



091 - PREVISÃO DO CONSUMO DE ÁGUA NA REDE DE ABASTECIMENTO DE CUIABÁ-MT: UMA ABORDAGEM INTEGRADA UTILIZANDO SMART METERS E APRENDIZADO DE MÁQUINA

Luiz Ricardo Adriano Ferreira ⁽¹⁾

Engenheiro ambiental e sanitaria (UFPR), Esp. Em ciências de dados e big data (UFPR).

Aline Barbosa de Carvalho ⁽²⁾

Engenheira Sanitarista e Ambiental (UFMT), MBA em perícia, auditoria e gestão ambiental (IPOG)

Ricardo Schleder Tozetto ⁽³⁾

Engenheiro eletricista, MSc em engenharia (UPF).

Murillo Oliveira Borges ⁽⁴⁾

Engenheiro eletricista (UB).

Wesley Henrique Alves Barbosa ⁽⁵⁾

Engenheiro civil (UFMS).

Endereço⁽¹⁾: Rua. Rocha Pombo, 690 – Vila Cruzeiro – Paranaguá – Paraná - CEP: 83221-330 – Brasil - Tel: +55 (41) 9 9573-1953 - e-mail: luiz.ricardo@igua.com.br.

RESUMO

A crescente demanda por água, impulsionada pelo crescimento populacional, tem tornado a gestão eficiente dos recursos hídricos uma preocupação constante. Este trabalho objetiva prever o consumo diário de água de um conjunto de clientes da rede de abastecimento de Cuiabá-MT, utilizando dados de micromedição e variáveis climáticas. A metodologia incluiu a coleta de registros diários de consumo, filtragem dos dados, seleção de matrículas mais frequentes, tratamento de consumos acumulados e a criação de variáveis relacionadas ao consumo das ligações mais próximas e condições climáticas. Foi utilizado o modelo *Light Gradient Boosting Machine (LightGBM)*, um algoritmo de aprendizado de máquina baseado no método de *Gradient Boosting*. O modelo foi treinado com dados históricos de consumo e variáveis climáticas, e avaliado por meio de métricas como MAE, RMSE e R². Os resultados mostraram que o *LightGBM* é capaz de prever o consumo diário de água com um MAE de 0.4266 m³, RMSE de 0.7114 m³ e R² de 0.6089. A análise de importância das variáveis indicou que o histórico de consumo médio e as condições climáticas são fatores cruciais na previsão. Os resultados indicam que, embora o modelo tenha capturado a maior parte da variabilidade dos dados, há espaço para melhorias, especialmente em relação aos grandes desvios observados nos resíduos. Recomenda-se a exploração de técnicas avançadas de aprendizado de máquina e a inclusão de mais variáveis para aprimorar a precisão do modelo. Em conclusão, a aplicação de modelos de aprendizado de máquina para a previsão de consumo de água mostrou-se promissora, mas diversas oportunidades de melhoria permanecem.

PALAVRAS-CHAVE: previsão de consumo de água, gestão hídrica, aprendizado de máquina, recursos hídricos, tecnologias de monitoramento.

INTRODUÇÃO

A gestão eficiente dos recursos hídricos tornou-se crucial diante do aumento da demanda por água, impulsionada pelo crescimento populacional. Nesse contexto, a previsão de consumo de água em redes de abastecimento emerge como uma ferramenta vital para otimizar a distribuição e utilização desse recurso valioso (STANCZYK et al., 2022).

O crescimento populacional e as mudanças nos padrões de consumo têm exercido uma pressão sobre os sistemas de abastecimento de água em todo o mundo. A gestão eficaz desses sistemas precisa equilibrar o fornecimento contínuo de água potável com a preservação dos recursos hídricos. Historicamente, o planejamento e a gestão dos recursos hídricos baseavam-se em dados históricos e projeções simples, muitas vezes incapazes de acompanhar a dinâmica da crescente demanda (BORETTI; ROSA, 2019; SALEHI, 2022).

Estudos como os de Boretti e Rosa (2019) e He et al. (2021) destacam a importância do uso sustentável da água e a necessidade de soluções inovadoras para a escassez hídrica urbana. Neste contexto, a precisão na previsão de consumo



de água pode ser uma ferramenta vital para uma gestão eficiente dos recursos hídricos. Gil-Gamboa et al. (2024) demonstram o uso de modelos de aprendizado profundo para prever o consumo a médio prazo, enquanto García-Soto et al. (2024) focam na previsão de séries temporais do consumo de água em centros urbanos. Arsene et al. (2022) discutem estratégias avançadas para monitorar o consumo de água em residências, integrando IoT e aprendizado de máquina. Pesantez, Berglund e Kaza (2020) desenvolvem um modelo de previsão de demanda de água utilizando dados de medidores inteligentes, enfatizando a importância da micromedição e da coleta de dados em tempo real.

A micromedição e a telemetria, com coleta de dados em tempo real, permitem uma compreensão precisa dos padrões de consumo e a identificação de anomalias (PESANTEZ; BERGLUND; KAZA, 2020). A aplicação de modelos de aprendizado de máquina aprimora as previsões, incorporando variáveis dinâmicas para fornecer previsões mais acuradas (HARTANTO; DARMAWAN, 2019).

A Iguá Saneamento, operadora do sistema de abastecimento de água de Cuiabá-MT, tem investido em hidrômetros *smart meter*, atingindo em dezembro de 2022 a marca de 10 mil hidrômetros com leitura remota conectados (IGUA SANEAMENTO, 2022). Atualmente, mais de 40% de todo o volume da companhia está sendo monitorado diariamente por hidrômetros inteligentes. Esta tecnologia baseada em *IoT* permite monitorar o consumo de água com agilidade e precisão, identificando vazamentos e fraudes, resultando em economia de custos e redução de perdas. Atualmente, mais de 6000 clientes em Cuiabá são monitorados diariamente por essa tecnologia, onde a companhia tem investido na previsão de consumo, foco deste trabalho.

OBJETIVO

Este trabalho tem como objetivo principal realizar a previsão do consumo diário de água na rede de abastecimento de Cuiabá-MT, integrando dados provenientes de micromedição e variáveis climáticas.

METODOLOGIA UTILIZADA

A metodologia foi estruturada em várias etapas, desde a coleta e pré-processamento dos dados até o treinamento e avaliação do modelo de regressão. A seguir, serão detalhadas as etapas de metodologias utilizadas neste trabalho:

Coleta e Pré-processamento dos Dados

1. **Coleta de Dados:** Foram coletados registros diários de consumo de hidrômetros da rede de abastecimento de Cuiabá-MT cedidos pela empresa Iguá Saneamento S.A.
2. **Filtragem temporal e de consumo:** Os dados foram filtrados para um período específico (01/10/2023 a 22/02/2024) e para valores de consumo diário por cliente entre 0.1 m^3 e 10 m^3 .
3. **Seleção de matrículas:** Foram selecionadas as 1000 matrículas com maior frequência de leitura e com coordenadas de latitude e longitude válidas.
4. **Tratamento de consumos acumulados:** Foram excluídos possíveis consumos acumulados, filtrando registros sem consumo no dia anterior.
5. **Criação de variáveis:** Foram adicionadas variáveis relacionadas ao consumo das 10 ligações mais próximas, sendo a mediana e a média de consumo dessas ligações. Foram utilizadas técnicas de geocodificação para calcular a proximidade entre diferentes hidrômetros.
6. **Dados climáticos:** Foram adicionados dados de temperatura média e temperatura média móvel de 7 dias ao conjunto de dados de consumo, extraídos da plataforma Climatempo.



Conjunto de dados

O conjunto de dados inclui cinco variáveis principais, todas com 63.675 entradas não-nulas e do tipo numérico. Estas variáveis são: Consumo médio mensal dos últimos três meses (m^3), Consumo médio do dia das 10 ligações mais próximas (m^3), Consumo mediano do dia das 10 ligações mais próximas (m^3), Temperatura média e Temperatura média móvel de sete dias ($^{\circ}C$).

Treinamento do modelo de aprendizagem de máquina

Métodos de aprendizado de máquina podem ser adotados para estimativa/resposta de um conjunto de entradas. Esses requerem uma fase de treinamento, chamada de treinamento supervisionado, que considera um conjunto de dados extraído de variáveis selecionadas no domínio do problema.

Neste estudo, os algoritmos foram programados em Linguagem de programação *Python* 3.8 usando bibliotecas de computação científica *NumPy*. O pacote *Pandas* foi usado para processamento de dados e análise. Os algoritmos de regressão e validação cruzada foram implementados usando a biblioteca de aprendizado de máquina *Scikit-learn* (PEDREGOSA, 2011) através do pacote *Pycaret* (SAGAR, 2021).

Para realizar a previsão de consumo foi treinado o modelo *Light Gradient Boosting Machine (LightGBM)*, considerando um conjunto de 1000 matrículas selecionadas, onde 70% foram separadas para treinamento do modelo e 30% para teste.

O *Light Gradient Boosting Machine (LightGBM)* é um algoritmo de aprendizado de máquina baseado no método de *Gradient Boosting*, que se destaca por sua eficiência e rapidez na construção de modelos preditivos. Utilizando uma abordagem de aprendizado por árvores, o *LightGBM* combina os resultados de várias árvores de decisão para melhorar a precisão das previsões. Entre suas otimizações estão o *Gradient-Based One-Side Sampling (GOSS)*, que seleciona amostras de dados mais relevantes para treinamento, e a técnica de *binning*, que otimiza o processo de divisão nos nós das árvores. No contexto deste estudo, o *LightGBM* foi escolhido devido à sua capacidade de lidar com a variabilidade dos dados de consumo de água e de incorporar variáveis dinâmicas, como condições climáticas e eventos sazonais, proporcionando previsões mais precisas e adaptáveis às flutuações na demanda.

Métricas de Avaliação de Modelos de Regressão

Múltiplos critérios de avaliação foram usados para comparar o desempenho do modelo treinado. Dado um conjunto de dados composto por N observações, o desempenho foi medido pelo Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE), Erro Percentual Médio Absoluto (MAPE) e Coeficiente de Determinação (R^2).

1. **MAE (Mean Absolute Error):** O Erro Médio Absoluto (MAE) é a média das diferenças absolutas entre os valores previstos e os valores reais.
2. **RMSE (Root Mean Squared Error):** O Erro Quadrático Médio da Raiz (RMSE) é a raiz quadrada da média das diferenças quadradas entre os valores previstos e os valores reais.
3. **R^2 (Coeficiente de Determinação):** O Coeficiente de Determinação (R^2) mede a proporção da variância na variável dependente que é previsível a partir das variáveis independentes.

RESULTADOS OBTIDOS

Na Figura 1 está o gráfico de distribuição do consumo diário de água em m^3 . Observa-se que a maioria dos registros de consumo diário está concentrada em níveis baixos, com um pico acentuado próximo a $0 m^3$. À medida que o consumo diário aumenta, a frequência diminui rapidamente, indicando que altos consumos são menos comuns.

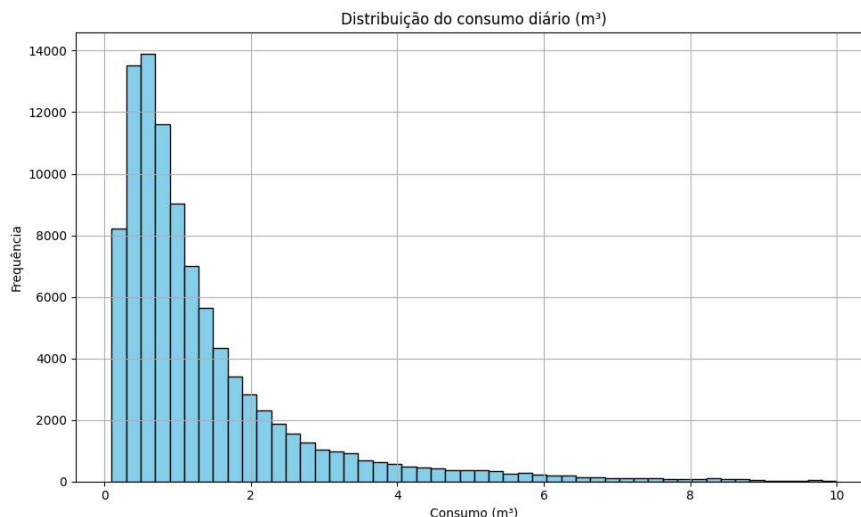


Figura 1 – Distribuição de consumo diário (m³)

No gráfico da Figura 2 está a soma de consumo diário em m³ ao longo do tempo, com uma variação relativamente estável ao longo dos meses analisados.

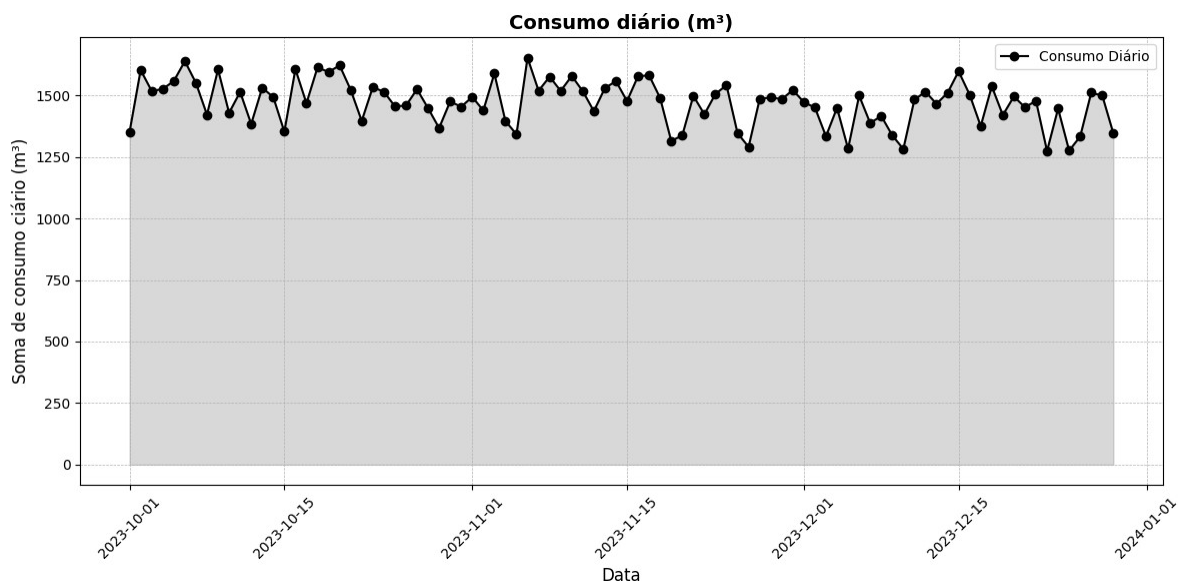


Figura 2 – Soma de consumo diário (m³).

O estudo utilizou o modelo *Light Gradient Boosting Machine (LightGBM)* para prever o consumo diário de hidrômetros com telemetria na cidade de Cuiabá-MT. A Tabela 1 abaixo apresenta as métricas de desempenho do modelo.

Tabela 1 – Avaliação do modelo de aprendizagem de máquina *LightGBM*.

| Model | MAE | RMSE | R ² |
|--|--------|--------|----------------|
| <i>Light Gradient Boosting Machine</i> | 0.4266 | 0.7114 | 0.6089 |



No gráfico da Figura 3 está a comparação entre o consumo real e previsto para o conjunto de teste do modelo LIGHTGBM.

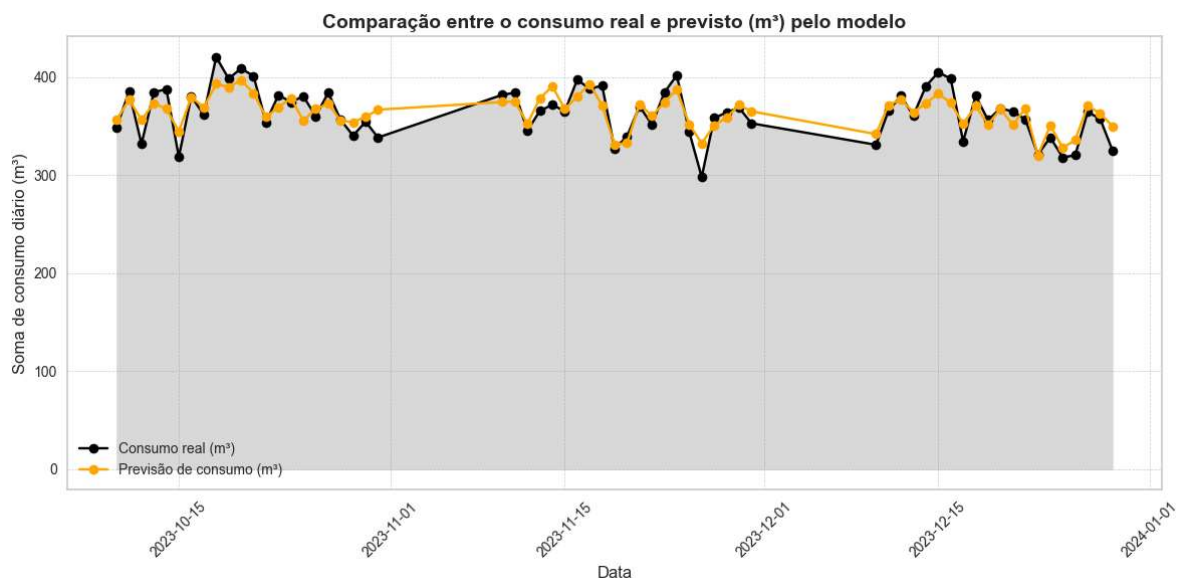


Figura 3 – Comparação entre o consumo real e previsto (m^3) pelo modelo.

No gráfico da Figura 4 estão os resíduos do modelo *LightGBM*, evidenciando a distribuição dos resíduos (diferença entre os valores previstos e reais) tanto para o conjunto de treinamento quanto para o de teste. Observa-se que os resíduos estão mais concentrados próximos a zero, embora existam alguns desvios maiores.

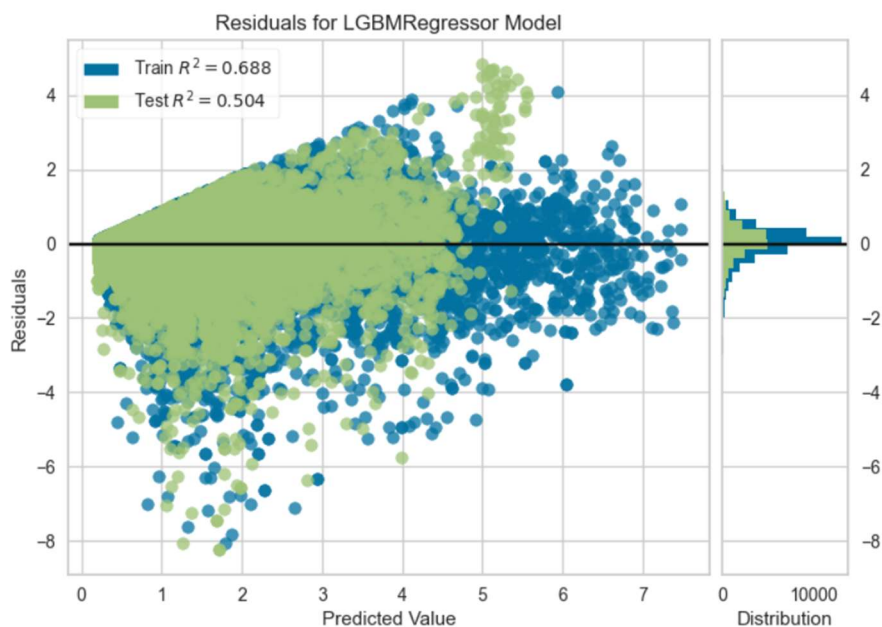


Figura 4 – Distribuição de resíduo do modelo *LIGHTGBM*.

Na Figura 5 está o gráfico de importância das variáveis que destaca que o "Consumo médio mensal dos últimos 3 meses (m^3)" é a variável mais importante no modelo, seguida pela "Temperatura média móvel de 7 dias ($^{\circ}C$)" e "Consumo médio das 10 ligações mais próximas (m^3)".

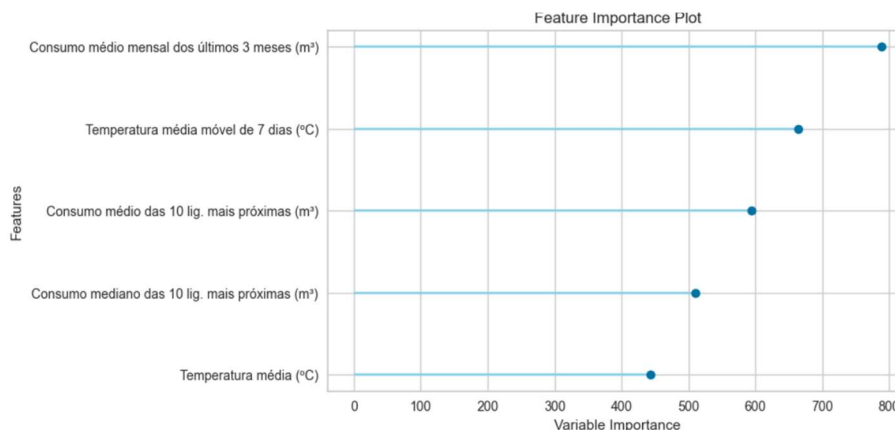


Figura 5 – Gráfico de importância das variáveis utilizadas.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os resultados mostram que o modelo *LightGBM* é capaz de prever o consumo diário de água com um MAE de 0.4266 m³, indicando que, em média, as previsões desviam-se dos valores reais em aproximadamente 0.4266 m³. O RMSE de 0.7114 m³, que é maior que o MAE, sugere a presença de alguns erros maiores nas previsões, provavelmente devido a outliers. O coeficiente de determinação (R²) de 0.6089 indica que o modelo explica cerca de 60.89% da variabilidade no consumo diário, sugerindo uma boa capacidade de captura da tendência geral dos dados, embora ainda haja espaço para melhorias.

A análise da distribuição do consumo diário revelou uma concentração de registros em baixos níveis de consumo, o que pode influenciar as métricas de desempenho do modelo. O gráfico de importância das variáveis indicou que o histórico de consumo médio dos últimos três meses é o fator mais influente na previsão, seguido de variáveis climáticas como a temperatura média móvel de 7 dias. A análise dos resíduos do modelo mostrou que, embora a maioria dos resíduos esteja concentrada em torno de zero, há alguns desvios maiores, indicando áreas para potencial melhoria no modelo.

A comparação entre os valores reais e previstos de consumo diário demonstra que o modelo é capaz de capturar bem as flutuações diárias, mas as divergências em certos pontos sugerem que o modelo pode ser melhorado com a inclusão de mais variáveis explicativas ou técnicas avançadas de modelagem.

Em comparação com o trabalho de Ferreira (2023), que também utilizou o modelo *LightGBM* para prever o consumo de água em Paranaguá-PR, observa-se que autor obteve um desempenho ligeiramente superior, com um MAE de 0.1807 m³, um RMSE de 0.3131 m³ e um R² de 0.462. A metodologia de Ferreira (2023) destacou a importância da micromedição e características específicas dos consumidores, como o consumo médio mensal e mediano, e foi eficaz para prever padrões de consumo detalhados.

Além disso, comparações com outros estudos mostram variações nas abordagens e resultados. Adamowski et al. (2012) compararam múltiplos modelos, incluindo redes neurais artificiais e métodos autorregressivos, para previsão de demanda de água em Montreal, Canadá, e encontraram que modelos híbridos de redes neurais e análise de ondas proporcionaram previsões precisas com um R² superior a 0.7. Este estudo sugere que métodos híbridos podem oferecer melhorias adicionais na precisão preditiva. Rai & Khadka (2020) revisaram vários fatores que influenciam o consumo residencial de água, destacando a importância das características socioeconômicas e comportamentais, como tamanho do domicílio e número de habitantes, na melhoria da precisão das previsões. Cheng et al. (2021) afirmam que a combinação de hidrômetros inteligentes com técnicas de aprendizagem de máquina resulta em previsões mais precisas e detalhadas, corroborando a eficácia da abordagem usada neste trabalho e no estudo de Ferreira.



CONCLUSÕES/RECOMENDAÇÕES

Neste trabalho, foi utilizado o modelo *Light Gradient Boosting Machine (LightGBM)* para realizar a previsão do consumo diário de água em Cuiabá-MT, utilizando dados de micromedição e variáveis climáticas. Os resultados indicam que o modelo é capaz de capturar a variabilidade dos dados de consumo de água com um MAE de 0.4266 m³, RMSE de 0.7114 m³ e um R² de 0.6089. Observou-se que há espaço para melhorias, especialmente em relação aos grandes desvios observados nos resíduos. A análise da importância das variáveis reforça a relevância do histórico de consumo e das condições climáticas na previsão do consumo de água.

A análise da distribuição do consumo diário revelou uma concentração de registros em baixos níveis de consumo, o que pode influenciar as métricas de desempenho do modelo. Comparando com outro estudo que também utilizou o *LightGBM* para prever o consumo de água em uma cidade diferente, observa-se que a metodologia que incluiu micromedição e características específicas dos consumidores proporcionou previsões ligeiramente mais precisas. Outros estudos também destacam a eficácia de modelos híbridos e a inclusão de características socioeconômicas e comportamentais para melhorar a precisão das previsões de consumo de água.

Para melhorar a precisão do modelo, recomenda-se a inclusão de variáveis adicionais que podem influenciar o consumo de água, como dados socioeconômicos (renda, número de moradores por residência), comportamento de uso da água e eventos especiais (feriados, festivais).

Além disso, a exploração de técnicas avançadas de aprendizado de máquina, como redes neurais profundas e algoritmos de aprendizado reforçado, pode oferecer melhorias adicionais na precisão das previsões.

Em suma, a aplicação de modelos de aprendizagem de máquina para a previsão de consumo de água mostrou-se promissora, mas há diversas oportunidades para aprimorar a precisão e a utilidade dos modelos através da inclusão de mais dados. Essas recomendações visam não apenas melhorar a previsão de consumo, mas também promover uma gestão mais eficiente e sustentável dos recursos hídricos.

AGRADECIMENTOS

Gostaria de agradecer à Iguá Saneamento pelo valioso apoio e generosa contribuição de dados para esta pesquisa. Reconheço e admiro a posição de liderança da empresa ao investir em análise de dados de hidrômetros inteligentes com tecnologia *IoT* para previsão de consumo de água. Sua parceria estratégica foi fundamental para o sucesso deste trabalho, permitindo-nos explorar novas abordagens e insights relevantes. Agradeço pela oportunidade de trabalhar em conjunto e pela contribuição para a construção de um futuro mais sustentável.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ADAMOWSKI, J.; CHAN, H. F.; PRASHER, S. O. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, v. 48, n. 1, 2012.
2. ARSENE, D. et al. Advanced strategies for monitoring water consumption patterns in households based on IoT and machine learning. *Water*, MDPI, v. 14, n. 14, p. 2187, 2022.
3. BORETTI, A.; ROSA, L. Reassessing the projections of the world water development report. *NPJ Clean Water*, Nature Publishing Group UK London, v. 2, n. 1, p. 15, 2019.
4. CHANG, Fi-John; CHANG, Li-Chiu; CHEN, Jui-Fa. Artificial Intelligence Techniques in Hydrology and Water Resources Management. *MDPI*, v. 15, p. 1846, 2023.
5. CHENG, M.; et al. Smart metering for water distribution systems: a review on technology, applications, and challenges. *Water Research*, v. 195, 2021.
6. CIENTÍFICA, Metodologia. Fundamentos de metodologia científica. São Paulo: Atlas, 2008.
7. FERREIRA, L. R. A. Previsão do consumo de água da rede de abastecimento através de modelos de aprendizagem de máquina. 2023. Monografia (Especialização em Data Science e Big Data) – Universidade Federal do Paraná, Curitiba, 2023. Orientador: Luiz Eduardo S. Oliveira.
8. GARCIA-SOTO, CG et al. Water consumption time series forecasting in urban centers using deep neural networks. *Applied Water Science*, Springer, v. 14, n. 2, p. 1–14, 2024.



9. GIL, Antonio Carlos. Métodos e técnicas de pesquisa social. 6. ed. São Paulo: Atlas, 2008.
10. GIL-GAMBOA, A et al. Medium-term water consumption forecasting based on deep neural networks. *Expert Systems with Applications*, Elsevier, v. 247, p. 123234, 2024.
11. HE, C. et al. Future global urban water scarcity and potential solutions. *Nature Communications*, Nature Publishing Group UK London, v. 12, n. 1, p. 4667, 2021.
12. IGUÁ Saneamento. Iguá tem o maior percentual de clientes com medição remota no Brasil. Disponível em: <https://igua.com.br/noticias/igua-tem-o-maior-percentual-de-clientes-com-medicao-remota-no-brasil>. Acesso em: 17 jun. 2024.
13. PANDEY, P. et al. Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, v. 147, n. 2, p. 04020106, 2021.
14. PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, JMLR. org, v. 12, p. 2825–2830, 2011.
15. PESANTEZ, J. E.; BERGLUND, E. Z.; KAZA, N. Smart meters data for modeling and forecasting water demand at the user-level. *Environmental Modelling & Software*, Elsevier, v. 125, p. 104633, 2020.
16. RAI, P.; KHADKA, A. Factors Influencing Residential Water Consumption: A Review and Meta-Analysis. *Water*, v. 12, n. 5, p. 1470, 2020.
17. SAGAR, A. PyCaret: An Open Source, Low-Code Machine Learning Library in Python. arXiv preprint arXiv:2102.10377, 2021.
18. SALEHI, M. Global water shortage and potable water safety; Today's concern and tomorrow's crisis. *Environment International*, Elsevier, v. 158, p. 106936, 2022.
19. SOARES ASCENÇÃO, É. et al. Applications of Smart Water Management Systems: A Literature Review. *Water*, MDPI, v. 15, n. 19, p. 3492, 2023.
20. STAŃCZYK, J. et al. Improving short-term water demand forecasting using evolutionary algorithms. *Scientific Reports*, Nature Publishing Group UK London, v. 12, n. 1, p. 13522, 2022.