

Unsupervised Water Leakage Detection Using IoT Devices and Anomaly Detection Algorithms

Robert A. Cabral¹, Robert V. O. Gonçalves², Maria V. Menezes², Paulo A. C. Aguiar²,
Elvis M. G. Stancanelli², Arthur C. Callado², Enyo J. T. Gonçalves²

¹ Instituto de Pesquisas Eldorado, Campinas, SP, Brazil

robertcabral.qx@gmail.com

² PCOMP – Universidade Federal do Ceará, Quixadá, CE, Brazil

robertvinicius@alu.ufc.br, {vivianemenezes, pauloaguiar,

elvis.stancanelli, arthur, enyo}@ufc.br

Abstract. This work proposes a water leakage detection system for residential settings using low-cost sensors and unsupervised machine learning algorithms. Water consumption data is collected by an STM32 microcontroller connected to a pulse-output water meter and transmitted via LoRaWAN for remote analysis. The models used Isolation Forest and Elliptic Envelope require no labeled data. Evaluation was conducted using three datasets, including one collected by the authors. The models achieved an average F1-score of 0.8579, performing well even in high-variability scenarios. The proposed solution stands out for its scalability, low cost, and adaptability to different household consumption patterns.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: *iot, machine learning, unsupervised*

1. INTRODUÇÃO

O Brasil enfrenta um cenário crítico de perdas de água tratada. Segundo o Instituto Trata Brasil, cerca de 37,78% da água é perdida durante a distribuição, volume suficiente para abastecer toda a população do Nordeste por um ano. As causas incluem vazamentos, redes antigas e falhas na medição, enquanto 32 milhões de brasileiros ainda não têm acesso regular à água tratada, evidenciando a dualidade entre escassez e desperdício [Instituto Trata Brasil 2024].

Essas perdas geram impactos ambientais, econômicos e sociais. Ambientalmente, elevam a pressão sobre mananciais e o uso de energia no tratamento [Instituto Trata Brasil 2024]. No aspecto econômico, a redução das perdas para 25% geraria um ganho líquido de R\$ 20,4 bilhões até 2034, mesmo considerando os investimentos necessários para tal [Ministério do Desenvolvimento Regional 2021]. A Agência Nacional de Águas destaca que reduzir perdas é essencial para a segurança hídrica e a resiliência climática [ANA 2024].

Diante desse contexto, soluções baseadas em Internet das Coisas (IoT) e aprendizado de máquina têm se mostrado promissoras para o monitoramento do consumo e a detecção de vazamentos. Estudos como [Di Mauro et al. 2019] e [Guezouli et al. 2024] apresentam arquiteturas com sensores e modelos supervisionados, porém, geralmente dependentes de dados rotulados ou estruturas complexas, o que limita sua adoção em residências.

Este trabalho propõe uma abordagem escalável, de baixo custo e baseada em aprendizado não supervisionado. Um sistema embarcado acoplado ao hidrômetro envia, via LoRaWAN, leituras de

2 • Robert A. Cabral, Robert G., Maria V. Menezes, Paulo A. C., Elvis M. G. S., Arthur Callado, Enyo G.

consumo para um servidor remoto, onde algoritmos detectam padrões anômalos sem necessidade de dados rotulados. Segundo [Liu et al. 2008] e [Khan et al. 2023], algoritmos não supervisionados são eficazes na identificação de anomalias em dados ruidosos e desbalanceados, como os de consumo residencial.

A proposta é validada por meio da comparação de diferentes modelos aplicados a múltiplos conjuntos de dados, avaliando *precision*, *recall*, *F1-score* e tempo de processamento. A principal vantagem está na capacidade de adaptação ao perfil de consumo de cada residência, permitindo detecção eficiente de vazamentos em cenários com escassez de dados históricos.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Internet das Coisas (IoT)

A Internet das Coisas (IoT) conecta objetos físicos à Internet, viabilizando a coleta e o compartilhamento autônomo de dados por meio de sensores embarcados [Gubbi et al. 2013]. Essa integração permite o monitoramento contínuo de ambientes e a automação inteligente de processos [Santos et al. 2021]. Entre as tecnologias de conectividade, destaca-se o LoRaWAN (*Long Range Wide Area Network*), que oferece comunicação de longa distância com baixo consumo energético [Mohammed et al. 2024]. Baseado na modulação LoRa, esse protocolo permite a transmissão de dados em áreas urbanas e rurais, mesmo com infraestrutura limitada [Centenaro and Vangelista 2022].

Na gestão hídrica, essas soluções tornam-se essenciais diante da escassez de recursos e dos altos índices de perdas de água no Brasil [ANA 2024], permitindo monitorar o consumo, detectar vazamentos e otimizar o uso da água.

2.2 Aprendizado Supervisionado e Não Supervisionado

O Aprendizado de Máquina (AM) é uma subárea da Inteligência Artificial focada no desenvolvimento de algoritmos capazes de extrair padrões dos dados e aprimorar seu desempenho em tarefas específicas sem programação explícita [Baranauskas 2003]. Os métodos de AM são classificados em supervisionados e não supervisionados, conforme a presença ou ausência de rótulos nos dados [Ludermir 2021]. Outros paradigmas, como o aprendizado por reforço, não serão abordados neste trabalho, dado o escopo e a natureza do problema.

No aprendizado supervisionado, o modelo aprende a partir de exemplos rotulados, sendo eficaz em tarefas como classificação e regressão. Já o aprendizado não supervisionado busca estruturar os dados com base em similaridade, agrupamentos ou detecção de anomalias [Hastie et al. 2009], sendo particularmente útil quando a obtenção de rótulos é inviável ou custosa, como na detecção de vazamentos, onde não há histórico confiável de ocorrências anotadas.

2.3 Técnicas de Detecção de Anomalias

Entre os métodos de aprendizado não supervisionado, destacam-se os algoritmos de detecção de anomalias, que identificam instâncias que se desviam significativamente do padrão esperado. Este trabalho compara dois modelos com abordagens distintas: *Isolation Forest* e *Elliptic Envelope*.

O *Isolation Forest*, proposto por [Liu et al. 2008] e aprimorado por [Hariri et al. 2021], assume que anomalias são mais fáceis de isolar devido à sua baixa frequência e afastamento dos demais pontos. O modelo constrói múltiplas árvores binárias com partições aleatórias, e a profundidade média necessária para isolar um ponto indica seu grau de anomalia. Já o *Elliptic Envelope* pressupõe uma distribuição normal multivariada, estimando uma elipse que delimita a região de maior densidade; pontos fora dessa fronteira são considerados anômalos [Pedregosa et al. 2011]. Embora mais sensível a dados gaussianos, é eficaz em cenários com baixa variabilidade.

3. TRABALHOS RELACIONADOS

A crescente demanda por soluções sustentáveis no gerenciamento de recursos hídricos tem impulsionado pesquisas sobre detecção de vazamentos com IA, aprendizado de máquina e IoT. Esta seção revisa trabalhos relevantes, destacando suas principais contribuições e diferenças em relação à proposta deste estudo.

[Guezouli et al. 2024] apresenta o QoW-Pro, uma arquitetura IoT baseada no algoritmo *Random Forest* para monitoramento da qualidade da água e detecção de vazamentos. Embora escalável, sua abordagem supervisionada depende de dados rotulados, o que pode limitar sua adoção em larga escala. O sistema utiliza dados reais obtidos de um protótipo residencial com sensores conectados a um ESP8266.

[Kulikovas et al. 2023] investiga métodos não supervisionados e semi-supervisionados, como *Isolation Forest*, PCA e SSDO, aplicados a dois conjuntos de dados reais: o *dataset DAIAD*¹, composto por dados horários de 92 residências em Alicante - Espanha, e outro com medições da residência de um dos autores. Como não havia registros de vazamentos, foram simulados cenários de anomalias para validação dos modelos.

[Selvaprasath et al. 2025] propõe um sistema supervisionado com sensores de pressão e vazão conectados a um NodeMCU via Wi-Fi. A detecção é feita por uma rede neural do tipo *Long Short-Term Memory* (LSTM), especializada no processamento de dados sequenciais, combinada com um mecanismo de atenção para focar nas informações mais relevantes. O sistema apresentou alto desempenho em testes simulados, mas sua dependência de dados rotulados limita a escalabilidade.

Diferentemente dos trabalhos anteriores, a proposta deste artigo combina características das abordagens citadas. Assim como [Guezouli et al. 2024] e [Selvaprasath et al. 2025], propomos um pipeline completo de coleta, transmissão e análise de dados em tempo real. No entanto, utilizamos algoritmos não supervisionados (*Isolation Forest* e *Elliptic Envelope*), que não exigem dados rotulados, tornando o sistema escalável para múltiplas residências com padrões de consumo distintos.

4. METODOLOGIA

A Figura 1 ilustra o fluxo geral do sistema proposto. O processo inicia com a leitura do consumo por um hidrômetro pulsante conectado a um microcontrolador, que registra os dados localmente a cada 10 minutos. Em seguida, os dados são transmitidos para a nuvem, onde são acumulados continuamente.

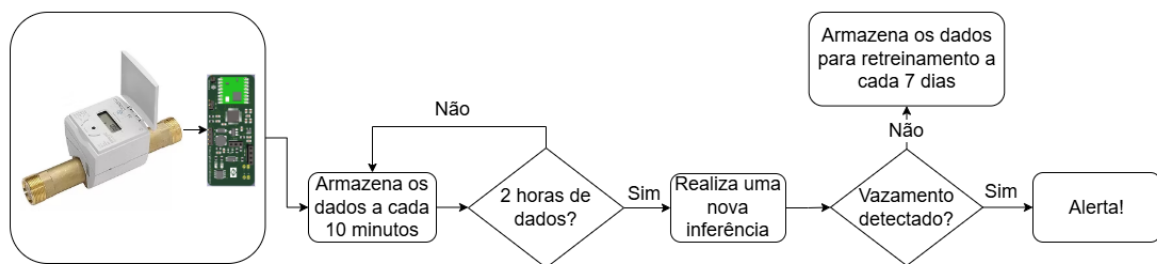


Fig. 1: Arquitetura do sistema.

A cada novo lote de 2 horas de consumo, o sistema realiza uma inferência com modelos de detecção de anomalias previamente treinados. Se não houver anomalias, os dados seguem sendo acumulados. Caso um padrão compatível com vazamento seja identificado, um alerta é gerado e enviado à concessionária de águas responsável.

¹Dataset DAIAD

4 • Robert A. Cabral, Robert G., Maria V. Menezes, Paulo A. C., Elvis M. G. S., Arthur Callado, Enyo G.

O sistema também armazena os dados para reprocessamento e readequação dos modelos, realizados a cada 7 dias. Essa estratégia permite a adaptação às variações de consumo e preserva a acurácia ao longo do tempo. O uso de modelos não supervisionados ainda favorece a escalabilidade, viabilizando sua rápida aplicação em novas residências, mesmo sem dados históricos.

4.1 Protótipo do sistema

O sistema foi implementado a partir do mesmo protótipo descrito por [Cabral et al. 2025], baseado no microcontrolador STM32F030C8T6. A placa de circuito impresso foi projetada para ser compacta, eficiente e de baixo consumo energético, sendo especialmente adequada para aplicações em Internet das Coisas (IoT).

A Figura 2 apresenta o protótipo, destacando a placa eletrônica desenvolvida, além da bateria e do painel solar conectados a um hidrômetro ultrassônico com saída pulsada. Este mesmo protótipo foi utilizado para a captação dos dados que compõem um dos conjuntos analisados neste estudo.

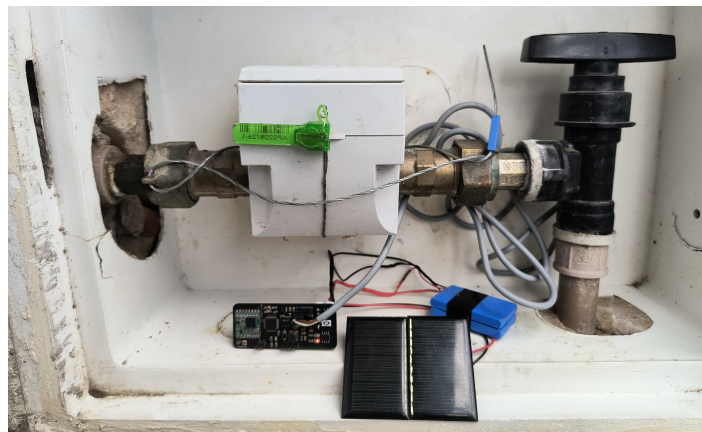


Fig. 2: Protótipo do sistema. [Cabral et al. 2025]

4.2 Aquisição do *Dataset* Próprio

A coleta do *dataset* próprio foi realizada em ambiente residencial, utilizando o protótipo descrito na Seção 4.1. O dispositivo foi instalado no ponto de medição principal da residência, conectado a um hidrômetro pulsado que gera um pulso por litro consumido.

O microcontrolador registrou, a cada 10 minutos, a quantidade acumulada de pulsos, correspondente ao volume de água consumido, e transmitiu os dados via LoRaWAN para um servidor remoto. O uso do LoRaWAN permite o envio de dados com um consumo energético baixíssimo. A descrição técnica da arquitetura LoRaWAN não será detalhada, por se tratar de uma camada de suporte e não do foco deste trabalho, que aborda a detecção de vazamentos por meio de aprendizado não supervisionado.

A residência pertence a um dos autores, o que possibilitou controle contínuo das condições experimentais. A coleta ocorreu por 40 dias, resultando em cerca de 57 mil amostras, com a rotina de consumo mantida, assegurando a representatividade dos padrões observados.

4.3 Conjuntos de Dados Utilizados

Este estudo utilizou três conjuntos de dados para avaliar a eficácia dos modelos de detecção de vazamentos. O primeiro é um *dataset* próprio, coletado a partir do protótipo embarcado desenvolvido

(Seção 4.2), representando fielmente os padrões de uso doméstico, com maior controle sobre as variáveis.

O segundo conjunto, com aproximadamente 260 mil registros, é o mesmo utilizado por [Kulikovas et al. 2023], possibilitando uma comparação direta com resultados da literatura. O terceiro foi extraído de [Di Mauro et al. 2019] e contém cerca de 360 mil registros, resultantes da concatenação dos arquivos individuais de consumo por cômodo, consolidando o uso total da residência..

Cada conjunto de dados manteve sua granularidade temporal original, com padronização das variáveis de entrada. Essa diversidade foi crucial para avaliar variados contextos de consumo, ampliando a validade externa e reforçando a escalabilidade do sistema proposto. Os datasets utilizados estão disponíveis publicamente no Kaggle.

4.4 Modelos de Aprendizado Não Supervisionado

Este estudo utiliza dois modelos de aprendizado não supervisionado na detecção de vazamentos: *Isolation Forest* e *Elliptic Envelope*. Ambos dispensam rótulos no treinamento, tornando o sistema escalável e aplicável em cenários sem dados anotados.

No *Isolation Forest*, a contaminação (*contamination*) foi ajustada para 0.007, mesma configuração aplicada ao *Elliptic Envelope*, permitindo comparação direta. O número de árvores no *Isolation Forest* seguiu o valor padrão (100).

Ambos os algoritmos buscam identificar anomalias, mas variações naturais no consumo, como o uso de chuveiro, podem gerar falsos positivos isolados, que cessam rapidamente quando o consumo retorna ao normal. Para mitigar esse efeito e aumentar a robustez, aplicou-se um pós-processamento com o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). Esse agrupamento permite distinguir eventos prolongados que caracterizam vazamentos reais. Definiu-se como evento significativo a ocorrência de pelo menos três anomalias dentro de uma janela de cinco minutos.

A Figura 3 apresenta um exemplo prático da aplicação do *Elliptic Envelope* com pós-processamento por DBSCAN sobre o *dataset* próprio. Observa-se que, embora diversas anomalias tenham sido sinalizadas ao longo do período, apenas o vazamento simulado no final do gráfico possui densidade suficiente para ser reconhecido como um evento real.

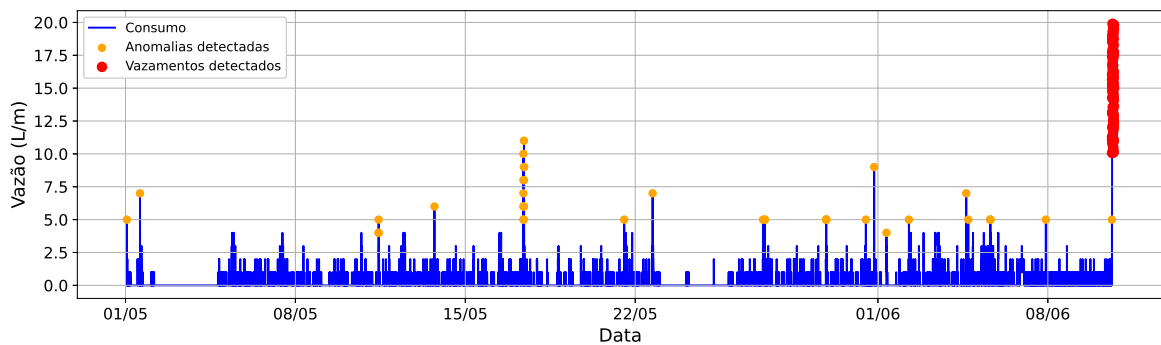


Fig. 3: Aplicação do *Elliptic Envelope* com DBSCAN no *dataset* próprio

Também foram testados outros algoritmos, como *One-Class SVM* e *Local Outlier Factor (LOF)*. No entanto, ambos apresentaram desempenho insatisfatório na detecção deste tipo de anomalia, possivelmente impactado pela distribuição dos dados, caracterizada por consumo baixo e intermitência.

6 • Robert A. Cabral, Robert G., Maria V. Menezes, Paulo A. C., Elvis M. G. S., Arthur Callado, Enyo G.

4.5 Métricas

Os modelos foram avaliados com base nas métricas de *precision*, *recall*, *F1-score*, além dos tempos médios de treinamento e inferência, medidos durante cem execuções consecutivas para maior robustez estatística.

A métrica de *precision* mede a proporção de verdadeiros positivos entre todas as instâncias classificadas como positivas pelo modelo, indicando sua capacidade de evitar falsos positivos. A *recall* avalia a proporção de verdadeiros positivos identificados em relação ao total de positivos reais, refletindo a capacidade de detectar vazamentos reais. Já o *F1-score* é a média harmônica entre *precision* e *recall*, proporcionando uma medida única que equilibra ambas as métricas, especialmente útil em cenários com classes desbalanceadas.

4.6 Procedimentos de Testes e Simulações

Para validar a eficácia dos modelos escolhidos na detecção de vazamentos de água, foi conduzido um conjunto sistemático de testes e simulações, utilizando três diferentes conjuntos de dados. Cada *dataset* foi dividido em 80% para treinamento e 20% para inferência, respeitando a ordem temporal, de modo que a inferência simulasse um cenário futuro, condição essencial para aplicações reais.

Na etapa de inferência, foram simulados vazamentos de diferentes intensidades, refletindo padrões típicos de consumo anômalo. A tipologia adotada segue [Lambert 2000], que classifica vazamentos como visíveis, não visíveis e inerentes (inferiores a 4,17 litros por minuto). Com base nisso, estabeleceu-se a seguinte categorização para este estudo:

- Leve: vazamentos com taxa entre 1 a 3 litros por minuto;
- Moderado: vazamentos com taxa entre 4 a 9 litros por minuto;
- Severo: vazamentos com taxa entre 10 a 30 litros por minuto.

Para cada categoria, foram inseridas 10 ocorrências com duração de 120 minutos, intercaladas por blocos reais de dados normais, preservando o padrão de consumo e garantindo um cenário realista.

Apesar dos modelos serem não supervisionados, a avaliação adotou uma abordagem supervisionada, viabilizada pela rotulagem dos dados simulados (normal ou vazamento). Essa prática, comum em estudos experimentais, permite calcular métricas como *precision*, *recall* e *F1-score* mesmo em métodos não supervisionados.

Os experimentos foram conduzidos com scripts desenvolvidos em *Python*, utilizando bibliotecas como *Scikit-Learn*, *Pandas* e *Matplotlib*. O código-fonte está disponível publicamente no Kaggle para fins de reprodutibilidade e futuras pesquisas.

5. RESULTADOS

Nesta seção, são apresentados os resultados dos modelos *Elliptic Envelope* e *Isolation Forest* aplicados aos três conjuntos de dados: *dataset* próprio, [Kulikovas et al. 2023] e [Di Mauro et al. 2019].

A Figura 4 mostra que o *Elliptic Envelope* obteve os melhores resultados médios, com *F1-score* superior e maior estabilidade. O menor *Recall* ocorreu no *dataset* de [Kulikovas et al. 2023] (0,7983 no *Elliptic* e 0,7959 no *Isolation*), indicando perda de cerca de 20% dos verdadeiros positivos, reflexo do baixo volume de consumo, dificultando a detecção de vazamentos leves. Ainda assim, ambos os modelos mantiveram boa eficácia na detecção de vazamentos moderados e severos. O *dataset* próprio teve comportamento semelhante, devido os longos períodos de consumo muito baixo. No conjunto de [Di Mauro et al. 2019], com maior variabilidade e volumes mais altos, os modelos, especialmente o *Elliptic Envelope*, alcançaram seus melhores desempenhos, com *Recall* de 0,8974 e *F1-score* de 0,8815.

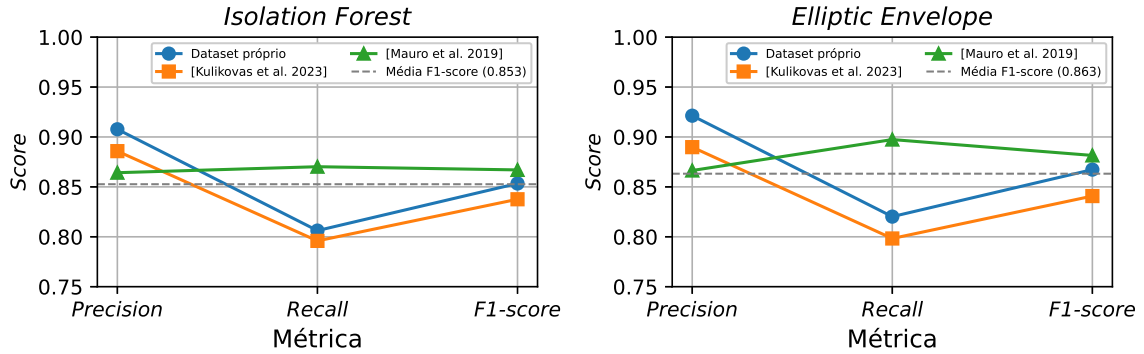


Fig. 4: Comparação dos modelos.

A Tabela I mostra que o *Isolation Forest* teve treinamento mais rápido (0,53 a 0,70s), porém com tempo de inferência que cresce com o tamanho dos dados (até 0,1748s). Por outro lado, o *Elliptic Envelope* demandou mais tempo para treinamento (até 20,46s), mas manteve inferência praticamente instantânea (menos de 0,005s), sendo mais indicado para aplicações em tempo real, como cenários em que o modelo é treinado a cada sete dias, mas realiza inferências constantes a cada poucas horas.

Tabela I: Tempo de treino e inferência dos modelos

Dataset	Amostras	Modelo	Treino (s)	Inferência (s)
Próprio	57.946	Isolation Forest	0.5380	0.1437
		Elliptic Envelope	4.4144	0.0026
[Kulikovas et al. 2023]	263.813	Isolation Forest	0.5463	0.1361
		Elliptic Envelope	15.0991	0.0037
[Di Mauro et al. 2019]	358.569	Isolation Forest	0.7061	0.1748
		Elliptic Envelope	20.4623	0.0049

Em termos gerais, o *Elliptic Envelope* apresentou o melhor desempenho médio, com *F1-score* superior a 0,86 no *dataset* próprio e no de [Di Mauro et al. 2019]. Sua alta velocidade de inferência o torna especialmente promissor para aplicações embarcadas em tempo real. O *Isolation Forest*, embora também apresente bom desempenho, mostrou-se mais sensível a variações nos dados, com quedas mais acentuadas no *Recall* em cenários de baixo consumo. Destaca-se ainda que ambos os modelos mantiveram bom desempenho frente a diferentes tamanhos de *datasets*, desde cenários com cerca de 40 dias de dados até outros com mais de 250 dias, evidenciando a escalabilidade e robustez da abordagem.

6. CONCLUSÃO

Este trabalho apresentou uma solução para detecção de vazamentos residenciais baseada em IoT e aprendizado não supervisionado, utilizando LoRaWAN e os modelos *Isolation Forest* e *Elliptic Envelope*. O *Elliptic Envelope* apresentou o melhor desempenho médio, com *F1-score* superior nos três conjuntos e inferência quase instantânea, ideal para aplicações em tempo real. O *Isolation Forest* destacou-se pelo treinamento mais rápido, mas com maior sensibilidade a cenários de baixo consumo.

A utilização de três conjuntos de dados foi essencial para validar os modelos em diferentes padrões de consumo, ampliando a robustez do estudo. Nos cenários de menor consumo, como no *dataset* próprio e em [Kulikovas et al. 2023], observou-se perda de aproximadamente 20% dos verdadeiros positivos, reflexo da dificuldade na detecção de vazamentos leves. Contudo, a detecção de vazamentos moderados e severos manteve-se eficiente. No conjunto de [Di Mauro et al. 2019], os modelos, especialmente o *Elliptic Envelope*, apresentaram desempenho superior.

Como trabalhos futuros, propõe-se investigar a adaptação dinâmica ao consumo do protótipo. Será crucial avaliar a viabilidade de sua instalação e o alcance em ambientes urbanos densos. Para validação

8 • Robert A. Cabral, Robert G., Maria V. Menezes, Paulo A. C., Elvis M. G. S., Arthur Callado, Enyo G.

dos modelos, utilizaremos a silhueta. Focaremos também na integração com sistemas de notificação, validação em ambientes diversos e otimizações energéticas. A análise visual (Figura 3) também identificou consumos pequenos e contínuos durante a madrugada, evidenciando possíveis microvazamentos. O uso do mesmo protótipo aliado a modelos de aprendizado de máquina pode contribuir positivamente na detecção de padrões de consumo contínuo.

Os resultados reforçam o potencial da abordagem como uma solução escalável, de baixo custo e aplicável tanto em residências quanto em políticas públicas de gestão hídrica.

REFERENCES

- ANA. *Conjuntura dos Recursos Hídricos no Brasil 2024*. ANA, 2024.
- BARANAUSKAS, J. A. Sistemas inteligentes: fundamentos e aplicações. In *Sistemas Inteligentes*. Departamento de Computação e Matemática, FFCLRP-USP, Ribeirão Preto, 4, 2003. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em: 10 maio 2025.
- CABRAL, R. A., C., P. A., BONFIM, M. S., AND COUTINHO, E. F. Desempenho energético no freertos com paralelismo em relação a sistema sequencial em um protótipo de hidrômetro lora. In *LII Seminário Integrado de Software e Hardware (SEMISH)*. Brasil, 2025. Disponível em: <https://sol.sbc.org.br/index.php/semish/article/view/36808/36594>.
- CENTENARO, M. AND VANGELISTA, L. Lorawan: A low power wide area network for the internet of things: A survey. *Electronics* 11 (1): 164, 2022.
- DI MAURO, A., DI NARDO, A., SANTONASTASO, G. F., AND VENTICINQUE, S. An iot system for monitoring and data collection of residential water end-use consumption. In *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*. IEEE, pp. 1–6, 2019.
- GUBBI, J., BUYIA, R., MARUSIC, S., AND PALANISWAMI, M. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems* 29 (7): 1645–1660, 2013.
- GUEZOULI, L., GUEZOULI, L., DJEGHABA, M. B. E., AND BENTAHROUR, A. Iot and ai for real-time water monitoring and leak detection. *Journal of Renewable Energies*, 2024.
- HARIRI, S., CARRASCO KIND, M., AND BRUNNER, R. J. Extended isolation forest. *IEEE Transactions on Knowledge and Data Engineering* 33 (4): 1479–1489, 2021.
- HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.
- INSTITUTO TRATA BRASIL. Estudo de Perdas de Água 2024: Desafios e oportunidades para o avanço do saneamento no Brasil, 2024. Disponível em: <https://www.tratabrasil.org.br/perdas-agua-2024>. Acesso em: maio de 2024.
- KHAN, M. A., SILVA, R., AND TRAN, L. Recent advances on anomaly detection using isolation forest variants, 2023.
- KULIKOVAS, LUKAS, PACKEVIČIUS, AND ŠARŪNAS. Anomalous water use detection using machine learning. In *Information Society and University Studies (IVUS)*. Kaunas, Lithuania, 2023.
- LAMBERT, A. What do we know about pressure: Leakage relationships in distribution systems? International Water Association (IWA), 2000. Disponível em: <https://repositorio.ufsc.br/handle/123456789/124431>.
- LIU, F. T., TING, K. M., AND ZHOU, Z.-H. Isolation forest. In *IEEE International Conference on Data Mining*. pp. 413–422, 2008.
- LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados* 35 (101): 99–118, jan-abr, 2021.
- MINISTÉRIO DO DESENVOLVIMENTO REGIONAL. Portaria nº 490, de 22 de março de 2021. <https://www.in.gov.br/en/web/dou/-/portaria-n-490-de-22-de-marco-de-2021-309988760>, 2021. Estabelece os procedimentos gerais para o cumprimento de índices de perda de água na distribuição. Diário Oficial da União, 23 de março de 2021.
- MOHAMMED, A.-W., ZHANG, Q., ALAZAB, M., RAZZAK, I., JOLFAEI, A., AND GADEKALLU, T. R. Low-power wide-area networks (lpwans) for internet of things: A comprehensive survey. *ACM Computing Surveys*, 2024.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETENTENHOFFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COUNAPEAU, D., BRUCHER, M., AND PERROT. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* vol. 12, 2011.
- SANTOS, M. M., FRANCO, M. L., ALVES, W. M., DE SOUZA, M. C., BARROS, G. F., AND DA SILVA, M. M. Internet das coisas: a busca do conceito e as perspectivas futuras sobre sua aplicabilidade. *Research, Society and Development* 10 (10): e140101018504, 2021.
- SELVAPRAKASH, J., ARASU, P. S. K., XAVIER, V. S., MARIA, O. R. R., MANIMEGALAI, M., AND SEBASTHIRANI, K. Water leakage detection and recognition system. In *3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2025.