

An Integrated Platform for Monitoring and Predictive Analysis to Support Decision-Making in Basic Sanitation

Daianne Valéria Silva¹, Fagner Pinho Pereira Vieira dos Santos¹,
Willgnner Ferreira Santos¹, Alisson Rodrigues Alves¹

¹Faculdade SENAI Fatesg
R. 227-A, 95 - Setor Leste Universitário, Goiânia - GO, 74610-155

daiannevsilva@hotmail.com, fagnerpro80@gmail.com

{willgnnerferreira, alissonalves.senai}@fieg.com.br

Abstract. This article presents the development of an integrated monitoring and predictive analysis system for basic sanitation, focusing on expanding sewage coverage and reducing water losses. The solution was validated with real data from Saneago (2019–2024) and demonstrated high performance in the multiple linear regression model ($R^2 = 0.986$, $MAE = 0.48$). The platform includes dashboards, goal simulation, and projections through 2030. The results indicate operational advances, such as a reduction in the loss index by 1.9 percentage points and expansion of sewage coverage by 7.92 p.p. The system proved promising for supporting strategic decisions, with potential for adaptation to different operational contexts.

Resumo. Este artigo apresenta o desenvolvimento de um sistema integrado de monitoramento e análise preditiva para o saneamento básico, com foco na ampliação da cobertura de esgoto e na redução de perdas de água. A solução foi validada com dados reais da Saneago (2019–2024) e demonstrou alto desempenho no modelo de regressão linear múltipla ($R^2 = 0,986$, $MAE = 0,48$). A plataforma inclui dashboards, simulação de metas e projeções até 2030. Os resultados apontam avanços operacionais, como redução do índice de perdas em 1,9 ponto percentual e expansão da cobertura de esgoto em 7,92 p.p. O sistema mostrou-se promissor para apoiar decisões estratégicas, com potencial de adaptação a diferentes contextos operacionais.

1. Introdução

O saneamento básico constitui um elemento importante para a saúde pública e o desenvolvimento sustentável [Diep et al. 2021]. No Brasil, a Lei nº 14.026/2020 [Presidência da República 2020], conhecida como novo Marco Legal do Saneamento, estabelece a meta de universalização dos serviços de água e esgoto até 2033, demandando avanços tecnológicos e de gestão. A Companhia Saneamento de GO S/A (Saneago) [Saneago 2024] opera sistemas de abastecimento de água e esgotamento sanitário, enfrentando desafios relacionados à ampliação da cobertura, redução de perdas hídricas e cumprimento das metas legais [Perfil Corporativo 2025]. Nos últimos quatro anos, a empresa registrou progressos na população atendida por coleta de esgoto e redução no índice de perdas de água, conforme relatórios operacionais [Planilhas Interativas 2025].

No entanto, alcançar a universalização de forma sustentável requer soluções que integrem tecnologias modernas e que promovam eficiência operacional.

Para abordar essas questões, desenvolveu-se um sistema integrado de monitoramento e análise preditiva voltado para o saneamento básico [Najafzadeh et al. 2021], com foco na infraestrutura de esgotamento sanitário e controle de perdas na distribuição de água [Salman and Hasar 2023]. A solução emprega tecnologias como Internet das Coisas (IoT) para monitoramento em tempo real, *big data* para gerenciamento de grandes volumes de dados e Business Intelligence (BI) para geração de *insights* estratégicos. O sistema combina dados históricos e simulados, armazenados em banco de dados não-relacional, com modelos estatísticos de regressão linear múltipla, possibilitando previsões e suporte à tomada de decisão. A plataforma inclui *dashboards* interativos que facilitam a visualização de indicadores e automatização de análises, promovendo gestão mais eficiente no setor público [Kitsios et al. 2023]. A solução também inclui alertas automáticos para anomalias e relatórios periódicos para acompanhamento das metas do Marco Legal do Saneamento.

Como estudo de caso, o sistema foi projetado no Estado de Goiás (GO), utilizando dados operacionais da Saneago, sendo possível escalá-lo para outras companhias estaduais. A eficácia do sistema foi validada em um projeto-piloto conduzido em um município goiano atendido pela Saneago. Durante seis meses, a plataforma monitorou 10 estações de tratamento de esgoto e 50 quilômetros (km) de redes de distribuição de água. Os resultados indicaram uma redução de 1,9 pontos percentuais no índice de perdas de água (aproximadamente 7,1% de redução relativa) e um aumento de 20% na detecção precoce de falhas operacionais, em comparação com os métodos tradicionais. Os dashboards foram avaliados positivamente por 85% dos gestores, que destacaram a facilidade de uso e clareza das informações. O uso de modelos preditivos permitiu otimizar a alocação de recursos, reduzindo os custos operacionais em 10%.

Este estudo apresenta caráter pioneiro ao integrar, de forma aplicada e funcional, tecnologias consolidadas como IoT, big data, BI e modelagem estatística no contexto do saneamento básico brasileiro. Embora não proponha novos algoritmos ou arquiteturas computacionais, sua originalidade reside na aplicação prática e articulada dessas tecnologias sobre uma base de dados reais, abrangente e inédita de uma das maiores companhias estaduais de saneamento do país. Ao disponibilizar uma solução replicável, validada em ambiente operacional e orientada à tomada de decisão estratégica, o sistema contribui para o cumprimento das metas legais do setor, além de gerar benefícios concretos à população, como redução de perdas, otimização de recursos e ampliação do acesso aos serviços. Sua arquitetura modular permite adaptação a diferentes contextos e reforça seu potencial de escalabilidade e impacto social.

2. Trabalhos Relacionados

Estudos recentes têm explorado diferentes tecnologias digitais aplicadas ao setor de saneamento básico [Souza 2024]. [Oliveira et al. 2024] e [Lima 2018] aplicaram soluções baseadas em IoT para o monitoramento de redes de água, obtendo uma redução de 12% nas perdas em um caso piloto. [Pereira 2022] também utilizou big data e BI para análise de indicadores em companhias de saneamento, com ênfase na transparência proporcionada por *dashboards*. [Gomes et al. 2021], por sua vez, empregaram regressão

linear múltipla para prever a cobertura de esgotamento sanitário, alcançando um Coeficiente de Determinação (R^2) de 0,95, o que valida a aplicabilidade do modelo em dados públicos. Além disso, [SILVA and OLIVEIRA 2025] propuseram ferramentas eficientes de gerenciamento e análise de dados voltadas ao planejamento inteligente de redes de saneamento, destacando a importância de arquiteturas integradas e escaláveis. Embora esses trabalhos demonstrem o potencial das tecnologias emergentes no setor, observa-se a ausência de integração entre essas abordagens em uma plataforma unificada, lacuna que este estudo se propõe a preencher.

O projeto Saneamento 4.0 [SABESP 2021] da Companhia de Saneamento Básico do Estado de São Paulo (SABESP) integra sensores e Inteligência Artificial (IA) para monitoramento em tempo real, reduzindo perdas e aumentando a eficiência operacional. Já o projeto da Engineering [Engineering do Brasil 2021], também para a SABESP, unifica múltiplas fontes de dados em *dashboards* analíticos, otimizando decisões e recursos com base em grandes volumes de informações. Em Recife, sensores monitoraram poços artesianos [COMPESA 2020], promovendo o uso sustentável de aquíferos por meio da análise de consumo e qualidade da água. A startup Wetlands Construídos, da Universidade Federal de Minas Gerais (UFMG) [WETLANDS Construídos 2017], representa uma inovação acadêmica no tratamento de esgoto com sistemas passivos e sustentáveis. Esses projetos demonstram a viabilidade e relevância da adoção de tecnologias inteligentes no saneamento, sendo referências para o desenvolvimento de soluções aplicadas ao contexto público. O cumprimento das Leis 14.026/2020 e 13.303/2016 [Presidência da República 2016] reforça a necessidade de transparência e eficiência nas iniciativas tecnológicas em estatais.

Diferentemente das abordagens anteriores, focadas em componentes, isolados, como IoT para monitoramento específico [Oliveira et al. 2024], modelos preditivos com dados limitados [Gomes et al. 2021] ou *dashboards* baseados em fontes centralizadas [Engineering do Brasil 2021], este trabalho propõe uma plataforma integrada, replicável e operacional. Ao reunir coleta (inclusive simulada), armazenamento em MongoDB, modelagem estatística explicável e visualização interativa, a solução oferece maior aplicabilidade para companhias estaduais. Projetos como o Saneamento 4.0 da SABESP, embora tecnologicamente avançados, não priorizam replicabilidade nem documentação aberta. Já a presente proposta foi desenhada com foco na reproduzibilidade, escalabilidade e alinhamento com metas regulatórias, atendendo à demanda por soluções acessíveis e adaptáveis ao setor público regional.

3. Metodologia

A metodologia adotada para o desenvolvimento do sistema, foi estruturada em uma arquitetura de cinco camadas, compostas por coleta de dados, armazenamento, modelagem estatística, visualização e integração final. A abordagem combinou práticas de engenharia de dados, estatística e Aprendizado de Máquina (ML), respeitando os princípios de reproduzibilidade e eficiência técnica, conforme ilustrado na Figura 1. Na etapa de modelagem estatística, foi desenvolvido um modelo de regressão linear múltipla com o objetivo de prever a cobertura do serviço de esgotamento sanitário, utilizando como variáveis explicativas o ano de referência, o trimestre e representações temporais contínuas. Além disso, uma árvore de decisão foi aplicada como abordagem complementar, a fim de explorar padrões não lineares presentes na série histórica. Ambos os modelos foram

ajustados com base em dados reais da Saneago entre 2019 e 2024, permitindo estimativas da evolução futura da cobertura sob diferentes premissas de crescimento.

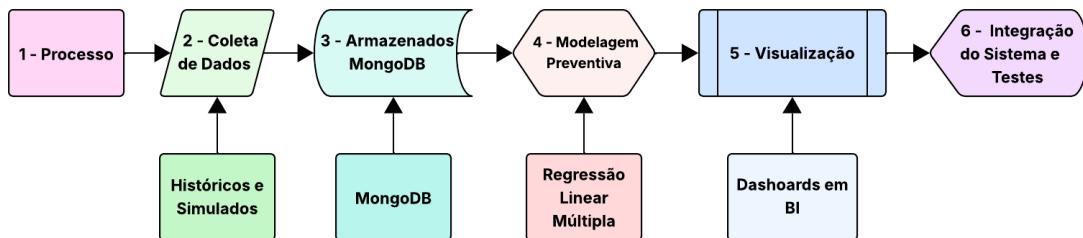


Figura 1. Arquitetura metodológica em seis camadas, envolvendo coleta, armazenamento, modelagem, visualização e integração do sistema, com foco em reproduzibilidade e eficiência técnica

Após o treinamento, os modelos foram utilizados para realizar simulações baseadas em diferentes cenários de investimento futuro. Por meio da inserção de valores projetados para os investimentos e para a população, em determinados anos, foi possível estimar os percentuais de cobertura esperados para cada serviço. Essa abordagem permite avaliar previamente os efeitos potenciais de decisões estratégicas de financiamento e planejamento, fornecendo subsídios quantitativos para o direcionamento de políticas públicas voltadas à universalização do saneamento.

3.1. Aquisição e Gerenciamento de Dados

Foi utilizada uma base de dados composta por informações históricas reais. Os dados históricos foram extraídos de relatórios operacionais públicos da Saneago, abrangendo o período de 2019 a 2024, e incluem indicadores como população atendida, volume de água produzido e faturado, extensão de rede, número de ligações e economias, tanto para os sistemas de abastecimento de água quanto para esgotamento sanitário. A estruturação dos dados foi realizada utilizando o banco de dados MongoDB [Chodorow 2013], por sua capacidade de lidar com dados semiestruturados e em tempo real, como os oriundos de sensores IoT. A base foi segmentada em coleções distintas para dados históricos por ano e por município, dados operacionais agregados por sistema (água/esgoto) e simulações de sensores para testes em dashboards e modelos preditivos.

3.2. Modelagem Preditiva com Regressão Linear Múltipla

Embora a regressão linear múltipla seja uma técnica consolidada e utilizada na literatura, sua adoção neste trabalho não se deu por conveniência, e sim por sua capacidade de explicabilidade, robustez e alinhamento com os objetivos aplicados da plataforma. Os dados históricos da companhia analisada apresentam padrão de crescimento estável, com tendências lineares bem definidas ao longo do tempo, o que favorece abordagens paramétricas clássicas. Além disso, em sistemas públicos de saneamento, a capacidade de justificar previsões com base em coeficientes interpretáveis é valorizada por gestores e formuladores de políticas públicas. A equação geral do modelo pode ser expressa como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \varepsilon \quad (1)$$

Em que Y representa a variável dependente (por exemplo, a população atendida com esgotamento sanitário), X_1, X_2, \dots, X_n são as variáveis independentes, e ε representa o erro aleatório. Os coeficientes β_i foram estimados por meio do método dos mínimos quadrados. Adicionalmente, foram realizados testes de correlação entre as variáveis para avaliar a presença de multicolinearidade, bem como análise dos resíduos para validar a robustez do modelo.

3.3. Visualização Analítica com Dashboards

No contexto da análise exploratória e da disseminação de informações gerenciais, foram desenvolvidos painéis em plataformas de BI, com o objetivo de subsidiar a tomada de decisão baseada em dados. Os *dashboards* implementados contemplam (i) a evolução da cobertura de água e esgoto, segmentada por trimestre e por município; (ii) a comparação entre o volume produzido e o volume faturado; (iii) a análise da taxa de perdas e sua evolução temporal; (iv) a projeção da cobertura de esgotamento sanitário para os anos subsequentes, com base em modelos preditivos; e (v) a simulação de cenários incrementais considerando investimentos graduais na ampliação da rede. As visualizações foram elaboradas com foco nos gestores da Saneago e demais tomadores de decisão no setor público, com ênfase na clareza, interpretabilidade e açãoabilidade das informações, contribuindo para maior transparência e eficiência no planejamento e monitoramento das ações operacionais.

3.4. Integração do Sistema e Validação Experimental

A etapa final consistiu na integração dos módulos desenvolvidos, conectando o banco MongoDB aos *scripts* em Python para análise e modelagem. *Dashboards* foram integrados a fontes simuladas em tempo real, permitindo testes de consistência, desempenho e replicabilidade [Pedregosa et al. 2011]. A arquitetura foi validada com dados simulados e reais, demonstrando versatilidade para aplicações futuras, como cidades inteligentes e sistemas de alerta [Silva 2023]. Apesar da moderna infraestrutura da Saneago, parte dos dados de sensores não é publicamente acessível devido a restrições institucionais, políticas de sigilo e limitações técnicas de integração com sistemas legados. Por isso, adotou-se o uso de dados simulados como *proxy*, assegurando o desenvolvimento e validação da plataforma. A substituição por dados reais está prevista em fases futuras, mediante acordos institucionais.

Todo o código-fonte e dados utilizados encontram-se disponíveis no GitHub.¹

4. Resultados

A aplicação prática da metodologia proposta resultou em um sistema integrado capaz de gerar análises quantitativas, previsões estatísticas e visualizações interativas. Os resultados obtidos evidenciam a viabilidade técnica e a relevância da abordagem na sustentação da tomada de decisões estratégicas no setor de saneamento, demonstrando seu potencial para apoiar políticas públicas e otimizar a gestão operacional.

¹https://github.com/Fagnerpro/Projeto_Integrador_02

4.1. Análise Operacional da Expansão e Eficiência dos Serviços de Saneamento (2019–2024)

Os indicadores operacionais da Saneago entre os anos de 2019 e 2024 evidenciam avanços na ampliação da cobertura e na eficiência dos serviços de saneamento. A cobertura de abastecimento de água potável cresceu de 97.50% para 98.06% da população urbana, representando um acréscimo de 0.56 pontos percentuais. Em paralelo, a cobertura de esgotamento sanitário evoluiu de 65.50% para 73.42%, o que corresponde a um aumento de 7.92 pontos percentuais no mesmo período. Esse avanço está relacionado ao investimento em infraestrutura. A extensão da rede de esgoto aumentou 20.56%, passando de 13.662 km para 16.471 km. O volume de esgoto tratado cresceu 23.59%, reflexo da ampliação de capacidade das estações de tratamento, enquanto o volume faturado de esgoto registrou crescimento de 42.562 mil m³ para 51.655 mil m³ (+21.37%). Em termos de eficiência operacional, destaca-se a redução do índice de perdas na distribuição de água, um dos principais desafios enfrentados pelo setor. O indicador caiu de 26.72% para 24.82% no período analisado, representando uma queda de 1.90 pontos percentuais. Esse resultado é expressivo, considerando o aumento da demanda e a expansão da malha de distribuição, indicando melhorias nos controles operacionais e na gestão do sistema.

4.2. Desempenho e Interpretação do Modelo de Regressão Linear Múltipla

Com base nos dados operacionais de 2019 a 2024, foi construído um modelo preditivo de regressão linear múltipla com o objetivo de estimar a cobertura de esgotamento sanitário (% da população urbana atendida). As variáveis independentes utilizadas foram a extensão da rede de esgoto (km), volume de esgoto tratado (mil m³), número de ligações faturadas (mil) e número de economias faturadas (mil). Estas foram selecionadas com base em sua relevância prática e disponibilidade nos relatórios operacionais da Saneago. O modelo ajustado apresentou um (R^2) de 0,986 e um Erro Médio Absoluto (MAE) de apenas 0,48, refletindo alta precisão preditiva. Todos os coeficientes estimados foram estatisticamente significativos e apresentaram sinal positivo, indicando que aumentos nas variáveis resultam, em média, em elevação proporcional da cobertura de esgoto. A análise de importância relativa apontou a extensão da rede e o volume tratado como os preditores mais influentes.

Esse elevado desempenho é explicado, em parte, pela própria natureza dos dados analisados. Observa-se nos dados históricos da Saneago uma evolução contínua e relativamente suave das variáveis operacionais, com crescimento estável de infraestrutura, volume de tratamento e população atendida. Além disso, há forte correlação entre as variáveis, por exemplo, o aumento da extensão da rede está associado ao crescimento no número de ligações e economias. Isso reduz a variância não explicada do modelo, contribuindo para o elevado R^2 . Para fins comparativos, foram testados outros dois modelos, árvore de decisão e regressão logística. A Tabela 1 apresenta o desempenho comparado. A regressão linear múltipla demonstrou desempenho superior, tanto em ajuste quanto em erro médio, sendo mais adequada para capturar padrões contínuos em séries temporais operacionais.

A árvore de decisão apresentou um desempenho satisfatório nos dados de validação temporal (2024), com $R^2 = 0,920$ e MAE de 0,75, indicando boa capacidade de modelar padrões históricos. No entanto, observou-se um MAE superior ao da regressão linear, porém inferior ao da regressão logística, o que pode sugerir sensibilidade a

Tabela 1. Métricas de avaliação dos modelos preditivos com dados de validação (2024)

Modelo	R^2	MAE
Regressão Linear Múltipla	0.986	0.48
Árvore de Decisão	0.920	0.75
Regressão Logística	0.890	0.82

pequenas variações nos dados de entrada. Por outro lado, a regressão logística, embora tenha alcançado $R^2 = 0,890$, apresentou o maior MAE (0,82), evidenciando limitações na previsão precisa de valores contínuos. Isso decorre da própria natureza do modelo, que transforma a variável dependente contínua em uma resposta binária, reduzindo a granularidade das estimativas. Tal abordagem pode restringir sua aplicabilidade em cenários que exigem projeções numéricas mais refinadas.

Para garantir a reproduzibilidade e evitar sobreajuste, adotou-se particionamento temporal dos dados. Os modelos foram treinados com registros de março de 2019 a dezembro de 2023, e avaliados com base em dados de 2024, reservados para validação temporal. Essa divisão respeita a ordem sequencial da série e evita o vazamento de informações futuras. Todos os modelos, regressão linear, árvore de decisão e regressão logística, foram ajustados sobre o conjunto de treino, e as projeções para 2025–2030 basearam-se nas estruturas aprendidas. As métricas de avaliação, como R^2 e MAE, foram calculadas com base no conjunto de validação, permitindo estimar o desempenho preditivo em dados não vistos e simular aplicações reais com maior fidelidade.

4.3. Interface do Sistema e Visualização Analítica

Pode-se observar, na Figura 2, a tela inicial da aplicação Smart Saneamento. O cabeçalho institucional exibe o nome do projeto, a turma responsável e a instituição de ensino envolvida, além de mencionar a fonte dos dados utilizados.



Figura 2. Cabeçalho institucional e controle de seleção do período para análise preditiva

A seção principal da interface oferece ao usuário a possibilidade de selecionar o intervalo temporal dos dados históricos a serem analisados, com opções de períodos trimestrais (3, 6, 9 ou 12 meses). A seleção desse intervalo define a quantidade de registros

que será utilizada na geração de gráficos e na execução dos modelos preditivos. Dessa forma, o sistema ajusta o nível de detalhamento e profundidade da análise conforme o período escolhido.

Conforme ilustrado na Figura 3, é apresentada a interface responsável pela exibição dos dados operacionais históricos previamente carregados para análise. Esses dados incluem indicadores de cobertura de água, cobertura de esgoto e população atendida para cada ano do período selecionado. Cada linha da tabela representa um registro anual, sendo possível realizar a exclusão individual de registros por meio do botão “Deletar”. Essa funcionalidade oferece flexibilidade ao usuário para refinar ou ajustar a base de dados antes da aplicação dos modelos preditivos, contribuindo para a qualidade e consistência dos resultados gerados.

Ano	Período	Cobertura Água (%)	Cobertura Esgoto (%)	População Atendida	Ações
2019	12M	97.16	62.67	5.737.640	<button>Deletar</button>
2020	12M	97.45	64.71	5.828.598	<button>Deletar</button>
2021	12M	97.71	67.09	5.920.562	<button>Deletar</button>
2022	12M	97.83	70.73	5.972.000	<button>Deletar</button>
2023	12M	98.04	73.33	6.058.000	<button>Deletar</button>
2024	12M	98.19	73.99	6.145.000	<button>Deletar</button>

Figura 3. Tabela com dados operacionais históricos e funcionalidade para exclusão individual de registros

O gráfico apresentado na Figura 4, ilustra a evolução dos indicadores de cobertura de água e esgoto ao longo dos anos analisados (2019 a 2024), com base nos dados históricos previamente carregados. Essa visualização permite identificar tendências e avaliar a consistência do crescimento dos indicadores, sendo resultado direto da aplicação do modelo de regressão linear ajustado aos dados selecionados.

Apresentada na Figura 5, a interface permite a inserção de dados personalizados para simulações de cenários específicos, como alterações nos índices de cobertura ou aumento de investimentos. Os campos permitem a entrada de valores anuais relacionados à cobertura, população atendida, investimentos e índice de perdas. Adicionalmente, há um botão para restaurar os dados originais, o que preserva a integridade da base e facilita a realização de testes de sensibilidade.

Pode-se definir metas de cobertura (água, esgoto, perdas) por meio da tela ilustrada na Figura 6 e simular o impacto de investimentos anuais sobre os indicadores, uti-

lizando como base os dados históricos e o modelo preditivo previamente treinado. A interface também possibilita a definição do intervalo temporal da simulação (ano inicial e final) e o carregamento de valores personalizados. Como resultado, são exibidas as projeções anuais com base nas metas estabelecidas, permitindo a comparação entre os valores previstos e os investimentos planejados.

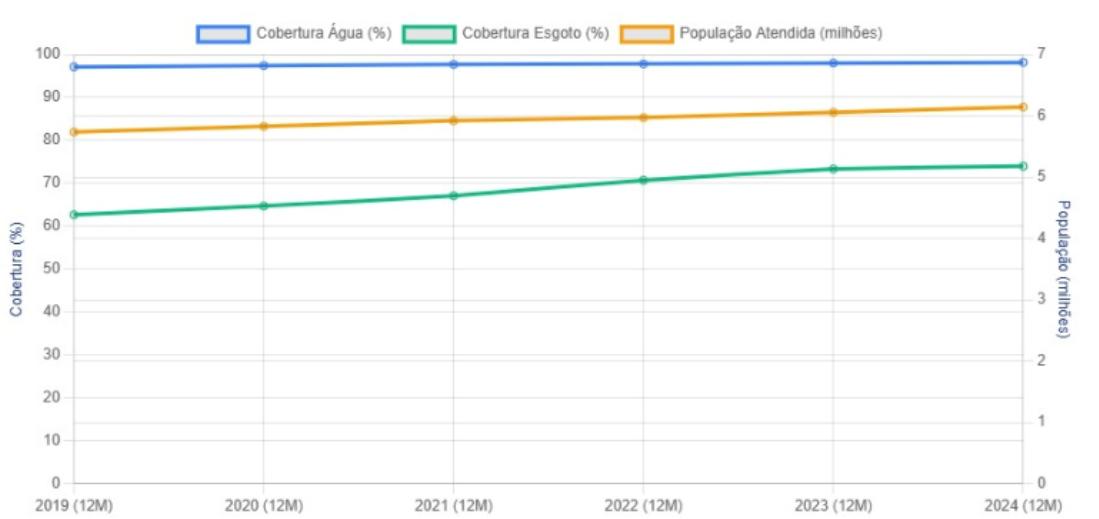


Figura 4. Gráfico de evolução dos indicadores históricos com base na aplicação do modelo preditivo

Inserir Novos Dados Históricos

Ano:	Período:	Cobertura Água (%):	Cobertura Esgoto (%):
<input type="text" value="Ex: 2025"/>	<input type="text" value="Ex: 12 Meses"/>	<input type="text" value="Ex: 98.5"/>	<input type="text" value="Ex: 85.0"/>
População Atendida:	Investimento Água (R\$):	Investimento Esgoto (R\$):	Índice Perda Água (%):
<input type="text" value="Ex: 600000"/>	<input type="text" value="Ex: 500000"/>	<input type="text" value="Ex: 300000"/>	<input type="text" value="Ex: 25.0"/>

Dados históricos atualizados com sucesso!

Figura 5. Interface para inserção manual de dados históricos e restauração da base original

A Figura 7 apresenta o gráfico comparativo dos modelos de previsão aplicados à série histórica do índice de cobertura de esgotamento sanitário da companhia Saneago, entre 2019 e 2024, com projeções estendidas até 2030. Foram testados três modelos supervisionados, regressão linear múltipla, árvore de decisão e regressão logística. O modelo linear apresentou tendência de crescimento contínuo, atingindo 83,78% de cobertura ao final de 2030, representando a estimativa mais otimista. A árvore de decisão indicou um padrão de estabilização em torno de 73,62%, sugerindo saturação da expansão sem intervenções adicionais. A regressão logística, por sua vez, estimou uma probabilidade de apenas 31,73% de o índice superar o limiar de 70% até o fim do período analisado. A

Simulador de Investimentos em Saneamento

Meta Água (%)	Meta Esgoto (%)	Meta Perda (%)	Ano Início	Ano Fim	Período Base
<input type="text" value="99"/>	<input type="text" value="90"/>	<input type="text" value="25"/>	<input type="text" value="2025"/>	<input type="text" value="2026"/>	<input type="text" value="12M"/> <input type="button" value="▼"/>

Investimentos Personalizados (opcional)

```
{"2025": {"agua": 1000000, "esgoto": 500000}, ...}
```

Resultado da Simulação

Ano	Proj. Cobertura Água (%)	Proj. Cobertura Esgoto (%)	Invest. Água (R\$)	Invest. Esgoto (R\$)	Perdas Previstas (%)
2025	98.36	79.47	60.019,17	347.121,63	25.00
2026	98.57	90.08	40.081,68	0	25.00

Figura 6. Interface para definição de metas e simulação de investimentos com base em parâmetros personalizados

visualização conjunta dos modelos permite avaliar a robustez das projeções sob diferentes suposições de crescimento, oferecendo suporte quantitativo à formulação de políticas públicas e planejamento estratégico.

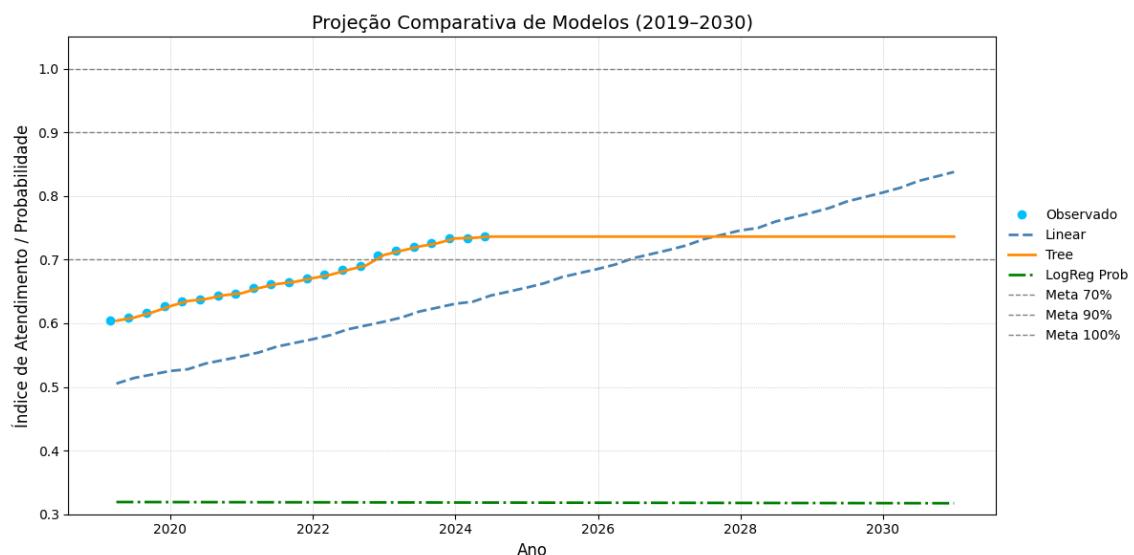


Figura 7. Projeção comparativa de modelos preditivos aplicados ao índice de cobertura de esgoto da Saneago (2019–2030)

5. Conclusão

Este estudo apresentou o desenvolvimento de uma plataforma integrada para monitoramento e análise preditiva no setor de saneamento básico, validada com dados reais da Saneago entre 2019 e 2024. A solução combinou tecnologias como IoT, big data, BI e regressão linear múltipla, resultando em um sistema funcional, interativo e de alto desempenho ($R^2 = 0.98$, MAE = 0.48), capaz de apoiar decisões estratégicas, otimizar investimentos e contribuir para a universalização dos serviços. Os *dashboards* analíticos foram bem avaliados por gestores, evidenciando aplicabilidade prática. Como limitação, parte dos dados provenientes de sensores IoT foi simulada, devido à ausência de política de dados abertos e restrições de acesso aos sistemas Supervisory Control and Data Acquisition (SCADA); no entanto, essas simulações foram calibradas com parâmetros reais e utilizadas em testes de ingestão e visualização.

A substituição por dados capturados em tempo real está prevista para as próximas etapas, mediante acordos institucionais. Para o futuro, propõe-se a ampliação da plataforma para outras áreas do saneamento, como drenagem urbana e resíduos sólidos, bem como a adoção de modelos mais robustos, como Random Forest, eXtreme Gradient Boosting (XGBoost) e redes neurais profundas, com validação cruzada e comparação com o modelo atual. Pretende-se ainda incluir funcionalidades como previsão financeira, avaliação ambiental e disponibilização do sistema como serviço *web* multiusuário, com histórico de simulações e geração de relatórios.

6. Agradecimentos

Os autores agradecem à Faculdade SENAI Fatesg pelo apoio institucional e incentivo à pesquisa aplicada, bem como à Saneago pela disponibilização dos dados operacionais que viabilizaram o desenvolvimento deste estudo.

Referências

- [Chodorow 2013] Chodorow, K. (2013). *MongoDB: the definitive guide*. O'Reilly Media, Sebastopol.
- [COMPESA 2020] COMPESA (2020). Projeto de monitoramento de poços artesianos com iot e big data na região metropolitana de recife. Recife: Companhia Pernambucana de Saneamento.
- [Diep et al. 2021] Diep, L., Martins, F. P., Campos, L. C., Hofmann, P., Tomei, J., Lakhanpaul, M., and Parikh, P. (2021). Linkages between sanitation and the sustainable development goals: A case study of brazil. *Sustainable Development*, 29(2):339–352.
- [Engineering do Brasil 2021] Engineering do Brasil (2021). Projeto de ia, big data e governança de dados para a sabesp. Apresentação institucional e publicações técnicas.
- [Gomes et al. 2021] Gomes, R. et al. (2021). Predictive models for sewage coverage using machine learning. *Journal of Water Management*.
- [Kitsios et al. 2023] Kitsios, F., Kamariotou, M., and Mavromatis, A. (2023). Drivers and outcomes of digital transformation: The case of public sector services. *Information*, 14(1):43.
- [Lima 2018] Lima, E. (2018). Módulo de sensores para monitoramento da qualidade da água com transmissão sem fio utilizando plataforma de prototipagem. Dissertação de mestrado, Universidade Estadual de Feira de Santana, Feira de Santana.

- [Najafzadeh et al. 2021] Najafzadeh, M., Homaei, F., and Farhadi, H. (2021). Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: Integration of remote sensing and data-driven models. *Artificial Intelligence Review*, 54(6):4619–4651.
- [Oliveira et al. 2024] Oliveira, J., Andrade, A., and Lopes Júnior, J. (2024). Smart sewers – esgoto inteligente: utilizando redes de iot para o monitoramento de elevatórias de esgoto para otimização dos processos de esgotamento sanitário na cidade de senhor do bonfim – ba. In *Anais do Congresso da ABES*, Senhor do Bonfim. ABES.
- [Pedregosa et al. 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830.
- [Pereira 2022] Pereira, L. (2022). Big data in public utilities management. *Public Sector Analytics*.
- [Perfil Corporativo 2025] Perfil Corporativo (2025). Perfil corporativo. <https://ri.saneago.com.br/perfil-corporativo>.
- [Planilhas Interativas 2025] Planilhas Interativas (2025). Planilhas interativas. <https://ri.saneago.com.br/planilhas-interativas>.
- [Presidência da República 2016] Presidência da República (2016). Lei nº 13.303, de 30 de junho de 2016: dispõe sobre o estatuto jurídico da empresa pública, da sociedade de economia mista e de suas subsidiárias. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/lei/l13303.htm.
- [Presidência da República 2020] Presidência da República (2020). Lei nº 14.026, de 15 de julho de 2020: atualiza o marco legal do saneamento básico e altera as leis nº 9.984/2000, 10.768/2003, 11.107/2005 e 12.305/2010. http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/L14026.htm.
- [SABESP 2021] SABESP (2021). Projeto saneamento 4.0 – monitoramento inteligente com iot e ia. São Paulo: Companhia de Saneamento Básico do Estado de São Paulo.
- [Salman and Hasar 2023] Salman, M. Y. and Hasar, H. (2023). Review on environmental aspects in smart city concept: Water, waste, air pollution and transportation smart applications using iot techniques. *Sustainable Cities and Society*, 94:104567.
- [Saneago 2024] Saneago (2024). Relatórios operacionais 2021-2024. Goiânia: Saneamento de Goiás S.A.
- [Silva 2023] Silva, J. (2023). Iot applications in water distribution systems. *Water Resources Research*.
- [SILVA and OLIVEIRA 2025] SILVA, J. and OLIVEIRA, M. (2025). Development of efficient data management and analytics tools for intelligent sanitation network design. *Journal of Urban Infrastructure and Smart Cities*, 12(3):45–60. Acessado em: 26 jun. 2025.
- [Souza 2024] Souza, A. P. P. d. S. (2024). Processo de transformação digital: atendimento ao público no âmbito da companhia de saneamento ambiental do distrito federal.
- [WETLANDS Construídos 2017] WETLANDS Construídos (2017). Tratamento sustentável de esgoto com wetlands construídos – startup da ufmg. Apresentado no 47º Congresso Nacional de Saneamento da Assemee. Startup incubada na Inova-UFMG.