries históricas. O REM_Q é a métrica comum para problemas de regressão, penalizando erros maiores, enquanto o MAE é menos sensível a valores discrepantes em comparação com o REM_Q. Além destas métricas, a validação cruzada é aplicada para avaliar o desempenho. No FR e RNA-MLP, utiliza-se o *k-fold cross-validation* (*k*-FCV), que divide o conjunto de dados em *k* partições, treina o modelo em *k* − 1 partições e avalia o desempenho na partição restante. Este processo é repetido *k* vezes, com cada partição sendo usada como conjunto de teste única vez. O XGBoost é semelhante ao FR e RNA-MLP, diferenciando-se na função específica para a validação cruzada.

Os critérios de convergência são importantes para determinar o momento em que os algoritmos devem encerrar o treinamento, visando evitar o sobreajuste ou garantir eficiência computacional. No contexto do algoritmo FR, os critérios de convergência incluem o número máximo de árvores n_a e o crescimento mínimo de folhas c_{mf} . Para o algoritmo XGBoost, os critérios compreendem o n_a , o tamanho mínimo da amostra por folha t_{maf} e o ganho mínimo γ , que define a quantidade mínima de redução na função de perda necessária para realizar nova divisão. Por fim, para o algoritmo RNA-MLP, os critérios envolvem o número máximo de épocas ϵ , a melhoria mínima na validação Δ_{min} e a paciência α , indicando o número de épocas sem melhoria que podem ocorrer antes de interromper o treinamento. O ajuste

nos valores destes critérios é realizado com base nas características específicas do conjunto de dados e nos requisitos do problema. A seleção apropriada dos critérios de convergência contribui para a eficiência dos modelos treinados.

Nos algoritmos de aprendizado de máquina, os hiperparâmetros, definidos pelo usuário antes do treinamento do modelo, não são ajustados automaticamente pela inteligência artificial. Esses hiperparâmetros regulam aspectos do treinamento, como a complexidade do modelo, a taxa de aprendizado e a regularização. A escolha adequada desses parâmetros influencia diretamente o desempenho do modelo. Para defini-los, podem ser empregados quatro métodos de otimização: i) otimização bayesiana (BO), que aplica técnicas estatísticas para modelar a relação entre os hiperparâmetros e a métrica de avaliação; ii) algoritmos genéticos (AG), abordagem evolutiva que manipula populações de combinações de hiperparâmetros; iii) pesquisa em grade, que realiza uma busca exaustiva com base em uma lista predefinida de hiperparâmetros e critérios de desempenho; e iv) pesquisa aleatória, que testa combinações de hiperparâmetros de forma aleatória para identificar o melhor ajuste.

4.6 Avaliação estatística dos modelos preditivos

A análise estatística é realizada para avaliar o desempenho dos modelos preditivos RNA-MLP, FR e XGBoost na estimativa das dosagens de cloro e flúor em ETA. A metodologia organiza-se em etapas que incluem a definição de métricas de avaliação, a aplicação de testes de hipótese e a realização de análises visuais. Utilizam-se três métricas principais para quantificar a precisão e a estabilidade dos modelos: a REMQ, o R^2 e o EMA. Para verificar diferenças significativas entre os erros das previsões e a referência ideal de erro nulo, aplica-se o teste de Mann-Whitney U, escolhido devido à sua resiliência a distribuições não normais dos dados. Para garantir representatividade e minimizar vieses, realizam-se extrações aleatórias de subamostras de tamanho N em cada rodada de análise. Os p -valores são avaliados com base no nível de significância $\alpha = 0.05$, sendo valores inferiores a este limiar indicativos de diferenças estatisticamente significativas, o que indica que os erros não são atribuíveis ao acaso.

São elaborados gráficos complementares para analisar a distribuição dos erros e avaliar o comportamento preditivo dos modelos. Empregam-se boxplots para ilustrar a dispersão dos erros, com as caixas representando o intervalo interquartil e a linha central indicando a mediana. Gráficos de resíduos são utilizados para identificar padrões ou desvios sistemáticos nas diferenças entre valores reais e previstos ao longo das observações. Em adicional, a dispersão é mapeada para visualizar a proximidade entre valores reais e previstos, tendo como referência a linha ideal $p = r$

(previsto igual ao real). Os critérios para seleção do modelo mais adequado incluem o desempenho geral nas métricas avaliadas, a consistência das previsões ao longo de diferentes períodos e subamostras, a ausência de padrões de tendência nos resíduos e a capacidade de realizar previsões simultâneas de cloro e flúor, garantindo maior flexibilidade e aplicabilidade.

4.7 Considerações finais

Neste capítulo foram apresentadas as características desejáveis da ETA para aplicação da metodologia proposta, bem como aspectos inerentes à coleta de dados. Além disto, são realizadas observações acerca da segmentação dos dados em conjunto de treinamento e validação interna, com intuito de garantir que os modelos não estarão contaminados por dados que os constituíram. Após a definição dos critérios para segmentação dos dados na construção dos modelos, são definidas métricas de avaliação da acurácia dos modelos, e os critérios de convergência para limitar o treinamento a parâmetros que atendam às restrições do problema analisado. No próximo capítulo serão apresentados os resultados da aplicação da metodologia proposta.

Resultados

Esta seção apresenta os resultados obtidos com a aplicação da metodologia, abrangendo a análise das variáveis de entrada, o desempenho dos modelos RNA-MLP, FR e XGBoost, e a comparação entre suas versões otimizadas e não otimizadas. As métricas estatísticas REMQ, R^2 , e EMA foram empregadas para avaliar a precisão e a estabilidade das previsões. Além disto, a distribuição dos erros foi analisada para identificar possíveis padrões ou desvios. A análise comparativa indicou os ganhos de desempenho proporcionados pelas versões otimizadas, destacando os modelos com maior precisão e consistência.

5.1 Escolha do sistema para aquisição, tratamento e processamento de dados

Para esta pesquisa, os dados foram obtidos da ETA operada pela Saneamento de Goiás S.A (Saneago S.A), localizada no Município de Brazabrantes, Goiás/Brasil. As informações foram solicitadas com base na Lei de Acesso à Informação (Perritt, 2001). O Município de Brazabrantes têm população de 3.981 *hab*, e pertence à Região Metropolitana de Goiânia, Capital do Estado de Goiás. Aproximadamente 67% da população é urbana e o Município de Brazabrantes está geograficamente localizado na bacia do Paraná, e integra o Comitê de Bacia Hidrográfica do Rio Meia Ponte, na qual a Cidade de Brazabrantes representa $\approx 1\%$ do território do município. A Cidade de Brazabrantes possui índice de atendimento urbano de água de 93,55%, e índice de hidrometração¹ de 97,67%, o que reflete realidade acima da média nacional brasileira, que é de atendimento urbano de 93,46% e hidrometração de 91,27%. A captação é realizada pelo Sistema Ribeirão Cachoeira, afluente do Rio Meia Ponte, situado no estado de Goiás/Brasil. A Saneago S.A detém a outorga

¹Processo de instalação de medidores/registradores de água para mensurar o consumo e emitir faturas com base nestes registros (Veiga; Kalbusch; Henning, 2023).

de $20L/s$, embora a capacidade produtiva da planta instalada seja de $14L/s$. A Figura 4 apresenta, na parte inferior, o ponto de captação de água nas coordenadas geográficas $16^{\circ}26'57.3''S$ e $49^{\circ}22'15.9''W$. Na parte superior, encontra-se a ETA localizada a aproximadamente $450m$ de distância da captação e em local mais alto na proximidade da cidade.

Figura 4 – Mapa georreferenciado da captação e ETA de Brazabrantes.

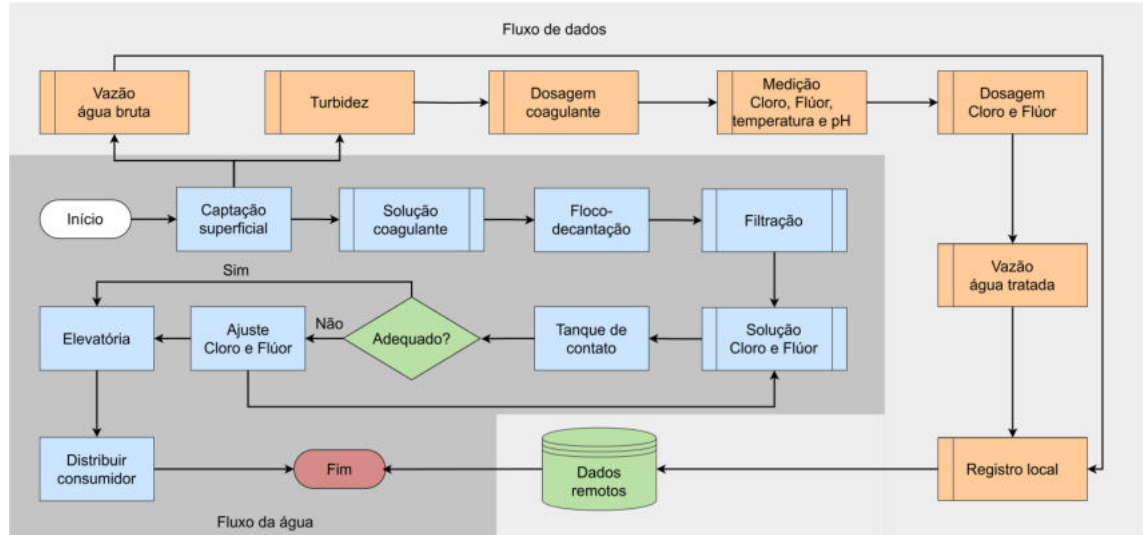


Os valores de cloração e fluoretação do sistema avaliado encontram-se dentro das recomendações internacionais de potabilidade da água (Fawell *et al.*, 2006; World Health Organization, 2022). Para águas superficiais sem fonte natural de flúor, o valor recomendado é de $1.5mg/L$, sendo indicada a fluoretação artificial na faixa de $0.5mg/L$ a $1.0mg/L$. A desinfecção com derivados de cloro, como hipocloritos, apresenta o valor recomendado de $5mg/L$. Embora seja possível reduzir sua concentração a zero após o tratamento, recomenda-se manter o valor residual entre $0.2mg/L$ e $1.0mg/L$ para garantir proteção contra recontaminação e controlar o crescimento de microrganismos na rede de distribuição. Caso sejam observadas grandes discrepâncias em relação aos parâmetros recomendados, torna-se necessário ajustar os valores de doses para alcançar os limites indicados.

A planta industrial da ETA da Cidade de Brazabrantes encontra-se segmentada em várias etapas, incluindo: i) captação superficial do manancial, ii) aferição da turbidez, iii) dosagem proporcional de coagulante de acordo com a turbidez medida, iv) decantação em tanque floccodecantador, v) filtração para remoção de partículas restantes, vi) fluoretação e cloração em tanques de contato e vii) aferição de T , pH , níveis de concentração de cloro e flúor a jusante para ajustes nas dosagens. Todo este processo culmina no bombeamento para reservatório elevado e distribuição à população. O diagrama operacional da ETA da Cidade de Brazabrantes é apresentado na Figura 5. Os produtos químicos empregados são hipoclorito de sódio para cloração, fluorsilicato de sódio para fluoretação e cloreto férrico para floccodecantação. A esco-

lha desta ETA baseia-se na adequada monitoração de parâmetros, com registro de dados históricos e representatividade semelhante a aproximadamente 80% das ETA operadas pela Saneago S.A.

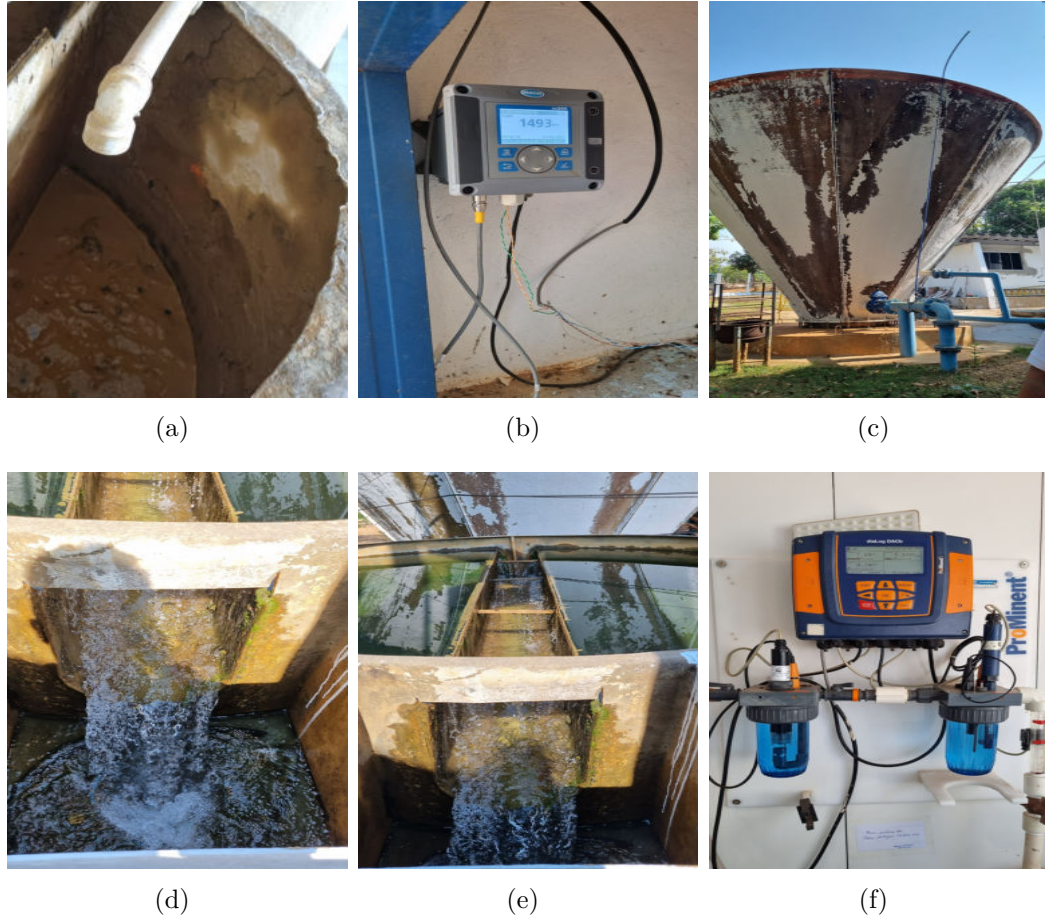
Figura 5 – *ETA Brazabrantes, Goiás/Brasil.*



Na Figura 5, os retângulos identificados na cor laranja com uma linha vertical indicam cálculo e registro dos valores das variáveis e os identificados na cor azul indicam o caminho da água. Na Figura 6 são apresentados alguns equipamentos, ambientes e indícios de processo realizados na ETA da Cidade de Brazabrantes. A operação desta ETA consiste em captação de manancial superficial, na qual o V_{ab} é medido e aduzido ao sistema de tratamento. A primeira etapa do tratamento é a adequação da turbidez da água bruta T_{ab} , através da aferição, Figura 6(a), e correção com coagulante, Figura 6(b). Os dados são registrados e a água coagulada segue para o floco-decantador, Figura 6(c), no qual a cinética é reduzida e os flocos sedimentam por gravidade. A água decantada é direcionada para o filtro de fluxo ascendente, Figura 6(d) e Figura 6(e) para remoção de partículas excedentes e correção de sabor e odor utilizando carvão. Na sequência, a água filtrada é enviada ao tanque de contato, no qual é realizada a dosagem de cloro, flúor e medido M_T e M_{pH} , Figura 6(f). Caso os parâmetros estejam adequados, a água desinfetada é enviada por recalque para o reservatório elevado, no qual é distribuída por gravidade para a população. Os dados de monitoramento destas etapas são armazenados localmente e posteriormente enviados ao repositório remoto.

Todas as informações registradas na ETA da Cidade de Brazabrantes do período de 28/outubro/2022 até 23/julho/2024 foram disponibilizadas, incluindo: i) turbidez da água bruta T_{ab} , ii) nível do reservatório apoiado N_{RA} e elevado N_{RE} , iii) informações elétricas de corrente, tensão, potência e frequência da bomba dosadora de coagulante, iv) dosagem de coagulante D_S , v) nível do reservatório

Figura 6 – Equipamentos, ambientes e processo: (a) tanque de solubilização de cloreto férrico em batelada, (b) medidor de turbidez em NTU, (c) floco-decantador, (d) tela de retenção de sólidos flutuantes/grosseiros na saída do floco-decantador, antes de entrar no filtro de fluxo ascendente, (e) calha de coleta de água superficial no floco-decantador e (f) medidor automatizado de cloro, flúor, M_T e M_{pH} .



do subdistrito de Deuslândia, Goiás, vi) informações elétricas de corrente, tensão, potência e frequência da bomba dosadora de flúor, vii) dosagem de flúor D_F , viii) vazão de água bruta V_{ab} , ix) volume histórico acumulativo de água bruta, x) vazão de água tratada V_{at} , xi) volume histórico de água tratada, xii) informações elétricas de corrente, tensão, potência e frequência da bomba dosadora de cloro, xiii) dosagem de cloro D_{Cl} , e xiv) medições de pH , M_{pH} , flúor M_F , cloro M_{Cl} e temperatura T da água tratada M_T . Destes dados, foram utilizados os que apresentaram maior correlação com a dosagem de cloro e flúor, como disposto na Tabela 2.

Com base no comportamento do sistema e na coerência das informações

Tabela 2 – *Correlação entre algumas variáveis observadas e a dosagem de cloro e flúor.*

	M_{Cl}	M_F	M_{pH}	M_T	V_{at}	V_{ab}	D_{Cl}	D_F
M_{Cl}	1,00							
M_F	$-1,00 \times 10^{-1}$	1,00						
M_{pH}	$5,00 \times 10^{-2}$	$2,20 \times 10^{-1}$	1,00					
M_T	$-2,00 \times 10^{-2}$	$2,10 \times 10^{-1}$	$-1,60 \times 10^{-1}$	1,00				
V_{at}	$4,80 \times 10^{-1}$	$-2,40 \times 10^{-1}$	$3,00 \times 10^{-2}$	0,00	1,00			
V_{ab}	$4,60 \times 10^{-1}$	$-2,40 \times 10^{-1}$	$3,00 \times 10^{-2}$	0,00	$9,20 \times 10^{-1}$	1,00		
D_{Cl}	$4,30 \times 10^{-1}$	$-1,80 \times 10^{-1}$	$4,00 \times 10^{-2}$	$-2,00 \times 10^{-2}$	$8,90 \times 10^{-1}$	$8,20 \times 10^{-1}$	1,00	
D_F	$4,30 \times 10^{-1}$	$-1,80 \times 10^{-1}$	$4,00 \times 10^{-2}$	$-2,00 \times 10^{-2}$	$8,90 \times 10^{-1}$	$8,20 \times 10^{-1}$	$9,80 \times 10^{-1}$	1,00

fornechas pelos técnicos, foram descartados os valores atípicos. O critério de descarte considerou valores de $2 \leq M_{pH} < 8$, $0 \leq D_{Cl} < 5mg/L$, $0 \leq D_F < 3mg/L$, $10^\circ C \leq M_T < 50^\circ C$, $0 \leq V_{ab} < 1m^3/min$ e $0 \leq V_{at} < 1m^3/min$, $0 \leq M_{Cl} < 50mg/L$ e $0 \leq M_F < 10mg/L$. Os parâmetros estatísticos do conjunto de dados são dispostos na Tabela 3, incluindo média μ , desvio padrão σ e os valores máximo V_{max} e mínimo V_{min} .

Tabela 3 – *Estatística descritiva dos parâmetros observados.*

Parâmetro	unidade	μ	σ	V_{max}	V_{min}
D_{Cl}	mg/L	1,45	1,61	4,82	0,00
M_{Cl}	mg/L	$6,20 \times 10^{-1}$	$4,93 \times 10^{-1}$	$2,83 \times 10^1$	$1,00 \times 10^{-2}$
M_{pH}	pH	6,62	$4,04 \times 10^{-1}$	7,93	2,67
M_T	$^\circ C$	$2,58 \times 10^1$	2,64	$4,51 \times 10^1$	$1,72 \times 10^1$
M_F	mg/L	$9,07 \times 10^{-1}$	$3,81 \times 10^{-1}$	9,14	0,00
D_F	mg/L	$8,60 \times 10^{-1}$	$9,50 \times 10^{-1}$	2,82	0,00
V_{ab}	m^3/min	$4,00 \times 10^{-1}$	$3,92 \times 10^{-1}$	1,00	0,00
V_{at}	m^3/min	$3,65 \times 10^{-1}$	$3,84 \times 10^{-1}$	1,00	0,00
T_{ab}	NTU	$1,31 \times 10^2$	$1,55 \times 10^2$	$3,00 \times 10^3$	2,00

As variáveis dispostas na Tabela 2 foram as utilizadas na construção do modelo. Inicialmente, o conjunto de dados fornecidos continha 599.386 linhas de valores históricos do período analisado, com aproximadamente 81Mb de dados estruturados no formato de banco de dados relacional, que foram exportados para o formato de *dataframe* em *.csv. Estes representavam as medições históricas de diversas variáveis que permitiam calcular o IQA. Observou-se que os dados tinham periodicidade de coleta de 15 minutos, e na data de 11 de maio de 2023, às 14h42min,

a periodicidade foi alterada para cada minuto. Devido a esta alteração, foi necessário ajustar a base do tempo de amostragem de 15 minutos para 1 minuto. Após a aplicação dos métodos para preenchimento das lacunas, o conjunto de dados passou a ter 909.250 linhas de valores históricos no período compreendido entre 31 de outubro de 2022 e 23 de julho de 2024.

Para o preenchimento das lacunas, foram aplicadas as técnicas PFS e STL. Com o objetivo de avaliar a precisão dos dois métodos, foi selecionado um trecho de dados conhecidos, no qual 20% das observações foram suprimidas. Utilizando os 80% restantes, ambas as técnicas foram empregadas para estimar os valores ausentes. A Tabela 4 dispõe os resultados da análise com as respectivas métricas de avaliação, enquanto a Figura 7 apresenta o comparativo gráfico referente ao parâmetro M_{Cl} . No gráfico, a linha amarela representa os valores medidos pelo sistema por meio do equipamento apresentado na Figura 6(f), a linha vermelha indica o preenchimento das lacunas utilizando PFS, e a linha azul apresenta o preenchimento utilizando STL. O eixo das abscissas corresponde às datas no formato mês, dia, ano, hora e minuto, com intervalos de um minuto, enquanto o eixo das ordenadas representa os valores medidos de cloro M_{Cl} .

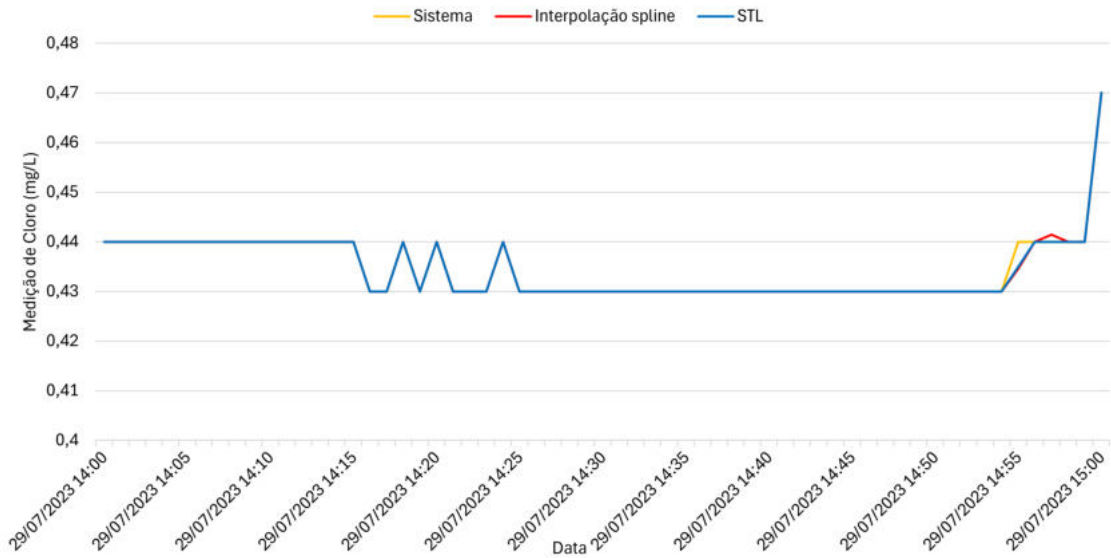
Tabela 4 – *Crítérios de avaliação dos métodos utilizados para preenchimento de lacunas.*

Parâmetro	REMQ PFS	REMQ STL	R^2 PFS	R^2 STL	EMA PFS	EMA STL
D_{Cl}	$1,04 \times 10^{-1}$	$9,11 \times 10^{-2}$	$9,96 \times 10^{-1}$	$9,97 \times 10^{-1}$	$8,08 \times 10^{-3}$	$5,57 \times 10^{-3}$
M_{Cl}	$2,27 \times 10^{-2}$	$1,84 \times 10^{-2}$	$9,98 \times 10^{-1}$	$9,99 \times 10^{-1}$	$1,34 \times 10^{-3}$	$1,16 \times 10^{-3}$
M_{pH}	$1,97 \times 10^{-2}$	$1,47 \times 10^{-2}$	$9,98 \times 10^{-1}$	$9,99 \times 10^{-1}$	$1,47 \times 10^{-3}$	$1,09 \times 10^{-3}$
M_T	$7,34 \times 10^{-2}$	$6,54 \times 10^{-2}$	$9,99 \times 10^{-1}$	$9,99 \times 10^{-1}$	$4,75 \times 10^{-3}$	$3,39 \times 10^{-3}$
M_F	$2,34 \times 10^{-2}$	$2,02 \times 10^{-2}$	$9,96 \times 10^{-1}$	$9,97 \times 10^{-1}$	$2,71 \times 10^{-3}$	$2,12 \times 10^{-3}$
D_F	$6,11 \times 10^{-2}$	$5,35 \times 10^{-2}$	$9,96 \times 10^{-1}$	$9,97 \times 10^{-1}$	$4,78 \times 10^{-3}$	$3,29 \times 10^{-3}$
V_{ab}	$2,74 \times 10^{-2}$	$2,27 \times 10^{-2}$	$9,95 \times 10^{-1}$	$9,97 \times 10^{-1}$	$4,90 \times 10^{-3}$	$3,94 \times 10^{-3}$
V_{at}	$2,69 \times 10^{-2}$	$2,29 \times 10^{-2}$	$9,95 \times 10^{-1}$	$9,96 \times 10^{-1}$	$5,31 \times 10^{-3}$	$3,68 \times 10^{-3}$
T_{ab}	$6,86 \times 10^{-5}$	$5,07 \times 10^{-5}$	$9,99 \times 10^{-1}$	$9,99 \times 10^{-1}$	$6,26 \times 10^{-8}$	$5,32 \times 10^{-8}$

5.2 Construção do modelo e validação interna

A partir das amostras obtidas da ETA, foi realizada uma análise estatística para comparar os valores dos parâmetros com os limites recomendados internacionalmente (World Health Organization, 2002). Do total de 909.250 registros do sistema, 908.054 atenderam aos critérios de cloração, correspondendo a $\approx 99\%$ dos dados.

Figura 7 – *Comparativo de métodos de preenchimento de dados faltantes.*



Em relação à fluoretação, 866.621 registros estavam dentro dos limites recomendados, representando $\approx 95\%$ do total. Assim, considera-se que a operação desta ETA segue os padrões internacionais para desinfecção com hipocloritos e suplementação artificial de flúor, sendo possível reproduzir seu comportamento com confiança.

Para a construção do modelo e validação interna, os dados foram segmentados entre 01 de janeiro de 2023 e 30 de março de 2023 (89 dias) para testes e validação interna, totalizando 131.041 registros, o que corresponde a $\approx 25\%$ dos dados. Os registros do período de 01 de abril de 2023 a 31 de dezembro de 2023 (275 dias) foram utilizados para o treinamento dos modelos, totalizando 393.121 registros, equivalente a $\approx 75\%$ dos dados. Os modelos foram construídos na linguagem R utilizando o pacote `caret`. O método de validação cruzada k -folds foi empregado com $k = 10$. A função `seed` foi fixada com valor constante ao longo de todo o processo, garantindo a reprodutibilidade dos geradores de números pseudoaleatórios e viabilizando a comparação entre execuções. Foram aplicados três métodos de inteligência artificial: RNA-MLP, FR e XGBoost.

São produzidos dois resultados de simulação, um com os hiperparâmetros padrão para inicialização do modelo de IA e outro com a otimização dos hiperparâmetros. No treinamento do modelo RNA-MLP com 1-camada oculta e sem otimização de hiperparâmetro, para cada cenário, o modelo foi treinado com 1, 3 e 5 neurônios na camada oculta S_z , e foi aplicado o decaimento de peso W_D no intervalo de $[0; 10^{-4}]$ para regularização e evitar o sobreajuste. A Tabela 5 dispõe os melhores valores obtidos na validação interna para a cloração, com as métricas REMQ, R^2 e EMA. O menor valor de REMQ foi obtido no modelo com $S_z = 5$, enquanto o me-

nor valor de EMA e o maior valor de R^2 foram observados no modelo com $S_z = 3$, ambos com $W_D = 0$. Para a fluoretação, os menores valores de REMQ e EMA, e o maior valor de R^2 foram obtidos com o modelo com $S_z = 5$ e $W_D = 0$. Os modelos de RNA-MLP com 2-camadas e 3-camadas, sem otimização de hiperparâmetros, apresentaram os mesmos resultados que o modelo de RNA-MLP com 1-camada, não sendo desejáveis devido à maior necessidade de recursos computacionais.

Tabela 5 – *Valores dos hiperparâmetros e das validações internas utilizando os métodos RNA-MLP, FR e XGBoost.*

Hiperparâmetros (RNA-MLP)				Cloração			Fluoretação		
S_z	–	–	W_D	REMQ	R^2	EMA	REMQ	R^2	EMA
3	–	–	0,00	$5,31 \times 10^{-1}$	$8,98 \times 10^{-1}$	$2,05 \times 10^{-1}$	$3,18 \times 10^{-1}$	$8,96 \times 10^{-1}$	$1,39 \times 10^{-1}$
5	–	–	0,00	$5,26 \times 10^{-1}$	$8,97 \times 10^{-1}$	$2,06 \times 10^{-1}$	$2,94 \times 10^{-1}$	$9,09 \times 10^{-1}$	$1,28 \times 10^{-1}$
Hiperparâmetros (FR)				Cloração			Fluoretação		
N_{rc}	–	–	–	REMQ	R^2	EMA	REMQ	R^2	EMA
2	–	–	–	$8,46 \times 10^{-2}$	$9,97 \times 10^{-1}$	$1,18 \times 10^{-2}$	$5,05 \times 10^{-2}$	$9,97 \times 10^{-1}$	$7,14 \times 10^{-3}$
4	–	–	–	$8,81 \times 10^{-2}$	$9,96 \times 10^{-1}$	$1,08 \times 10^{-2}$	$5,22 \times 10^{-2}$	$9,96 \times 10^{-1}$	$6,32 \times 10^{-3}$
Hiperparâmetros (XGBoost)				Cloração			Fluoretação		
η	D_M	C_{ST}	S_{sa}	REMQ	R^2	EMA	REMQ	R^2	EMA
0,4	3	0,8	0,75	$2,78 \times 10^{-1}$	$9,70 \times 10^{-1}$	$1,12 \times 10^{-1}$	$1,72 \times 10^{-1}$	$9,66 \times 10^{-1}$	$7,34 \times 10^{-2}$
0,4	3	0,8	1,00	$2,81 \times 10^{-1}$	$9,69 \times 10^{-1}$	$1,15 \times 10^{-1}$	$1,72 \times 10^{-1}$	$9,67 \times 10^{-1}$	$7,19 \times 10^{-2}$

No treinamento do modelo FR sem otimização de hiperparâmetros, o modelo foi treinado com diferentes números de variáveis amostradas aleatoriamente como candidatas em cada divisão, sendo $N_{rc} = [2, 4, 7]$. Os valores obtidos na validação interna para a cloração e para a fluoretação estão dispostos na Tabela 5, na qual o modelo com $N_{rc} = 2$ apresentou o menor valor de REMQ e o maior valor de R^2 , enquanto o modelo com $N_{rc} = 4$ obteve o menor valor de EMA para cloração e fluoretação. No treinamento do modelo XGBoost sem otimização dos hiperparâmetros, foram utilizadas as mesmas quantidades de amostras do modelo FR, e foi mantido o parâmetro $\gamma = 0$ para evitar a aplicação de penalidades no controle de complexidade do modelo. Embora esta configuração possa levar ao sobreajuste, não foi estabelecida a redução mínima de perda necessária para realizar divisões adicionais em um nó da árvore. Além disto, foi definido o peso mínimo de instância $M_{cw} = 1$, indicando que a quantidade mínima total de instâncias necessárias em um nó para realizar nova divisão seja sempre 1. Esta configuração controla a profundidade mínima da árvore, evitando que nós com número pequeno de instâncias sejam divididos, tornando o algoritmo menos conservador.

A Tabela 5 dispõe os valores dos hiperparâmetros e as métricas de convergência da validação interna para o modelo XGBoost. Os menores valores de REMQ e EMA e o maior valor de R^2 foram obtidos com $N_{rn} = 150$ (número de rodadas), $D_M = 3$ (profundidade máxima das árvores), $\eta = 0.4$ (taxa de aprendizado) e $C_{ST} = 0.8$ (proporção de colunas amostradas). A proporção de observações (linhas) amostradas em cada árvore apresentou melhor desempenho com $S_{sa} = 0.75$ para a cloração e $S_{sa} = 1.00$ para a fluoretação.

5.3 Otimização de hiperparâmetros

Para a otimização dos hiperparâmetros, foi utilizada a pesquisa aleatória devido à ausência de conhecimento prévio sobre o espaço de busca e ao alto custo computacional de alternativas exaustivas. Após a amostragem inicial, foram realizados ajustes em torno dos valores ótimos encontrados, utilizando pesquisa em grade para refinar o ajuste de cada modelo treinado. O modelo RNA-MLP foi treinado com três camadas ocultas e otimização de hiperparâmetros, configurado com 1 a 15 neurônios em cada camada oculta S_1 , S_2 e S_3 e aplicado o decaimento de peso W_D no intervalo $[0; 10^{-6}]$ para regularização e prevenção de sobreajuste. A Tabela 6 dispõe os melhores valores de avaliação para a cloração após a validação interna, incluindo as métricas REMQ, R^2 e EMA. O menor REMQ, o menor EMA e o maior R^2 foram obtidos com a configuração $S_1 = 11$, $S_2 = 4$, $S_3 = 8$ e $W_D = 10^{-5}$. Para a fluoretação, o menor REMQ e o maior R^2 foram alcançados com a configuração $S_1 = 10$, $S_2 = 12$, $S_3 = 2$ e $W_D = 10^{-6}$, enquanto o menor EMA foi obtido com $S_1 = 10$, $S_2 = 11$, $S_3 = 2$ e $W_D = 10^{-6}$.

O modelo FR foi treinado com diferentes quantidades de variáveis amostradas aleatoriamente como candidatas em cada divisão, sendo N_{rc} variando no intervalo $[1; 10]$. Os resultados da validação interna para cloração e fluoretação estão dispostos na Tabela 6. O modelo com $N_{rc} = 4$ registrou o menor valor de REMQ e o maior valor de R^2 para cloração e fluoretação, enquanto o modelo com $N_{rc} = 5$ obteve o menor valor de EMA em ambos os casos.

No treinamento do modelo XGBoost após a otimização dos hiperparâmetros, o parâmetro $\gamma = 0$ foi mantido e o peso mínimo de instância foi definido como $M_{cw} = 1$. A Tabela 6 dispõe os valores dos hiperparâmetros e as métricas de convergência otimizadas obtidas na validação interna para o modelo XGBoost. Os menores valores de REMQ e EMA foram alcançados com número de rodadas $N_{rn} = 300$, profundidade máxima $D_M = 8$, taxa de aprendizado $\eta = 0,2$, proporção de colunas amostradas $C_{ST} = 0,8$ e proporção de instâncias amostradas $S_{sa} = 0,75$, tanto para

Tabela 6 – *Valores otimizados de hiperparâmetros e validações para os métodos RNA-MLP, FR e XGBoost.*

Hiperparâmetros (RNA-MLP)				Cloração			Fluoretação		
S_1	S_2	S_3	W_D	REMQ	R^2	EMA	REMQ	R^2	EMA
11	4	8	10^{-5}	$1,39 \times 10^{-1}$	$9,94 \times 10^{-1}$	$4,12 \times 10^{-2}$	$1,29 \times 10^{-1}$	$9,83 \times 10^{-1}$	$3,51 \times 10^{-2}$
10	11	2	10^{-6}	$3,02 \times 10^{-1}$	$9,60 \times 10^{-1}$	$1,02 \times 10^{-1}$	$8,99 \times 10^{-2}$	$9,92 \times 10^{-1}$	$1,79 \times 10^{-2}$
10	12	2	10^{-6}	$3,01 \times 10^{-1}$	$9,61 \times 10^{-1}$	$1,04 \times 10^{-1}$	$8,45 \times 10^{-2}$	$9,93 \times 10^{-1}$	$1,89 \times 10^{-2}$
Hiperparâmetros (FR)				Cloração			Fluoretação		
N_{rc}	-	-	-	REMQ	R^2	EMA	REMQ	R^2	EMA
4	-	-	-	$7,38 \times 10^{-2}$	$9,98 \times 10^{-1}$	$7,93 \times 10^{-3}$	$5,41 \times 10^{-2}$	$9,97 \times 10^{-1}$	$5,56 \times 10^{-3}$
5	-	-	-	$7,42 \times 10^{-2}$	$9,98 \times 10^{-1}$	$7,63 \times 10^{-3}$	$5,43 \times 10^{-2}$	$9,97 \times 10^{-1}$	$5,46 \times 10^{-3}$
Hiperparâmetros (XGBoost)				Cloração			Fluoretação		
η	D_M	C_{ST}	S_{sa}	REMQ	R^2	EMA	REMQ	R^2	EMA
0,2	6	0,8	0,75	$8,86 \times 10^{-2}$	$9,97 \times 10^{-1}$	$1,58 \times 10^{-2}$	$8,32 \times 10^{-2}$	$9,98 \times 10^{-1}$	$1,40 \times 10^{-2}$
0,2	8	0,8	0,75	$5,61 \times 10^{-2}$	$9,97 \times 10^{-1}$	$9,31 \times 10^{-3}$	$5,42 \times 10^{-2}$	$9,97 \times 10^{-1}$	$8,87 \times 10^{-3}$

a cloração quanto para a fluoretação. O maior valor de R^2 em ambos os casos foi obtido com $N_{rn} = 150$ e $D_M = 6$, mantendo os demais hiperparâmetros inalterados.

5.3.1 Análise da solução com hiperparâmetros otimizados

Para a validação da solução, foram utilizados os dados do período de 31 de outubro de 2022 a 31 de dezembro de 2022 (62 dias), totalizando 129.601 registros, denominados amostra de 2022 (A22), e os dados de 01 de abril de 2024 a 12 de abril de 2024 (12 dias), totalizando 25.084 registros, denominados amostra de 2024 (A24). Inicialmente, foram realizadas validações dos modelos construídos sem a otimização dos hiperparâmetros. Os resultados das métricas aplicadas nesta validação externa estão dispostas na Tabela 7. Observa-se que, para a cloração, o método de predição mais eficiente foi a Floresta Aleatória. Para a fluoretação, o melhor desempenho foi obtido pelo RNA-MLP na A22 e pelo XGBoost na A24. A Figura 8(a), Figura 8(b), Figura 8(c) e Figura 8(d) apresentam os comportamentos das dosagens de cloro e flúor ao longo de 24 horas da A22, comparando os valores aferidos no sistema com os valores preditos pelos modelos.

Os testes de validação externa com hiperparâmetros otimizados foram realizados utilizando as mesmas amostras A22 e A24. Os resultados das métricas aplicadas nesta validação estão dispostas na Tabela 8. Para a cloração, na amostra A22, o menor valor de REMQ e o maior valor de R^2 foram obtidos pelo modelo FR, enquanto o RNA-MLP apresentou o menor valor de EMA. Na amostra A24, os menores valores de REMQ e EMA também foram alcançados pelo modelo FR, mas o maior valor de R^2 foi obtido pelo RNA-MLP. Para a fluoretação, na amostra

Tabela 7 – *Validação externa comparando os resultados dos modelos com os valores reais coletados no sistema.*

Amostra	Modelos	Cloração			Fluoretação		
		REMQ	R^2	EMA	REMQ	R^2	EMA
2022	RNA-MLP	$4,21 \times 10^{-1}$	$9,64 \times 10^{-1}$	$2,31 \times 10^{-1}$	$2,07 \times 10^{-1}$	$9,60 \times 10^{-1}$	$1,04 \times 10^{-1}$
2022	FR	$3,10 \times 10^{-1}$	$9,87 \times 10^{-1}$	$2,47 \times 10^{-1}$	$3,37 \times 10^{-1}$	$9,54 \times 10^{-1}$	$2,40 \times 10^{-1}$
2022	XGBoost	$8,11 \times 10^{-1}$	$8,14 \times 10^{-1}$	$5,90 \times 10^{-1}$	$2,51 \times 10^{-1}$	$9,54 \times 10^{-1}$	$1,61 \times 10^{-1}$
2024	RNA-MLP	$3,25 \times 10^{-1}$	$8,54 \times 10^{-1}$	$1,66 \times 10^{-1}$	$2,72 \times 10^{-1}$	$6,50 \times 10^{-1}$	$1,35 \times 10^{-1}$
2024	FR	$2,68 \times 10^{-1}$	$8,47 \times 10^{-1}$	$1,42 \times 10^{-1}$	$1,75 \times 10^{-1}$	$8,65 \times 10^{-1}$	$9,88 \times 10^{-2}$
2024	XGBoost	$3,05 \times 10^{-1}$	$8,12 \times 10^{-1}$	$1,62 \times 10^{-1}$	$1,60 \times 10^{-1}$	$8,71 \times 10^{-1}$	$8,61 \times 10^{-2}$

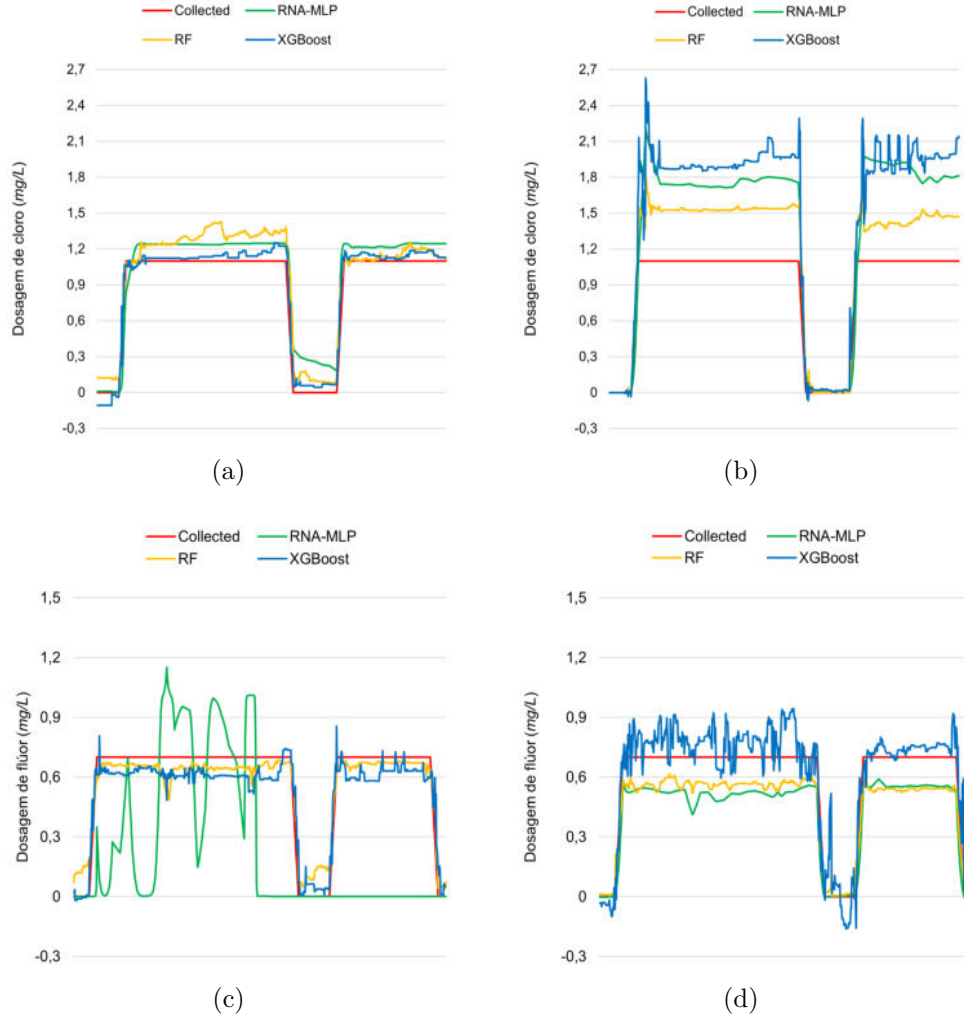
A22, os menores valores de REMQ e EMA, bem como o maior valor de R^2 , foram obtidos pelo modelo RNA-MLP. Na amostra A24, todos os melhores valores (REMQ, EMA e R^2) foram alcançados pelo modelo XGBoost. A Figura 9(a), Figura 9(b), Figura 9(c) e Figura 9(d) apresentam os comportamentos das dosagens ao longo de 24 horas da amostra A24, comparando os valores reais do sistema com as previsões geradas pelos modelos propostos.

Tabela 8 – *Validação externa comparando os resultados dos modelos otimizados com os valores reais coletados no sistema.*

Amostra	Modelos	Cloração			Fluoretação		
		REMQ	R^2	EMA	REMQ	R^2	EMA
2022	RNA-MLP	$1,90 \times 10^{-1}$	$9,89 \times 10^{-1}$	$6,37 \times 10^{-2}$	$6,11 \times 10^{-2}$	$9,96 \times 10^{-1}$	$1,39 \times 10^{-2}$
2022	FR	$1,49 \times 10^{-1}$	$9,93 \times 10^{-1}$	$3,93 \times 10^{-2}$	$1,28 \times 10^{-1}$	$9,85 \times 10^{-1}$	$3,17 \times 10^{-2}$
2022	XGBoost	$1,87 \times 10^{-1}$	$9,89 \times 10^{-1}$	$5,76 \times 10^{-2}$	$1,07 \times 10^{-1}$	$9,89 \times 10^{-1}$	$3,85 \times 10^{-2}$
2024	RNA-MLP	$4,17 \times 10^{-1}$	$9,44 \times 10^{-1}$	$9,64 \times 10^{-2}$	$2,47 \times 10^{-1}$	$9,44 \times 10^{-1}$	$4,40 \times 10^{-2}$
2024	FR	$4,03 \times 10^{-1}$	$9,48 \times 10^{-1}$	$7,79 \times 10^{-2}$	$2,45 \times 10^{-1}$	$9,45 \times 10^{-1}$	$5,39 \times 10^{-2}$
2024	XGBoost	$4,11 \times 10^{-1}$	$9,46 \times 10^{-1}$	$9,43 \times 10^{-2}$	$2,38 \times 10^{-1}$	$9,48 \times 10^{-1}$	$6,05 \times 10^{-2}$

Ao comparar os resultados da Tabela 8, com hiperparâmetros otimizados, com os da Tabela 7, sem otimização, observa-se que o ajuste dos hiperparâmetros aprimorou o desempenho dos modelos com base nas métricas R^2 e EMA. Para a cloração, após a otimização, os valores de REMQ de todos os modelos foram superiores quando as previsões foram comparadas com os dados das amostras A22 e A24, como disposto na Tabela 6. Para a fluoretação, com exceção do modelo RNA-MLP, foi observado o mesmo comportamento em relação ao REMQ. Embora este

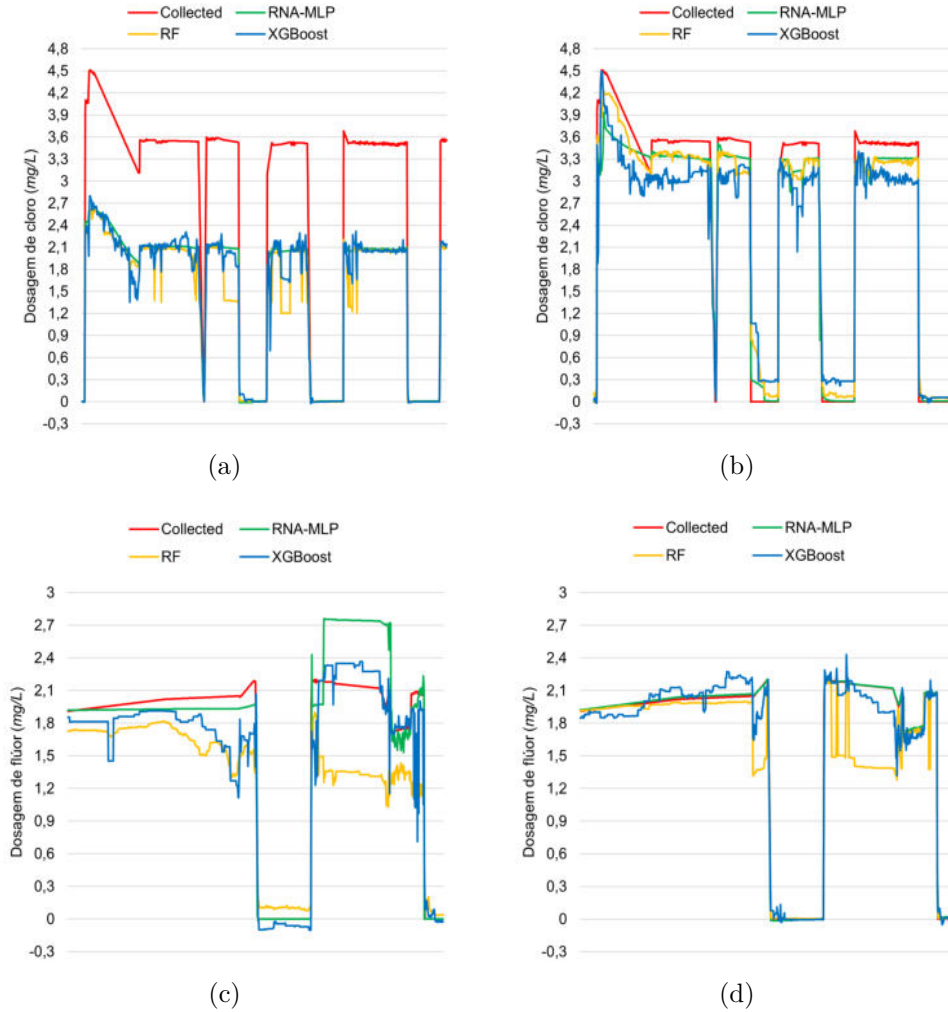
Figura 8 – *Valores reais do sistema e previsões dos modelos com amostra de 2022: (a) cloração sem otimização de hiperparâmetros, (b) cloração com otimização de hiperparâmetros, (c) fluoretação sem otimização de hiperparâmetros e (d) fluoretação com otimização de hiperparâmetros.*



resultado possa indicar sobreajuste, a validação externa demonstra que as previsões mantêm conformidade com os valores aferidos durante a operação do sistema.

Ao analisar os valores de REMQ, R^2 e EMA após a otimização dos hiperparâmetros, observa-se que a diferença de precisão entre os modelos, considerando as amostras A22 e A24, está na ordem de 10^{-2} . Os valores obtidos tanto nas coletas do sistema quanto nas previsões estão dentro dos limites recomendados internacionalmente, como apresentado nas Figura 8(b) e Figura 8(d). Com relação aos modelos, o FR exige quantidade significativa de processamento e memória, especialmente em grandes conjuntos de dados, devido ao armazenamento e ao treinamento de múltiplas árvores de decisão. Embora possua poucos hiperparâmetros, seu tempo de execu-

Figura 9 – *Valores reais do sistema e previsões dos modelos com amostra de 2024: (a) cloração sem otimização de hiperparâmetros, (b) cloração com otimização de hiperparâmetros, (c) fluoretação sem otimização de hiperparâmetros e (d) fluoretação com otimização de hiperparâmetros.*



ção pode ser elevado pela necessidade de construir e armazenar elevado número de árvores para garantir a robustez do modelo.

Os modelos RNA-MLP e XGBoost apresentam rápida execução na fase de predição, porém o tempo de treinamento pode ser elevado em configurações com vários hiperparâmetros e grandes volumes de dados. O RNA-MLP requer ajustes de parâmetros, como o número de neurônios por camada, a taxa de aprendizado e a função de ativação, enquanto o XGBoost demanda a configuração de hiperparâmetros relacionados à estrutura das árvores de decisão, como taxa de aprendizado, número de árvores e profundidade máxima. Devido à alta quantidade de hiperparâmetros, é necessário conhecimento técnico adequado para configurar estes modelos de forma

eficaz e evitar sobreajuste, uma vez que escolhas inadequadas podem aumentar a complexidade e comprometer a generalização das previsões. Embora o RNA-MLP e o XGBoost possuam mecanismos de regularização para reduzir este problema, o risco de adaptação excessiva aos dados de treinamento é maior em modelos complexos com vários parâmetros.

5.3.2 Avaliação de desempenho dos modelos preditivos

De posse dos dados processados, realiza-se a análise estatística para avaliar o desempenho dos modelos preditivos empregados no estudo. Para isto, comparam-se os valores previstos pelos modelos RNA-MLP, FR e XGBoost aos valores reais coletados nas amostras A22 e A24 durante o período de observação. A análise abrange o cálculo das seguintes métricas de desempenho: i) REMQ, ii) R^2 e iii) EMA, além da aplicação do teste estatístico de Mann-Whitney U com subamostras de tamanho $N = 1000$. A Tabela 9 dispõe os valores de REMQ, R^2 , EMA e p -valor para cada modelo e parâmetro avaliado, destacando os desempenhos individuais de cada abordagem.

Tabela 9 – *Resultados das métricas de avaliação para os modelos preditivos de cloro e flúor.*

Parâmetro	Modelo	REMQ	R^2	EMA	p -valor
Cloro	RNA-MLP	0,93	0,44	0,68	$1,11 \times 10^{-8}$
	FR	0,99	0,38	0,68	$5,33 \times 10^{-29}$
	XGBoost	0,96	0,41	0,65	$3,40 \times 10^{-28}$
	RNA-MLP Otimizado	0,66	0,72	0,56	$9,99 \times 10^{-5}$
	FR Otimizado	0,40	0,90	0,36	$1,07 \times 10^{-10}$
	XGBoost Otimizado	0,79	0,60	0,70	$1,11 \times 10^{-8}$
Flúor	RNA-MLP	0,51	0,45	0,44	$3,17 \times 10^{-225}$
	FR	0,30	0,81	0,21	$3,38 \times 10^{-232}$
	XGBoost	0,25	0,87	0,16	$2,37 \times 10^{-163}$
	RNA-MLP Otimizado	0,18	0,93	0,13	$7,17 \times 10^{-71}$
	FR Otimizado	0,19	0,92	0,13	$3,70 \times 10^{-207}$
	XGBoost Otimizado	0,15	0,95	0,09	$7,32 \times 10^{-18}$

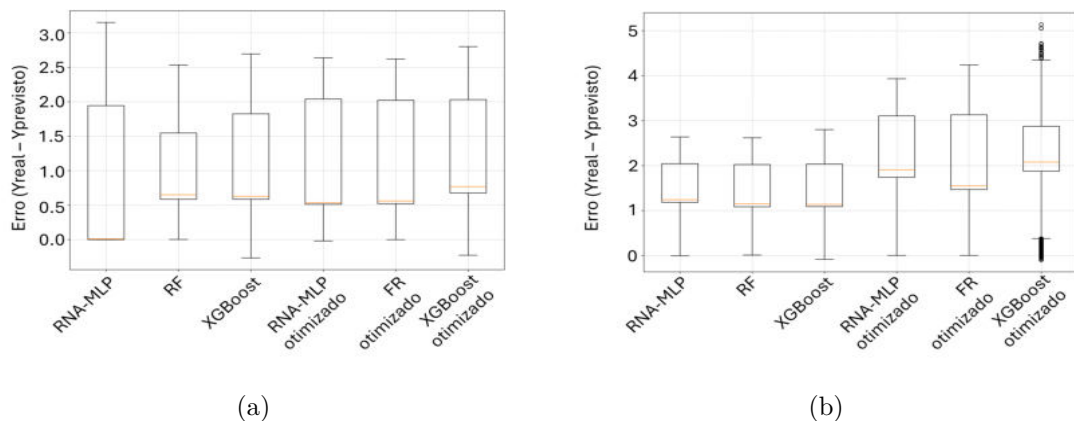
Dos valores dispostos na Tabela 9 o REMQ quantifica o desvio médio das previsões, penalizando mais intensamente erros de maior magnitude. Para o parâmetro de cloro, o modelo FR otimizado apresenta $\text{REMQ} = 0,40$, $R^2 = 0,90$ e $\text{EMA} = 0,36$, indicando desempenho superior. No caso do flúor, o modelo XGBoost

otimizado se destaca com $REMQ = 0,15$, $R^2 = 0,95$ e $EMA = 0,09$, demonstrando alta precisão nas previsões. O coeficiente R^2 representa a proporção da variabilidade explicada pelo modelo, sendo $R^2 \approx 1$ indicativos de melhor desempenho. O EMA fornece medida direta do erro médio das previsões em relação aos valores reais.

O p -valor avalia a significância estatística, sendo p -valor < 0.05 indicativos de diferenças significativas entre as previsões e os valores reais. Todos os modelos apresentaram p -valores baixos, indicando que os erros não são aleatórios. Os modelos otimizados, em especial FR e XGBoost, demonstraram melhorias expressivas nas métricas de desempenho, com redução do REMQ e do EMA e aumento do R^2 . Embora o RNA-MLP otimizado tenha apresentado progresso, seu desempenho permanece inferior ao dos outros modelos otimizados. Observa-se que, devido ao grande número de amostras, os baixos p -valores devem ser interpretados em conjunto com as demais métricas para garantir uma análise precisa.

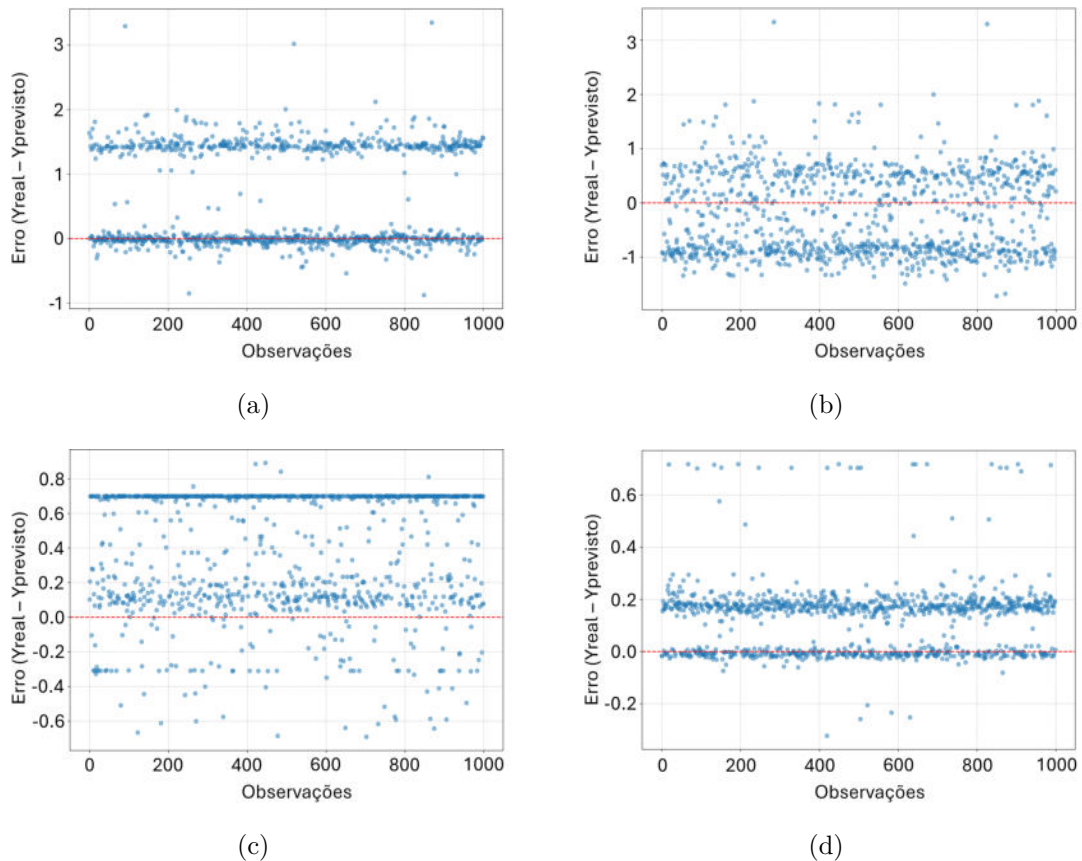
Para avaliar a distribuição dos erros e identificar possíveis pontos fora da curva, foram elaboradas a Figura 10(a) e a Figura 10(b). A linha central de cada *boxplot* representa a mediana dos erros, enquanto as extremidades das caixas correspondem ao primeiro e ao terceiro quartis, definindo o intervalo interquartil. Os *boxplot* apresentam a dispersão dos erros para cada modelo, complementando a análise quantitativa das métricas de desempenho. Observa-se que os modelos otimizados apresentam menor dispersão, sugerindo previsões mais precisas. No caso do cloro, os modelos otimizados apresentam caixas mais compactas e linhas de alcance menores, indicando distribuição de erros mais concentrada. O modelo FR otimizado destaca-se por apresentar a mediana próxima de zero e número reduzido de valores atípicos, sugerindo maior consistência nas previsões. Para o flúor, o modelo XGBoost otimizado demonstra menor dispersão e valores próximos de zero, reforçando sua precisão e estabilidade.

Figura 10 – Distribuição dos erros dos modelos preditivos para: (a) cloro e (b) flúor.



Foram realizadas análises visuais dos resíduos para identificar possíveis padrões ou desvios sistemáticos em relação aos valores reais. A Figura 11 apresenta, no eixo das abscissas, a sequência das previsões ao longo das observações, enquanto o eixo das ordenadas apresenta a diferença entre os valores reais e previstos pelos modelos. Os pontos azuis representam os erros individuais de cada observação, e a linha horizontal vermelha indica a referência de erro nulo, em que as previsões coincidiriam exatamente com os valores reais. Observa-se que a versão otimizada do XGBoost apresenta menor dispersão ao redor da linha de erro nulo, evidenciando maior precisão nas previsões. A redução na amplitude dos erros e a menor quantidade de pontos distantes da linha vermelha indicam uma menor ocorrência de erros extremos. O RNA-MLP otimizado apresenta menor variabilidade nos erros em comparação à versão não otimizada, entretanto, sua dispersão ainda é superior à dos modelos XGBoost, o que sugere menor consistência nas previsões.

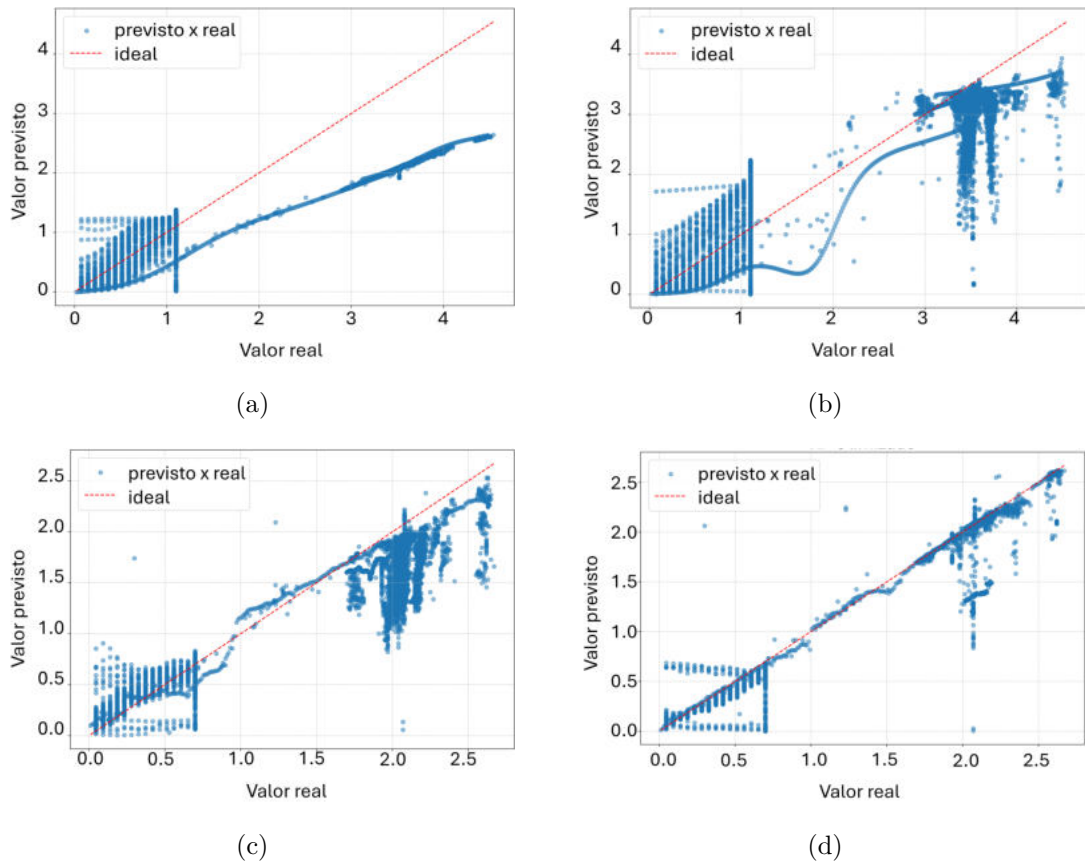
Figura 11 – *Resíduos dos modelos preditivos ao longo das observações: (a) XGBoost sem otimização cloro, (b) XGBoost otimizado cloro, (c) RNA-MLP sem otimização flúor e (d) RNA-MLP otimizado flúor.*



Esta análise visual dos resíduos indica que os modelos otimizados apresentam maior concentração de erros próximos de zero, o que reforça sua precisão. A

ausência de padrões sistemáticos nos resíduos sugere que os erros estão distribuídos aleatoriamente, indicando ausência de viés e comportamento adequado do modelo. Além disto, a menor ocorrência de valores atípicos nos modelos XGBoost aponta para maior estabilidade e melhor capacidade de generalização, corroborando a ausência de sobreajuste nas previsões. Para apresentar a proximidade entre os valores previstos e os valores reais, foi mapeada a dispersão, apresentada na Figura 12. Cada ponto azul representa uma previsão em relação ao valor real correspondente, enquanto a linha vermelha indica os valores ideais, em que $p = r$.

Figura 12 – *Dispersão com valores reais $\times$ previstos:*
 (a) RNA-MLP sem otimização cloro, (b) RNA-MLP otimizado cloro, (c) FR sem otimização flúor e (d) FR otimizado flúor.



Para o modelo RNA-MLP, observa-se dispersão considerável dos pontos em torno da linha ideal, com tendência de subestimar ou superestimar os valores em diferentes intervalos, sugerindo menor precisão nas previsões. No modelo RNA-MLP otimizado, apesar das melhorias, ainda são perceptíveis padrões não lineares e maior concentração de pontos distantes da linha vermelha, indicando limitações na captura precisa dos valores reais. O modelo FR apresenta distribuição mais próxima da linha ideal, com menor dispersão dos pontos. No entanto, a presença de alguns

pontos afastados indica dificuldades do modelo em prever determinados valores com alta precisão. No modelo FR otimizado, observa-se distribuição mais uniforme e concentrada ao redor da linha ideal, com redução significativa dos desvios e maior proximidade dos pontos em relação à linha vermelha. Estes resultados indicam que a otimização aprimorou a capacidade de generalização e previsibilidade do modelo. A ausência de concentrações significativas de pontos distantes da linha ideal reforça que os modelos otimizados não apresentam indícios de sobreajuste ou subajuste.

Todos os modelos apresentaram melhorias após a otimização. Entretanto, para o cloro, o modelo FR otimizado obteve o melhor desempenho, com $REMQ = 0,40$, $R^2 = 0,91$, $EMA = 0,36$ e $p\text{-valor} = 2,17 \times 10^{-18}$. Para o flúor, o modelo XGBoost otimizado destacou-se com os melhores resultados: $REMQ = 0,15$, $R^2 = 0,95$, $EMA = 0,09$ e $p\text{-valor} = 7,32 \times 10^{-18}$. Ao considerar único modelo para prever ambos os parâmetros, o XGBoost otimizado apresentou o melhor desempenho. Os critérios que embasaram esta escolha foram: i) menor valor de REMQ para ambos os parâmetros, ii) elevado R^2 , indicando alta capacidade de explicação da variância dos dados, iii) resíduos sem padrões de tendência, reforçando a ausência de sobreajuste e iv) consistência nas previsões em diferentes períodos e subamostras.

5.4 Discussão

A aplicação de técnicas de inteligência artificial em processos de tratamento de água tem se apresentado promissora na otimização das dosagens de cloro e flúor. Este estudo observou que algoritmos como RNA-MLP, FR e XGBoost fornecem resultados consistentes, com ênfase nas versões otimizadas dos modelos. Para a predição de cloro, o modelo FR otimizado apresentou o menor erro médio quadrático e o maior coeficiente de determinação, indicando elevada precisão. No caso do flúor, o modelo XGBoost otimizado apresentou desempenho superior, refletindo sua capacidade de lidar com padrões não lineares e variações nos dados analisados.

Os resultados indicam convergência com estudos anteriores, como o de Solaimany-Aminabad, Maleki e Hadi (2013), que analisaram a eficiência da FR em sistemas com múltiplas variáveis. Contudo, em comparação com pesquisas que utilizam regressão linear ou métodos convencionais, os modelos propostos apresentaram ganhos expressivos em termos de precisão e estabilidade. A análise indicou que o RNA-MLP, mesmo na versão otimizada, apresentou maior dispersão dos resíduos, sugerindo limitação na identificação de padrões complexos, possivelmente devido à variabilidade dos dados ou à ausência de normalização em determinadas etapas. A superioridade do FR e do XGBoost pode ser atribuída à capacidade destes algoritmos de ponderar atributos com diferentes escalas e lidar com dados não homogêneos.

A Estação de Tratamento de Água analisada neste estudo dispõe de sistema supervisorio que permite ao operador monitorar as medições e controlar o funcionamento de determinados equipamentos, além de possibilitar a operação em modo automático. Os dispositivos para medição de cloração e fluoretação são complementados por testes em bancada, realizados pelo operador para verificação adicional e aferição de parâmetros não automatizados, como a cor.

A aplicação do método proposto apresenta potencial não apenas para monitoramento, mas também para a automação integral do processo de dosagem de produtos químicos na estação de tratamento de água por meio de atuadores inteligentes. Com a integração de sensores em tempo real e válvulas automatizadas controladas pelo modelo preditivo, seria possível ajustar dinamicamente as dosagens de cloro e flúor conforme as variações nas características da água bruta. Esta abordagem reduz a necessidade de intervenções manuais e aprimora a precisão do controle, aumentando a resiliência operacional da ETA. Adicionalmente, a automatização do processo possibilita respostas mais rápidas a flutuações inesperadas na qualidade da água, promovendo maior segurança e eficiência no uso de insumos químicos, além de contribuir para a sustentabilidade das operações, especialmente em regiões com recursos limitados.

Embora alguns estudos indiquem que redes neurais profundas apresentam maior precisão, neste trabalho observou-se que os modelos com estrutura com base em árvores proporcionaram interpretações mais claras e ajustes mais eficientes. Este resultado converge com a literatura, que aponta o XGBoost como referência para sistemas com alta dimensionalidade e dados esparsos (Chen; Guestrin, 2016). A principal lacuna identificada neste estudo refere-se à ausência de métodos computacionais otimizados para estações de pequeno e médio porte com captação superficial. A abordagem proposta oferece uma alternativa resiliente e de baixo custo computacional, contribuindo para suprir esta lacuna ao aplicar inteligência artificial em estações com recursos limitados.

Entre as principais dificuldades identificadas, destaca-se a necessidade de grandes volumes de dados consistentes e bem distribuídos ao longo do tempo. A interpolação por funções spline e a decomposição sazonal são eficientes para preencher lacunas, embora possam introduzir distorções em situações de alta variabilidade nos dados. Adicionalmente, o treinamento dos modelos apresentou elevado custo computacional, com tempo médio de processamento superior a 48 horas, indicando a necessidade de infraestrutura de processamento mais resiliente. Outra limitação corresponde à ausência de mecanismo de retroalimentação automática, capaz de armazenar os dados coletados e realizar atualizações periódicas de forma autônoma. Esta limitação restringe a capacidade do modelo de adaptar-se continuamente às mu-

danças nas condições da água bruta durante a sazonalidade, reduzindo sua precisão preditiva em cenários dinâmicos.

Estas demandas computacionais foram atendidas por meio da utilização de servidor equipado com 44 CPU Intel(R) Xeon(R) Gold 6152, operando a 2,10 GHz, e 512 GB de memória RAM, sem utilização de GPU, destacando os requisitos de infraestrutura necessários para tarefas desta natureza. A complexidade da modelagem de tratamento de água decorre do dilema da dimensionalidade, no qual cada variável adicional monitorada introduz um novo parâmetro ao modelo, aumentando a carga computacional necessária para identificar o modelo preditivo mais adequado. Consequentemente, os elevados custos computacionais observados neste estudo refletem a intrincada natureza da otimização dos processos de tratamento de água, especialmente à medida que cresce o número de parâmetros monitorados.

Como solução para as limitações, recomenda-se a implementação de processo de aprendizado contínuo, permitindo que o modelo ajuste automaticamente seus parâmetros com base nos dados mais recentes, aumentando sua resiliência e precisão preditiva. Este mecanismo reduziria a necessidade de intervenções manuais para reconfiguração e novo treinamento, otimizando os recursos computacionais ao concentrar os esforços em atualizações incrementais. Além disto, a aplicação de técnicas de seleção de atributos pode contribuir para reduzir a dimensionalidade sem comprometer a precisão das previsões. Com estas estratégias, reforça-se a aplicabilidade de modelos de inteligência artificial na gestão de recursos hídricos, promovendo maior segurança no monitoramento da qualidade da água. Os resultados indicam que a integração de inteligência artificial em processos de tratamento é viável e contribui para o cumprimento dos padrões regulatórios, além de promover a sustentabilidade ao reduzir o uso de reagentes químicos.

Conclusão

Este estudo investigou a aplicação de modelos de inteligência artificial para otimizar os processos de cloração e fluoretação em estações de tratamento de água, com ênfase na precisão preditiva e na eficiência operacional. As técnicas de aprendizado de máquina utilizadas, incluindo RNA-MLP, FR e XGBoost, indicaram resultados promissores, com destaque para as versões otimizadas destes modelos. A hipótese, que previa a melhoria das previsões de dosagem de produtos químicos por meio de algoritmos supervisionados, foi corroborada. Os objetivos estabelecidos foram alcançados, contemplando a redução dos erros de previsão e a maior consistência dos resultados em diferentes amostras e períodos.

O modelo FR otimizado apresentou o melhor desempenho na previsão das dosagens de cloro, com menores valores de REMQ e EMA e coeficiente de determinação elevado, refletindo sua alta capacidade de generalização. O modelo XGBoost otimizado destacou-se na predição de flúor, e na previsão de ambos os parâmetros (cloro e flúor), demonstrando maior precisão e estabilidade nas previsões. Estes resultados indicam que os algoritmos otimizados são eficientes no manejo da variabilidade das condições da água tratada, assegurando a conformidade com os padrões de potabilidade e contribuindo para a segurança da saúde pública.

Este estudo demonstrou, por meio de simulação, a viabilidade da aplicação de modelos de inteligência artificial para a predição dos valores de cloração e fluoretação em Estações de Tratamento de Água, utilizando regressão supervisionada. O objetivo principal foi alcançado, assim como os objetivos específicos, incluindo a potencial redução de custos operacionais e o aumento da eficiência no processo. Além disto, o modelo desenvolvido possibilita um ambiente de simulação que aprimora a precisão das intervenções na planta industrial, fortalece o controle de qualidade do produto final e contribui para o cumprimento do ODS 6.

Portanto, conclui-se que a integração de inteligência artificial nas operações de tratamento de água é uma abordagem viável e eficiente, capaz de aumentar a operacionalidade do processo, reduzir custos e garantir o atendimento aos padrões regulatórios. Os resultados indicam que esta estratégia pode ser adaptada a diferen-

tes cenários de monitoramento e controle, com potencial de aplicação em estações com recursos limitados, promovendo sustentabilidade, automação e inovação tecnológica no setor.

6.1 Contribuições diretas do trabalho

6.1.1 Rotinas desenvolvidas para aplicação na ETA

O desenvolvimento de rotinas programadas em linguagem R possibilitou a obtenção dos resultados neste estudo, sendo fundamental para a coleta, pré-processamento, modelagem e avaliação dos dados da ETA. Foram implementadas funções específicas para segmentação de séries temporais, preenchimento de lacunas nos dados, normalização de variáveis e divisão dos conjuntos de treino e teste, assegurando que os modelos fossem treinados adequadamente. Além disto, os códigos otimizados permitiram a execução eficiente de algoritmos de aprendizado de máquina, como redes neurais artificiais do tipo perceptron multicamadas (RNA-MLP), florestas aleatórias (FR) e XGBoost, incluindo ajustes automáticos de hiperparâmetros por meio de busca em grade e validação cruzada. A estrutura modular destas rotinas possibilita sua reutilização e adaptação a diferentes cenários operacionais, tornando-as ferramenta versátil para a otimização da previsão das dosagens de cloro e flúor.

A aplicação destas rotinas na ETA de Brazabranes, operada pela Saneago, fornece solução resiliente para a automação do controle das dosagens, permitindo o ajuste dinâmico dos parâmetros operacionais com base em previsões mais precisas. Os códigos desenvolvidos podem ser integrados a sistemas supervisórios e dispositivos de controle, como atuadores e reguladores ajustáveis, otimizando o uso de produtos químicos e minimizando desperdícios. Além disto, a estrutura modular das rotinas facilita sua replicação em outras estações de tratamento de água, aprimorando a confiabilidade na tomada de decisão e auxiliando no atendimento aos padrões de potabilidade. Desta forma, este estudo confirma a viabilidade da aplicação da inteligência artificial no tratamento de água e disponibiliza ferramentas computacionais eficientes que podem ser incorporadas em ambientes operacionais, garantindo processo mais preciso e sustentável.

6.1.2 Publicação científica

Como desdobramento desta pesquisa, foi elaborado o artigo intitulado: *Modeling and control of critical parameters in water treatment plants using artificial intelligence*, como apresentado na Figura 13, e submetido ao periódico Water

Resources Management (ISSN 0920-4741), classificado no estrato A1 do Qualis CAPES. A submissão deste artigo fortalece a contribuição do estudo para a aplicação de inteligência artificial no controle de parâmetros críticos em estações de tratamento de água, além de ampliar a disseminação dos resultados entre pesquisadores e profissionais da área.

Figura 13 – *Artigo submetido.*

Modeling and control of critical parameters in water treatment plants using artificial intelligence

Diego Takashi Sato^{a,d,*}, Orlando M. Oliveira Belo^c, Antonio P. Castro Junior^f, Clóves Gonçalves Rodrigues^e, Antonio Paulo Coimbra^b and Wesley Pacheco Calixto^{a,b,d,*}

^aTechnology Research and Development Center (GCITE), Federal Institute of Goiás (IFG), Goiania, Goiás/Brasil. PC 74.055-110,

^bInstitute of Systems and Robotics (ISR), Coimbra University, Coimbra/Portugal. PC 3030-290,

^cDepartment of Informatics, School of Engineering, University of Minho, Braga/Portugal. PC 4710-057,

^dElectrical, Mechanical & Computer Engineering School (EMC), Federal University of Goiás (UFG), Goiania, Goiás/Brasil. PC 74.605-220,

^eGraduate Program in Production and Systems Engineering, Pontifical Catholic University of Goiás, Goiania, Goiás/Brazil. PC 74.605-010,

^fDirectorate of Data Science, Court of Justice of the State of Goiás, Goiania, Goiás/Brazil. PC 74.130-011,

ARTICLE INFO

Keywords:

System modeling
Water treatment
Artificial intelligence
Process optimization
Sustainability

ABSTRACT

This study proposes a predictive model to optimize chlorination and fluoridation processes in water treatment plants by applying machine learning algorithms such as multilayer artificial neural networks, random forests, and XGBoost. The data analyzed, obtained from a plant with surface water intake, undergo pre-processing techniques and hyperparameter optimization to ensure accurate predictions. The findings indicate significant performance improvements after the optimization process, with reduced errors and improved generalization capacity of the models. The optimized models present RMSE ≈ 0.40 for chlorine dosage, with a coefficient of determination ≈ 0.90 , and RMSE ≈ 0.15 and ≈ 0.95 for fluoridation. These observations suggest that the integration of artificial intelligence in water treatment operations allows efficient and automated monitoring while ensuring compliance with regulatory drinking water standards. The results highlight the potential of the approach to contribute to the sustainability of the sector by reducing chemical inputs and promoting more precise and safer operations.

Acknowledgements:

The authors acknowledge national funding from FCT - Fundação para a Ciência e a Tecnologia, I.P., under project UIDB/00048/2020 (DOI: 10.54499/UIDB/00048/2020), and from the National Council for Scientific and Technological Development (CNPq/Brazil) for its support through a Research Productivity Scholarship (process no. 301644/2022-5). The authors also extend their gratitude to Saneago S.A. for providing the data and to LaMCAD/UFG for supporting the research with computational resources.

Diego Takashi Sato: Master's Student
Orlando M. Oliveira Belo: Visiting Researcher
Antonio P. Castro Junior: Co-Supervisor
Clóves Gonçalves Rodrigues: Visiting Researcher
Antonio Paulo Coimbra: Visiting Researcher
Wesley Pacheco Calixto: Supervisor

* Corresponding authors

✉ diegotakashi@saneago.com.br (D.T. Sato); wesley.pacheco@ufg.br (W. Calixto)

ORCID(s): 0009-0004-1356-7265 (D.T. Sato); 0000-0003-2157-8891 (O.M.O. Belo); 0000-0002-6127-3139 (A.P. Castro Junior); 0000-0003-0140-9847 (C. Rodrigues); 0000-0003-3780-2405 (A.P. Coimbra); 0000-0002-1928-4432 (W. Calixto)

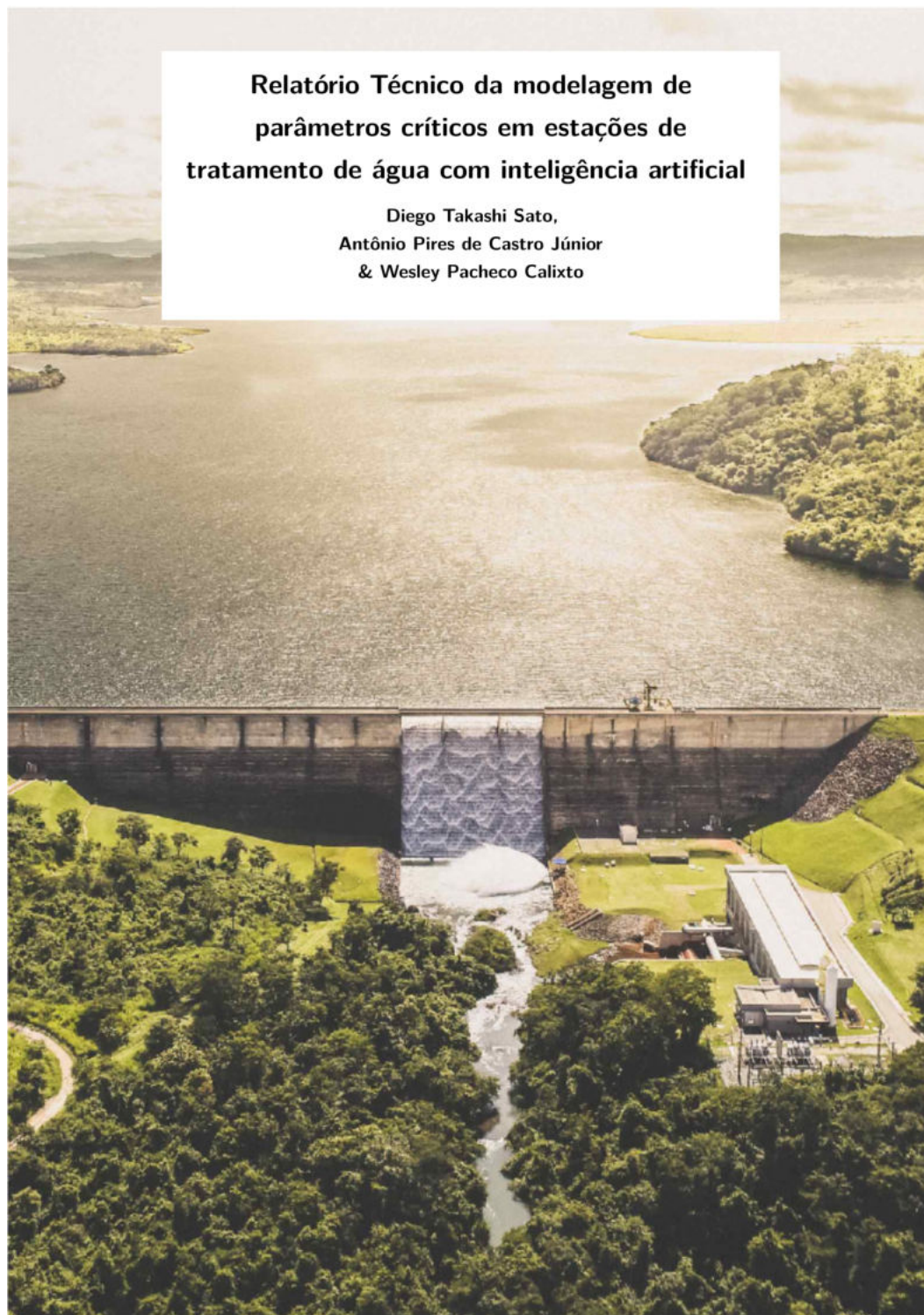
6.1.3 Relatório técnico

Como contribuição direta desta pesquisa, foi elaborado o relatório técnico científico, conforme apresentado na Figura 14, documentando os procedimentos aplicados aos dados da ETA de Brazabrantes. O documento abrange desde a aquisição e o tratamento dos dados até a implementação e avaliação dos modelos de inteligência artificial, oferecendo base metodológica estruturada para futuras aplicações. O relatório foi disponibilizado à Saneago, permitindo a replicação das análises em outras ETA e fornecendo suporte técnico para a otimização dos processos de cloração e fluoretação. Assim, este estudo não apenas contribui para o avanço do conhecimento científico, mas também gera impacto prático na gestão e na eficiência operacional da empresa.

6.2 Trabalhos futuros

- Integração de diferentes algoritmos, como RNA-MLP, FR e XGBoost, em abordagens de aprendizado de máquina do tipo *ensemble*.
- Investigação do potencial de redes neurais profundas, como LSTM (*Long Short-Term Memory*) e *Transformers*, para modelagem de padrões temporais complexos.
- Ampliação do estudo para a previsão e controle da dosagem de outros produtos químicos no tratamento da água, como coagulantes e alcalinizantes.
- Implementação prática dos modelos preditivos em sistemas de controle em tempo real, com conexão a atuadores inteligentes para ajuste dinâmico e autônomo da aplicação de produtos químicos.
- Replicação da metodologia proposta em estações de tratamento de água com características distintas para validação da capacidade de generalização dos modelos.

Figura 14 – *Relatório técnico-científico.*



Referências

AL-HAIJA, Qasem Abu. Exploration of tools for data science. **Data Science with Semantic Technologies: Theory, Practice, and Application**, Wiley Online Library, p. 31–69, 2022. 32

ALIASHRAFI, Atefeh; ZHANG, Yirao; GROENEWEGEN, Hannah; PELEATO, Nicolas M. A review of data-driven modelling in drinking water treatment. **Reviews in Environmental Science and Bio/Technology**, v. 20, n. 4, p. 985–1009, 2021. ISSN 1572-9826. 34

ALLEN, Erik W. Process water treatment in canada’s oil sands industry: I. target pollutants and treatment objectives. **Journal of Environmental Engineering and Science**, v. 7, n. 2, p. 123–138, 2008. 19

AOUN, Antoine; DARWICHE, Farah; HAYEK, Sibelle Al; DOUMIT, Jacqueline. The fluoride debate: The pros and cons of fluoridation. **Preventive nutrition and food science**, v. 23, n. 3, p. 171–180, 2018. 23

ARAÚJO, W. R. H. **Potencialização do rendimento do gerador a relutância chaveado empregando técnica de rastreamento associada a controle de tensão otimizado**. Dissertação (Mestrado) — Universidade Federal de Goiás, Goiânia, Goiás, BR, 2020. 37

BLEI, David M.; SMYTH, Padhraic. Science and data science. **Proceedings of the National Academy of Sciences**, v. 114, n. 33, p. 8689–8692, 2017. 32

BLÁZQUEZ-GARCÍA, Ane; CONDE, Angel; MORI, Usue; LOZANO, Jose A. A review on outlier/anomaly detection in time series data. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 3, p. 1–33, 2021. 35

BREIMAN, Leo. Random forests. **Machine Learning**, Netherlands, v. 45, p. 5–32, October 2001. 34, 35

BULHOES, Junio S; MARTINS, Cristiane L; OLIVEIRA, Marcia D; CALHEIROS, Debora F; CALIXTO, Wesley P. Indirect prediction system for variables that have gaps in their time series. **Chaos, Solitons & Fractals**, Elsevier, v. 131, p. 109509, 2020. 36

CESCON, Anna; JIANG, Jia-Qian. Filtration process and alternative filter media material in water treatment. **Water**, v. 12, n. 12, 2020. ISSN 2073-4441. 28

CHANDRA, Shreeram Suresh; UPADHYE, Akshay; SARAVANAN, Purushothaman; GURUGOPINATH, Sanjeev; MURALISHANKAR, R. Deep neural network architectures for spectrum sensing using signal processing features. *In: IEEE. 2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics*. DC, USA, 2021. p. 129–134. 34

CHARROIS, Jeffrey WA. Private drinking water supplies: challenges for public health. **Canadian Medical Association Journal**, Canadian Medical Association, v. 182, n. 10, p. 1061–1064, 2010. 20, 31

CHEN, Dechao; QIAO, Tianjian; TAN, Hongzhou; LI, Mingming; ZHANG, Yunong. Solving the problem of runge phenomenon by pseudoinverse cubic spline. *In: IEEE. 2014 IEEE 17th International Conference on Computational Science and Engineering*. Chengdu, China, 2014. p. 1226–1231. 36

CHEN, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. *In: . New York, NY, USA: Association for Computing Machinery, 2016. p. 785–794. 35, 64*

CHHETRI, Ravi Kumar; KLUPSCH, Ewa; ANDERSEN, Henrik Rasmus; JENSEN, Pernille Erland. Treatment of arctic wastewater by chemical coagulation, uv and peracetic acid disinfection. **Environmental Science and Pollution Research**, Springer, v. 25, p. 32851–32859, 2018. 29

CHOI, Junhyuk; LIM, Kyoung Jae; JI, Bongjun. Robust imputation method with context-aware voting ensemble model for management of water-quality data. **Water Research**, Elsevier BV, v. 243, p. 120369, 9 2023. ISSN 0043-1354. 36

CLEVELAND, Robert B; CLEVELAND, William S; MCRAE, Jean E; TERPENNING, Irma *et al.* Stl: A seasonal-trend decomposition. **J. off. Stat**, v. 6, n. 1, p. 3–73, 1990. 36, 40

DAYHOFF, Judith E. **Neural network architectures: an introduction**. New York, NY: Van Nostrand Reinhold Co., 1990. 34

DESTIANI, R; TEMPLETON, MR. Chlorination and ultraviolet disinfection of antibiotic-resistant bacteria and antibiotic resistance genes in drinking water. **AIMS Environmental Science**, v. 6, n. 3, 2019. 29

DORIA, Miguel França. Factors influencing public perception of drinking water quality. **Water policy**, IWA Publishing, v. 12, n. 1, p. 1–19, 2010. 20

DURRLEMAN, S.; SIMON, R. Flexible regression models with cubic splines. **Statistics in Medicine**, v. 8, n. 5, p. 551–561, 1989. 40

FAUST, S. M.; ALY, O. M. **Chemistry of water treatment**. 2nd. ed. Boca Raton, FL: Taylor and Francis Group, 1998. 485-565 p. 28

FAWELL, J.; BAILEY, K.; CHILTON, J.; DAHI, E.; FEWTRELL, L.; MAGARA, Y. **Fluoride in drinking-water**. London, UK: World Health Organization, 2006. 47

FEDORKO, Richard; KRÁL', Štefan; FEDORKO, Igor. Artificial intelligence and machine learning in the context of e-commerce: A literature review. **Communication and Intelligent Systems: Proceedings of ICCIS 2021**, Springer, p. 1067–1082, 2022. 33

FERNÍCOLA, Nilda Gallego Gándara de; AZEVEDO, Fausto Antonio de. Metemoglobinemia e nitrato nas águas. **Revista de Saúde Pública**, Faculdade de Saúde Pública da Universidade de São Paulo, v. 15, n. 2, p. 242–248, Apr 1981. 29

FORNBERG, Bengt; ZUEV, Julia. The runge phenomenon and spatially variable shape parameters in rbf interpolation. **Computers & Mathematics with Applications**, Elsevier, v. 54, n. 3, p. 379–398, 2007. 36

FUNASA; UFF; IBAM. **Cadernos temáticos: Abastecimento de água potável**. Rio de Janeiro, RJ: Grafitto, 2016. 27

GALAL-GORCHEV, Hend. Chlorine in water disinfection. **Pure and Applied chemistry**, De Gruyter, v. 68, n. 9, p. 1731–1735, 1996. 19

GANZAROLI, Cleber Asmar; CARVALHO, Douglas Freire de; COIMBRA, Antonio Paulo; COUTO, Luiz Alberto do; CALIXTO, Wesley Pacheco. Comparative analysis of the optimization and implementation of adjustment parameters for advanced control techniques. **Energies**, v. 15, n. 11, 2022. ISSN 1996-1073. 37

GARBIN, Cléa Adas Saliba; SANTOS, Luis Felipe Pupim dos; GARBIN, Artênio José Isper; MOIMAZ, Suzely Adas Saliba; SALIBA, Orlando. Fluoridation of public

water supply: bioethical, legal and political approach. **Revista Bioética**, SciELO Brasil, v. 25, p. 328–337, 2017. 30

GHAHRAMANI, Zoubin. Probabilistic machine learning and artificial intelligence. **Nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 452–459, 2015. 33

GHOBADI, Azadeh; CHERAGHI, Mehrdad; SOBHANARDAKANI, Soheil; LORESTANI, Bahareh; MERRIKHPOUR, Hajar. Groundwater quality modeling using a novel hybrid data-intelligence model based on gray wolf optimization algorithm and multi-layer perceptron artificial neural network: a case study in asadabad plain, hamedan, iran. **Environmental Science and Pollution Research**, Springer, p. 1–15, 2022. 34

GIRI, Subhasis; QIU, Zeyuan. Understanding the relationship of land uses and water quality in twenty first century: A review. **Journal of environmental management**, Elsevier, v. 173, p. 41–48, 2016. 21

GLEICK, Peter H. The human right to water. **Water policy**, Elsevier, v. 1, n. 5, p. 487–503, 1998. 19

GOURMELON, M; LAZURE, P; HERVIO-HEATH, D; SAUX, JC Le; CAPRAIS, MP; GUYADER, F Le; CATHERINE, M; POMMEPUY, M; REES, G; POND, K *et al.* Safe management of shellfish and harvest waters. WHO, IWA Publishing, London, p. 297, 2010. 34

GUDINA, Djair Felix da Silva Natascha Barbosa. Restrições no abastecimento de água: atual situação da rede de distribuição em região de maceió. **Cadernos de Graduação - Ciências exatas e tecnológicas**, v. 4, n. 2, p. 173–186, NOV 2017. ISSN 1980-1777. 27

GUIDI, José Marcius Marson. Análise da variação do consumo de água utilizando dados obtidos por sistemas supervisores. *In: Congresso ABES FENASAN 2017*. São Paulo, SP: [s.n.], 2017. 27

HOFKES, EH. **Small community water supplies**. Netherlands: International Reference Centre for Community Water Supply and Sanitation, 1981. 28

HRUDEY, Steve E. Chlorination disinfection by-products, public health risk trade-offs and me. **Water research**, Elsevier, v. 43, n. 8, p. 2057–2092, 2009. 19

HUO, Yonghua; YAN, Yu; DU, Dan; WANG, Zhihao; ZHANG, Yixin; YANG, Yang. Long-term span traffic prediction model based on stl decomposition and lstm.

In: IEEE. 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS). Matsue, Japan, 2019. p. 1–4. 36

JAYAWEERA, CD; AZIZ, N. Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment. *In: IOP PUBLISHING. Journal of physics: conference series.* United Kingdom, 2018. v. 1123, n. 1, p. 012032. 34

JENA, Pramod Kumar; RAHAMAN, Sayed Modinur; MOHAPATRA, Pradeep Kumar Das; BARIK, Durga Prasad; PATRA, Dikshya Surabhi. Surface water quality assessment by Random Forest. **Water Practice and Technology**, v. 18, n. 1, p. 201–214, 12 2022. 34

JENKINS, J. M.; KOWALSKI, M.; ALVARENGA, E. F. S. Predictive modelling of water losses using random forests on weather covariates. **Water Supply**, v. 18, n. 6, p. 2180–2187, 02 2018. 34

JOHNSON, Sarah Fossen. Methemoglobinemia: Infants at risk. **Current problems in pediatric and adolescent health care**, Elsevier, v. 49, n. 3, p. 57–67, 2019. 29

JOLLY, Kevin. **Machine learning with scikit-learn quick start guide: classification, regression, and clustering techniques in Python.** Birmingham, Reino Unido: Packt Publishing Ltd, 2018. 34

JUN, Soojin; IRUDAYARAJ, Joseph; DEMIRCI, Ali; GEISER, David. Pulsed uv-light treatment of corn meal for inactivation of aspergillus niger spores. **International Journal of Food Science & Technology**, Wiley Online Library, v. 38, n. 8, p. 883–888, 2003. 40

KAUR, Amandeep; BALA, Ritu; BHINDER, Surinder Singh; KANSAL, Sushil Kumar. Health impacts due to fluoride contamination in water: current scenario. *In: Contamination of Water.* New Delhi, India: Academic Press, 2021. p. 65–84. ISBN 978-0-12-824058-8. 30

KHAN, Farhan Mohammad; GUPTA, Rajiv; SEKHRI, Sheetal. A novel pca-fa-ann based hybrid model for prediction of fluoride. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 35, p. 2125–2152, 2021. 23

KHAN, Md; ISLAM, Nazrul; UDDIN, Jia; ISLAM, Sifatul; NASIR, Mostofa. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. **Journal of King Saud University - Computer and Information Sciences**, v. 34, 06 2021. 22, 23

KIM, Moon-Kyung; ZOH, Kyung-Duk. Occurrence and removals of micropollutants in water environment. **Environmental Engineering Research**, Korean Society of Environmental Engineers, v. 21, n. 4, p. 319–332, 2016. 28

KUČERA, Tomáš; TUHOVČÁK, Ladislav; BIELA, Renata. Technical auditing of water supply systems – part 2: water treatment plant. **Water Supply**, v. 18, n. 4, p. 1141–1150, 06 2017. ISSN 1606-9749. 20

LAINÉ, J. M.; VIAL, D.; MOULART, P. Status after 10 years of operation - overview of uf technology today. **Desalination**, v. 131, p. 17–25, 2000. 28

LASHARI, Shan e Zehra; TAKBIRI-BORUJENI, Ali; FATHI, Ebrahim; SUN, Ting; RAHMANI, Reza; KHAZAELI, Mehdi. Drilling performance monitoring and optimization: a data-driven approach. **Journal of Petroleum Exploration and Production Technology**, Springer, v. 9, p. 2747–2756, 2019. 34

LAWSON, Anton E. The generality of hypothetico-deductive reasoning: Making scientific thinking explicit. **The American Biology Teacher**, JSTOR, v. 62, n. 7, p. 482–495, 2000. 38

LEVY, Dan Henry; SGAN-COHEN, Harold; SOLOMONOV, Michael; SHEMESH, Avi; ZIV, Eran; GLASSBERG, Elon; YAVNAI, Nirit. Association of nationwide water fluoridation, changes in dental care legislation, and caries-related treatment needs: A 9-year record-based cross-sectional study. **Journal of Dentistry**, v. 134, p. 104550, 2023. ISSN 0300-5712. 31

LI, Xiang; DONG, Donglin; LIU, Kun; ZHAO, Yi; LI, Minmin. Identification of mine mixed water inrush source based on genetic algorithm and xgboost algorithm: A case study of huangyuchuan mine. **Water**, v. 14, n. 14, 2022. 35

LIBRANTZ, André Felipe; SANTOS, Fábio Cosme Rodrigues dos; DIAS, Cleber Gustavo. Artificial neural networks to control chlorine dosing in a water treatment plant. **Acta Scientiarum. Technology**, v. 40, n. 1, p. e37275, set. 2018. ISSN 1807-8664. 22, 23

LUDERMIR, Teresa Bernarda. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, p. 85–94, 2021. 33

LUKENS, John N. The legacy of well-water methemoglobinemia. **Jama**, American Medical Association, v. 257, n. 20, p. 2793–2795, 1987. 29

MAKUNGO, R.; ODIYO, J. O.; TSHIDZUMBA, N. Performance of small water treatment plants: The case study of mutshedzi water treatment plant. **PHYSICS**

AND CHEMISTRY OF THE EARTH, v. 36, n. 14-15, p. 1151–1158, 2011. ISSN 1474-7065. 11th WaterNet/WARFSA/GWP-SA Annual Symposium, Victoria Falls, ZIMBABWE, OCT 27-29, 2010. 20

MARIN, Heimar de Fatima; ZINADER, Juliana Pereira de Souza; PIRES, Frank James da Silva; BARROS, Jacson Venancio de. The brazilian digital health system: building the digital transformation to engage country citizens. *In: Roadmap to Successful Digital Health Ecosystems*. London, UK: Elsevier, 2022. p. 473–487. 21

MARTINEZ, Terry T; LONG, Christopher. Explosion risk from swimming pool chlorinators and review of chlorine toxicity. **Journal of Toxicology: Clinical Toxicology**, Taylor & Francis, v. 33, n. 4, p. 349–354, 1995. 30

MASJOUDI, Mahsa; MOHSENI, Madjid; BOLTON, James R. Sensitivity of bacteria, protozoa, viruses, and other microorganisms to ultraviolet radiation. **Journal of Research of the National Institute of Standards and Technology**, Superintendent of Documents, v. 126, p. 1–77, 2021. 40

MEULEMANS, C.C.E. The basic principles of uv–disinfection of water. **Ozone: Science & Engineering**, Taylor & Francis, v. 9, n. 4, p. 299–313, 1987. 40

MIKLOS, David B.; REMY, Christian; JEKEL, Martin; LINDEN, Karl G.; DREWES, Joerg E.; HUEBNER, Uwe. Evaluation of advanced oxidation processes for water and wastewater treatment - a critical review. **WATER RESEARCH**, PERGAMON-ELSEVIER SCIENCE LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND, v. 139, p. 118–131, AUG 1 2018. ISSN 0043-1354. 28

MO, Tiexiang; LI, Shanshan; LI, Guodong. An interpretable machine learning model for predicting cavity water depth and cavity length based on xgboost-shap. **JOURNAL OF HYDROINFORMATICS**, v. 25, n. 4, p. 1488–1500, JUL 2023. 35

MOMBA, M. N. B.; OBI, C. L.; THOMPSON, P. Survey of disinfection efficiency of small drinking water treatment plants: Challenges facing small water treatment plants in south africa. **WATER SA**, v. 35, n. 4, p. 485–493, JUL 2009. ISSN 0378-4738. 19

MORALES, Eduardo F; ESCALANTE, Hugo Jair. A brief introduction to supervised, unsupervised, and reinforcement learning. *In: Biosignal processing and*

classification using computational learning and intelligence. Puebla, Mexico: Elsevier, 2022. p. 111–129. 34

MORE, Kagiso Samuel; WOLKERSDORFER, Christian. Application of machine learning algorithms for nonlinear system forecasting through analytics — a case study with mining influenced water data. **Water Resources and Industry**, v. 29, p. 100209, 2023. ISSN 2212-3717. 22, 23, 35

MULLANE, DM O; BAEZ, RJ; JONES, S; LENNON, MA; PETERSEN, PE; RUGG-GUNN, AJ; WHELTON, H; WHITFORD, Gary M. Fluoride and oral health. **Community dental health**, Dennis Barber & The British Association for the Study of Community Dentistry, v. 33, n. 2, p. 69–99, 2016. 30

MULLEN, Joe. History of water fluoridation. **British Dental Journal**, Nature Publishing Group, v. 199, n. 7, p. 1–4, 2005. 30

MURRAY, John J; World Health Organization *et al.* **Appropriate use of fluorides for human health.** Geneve, Switzerland: World Health Organization, 1986. 20, 30

NADIRI, Ata Allah; FIJANI, Elham; TSAI, Frank T.-C.; MOGHADDAM, Asghar Asghari. Supervised committee machine with artificial intelligence for prediction of fluoride concentration. **Journal of Hydroinformatics**, v. 15, n. 4, p. 1474–1490, 05 2013. 23

NAMOR, AF Danil de. Water purification: from ancient civilization to the xxi century. **Water Science and Technology: Water Supply**, IWA Publishing, v. 7, n. 1, p. 33–39, 2007. 27, 28

NEGOITESCU, Arina; TOKAR, Adriana. Water disinfection and the implementation of modern technologies at potable water treatment station. **Revista de Chimie**, v. 63, n. 10, p. 1079–1089, 2012. 28

NEWBRUN, Ernest. Effectiveness of water fluoridation. **Journal of Public Health Dentistry**, Wiley Online Library, v. 49, n. 5, p. 279–289, 1989. 20, 30, 31

OCAMPO-RODRÍGUEZ, Dulce Brigitte; VÁZQUEZ-RODRÍGUEZ, Gabriela Alejandra; MARTÍNEZ-HERNÁNDEZ, Sylvia; ITURBE-ACOSTA, Ulises; CORONEL-OLIVARES, Claudia. Water disinfection: a review of conventional and advanced treatments with chlorine and peracetic acid. **Ingeniería del agua**, IWA Publishing, v. 26, n. 3, p. 185–204, 2022. 20, 28, 29

OUYANG, Zuokun; JABLOUN, Meryem; RAVIER, Philippe. Stlformer: exploit stl decomposition and rank correlation for time series forecasting. *In: IEEE. 2023 31st European Signal Processing Conference (EUSIPCO)*. Helsinki, Finland, 2023. p. 1405–1409. 36

PAI, Swathi; ARUDRA, Annepu; V, Dankan Gowda; VADAR, Parashuram Shankar; JADHAV, Mahadeo Ramchandra; KAUR, Mandeep. Predictive modeling of dental health outcomes based on fluoride concentrations using ai. *In: 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. Bangalore, India: [s.n.], 2023. p. 1–7. 23

PERES, Marco Aurélio; FERNANDES, Liliane Simara; PERES, Karen Glazer. Inequality of water fluoridation in southern brazil—the inverse equity hypothesis revisited. *Social Science & Medicine*, Elsevier, v. 58, n. 6, p. 1181–1189, 2004. 20, 31

PERRITT, Henry H. **Law and the information superhighway**. New York, NY, USA: Wolters Kluwer, 2001. 46

PINTO, A.; FERNANDES, A.; VICENTE, H.; NEVES, J. Optimizing water treatment systems using artificial intelligence based tools. **WIT Transactions on Ecology and the Environment**, Southampton, Hampshire, UK, v. 125, p. 185–194, 2009. ISSN 1743-3541. 21, 23

RHEE, F Van; BEAUMONT, DM. Gastric stricture complicating oral ingestion of bleach. **British journal of clinical practice**, Medicom International, v. 44, n. 12, p. 681–682, 1990. 30

RIEPL, D; WOLF, L; KAEMPGEN, B; HEINZ, H. Making informed decisions—a collaborative and knowledge based iwrn planning exercise in wadishueib, jordan. **Proceedings IWRM**, 2010. 32

RODRIGUEZ, Rafael; PASTORINI, M.; ETCHEVERRY, L.; CHRETIES, C.; FOS-SATI, M.; CASTRO, A.; GORGOGNONE, A. Water-quality data imputation with high percentage of missing values: A machine learning approach. **Sustainability**, 2021. 36

ROY, Surendra; DASS, Gurcharan. Fluoride contamination in drinking water—a review. **Resour. Environ**, v. 3, n. 3, p. 53–58, 2013. 31

RUSSEL, SJ; NORVING, P. **Artificial Intelligence - a modern aproach**. New Jersey: Prentice Hall, 1995. 33

SADEQ, Mina; MOE, Christine L; ATTARASSI, Benaissa; CHERKAOUI, Imad; ELAOUAD, Rajae; IDRISSE, Larbi. Drinking water nitrate and prevalence of methemoglobinemia among infants and children aged 1–7 years in moroccan areas. **International journal of hygiene and environmental health**, Elsevier, v. 211, n. 5-6, p. 546–554, 2008. 29

SAFEER, Soma; PANDEY, Ravi P; REHMAN, Bushra; SAFDAR, Tuba; AHMAD, Iftikhar; HASAN, Shadi W; ULLAH, Asmat. A review of artificial intelligence in water purification and wastewater treatment: Recent advancements. **Journal of Water Process Engineering**, Elsevier, v. 49, p. 102974, 2022. 34

SCHEFFER, Judi. Dealing with missing data. **Research Letters in the Information and Mathematical Sciences**, Massey University, n. 3, p. 153–160, 2002. 35

SEBASTIAN, Kristian; GAO, Hui; XING, Xudong. Utilizing an ensemble stl decomposition and gru model for base station traffic forecasting. *In: IEEE. 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. Chiang Mai, Thailand, 2020. p. 314–319. 36

SILVA, Wesley Pereira da; CARLOS, Thayrine Dias; CAVALLINI, Grasielle Soares; PEREIRA, Douglas Henrique. Peracetic acid: Structural elucidation for applications in wastewater treatment. **Water Research**, Elsevier, v. 168, p. 115143, 2020. 29

SILVERIO, Astrid Merino; CUNHA, Maristela Casé Costa; ANDRADE, Wba-neide Martins de; SIMPLÍCIO, Luane dos Santos. Relationships of perception of health risks by use of pesticides: a man-environment look. **Environmental Monitoring and Assessment**, Springer, v. 195, n. 6, p. 773, 2023. 21

SLAUGHTER, Robin J.; WATTS, Martin; VALE, J. Allister; GRIEVE, Jacob R.; SCHEP, Leo J. The clinical toxicology of sodium hypochlorite. **CLINICAL TOXICOLOGY**, v. 57, n. 5, p. 303–311, MAY 4 2019. ISSN 1556-3650. 20, 29

SOLAIMANY-AMINABAD, Mehri; MALEKI, Afshin; HADI, Mahdi. Application of artificial neural network (ann) for the prediction of water treatment plant influent characteristics. **Journal of advances in environmental health research**, Kurdistan University of Medical Sciences, v. 1, n. 2, p. 89–100, 2013. 21, 63

TAHERSHAMSI, A; MENHAJ, MB; AHMADIAN, R. Sediment loads prediction using multilayer feedforward neural networks. **Amirkabir**, v. 16, n. 63, p. 103–110, 2006. 34

TSUTIYA, Milton Tomoyuki. **Abastecimento de água**. 4th. ed. São Paulo, SP: Departamento de Engenharia Hidráulica e Sanitária da Escola Politécnica da Universidade de São Paulo, 2006. 26

UN-WATER. Summary progress update 2021: Sdg 6 – water and sanitation for all. **UN-Water**, Geneva, Switzerland, July 2021. 24

VEIGA, Fabiano da; KALBUSCH, Andreza; HENNING, Elisa. Drivers of urban water consumption in brazil: a countrywide, cross-sectional study. **Urban Water Journal**, Taylor & Francis, v. 20, n. 10, p. 1462–1470, 2023. 46

WANG, Dongsheng; WANG, Yan; ZHOU, Rui; CAO, Yong; JIANG, Fuchun; ZHANG, Xue; LI, Jinghua. Water plant optimization control system based on machine learning. **DESALINATION AND WATER TREATMENT**, v. 222, p. 168–181, MAY 2021. ISSN 1944-3994. 22, 23

WEGMAN, E. J.; WRIGHT, I. W. Splines in statistics. **Journal of the American Statistical Association**, v. 78, n. 382, p. 351, 1983. 40

WONGBURI, Praewa; PARK, Jae K. Big data analytics from a wastewater treatment plant. **Sustainability**, v. 13, n. 22, 2021. ISSN 2071-1050. 33

World Health Organization. **Guidelines for drinking-water quality**. Geneve, Switzerland: World Health Organization, 2002. 51

_____. _____. Geneve, Switzerland: World Health Organization, 2004. 21

_____. Developing drinking-water quality regulations and standards: general guidance with a special focus on countries with limited resources. World Health Organization, Geneve, Switzerland, 2018. 21

_____. **Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda**. Geneve, Switzerland: World Health Organization, 2022. 47

WU, Rui; HAMSHAW, Scott D.; YANG, Lei; KINCAID, Dustin W.; ETHERIDGE, Randall; GHASEMKHANI, Amir. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data. **IEEE Sensors Journal**, Institute of Electrical and Electronics Engineers (IEEE), v. 22, n. 11, p. 10671–10683, jun 1 2022. ISSN 1530-437X. 36

ZAVADOVA, Z; GRESLAND, L; ROSENBERGOVA, M. Inactivation of single- and double-stranded ribonucleic acid of encephalomyocarditis virus by ultraviolet

light. **Acta Virol. (Prague)**, **12**: 515-22(Nov. 1968)., 01 1968. Disponível em: <https://www.osti.gov/biblio/4779305>. 40

ZHANG, Yanyang; GAO, Xiang; SMITH, Kate; INIAL, Goulven; LIU, Shuming; CONIL, Lenny B; PAN, Bingcai. Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. **Water research**, Elsevier, v. 164, p. 114888, 2019. 21

ZHAO, Lufei; LUO, Tonglin; JIANG, Xuchu; ZHANG, Biao. Prediction of soil moisture using bigru-lstm model with stl decomposition in qinghai–tibet plateau. **PeerJ**, PeerJ Inc., v. 11, p. e15851, 2023. 36

ZOU, Huasheng; TANG, Haoyuan. Comparison of different bacteria inactivation by a novel continuous-flow ultrasound/chlorination water treatment system in a pilot scale. **Water**, MDPI, v. 11, n. 2, p. 258, 2019. 29