

DETECÇÃO DE ANOMALIAS EM REDES DE DISTRIBUIÇÃO DE ÁGUA UTILIZANDO INTELIGÊNCIA ARTIFICIAL E MACHINE LEARNING

Yuri Crotti⁽¹⁾

Possui Mestrado em Tecnologia da Informação e Comunicação e Bacharelado em Engenharia da Computação pela Universidade Federal de Santa Catarina. Atualmente, lidera o time de Engenharia de Machine Learning na galax.ia, Também é professor na UniSATC, formando futuros líderes da indústria.

Giovani Gracioli⁽¹⁾

Professor na UFSC e Co-fundador da galax.ia, atuando como Diretor de P&D&I no SecCPS desde 2022. Possui experiência anterior como Professor Assistente e Pesquisador Associado na Technical University of Munich e University of Waterloo.

Felipe Willian Fraport⁽¹⁾

Cofundador da galax.ia desde setembro de 2021, onde trabalha para combater a escassez hídrica com soluções tecnológicas. Anteriormente, atuou como Engenheiro de Software Embarcado e como Desenvolvedor de Software Embarcado, desenvolvendo software para sistemas de automação e controle de motores.

Thiago Zimmermann Loureiro Chaves⁽¹⁾

Possui graduação em Ciência da Computação pela Universidade Federal de Santa Catarina (2023). Atualmente, é mestrando na mesma instituição (2024). Tem experiência na área de Ciência da Computação, com ênfase em Deep Learning, atuando principalmente nos seguintes temas: Deep learning. Atualmente atua como Engenheiro de Machine Learning na galax.ia.

Endereço⁽¹⁾: Av. Juscelino Kubitschek, n.º 350, 2º Andar - Centro - Joinville - SC - CEP: 89201-100 - Brasil
- Tel: +55 (48) 988469248 - e-mail: yuri.crotti@galax-ia.com

RESUMO

A escassez de água é uma questão global crítica, intensificada pelo aumento populacional. No Brasil, as perdas na rede de distribuição de água atingem 40%, causando prejuízos econômicos e tarifas mais altas. Este trabalho desenvolveu uma abordagem baseada em Inteligência Artificial (IA) e Machine Learning para detectar anomalias em redes de distribuição de água. Utilizando sensores IoT, os dados foram coletados e pré-processados, aplicando algoritmos de aprendizado de máquina, como Isolation Forest, para identificar vazamentos e irregularidades. Os resultados mostram que o Logger 1 apresentou a maior taxa de anomalias, seguido pelos Loggers 4, 3 e 2. Esta metodologia demonstrou eficácia na priorização de loggers para investigação, permitindo ações corretivas e preventivas. A aplicação de IA e Machine Learning na gestão hídrica pode reduzir perdas, melhorar a eficiência e promover a sustentabilidade do uso da água. Este estudo contribui significativamente para o desenvolvimento de soluções tecnológicas que enfrentam os desafios da escassez hídrica, promovendo a sustentabilidade e a eficiência no uso dos recursos hídricos.

PALAVRAS-CHAVE: Detecção de Anomalias, Aprendizado de Máquina, Inteligência Artificial, Gestão de Água, Sustentabilidade.

INTRODUÇÃO

A escassez de água limpa é uma das questões mais prementes enfrentadas pelo mundo moderno, que abriga uma população de 7,7 bilhões de pessoas. Com a previsão de aumento populacional global para entre 9,4 e 10,2 bilhões até 2050, representando um incremento de 22 a 34%, a pressão sobre os sistemas hídricos tende a intensificar-se consideravelmente. Atualmente, cerca de 47% da população mundial, aproximadamente 3,6 bilhões de pessoas, vivem em áreas que sofrem com a escassez de água por pelo menos um mês a cada ano. Algumas estimativas sugerem que esse número pode ser ainda maior, atingindo 52% da população global, ou seja, 4,0 bilhões de pessoas (Boretti e Rosa, 2019).

A Pesquisa Nacional de Saneamento Básico de 2017, conduzida pelo IBGE, trouxe à tona questões cruciais relacionadas ao abastecimento de água e ao esgotamento sanitário. No Brasil, as perdas na rede de distribuição de água são um problema crítico, com uma média de perdas de mais de 40%, equivalente a cerca de 7 bilhões de metros cúbicos por ano (IBGE, 2020). Essas perdas não só geram prejuízos econômicos significativos, dificultando o investimento em projetos de expansão e melhoria da distribuição de água, mas também resultam em tarifas mais altas para os consumidores.

Para enfrentar esses desafios, é essencial implementar estratégias eficazes de gestão integrada dos recursos hídricos, que incluam a redução das perdas nos sistemas de distribuição, a reutilização segura de águas residuais, a dessalinização e a alocação apropriada de água. Além disso, tecnologias de economia de água e campanhas de conscientização são cruciais para promover o uso sustentável dos recursos hídricos.

Uma das soluções tecnológicas promissoras para combater a escassez hídrica e reduzir as perdas é a implementação de sistemas avançados de monitoramento e gestão de água. Tecnologias como sensores IoT (Internet das Coisas) podem detectar vazamentos em tempo real, permitindo reparos rápidos e reduzindo desperdícios. Além disso, técnicas de inteligência artificial (IA) e aprendizado de máquina podem otimizar a distribuição de água, prevendo demandas e ajustando automaticamente os fluxos para minimizar perdas (Krishnan et al 2022).

Uma das soluções tecnológicas mais promissoras para combater a escassez hídrica e reduzir as perdas é a aplicação de inteligência artificial (IA) para detecção de anomalias e possíveis vazamentos. A IA pode analisar dados de sensores em tempo real para identificar padrões que indicam vazamentos ou outras irregularidades na rede de distribuição de água. Algoritmos de aprendizado de máquina são treinados para reconhecer sinais sutis de vazamentos, como mudanças na pressão da água, variações de fluxo inesperadas e consumo anômalo (Hu et al, 2021), (El-Shafeiy et al, 2023).

Além disso, conforme apresentado por Fan, Zhang e Yu (2021), a IA pode ser integrada a sistemas de modelagem preditiva, ajudando a antecipar falhas e a implementar medidas preventivas. Sensores IoT monitoram continuamente parâmetros como pressão, fluxo e qualidade da água, fornecendo dados essenciais para os algoritmos de IA. Métodos ativos de detecção de vazamentos, que incluem análises de sinais acústicos e vibrações, são mais eficazes para monitoramento contínuo do que os métodos passivos.

Estudos de caso, conforme apresentado por Sun et al. (2019) e Sousa et al. (2023) e mostraram que a aplicação de modelos de aprendizado de máquina em redes de distribuição de água pode reduzir significativamente as perdas e melhorar a eficiência da gestão hídrica. Esses avanços não só ajudam a identificar e corrigir vazamentos rapidamente, mas também contribuem para a sustentabilidade e eficiência dos sistemas de distribuição de água.

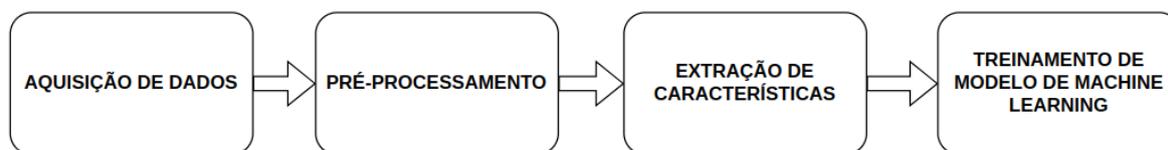
OBJETIVOS

O objetivo deste trabalho é desenvolver uma abordagem baseada em Inteligência Artificial, especificamente utilizando técnicas de Machine Learning, para a detecção de anomalias em dados de monitoramento de redes de água. A detecção precoce de anomalias, como vazamentos e irregularidades no fluxo, é essencial para melhorar a eficiência da distribuição de água, reduzir perdas e minimizar custos operacionais. Através da aplicação de algoritmos de aprendizado de máquina, este trabalho busca criar um sistema capaz de analisar dados em tempo real provenientes de sensores instalados na rede de distribuição, identificando padrões e desvios que possam indicar problemas. Com isso, pretende-se proporcionar uma ferramenta que auxilie na gestão proativa dos recursos hídricos, contribuindo para a sustentabilidade e a eficiência no uso da água.

METODOLOGIA

Neste trabalho, foram desenvolvidas quatro etapas metodológicas, conforme ilustrado na Figura 1. As etapas são: Aquisição de Dados, Pré-processamento, Extração de Características e Treinamento do Modelo.

Figura 1 – Etapas da metodologia realizada



Fonte: Autor.

AQUISIÇÃO DE DADOS

Neste trabalho, foram utilizados quatro dispositivos de aquisição de dados de pressão denominados Orion Alpha. Esses dispositivos têm como objetivo realizar leituras de pressão quando acoplados a hidrômetros. Os dados coletados foram obtidos no ano de 2023, em uma cidade de Rondônia, com o objetivo de capturar e analisar comportamentos de pressão. As aquisições de dados ocorrem a cada 10 minutos.

PRÉ-PROCESSAMENTO

No desenvolvimento de sistemas de machine learning, a qualidade dos dados é fundamental para o sucesso dos modelos. A etapa de pré-processamento desempenha um papel essencial nessa fase, garantindo que os dados estejam prontos para serem utilizados na extração de características.

Primeiramente, para cada série de dados, no pré-processamento verifica se ela possui 30% ou menos de valores nulos. Se essa condição for verdadeira, os valores nulos da série são substituídos pela mediana da própria coluna. Caso contrário, a série é eliminada. Este processo garante que apenas séries com uma quantidade aceitável de dados faltantes sejam mantidas, mitigando o impacto de dados ausentes no desempenho do modelo.

Além disso, os valores duplicados são removidos com base no índice, assegurando que cada entrada no conjunto de dados seja única. A remoção de duplicatas é crucial para evitar redundâncias que podem distorcer as análises e prejudicar a eficácia do modelo de machine learning.

EXTRAÇÃO DE CARACTERÍSTICAS

A extração de características é uma etapa fundamental no desenvolvimento de sistemas de machine learning, particularmente em problemas envolvendo séries temporais. A etapa do extrator de características foi projetada para transformar dados brutos em informações úteis para o treinamento de modelos. Foram desenvolvidos métodos para extração de características que serão apresentados a seguir:

1. O método de shifting desloca os valores de uma coluna para trás ou para frente por um número específico de períodos. Este método é utilizado para criar novas características que representam valores passados ou futuros, auxiliando na modelagem de dependências temporais entre os dados.
2. O método de percent change calcula a variação percentual em relação aos períodos anteriores. Este método captura a magnitude das mudanças ao longo do tempo, permitindo a análise da volatilidade ou estabilidade dos dados.
3. O método window-based envolve o cálculo de estatísticas móveis e expansivas, tais como média, mediana, soma, valor mínimo e valor máximo. Este método é utilizado para suavizar a série temporal, capturando tendências de longo prazo e variações de curto prazo.

4. O janelamento é uma técnica que divide a série temporal em segmentos menores, chamados janelas, para capturar padrões e tendências ao longo do tempo. Este método cria uma visão mais detalhada e informativa da série temporal, facilitando a detecção de padrões sazonais e tendências.
5. A técnica MSTL decompõe séries temporais em componentes distintas: sazonalidade múltipla, tendência e resíduos. A sazonalidade múltipla captura padrões sazonais recorrentes que ocorrem em múltiplas frequências, a tendência representa a direção geral da série temporal ao longo do tempo, e os resíduos capturam as variações aleatórias ou não explicadas pelos componentes sazonais e de tendência. A MSTL facilita a análise e a modelagem de séries temporais complexas, permitindo a identificação de padrões ocultos e a melhoria da precisão das previsões.

Os métodos fornecem uma base robusta para a extração de características em séries temporais. A utilização dessas técnicas permite transformar dados brutos em informações valiosas, aprimorando a construção de modelos machine learning.

TREINAMENTO DE MODELO DE MACHINE LEARNING

O modelo *Isolation Forest* é uma técnica de aprendizado de máquina amplamente utilizada para detecção de anomalias. Introduzido por Liu, Ting e Zhou (2008), o *Isolation Forest* (ou Floresta de Isolamento) se destaca por sua eficiência e simplicidade na identificação de dados atípicos.

Esse modelo cria várias árvores de decisões, onde cada árvore é uma estrutura binária formada pela divisão recursiva dos dados. Cada nó da árvore é gerado escolhendo aleatoriamente uma característica e, em seguida, selecionando aleatoriamente um valor de corte dentro do intervalo dos dados dessa característica. Esse processo continua até que cada ponto de dado esteja isolado em um nó folha.

O comprimento do caminho para isolar uma amostra, ou seja, o número de divisões necessárias para isolar o ponto, é usado como uma medida de anomalia. A intuição por trás disso é que as anomalias são mais fáceis de isolar e, portanto, têm comprimentos de caminho mais curtos nas árvores de isolamento.

A pontuação de anomalia de uma amostra é a média dos comprimentos de caminho em todas as árvores de isolamento, ajustada para o tamanho da amostra. A fórmula é dada pela equação 1.

$$s(x, n) = 2 \frac{E(h(x))}{c(n)} \quad \text{equação (1)}$$

onde $E(h(x))$ é o comprimento de caminho esperado de x e $c(n)$ é o comprimento de caminho esperado de uma amostra de n pontos. Uma pontuação próxima de 1 indica uma alta probabilidade de anomalia, enquanto uma pontuação próxima de 0 indica que a amostra é normal.

RESULTADOS

Para cada um dos quatro loggers de aquisição, foram gerados modelos de machine learning treinados para a detecção de anomalias, ou seja, comportamentos diferentes do normal. Na Tabela 1, pode-se observar o percentual de anomalias detectadas para cada um dos loggers. Este resultado mostra que o logger mais anômalo é o Logger 1, seguido pelos Loggers 4, 3 e 2. Isso evidencia que a abordagem é capaz para identificar os loggers que requerem maior atenção para análises aprofundadas.

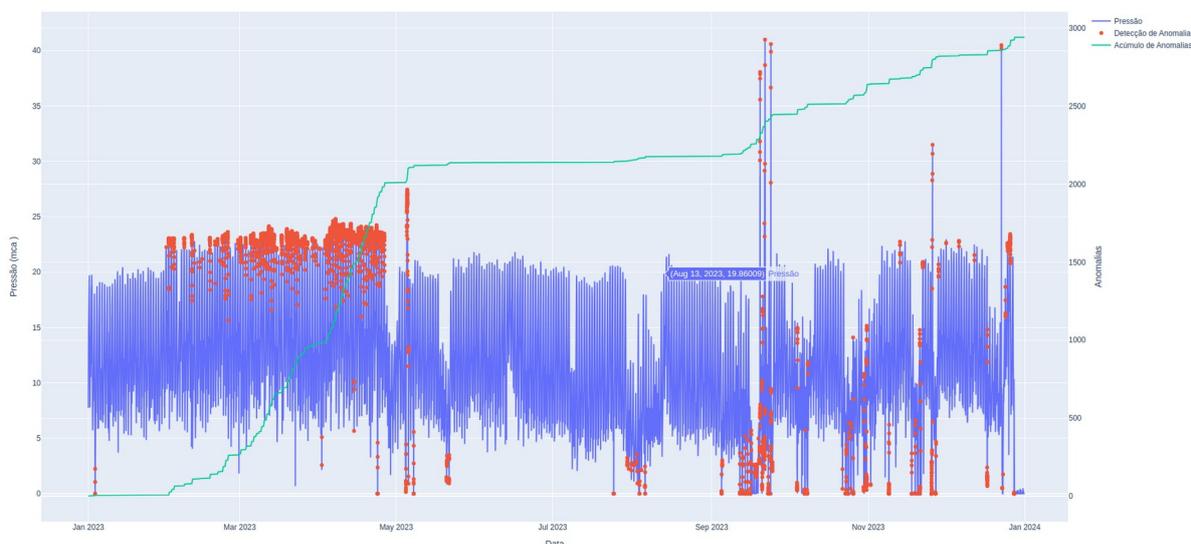
Tabela 1 – Ranqueamento de anomalias

Logger	Percentual de anomalias (%)
Logger 1	3,18 %
Logger 4	2,54 %
Logger 3	1,52 %
Logger 2	1,12 %

Fonte: elaboração própria.

Ao analisar os resultados do Logger 1, considerado o mais anômalo conforme apresentado na Figura 2, observa-se uma incidência relativamente baixa de anomalias entre janeiro e abril de 2023, com um acúmulo gradual (linha verde). A partir de maio de 2023, há um aumento considerável no número de anomalias detectadas, coincidente com uma maior variabilidade nas medições de pressão. Esse padrão pode indicar uma mudança significativa nas condições operacionais ou no comportamento do sistema. No entanto, em agosto de 2023, ocorre um pico extremo, onde tanto a pressão quanto as anomalias apresentam valores atípicos, sugerindo um evento anômalo significativo.

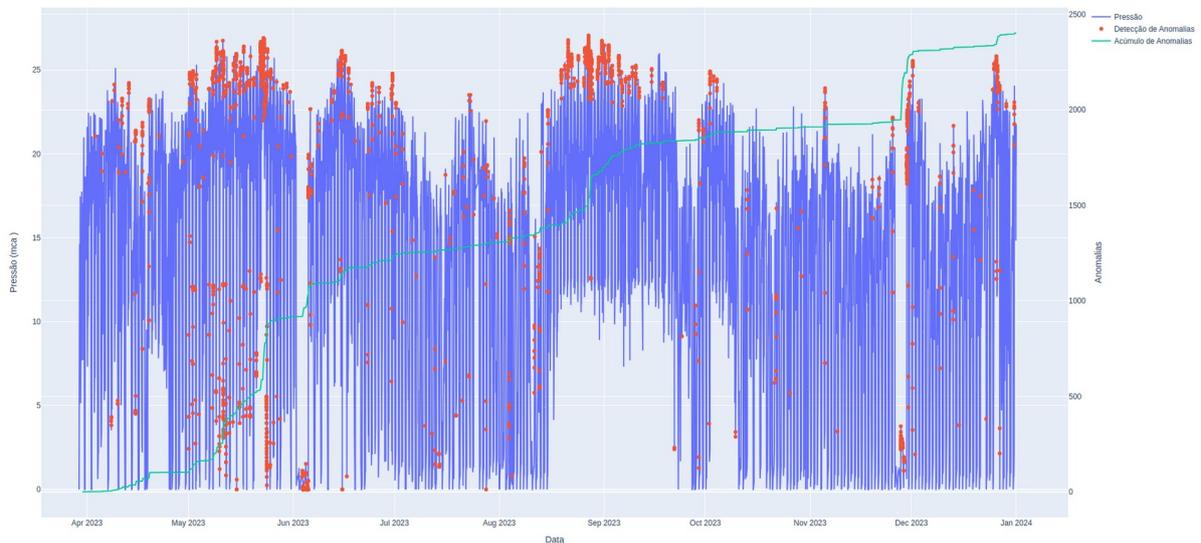
Figura 2 – Detecção de Anomalias Logger 1



Fonte: do Autor.

Agora para o Logger 4, conforme apresentado na Figura 3, entre abril e junho de 2023, observa-se um aumento gradual no número de anomalias detectadas, juntamente com flutuações na pressão. A partir de junho de 2023, há um aumento substancial na detecção de anomalias, coincidindo com um padrão de variabilidade maior nas leituras de pressão. Esse aumento nas anomalias pode indicar alterações nas condições operacionais ou problemas emergentes no sistema. Um pico notável de anomalias ocorre ao longo de vários períodos entre julho e agosto de 2023, onde a densidade de pontos vermelhos é alta, sugerindo eventos anômalos repetitivos ou persistentes.

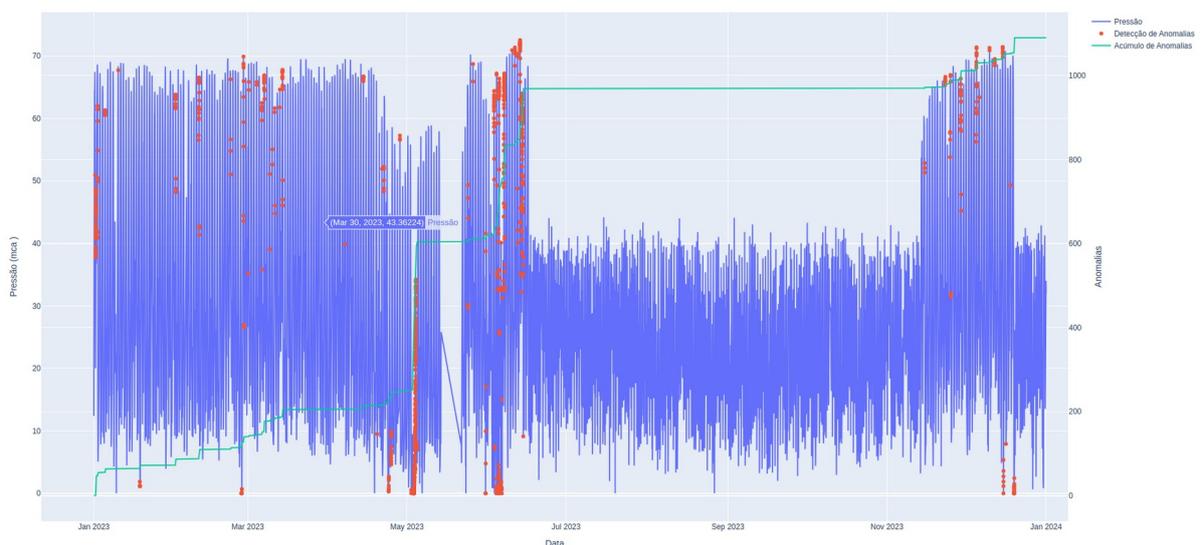
Figura 3– Detecção de Anomalias Logger 4



Fonte: do Autor.

Para o Logger 3, conforme mostrado na Figura 4, observa-se um aumento gradual no número de anomalias detectadas entre janeiro e maio de 2023, juntamente com flutuações na pressão. A partir de maio de 2023, há um aumento substancial na detecção de anomalias. No período de junho a novembro de 2023, não se observam detecções de anomalias, com uma pressão menos dinâmica.

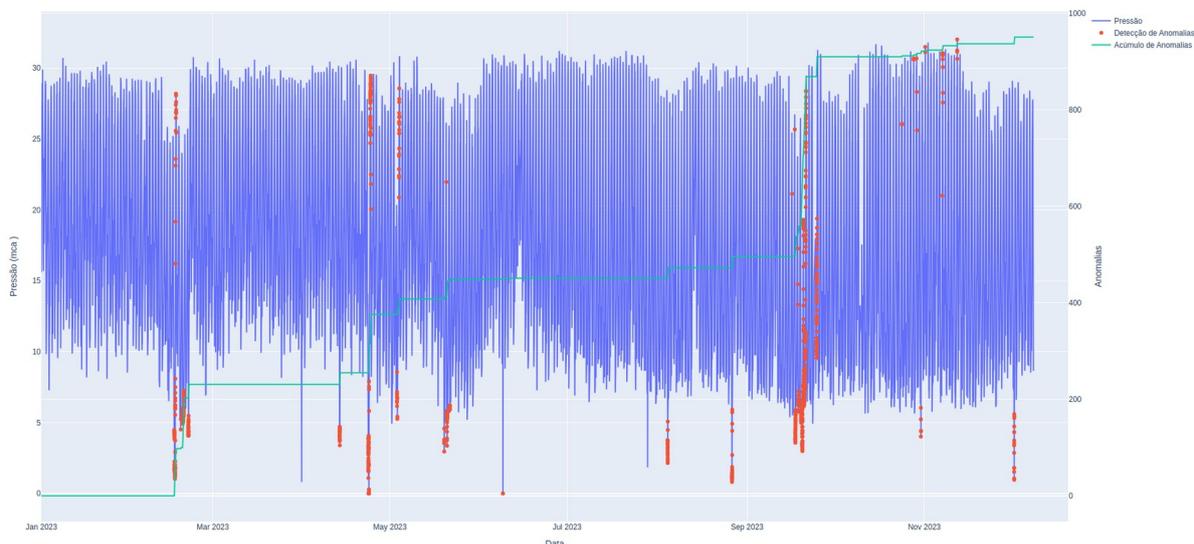
Figura 4 – Detecção de Anomalias Logger 3



Fonte: do Autor.

Por fim, para o Logger 2, considerado menos anômalo, conforme a Figura 5, observa-se que o índice de detecções é significativamente menor em comparação com os outros Loggers. Embora algumas inclinações sejam observadas, a quantidade final de anomalias não é elevada.

Figura 5 – Detecção de Anomalias Logger 2



Fonte: do Autor.

Esses resultados demonstram que a metodologia de detecção de anomalias é eficaz para identificar e priorizar os loggers que necessitam de maior vigilância e investigação, permitindo ações corretivas e preventivas mais direcionadas.

CONCLUSÕES

Este trabalho desenvolveu e aplicou uma abordagem baseada em Inteligência Artificial, utilizando técnicas de Machine Learning, para a detecção de anomalias em redes de distribuição de água. Com foco na eficiência da distribuição de água e na redução de perdas, foi possível criar um sistema capaz de analisar dados em tempo real provenientes de sensores, identificando padrões e desvios que indicam problemas como vazamentos e irregularidades no fluxo.

Os resultados demonstraram a eficácia da metodologia aplicada. A análise dos percentuais de anomalias detectadas revelou que o Logger 1 apresentou a maior taxa de anomalias, seguido pelos Loggers 4, 3 e 2. Esse ranking permitiu identificar os loggers que requerem maior atenção para análises aprofundadas. Em particular, o Logger 1 mostrou um aumento significativo de anomalias a partir de maio de 2023, culminando em um pico extremo em agosto de 2023, sugerindo eventos anômalos significativos.

A abordagem proposta mostrou-se capaz na identificação de anomalias, permitindo a priorização de loggers para investigação e manutenção preventiva. Esta capacidade de detecção precoce é crucial para melhorar a gestão dos recursos hídricos, reduzindo perdas e custos operacionais. Além disso, contribui para a sustentabilidade e eficiência no uso da água, um recurso cada vez mais escasso.

O uso de modelos de aprendizado de máquina, como o Isolation Forest, para a análise de dados de pressão em redes de distribuição de água, provou ser uma ferramenta capaz na identificação de anomalias. A implementação dessa tecnologia pode auxiliar operadores de sistemas de distribuição de água a tomar decisões informadas e proativas, melhorando a resiliência e a sustentabilidade dos sistemas de abastecimento.

Em suma, a aplicação de técnicas de Inteligência Artificial e Machine Learning na detecção de anomalias em redes de distribuição de água representa um avanço na gestão de recursos hídricos. Este trabalho contribui para o desenvolvimento de soluções tecnológicas que podem mitigar os desafios da escassez hídrica, promovendo a sustentabilidade e a eficiência no uso da água.

REFERÊNCIAS BIBLIOGRÁFICAS

- BORETTI, A., ROSA, L. *Reassessing the projections of the world water development report*. *NPJ Clean Water*, v.2, n.1, p.15, 2019.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Pesquisa Nacional de Saneamento Básico 2017: Abastecimento de água e esgotamento sanitário*. Rio de Janeiro: IBGE, 2020.
- KRISHNAN, S. R., NALLAKARUPPAN, M. K., CHENGODEN, R., KOPPU, S., IYAPPARAJA, M., SADHASIVAM, J., SETHURAMAN, S. *Smart water resource management using Artificial Intelligence —A review*. *Sustainability*, v.14, n.20, p.13384, 2022.
- HU, Z., CHEN, B., CHEN, W., TAN, D., SHEN, D. *Review of model-based and data-driven approaches for leak detection and location in water distribution systems*. *Water Supply*, v.21, n.7, p.3282-3306, 2021.
- EL-SHAFEIY, E., ALSABAAN, M., IBRAHEM, M. I., ELWAHSH, H. *Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique*. *Sensors*, v.23, n.20, p.8613, 2023.
- FAN, X., ZHANG, X., YU, X. B. *Machine learning model and strategy for fast and accurate detection of leaks in water supply network*. *Journal of Infrastructure Preservation and Resilience*, v.2, p.1-21, 2021.
- SUN, Congcong et al. *Leak localization in water distribution networks using pressure and data-driven classifier approach*. *Water*, v.12, n.1, p.54, 2019.
- SOUSA, D. P., DU, R., MAIRTON BARROS DA SILVA JR, J., CAVALCANTE, C. C., FISCHIONE, C. *Leakage detection in water distribution networks using machine-learning strategies*. *Water Supply*, v.23, n.3, p.1115-1126, 2023
- LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. *Isolation forest*. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE, p. 413-422, 2008.